



Vince In Bono Mafum

881201

UNIVERSIDAD ANAHUAC

ESCUELA DE ACTUARIA

CON ESTUDIOS INCORPORADOS A LA
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

2
2001

**"SIMULACION ESTOCASTICA DE MODELOS
DINAMICOS APLICADA A LA MANUFACTURA"**

T E S I S

QUE PARA OPTAR EL TITULO DE:

A C T U A R I O

P R E S E N T A :

MARIA ELVIRA GUTIERREZ LOPEZ

DIRECTOR DE TESIS: ACT. JESUS OROZCO TOPETE

MEXICO, D. F.

1995



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A DIOS

A MIS PAPAS

Por su Cariño, Apoyo y Comprensión

A MI HERMANA

Por su Compañía y Alegría

A MIS ABUELITOS

Los Pilares de la Familia

A JESUS

Por Todo su Apoyo y Gran Entusiasmo

A YAMEL

Por el Trabajo Siempre con Alegría

A ANGEL

Por su Amor

A MI NUEVA FAMILIA

Por su Cariño y Aceptación

A MIS TIOS, PRIMOS Y AMIGOS . . .

Que los Quiero Tanto

INDICE

INTRODUCCION

I	NOCIONES DE SIMULACION	1
1.1	Concepto	1
1.2	Ventajas del uso de la simulación	2
1.3	Modelos	5
1.4	Clasificación de los modelos para simulación	8
1.5	Simulación en manufactura	8
II	PROBABILIDAD Y ESTADISTICA	11
2.1	Introducción	11
2.2	Variables aleatorias y sus propiedades	11
2.2.1	Variable aleatoria discreta	12
2.2.2	Variable aleatoria continua	13
2.2.3	Variables aleatorias n-dimensionales	14
2.2.4	Variables aleatorias independientes	16
2.2.5	Parámetros	16
2.3	Resultados de la simulación y procesos estocásticos	19
2.4	Estimación de medias, varianzas y correlaciones	20
2.5	Intervalos de confianza y pruebas de hipótesis para la media	23
2.6	Distribuciones de probabilidad	27
2.6.1	Parametrización de distribuciones continuas	29
2.6.2	Distribuciones continuas	29
2.6.3	Distribuciones discretas	43
2.6.4	Distribuciones empíricas	51
2.7	Elección de distribuciones de probabilidad	52
2.7.1	Pruebas chi-cuadrada	54
2.7.2	Pruebas Kolmogorov-Smirnov	58

III	METODOLOGIA	62
3.1	Formulación del problema y planeación del estudio	63
3.2	Construcción del modelo	63
3.3	Recolección y procesamiento de datos	65
3.3.1	Recolección y almacenamiento	65
3.3.2	Validación	67
3.3.2	Conversión	68
3.3.4	Manipulación	68
3.4	Evaluación del modelo y estimación de las distribuciones de probabilidad	69
3.5	Formulación de un programa para la computadora	71
3.6	Pruebas piloto y validación	73
3.7	Diseño de experimentos	73
3.7.1	Diseños Factoriales 2^k	74
3.7.2	Tratando con muchos factores	77
3.8	Experimentación	80
3.9	Análisis de los resultados	80
3.9.1	Análisis de los datos de salida para un solo sistema	80
3.9.2	Técnicas de reducción de varianza	93
3.10	Documentación e implementación de los resultados	97
IV	CASO PRACTICO	98
4.1	Formulación del problema y planeación del estudio	98
4.2	Construcción del modelo	99
4.2.1	Flujo de materiales	102
4.2.2	Ciclos de producción	103
4.2.3	Generación de la demanda en la lavadora	104
4.2.4	Descompostura de las líneas de producción	105
4.2.5	Movimiento de la banda transportadora	105
4.2.6	Colgado en la banda transportadora	105
4.2.7	Turnos de producción	108
4.2.8	Asignación de espacios en el almacén	108
4.2.9	Operación y descompostura de los montacargas	110

4.3	Recolección y procesamiento de datos	112
4.3.1	Recolección y almacenamiento	112
4.3.2	Validación	113
4.3.3	Conversión y manipulación	113
4.4	Evaluación del modelo y estimación de las distribuciones de probabilidad	114
4.5	Formulación de un programa para la computadora	114
4.6	Pruebas piloto y validación	115
4.7	Diseño de experimentos	115
4.8	Experimentación	116
4.9	Análisis de los resultados	117
4.10	Documentación e implementación de los resultados	117
4.10.1	Retraso de la producción	118
4.10.2	Número de canastillas en el almacén	118
4.10.3	Número de partes en espera de ser colgadas	118

CONCLUSIONES

ANEXOS

BIBLIOGRAFIA

INTRODUCCION

INTRODUCCION

Actualmente, el incremento de la competencia extranjera y la firma del Tratado de Libre Comercio (TLC) entre México, Estados Unidos y Canadá ha propiciado que las empresas mexicanas busquen ser más competitivas tanto a nivel nacional como internacional.

Anteriormente, en un mundo totalmente protegido, las empresas daban por hecho que podían promover sus productos y servicios en todos los segmentos del mercado nacional, sin preocuparse demasiado por su competitividad frente a otros productores nacionales y menos frente a los fabricantes extranjeros.

La llegada de los competidores ha significado la quiebra o transformación de muchas empresas. En el proceso de globalización de la economía mundial hay cada vez menos lugar para quienes operan con costos mayores a los del promedio de sus competidores.

Se están abriendo las fronteras mexicanas en una época en que las economías de Asia y Europa tienen como objetivo incrementar aún más sus ya elevados niveles de productividad; y en que la creación de nuevos mercados requiere de inversiones considerables de capital y la mano de obra barata ya no representa la ventaja comparativa de hace algunos años (ventaja que las economías de Taiwan y Corea utilizaron con bastante éxito).

En el nuevo entorno competitivo las empresas más exitosas están revisando a fondo sus estrategias. Muchas empresas están buscando mejorar la calidad de sus productos a través de la introducción del control estadístico de los procesos, los círculos de calidad y el buen manejo de relaciones clientes - proveedores en todas las áreas de la empresa. Asimismo, se han establecido programas de reducción de costos y mejora en la productividad para asegurar una estructura de costos bajos. Las empresas más agresivas han cerrado líneas o plantas incosteables, ampliado y modernizado los procesos de producción en las plantas eficientes y optimizado la logística de distribución de los productos.

La tecnología, elemento al que no se le había prestado toda la importancia debida, ha incrementado su valor porque la innovación se ha convertido en un factor competitivo clave. Las empresas empiezan a reforzar su tecnología para desarrollar nuevos productos y optimizar los procesos productivos en forma permanente.

INDUCE (Investigación y Desarrollo Universitario para la Competitividad Empresarial) fué un proyecto conjunto entre la Universidad Anáhuac y Hewlett-Packard de México. Creado en 1986 como un centro de investigación y desarrollo que buscaba incrementar la competitividad de las empresas de manufactura por medio de sistemas computacionales con tecnología de punta. INDUCE, en su labor de apoyo a la empresa mexicana, detectó la necesidad del uso de la simulación como una herramienta para incrementar la productividad y competitividad de las empresas de manufactura; por lo que en 1991 se realizó este trabajo, siendo publicado en 1995.

En términos generales, la simulación es una técnica de análisis de modelos en la que a partir de una situación real o prevista construye un modelo para realizar una serie de experimentos (generalmente en computadora) con el fin de obtener conclusiones aplicables a la misma realidad.

Debido a que la simulación aplicada al área de manufactura emula en computadora (sin tener que experimentar en la realidad) los efectos de cambios en el diseño y operación de una planta, la dirección puede tomar decisiones oportunas que permitan el incremento de la productividad y de la competitividad de la empresa ante el mercado.

Existen diversos tipos de simulación:

- Financiera
- De Procesos
- Entrenamiento
- Estocástica de Procesos Dinámicos

Esta tesis está enfocada a la simulación estocástica de procesos dinámicos. La simulación es estocástica ya que utiliza variables aleatorias para describir los procesos reales, y dinámica ya que es cambiante en el tiempo.

La simulación estocástica de procesos dinámicos (a partir de ahora referida como simulación) es una herramienta muy difundida en otros países ya que permite la reducción de costos y de la inversión de capital. Sin embargo, el uso de la simulación es casi desconocido por las empresas de manufactura nacionales.

La simulación puede abarcar diversas áreas, sin embargo el objetivo de este trabajo consiste en mostrar una metodología de simulación para llevar a cabo básicamente operaciones de diseño y rediseño de la distribución de una planta que se aplique a la industria manufacturera nacional. Esta metodología muestra todos los pasos a seguir para efectuar un estudio de simulación, así como, las

limitaciones que tiene su aplicación en las empresas mexicanas.

Este trabajo fue realizado de la siguiente manera: revisión de la bibliografía existente, propuesta de metodología aplicada a México, caso práctico en planta de manufactura totalmente mexicana y afinación de la metodología para su aplicación óptima en la industria nacional.

La presente Tesis está estructurada de la siguiente manera:

En el capítulo I se explica el significado de la simulación. Se definen conceptos y se enumeran las ventajas del uso de la simulación, así como las diversas aplicaciones enfocadas al área de manufactura.

En el capítulo II se presentan los conceptos básicos de probabilidad y estadística necesarios para poder llevar a cabo un estudio de simulación adecuado y obtener resultados estadísticamente correctos.

En el capítulo III se expone la metodología propuesta para la industria manufacturera nacional. Se detallan los puntos en que es necesario hacer énfasis debido a la mentalidad del personal mexicano.

El capítulo IV se dedica a la aplicación de la metodología descrita en el capítulo anterior sobre un caso práctico en una empresa nacional que se encontraba en proceso de rediseño de la distribución de la planta. Este caso se llevó a cabo tanto en la empresa como en las instalaciones de INDUCE quien proporcionó el equipo de cómputo necesario para correr el modelo y calcular las distribuciones de probabilidad requeridas. Para un fácil seguimiento, el capítulo IV consta de las mismas secciones que el capítulo III.

CAPITULO I

NOCIONES DE SIMULACION

I. NOCIONES DE SIMULACION

Actualmente, el verbo *simular* se ha puesto de moda en numerosas disciplinas científicas para describir el viejo arte de la construcción de un modelo.

El uso moderno de la palabra *simulación* proviene de fines de 1940 cuando los científicos von Neumann y Ulam inventaron el término "análisis de Monte Carlo" para aplicarlo a una técnica matemática utilizada en la resolución de problemas de protección nuclear que eran demasiado costosos para resolverse experimentalmente y muy complejos para estudiarse analíticamente.

A principios de 1950 surgieron las primeras computadoras de alta velocidad gracias a las cuales la simulación tomó otro significado ya que surgió la posibilidad de experimentar con modelos matemáticos en computadora. Muchas aplicaciones se transportaron a las computadoras y, como consecuencia, se logró resolver un mayor número de problemas teóricos y prácticos.

En esta tesis se presenta una metodología aplicada a la industria manufacturera nacional para resolver problemas de simulación aplicados al diseño y rediseño de plantas manufactureras.

I.1. CONCEPTO

A pesar de que en los párrafos anteriores se ha utilizado la palabra *simulación*, ahora se buscará proporcionar una definición más precisa. Desafortunadamente, no existe un acuerdo universal sobre su significado exacto.

En términos generales, la simulación resuelve problemas cuando el análisis teórico no puede. Es una técnica de análisis que construye un modelo a partir de una situación real, realiza sobre éste una serie de experimentos y obtiene conclusiones aplicables a la realidad.

Sin embargo, si se busca una definición más formal y acorde a los objetivos de esta tesis, se tiene: Simulación es una técnica numérica para conducir experimentos en una computadora, los cuales requieren de ciertos tipos de modelos matemáticos que describan el comportamiento de un negocio o sistema económico (o algún componente de ellos) en periodos extensos de tiempo real.

1.2 VENTAJAS DEL USO DE LA SIMULACION

La investigación científica utiliza el llamado método científico que consta de los siguientes pasos:

- 1. Observación de un sistema físico.**
- 2. Formulación de una hipótesis (en este caso modelo matemático) que intenta explicar las observaciones efectuadas sobre el sistema.**
- 3. Predicción del comportamiento del sistema, en base a la hipótesis formulada y mediante la deducción matemática, esto es, por la obtención de soluciones del modelo matemático.**
- 4. Experimentación para probar la validez de las hipótesis o del modelo matemático.**

Sin embargo, debido a ciertos factores explicados posteriormente no siempre resulta recomendable o factible seguir estos pasos, en cuyo caso la simulación podría considerarse como un sustituto satisfactorio para el estudio del proceso.

En primer lugar, la observación de ciertos procesos en el mundo real puede resultar muy costosa o imposible de realizar, por lo que la simulación puede ser utilizada para generar datos numéricos que describan procesos.

En segundo lugar, si el sistema observado es tan complejo que resulta imposible describirlo en términos de un sistema de ecuaciones matemáticas, se pueden obtener soluciones analíticas por medio de simulación para ser utilizadas con propósitos predictivos.

En tercer lugar, aún cuando un modelo matemático logre formularse para describir algún sistema, puede no obtenerse una solución del modelo por medio de técnicas analíticas directas y, consecuentemente, tampoco se harán predicciones acerca del comportamiento futuro del sistema. Sin embargo, el modelo puede ser utilizado para la simulación del sistema; cabe aclarar que aunque este método no garantiza soluciones óptimas y exactas del modelo, es posible experimentar con diferentes soluciones y determinar las que resulten de mayor utilidad en la predicción del comportamiento del sistema.

En cuarto lugar, si la experimentación se lleva a cabo en el mundo real las condiciones bajo las cuales se lleva a cabo el experimento pueden variar de experimento a experimento. En cambio, si se utiliza la simulación las condiciones pueden ser fijadas para todos los experimentos por medio de variables exógenas.

Finalmente, puede resultar imposible o muy costoso la experimentación que valide los modelos matemáticos que describen un sistema, por lo que se recomienda el uso de la simulación para probar las hipótesis propuestas.

Otras razones que soportan la ventaja del uso de la simulación sobre el uso del método científico son:

1. La simulación permite estudiar las interacciones que se presentan en el interior de un sistema dado (empresa, industria, economía o un subsistema de cualquiera de ellos), así como experimentar con ellas.
2. Por medio de la simulación se estudian los efectos que puede sufrir la operación de un sistema real al producirse cambios (informativos, de organización y ambientales).
3. La observación detallada del sistema simulado, permite un mejor entendimiento del mismo y proporciona sugerencias para su optimización.
4. La simulación puede ser utilizada como un recurso pedagógico en la enseñanza de los conocimientos básicos del análisis teórico y estadístico, así como, en la toma de decisiones.
5. Frecuentemente, el conocimiento que se obtiene al diseñar un estudio de simulación sugiere cambios en el sistema en cuestión. Entonces, los efectos producidos por estos cambios pueden probarse por medio de la simulación antes de ser implantados en el sistema real.
6. La simulación ayuda a determinar las variables clave dentro de un sistema y la manera en que interactúan entre sí.
7. La simulación permite experimentar con nuevas situaciones sobre las que se tiene poca o ninguna información, con el objeto de estar preparados para cualquier eventualidad.

8. La simulación sirve como un *sistema de preservación* en el que se prueban nuevas políticas y reglas de decisión en la operación de un sistema, antes de tomar el riesgo de experimentar con el sistema real.

9. La simulación permite estudiar los sistemas dinámicos.

10. Al presentarse nuevos componentes de un sistema, la simulación puede ayudar a determinar los obstáculos y problemas que resulten de la operación del sistema.

11. La simulación convierte a los especialistas en técnicos generales. Obliga a los analistas a apreciar y entender todos los aspectos del sistema, por lo que las conclusiones serán menos susceptibles a la parcialidad por inclinaciones particulares y a volverse impracticables dentro de la configuración del sistema.

En una encuesta realizada en 1978 por el Instituto de Ciencias Administrativas (TIMS por sus siglas en inglés: The Institute of Management Sciences) y la Sociedad de Investigación de Operaciones (ORSA por sus siglas en inglés: Operations Research Society of America), se determinó que la simulación es utilizada frecuentemente por las organizaciones con bastante éxito. Actualmente, gracias a la existencia de computadoras cada vez más veloces y accesibles, el uso de la simulación ha ido incrementándose año con año.

Sin embargo, han existido diversos impedimentos para que la simulación sea completamente aceptada. En primer lugar, los modelos emulan sistemas reales sumamente complejos, por lo que la elaboración del código de programación puede resultar una tarea ardua; actualmente, esta labor ha sido simplificada gracias al desarrollo de software que proporciona automáticamente muchas de las características requeridas en un modelo de simulación. Un segundo problema consiste en que los modelos de simulación requieren de mucho tiempo de computadora; afortunadamente, el costo de las computadoras continua reduciéndose en la actualidad. Finalmente, muchos "estudios" de simulación han consistido únicamente en la construcción del modelo, programación y una sola corrida, por lo que las conclusiones obtenidas han sido incorrectas provocando recelo sobre la efectividad de la simulación.

1.3. MODELOS

Un *modelo* es la abstracción de un sistema real que tiene la posibilidad de emplearse para propósitos de predicción y control.

Debido a que la simulación involucra la creación de un modelo a partir de un sistema real, es conveniente explicar brevemente el significado de éste último. Un *sistema* puede ser descrito como el conjunto de elementos interrelacionados que busca cumplir con un objetivo. La mayoría de los sistemas no se encuentran aislados sino que funcionan dentro de un ambiente que afecta su comportamiento. El establecer las fronteras del sistema para la construcción del modelo no es sencillo, ya que se tienen problemas en la elección de los componentes debido a que:

- La mayor parte de los sistemas están compuestos por subsistemas, resultando difícil establecer los niveles de detalle de estos subsistemas.
- La mayoría de los sistemas forman parte de sistemas más grandes y la conceptualización del sistema en marcos muy pequeños puede conducir a problemas de suboptimización (se optimiza el subsistema pero el sistema se encuentra por debajo de su nivel óptimo).
- Los sistemas tienden a interactuar y a sobreponerse entre sí.
- Aún cuando las fronteras del sistema han sido establecidas es necesario crear ligas con el medio ambiente.

Se busca que el modelo sirva como una aproximación razonable del sistema real y que incorpore la mayor parte de los elementos importantes de éste; por otro lado, no es conveniente que el modelo sea muy complejo y resulte imposible entenderlo o manipularlo.

Un modelo consta de *variables* que se emplean para relacionar los componentes del mismo. Las variables se clasifican en:

a) Exógenas o de entrada.

Son independientes y definen al sistema, son predeterminadas y proporcionadas independientemente del sistema que se modela. "Puede considerarse que estas variables actúan sobre el sistema, pero no reciben acción alguna por parte de éste".

Las variables exógenas se pueden clasificar en:

- Controlables

Son aquellas variables o parámetros susceptibles de controlar por quienes toman decisiones o crean políticas para el sistema. Por ejemplo, una empresa es capaz de controlar la cantidad de materias primas que compra, horario, etc.

- Incontrolables

Es el medio ambiente en el cual el sistema modelado existe y no el sistema en sí o los encargados de tomar decisiones quien genera las variables no controlables.

b) De Estado.

Los modelos de simulación son dinámicos y describen el comportamiento del sistema sobre el tiempo. Como consecuencia, el estado del sistema en un período determinado de tiempo puede ser requerido como dato de entrada durante el siguiente período de tiempo. Las variables de estado llevan la información sobre el estado del sistema de un período de tiempo a otro.

Estas variables interaccionan con las variables exógenas y endógenas de acuerdo a las relaciones funcionales supuestas para el sistema. En un período particular de tiempo, el valor de una variable de estado puede depender no solamente de los valores de una o más variables exógenas en algún período precedente, sino también del valor de ciertas variables endógenas en períodos anteriores. Cuando un componente toma su entrada de una porción de su propia salida en un período anterior, se dice que ocurre retroalimentación.

Como ejemplos de variables de estado en una empresa se tiene el efectivo, inventario, pasivo, etc en un período de tiempo particular.

Los sistemas pueden ser clasificados en discretos y continuos. Un sistema *discreto* es aquel en el cual las variables de estado cambian instantáneamente en puntos separados de tiempo. Como ejemplo, se tiene un banco en donde la variable de estado (número de clientes) solamente varía cuando llega un cliente o cuando un cliente finaliza sus operaciones y parte. Un sistema *continuo* es aquel en el que la variable de estado cambia continuamente en el tiempo. Un avión volando puede ser un ejemplo de un sistema continuo, ya que las variables de estado: posición y velocidad pueden cambiar continuamente en el tiempo. En la práctica, muy pocos sistemas son totalmente discretos o continuos; sin embargo, como algún tipo de cambio debe predominar dentro del sistema, éste será clasificado como tal.

c) Endógenas o de Salida.

Las variables endógenas dependen del sistema y son generadas por la interacción de las variables exógenas con las de estado, de acuerdo con las características de operación del sistema. Por ejemplo, en la industria se pueden tener como variables endógenas las ventas, la mano de obra total, precios y producción total.

Es necesario observar que de acuerdo al propósito de la investigación una variable puede ser clasificada como exógena, de estado o endógena.

En los experimentos de simulación las variables exógenas pueden emplearse de dos maneras:

a) Parámetros dados o constantes

Son determinados por el medio ambiente o por aquellas personas que toman decisiones.

Son estimados con anterioridad y almacenados en la computadora para ser utilizados como datos de entrada.

b) Variables aleatorias.

Para describir adecuadamente el comportamiento de un sistema real, es necesario incluir la naturaleza aleatoria o probabilística del sistema por medio de estas variables. Se requiere calcular la distribución de probabilidad de las variables (ver capítulo II) y generarlas internamente en la computadora.

Cuando se dirigen experimentos de simulación sobre un sistema dado, el interés se centra en los efectos que tienen los diferentes niveles de diversos factores sobre las variables endógenas del sistema.

Es decir, un experimento de simulación consiste en una serie de corridas en computadora en las que se prueba empíricamente (con el empleo de datos de simulación), los efectos de niveles alternativos de los factores sobre los valores de las variables endógenas.

1.4. CLASIFICACION DE LOS MODELOS PARA SIMULACION

Resulta conveniente clasificar los modelos de simulación dentro de las siguientes dimensiones:

1) Modelos *Estáticos* vs. *Dinámicos*.

Un modelo estático es aquel modelo que no toma en cuenta, explícitamente, a la variable tiempo. Por otro lado, un modelo dinámico evoluciona a través del tiempo.

2) Modelos *Determinísticos* vs. *Estocásticos*.

Un modelo determinístico es aquel que no contiene elementos aleatorios. En este modelo el resultado es "determinado" una vez que han sido especificados los datos de entrada y las relaciones del modelo, aún cuando se requiere de una gran cantidad de tiempo de computadora para su evaluación. Un modelo estocástico tiene componentes aleatorios por lo que los resultados obtenidos serán también aleatorios y deben ser considerados como estimaciones de las características verdaderas del modelo.

3) Modelos *Continuos* vs. *Discretos*.

Los modelos discretos y continuos se definen de manera similar a los sistemas discretos y continuos descritos en la sección anterior. Es necesario mencionar que no siempre un modelo discreto es utilizado para modelar un sistema discreto y viceversa; un modelo será discreto o continuo dependiendo de los objetivos del estudio.

1.5. SIMULACION EN MANUFACTURA

Actualmente, el incremento de la competencia entre las empresas ha ocasionado un mayor énfasis en la automatización de las plantas, para así tener un aumento en la productividad y calidad, así como, reducción en los costos. Debido a que la mayor parte de los sistemas automatizados son complejos, solamente pueden ser analizados con cierto detalle por medio de la simulación. El uso de la simulación se ha incrementado notablemente en los últimos tiempos gracias a los siguientes factores:

los costos de las computadoras se han reducido, existe software enfocado a la simulación en manufactura que reduce el tiempo de programación y los paquetes animados permiten a los ingenieros de manufactura tener una mejor comprensión del funcionamiento de la simulación.

Probablemente el mayor beneficio del uso de la simulación en manufactura consiste en permitir al gerente de planta tener una visión global del sistema y apreciar los efectos producidos por cambios locales dentro del sistema. La simulación facilita el diseño de nuevas instalaciones, la investigación de aspectos operacionales en plantas ya existentes y la prueba de estrategias de control en fábricas automatizadas. En adición a estas ventajas, la simulación proporciona los siguientes beneficios específicos:

- Incremento en el volumen de producción
- Reducción del inventario de partes en proceso
- Incremento en el tiempo de uso de máquinas y empleados
- Mayor número de entregas a tiempo
- Reducción de requerimientos de capital (terreno, edificios, máquinas, etc) y de costos de operación
- Garantía de que el diseño de un sistema propuesto funcionará tal y como se espera
- La información obtenida durante la creación del modelo de simulación proporcionará un mejor entendimiento del sistema real que a su vez proporcionará otros beneficios adicionales

Por lo general la simulación en manufactura se ha enfocado exitosamente a:

Requerimientos de equipo y personal

- Número y tipo de máquinas que se requieren para cumplir con un determinado objetivo (por ejemplo: la producción de 1000 partes a la semana)
- Número, tipo y localización física del equipo de transporte y soporte
- Localización y tamaño de los inventarios en proceso
- Evaluación de cambios en el volumen o mezcla de productos (por ejemplo: el impacto de nuevos productos)
- Evaluación del efecto de una nueva pieza de equipo en una línea de manufactura existente
- Evaluación de la inversión del capital
- Planeación de los requerimientos laborales

Evaluación del funcionamiento del sistema

- Análisis del volumen de producción
- Análisis del tiempo en el sistema de una pieza particular
- Detección y análisis de cuellos de botella

Evaluación de procedimientos operacionales

- Calendarización de la producción (por ejemplo: evaluación de políticas que propongan el cargar piezas en determinada estación, producir las partes de manera secuencial a través de las estaciones de trabajo del sistema, etc)
- Políticas de niveles de inventario de materia prima y en proceso
- Estrategias de control (por ejemplo: para un transportador aéreo)
- Análisis de confiabilidad
- Políticas de control de calidad

Existen diversas medidas del funcionamiento del sistema obtenidas por medio de un estudio de simulación:

- Número de procesos producidos por unidad de tiempo
- Tiempo de los procesos en el sistema
- Tiempo de los procesos en cola
- Tiempo que esperan las partes un transporte
- Tiempo que tardan las partes en ser procesadas
- Retrasos en entregas (por ejemplo: proporción de órdenes retrasadas)
- Tamaño de los inventarios en proceso
- Uso del equipo y personal (por ejemplo: proporción de tiempo de ocupación)
- Proporción de tiempo en que una máquina permanece descompuesta, vacía (esperando partes de la estación anterior), bloqueada (esperando que una parte terminada sea removida) o en mantenimiento preventivo.
- Porcentaje de procesos defectuosos o que requieren ser reprocesados

CAPITULO II

PROBABILIDAD Y ESTADISTICA

II. PROBABILIDAD Y ESTADISTICA

II.1. INTRODUCCION

Como se mencionó en el capítulo anterior, los sistemas por simular generalmente están compuestos por uno o más elementos asociados con la incertidumbre por lo que evolucionan a través del tiempo de una manera que no es totalmente predecible. Estos sistemas son denominados sistemas estocásticos y para poder ser simulados se requiere que la variabilidad de sus elementos sea caracterizada por medio de conceptos probabilísticos. Por otro lado, los resultados de un modelo de simulación también son probabilísticos por lo que requieren de interpretación estadística.

El uso de la probabilidad y estadística es una parte importante en un estudio de simulación. En particular, son necesarias para entender la manera de modelar un sistema probabilístico, validar, escoger las distribuciones de probabilidad de los datos de entrada, generar muestras aleatorias de estas distribuciones, realizar análisis estadísticos de los resultados y diseñar experimentos.

En este capítulo se establece la notación estadística usada durante la tesis y se repasa la probabilidad y estadística básicas y necesarias en un estudio de simulación. Asimismo, se mencionan los peligros de aplicar estadística clásica basada en observaciones independientes a resultados de simulación (que en muy pocas ocasiones son independientes entre sí).

II.2. VARIABLES ALEATORIAS Y SUS PROPIEDADES

Un *experimento* es un proceso en el cual los resultados no se conocen con certeza. El conjunto de todos los resultados posibles es denominado *espacio muestral* y se denota por S ; los resultados son llamados *puntos muestrales* y se denotan por s . El espacio muestral puede ser finito, infinito numerable o infinito no numerable.

Una variable aleatoria es una función X que asigna a cada uno de los elementos $s \in S$ un número real $X(s)$.

En general, se denota a las variables aleatorias por letras mayúsculas tales como X, Y, Z y a los valores tomados por las variables aleatorias por letras minúsculas: x, y, z.

La función de distribución F(X) de una variable aleatoria X se define de la siguiente manera:

$$F(X) = P(X \leq x) \quad \text{para } -\infty < x < \infty$$

en donde $P(X \leq x)$ es la probabilidad de que ocurra el evento $\{X \leq x\}$.

La función de distribución F(X) tiene las siguientes propiedades:

1. $0 \leq F(X) \leq 1 \quad \forall x$
2. F(X) es no decreciente y continua por la derecha
3. $\lim_{x \rightarrow -\infty} F(X) = 0$ y $\lim_{x \rightarrow \infty} F(X) = 1$

II.2.1. VARIABLE ALEATORIA DISCRETA

Sea X una variable aleatoria. Si el número de valores posibles de X es finito o infinito numerable, se dice que X es una variable aleatoria *discreta*. Es decir, es posible anotar los valores de X en forma de lista: $x_1, x_2, \dots, x_n, \dots$. En el caso finito la lista termina y en el caso infinito numerable la lista continúa indefinidamente.

La probabilidad de que la variable aleatoria X tome el valor x_i está determinada por:

$$p(x_i) = P(X = x_i) \quad \text{para } i = 1, 2, \dots$$

y se debe cumplir que:

$$i) p(x_i) \geq 0 \quad \forall i$$

$$ii) \sum_{i=1}^{\infty} p(x_i) = 1$$

Todas las propiedades de la variable aleatoria discreta X pueden ser calculadas a partir de $p(x)$ que se denomina *probabilidad o función de masa de probabilidad*. Si $I=[a, b]$ es un intervalo tal que a y b son números reales y $a \leq b$, entonces:

$$P(X \in I) = \sum_{a \leq x_i \leq b} p(x_i)$$

donde la función de distribución $F(x)$ para la variable aleatoria discreta X está dada por:

$$F(x) = \sum_{x_i \leq x} p(x_i) \quad \forall -\infty < x < \infty$$

II.2.2. VARIABLE ALEATORIA CONTINUA

Considérense ahora variables aleatorias que pueden tomar un número infinito de valores.

Si ahora X puede tomar todos los valores posibles (digamos por ejemplo entre 0 y 1) Qué sucede con las probabilidades puntuales $p(x_i)$? Puesto que los valores posibles de X no son contables, no se puede hablar del i -ésimo valor de X y por lo tanto $p(x_i)$ pierde significado. Lo que se hará es sustituir la función p definida solo para x_1, x_2, \dots por una función f definida para todos los valores de x .

Entonces, se sustituye:

$$p(x_i) \geq 0 \quad \forall x \quad \text{por } f(x) \geq 0$$

$$\sum_{i=1}^{\infty} p(x_i) = 1 \quad \text{por } \int_0^1 f(x) dx = 1 \text{ (en este caso particular)}$$

Es decir, la probabilidad y la función de densidad son equivalentes y se aplican para casos discretos o continuos respectivamente.

Se dice que una variable aleatoria X es *continua*, si existe una función no negativa $f(x)$ que cumple con lo siguiente:

$$i) f(x) \geq 0 \quad \forall x$$

$$ii) \int_{-\infty}^{\infty} f(x) dx = 1$$

$$iii) P(X \in B) = \int_B f(x) dx \quad \text{siendo } B \text{ un conjunto continuo de números reales}$$

Todas las propiedades de la variable aleatoria continua X pueden ser calculadas a partir de $f(x)$ denominada *función de densidad de probabilidad*.

La función de distribución para una variable aleatoria continua está dada por:

$$F(x) = P(X \in [-\infty, x]) = \int_{-\infty}^x f(y) dy \quad \forall -\infty < x < \infty$$

más aún, si $I = [a, b]$ y a, b son números reales tal que $a < b$, entonces:

$$P(X \in I) = \int_a^b f(y) dy = F(b) - F(a)$$

derivándose esta fórmula del *teorema fundamental de cálculo* donde $F'(x) = f(x)$

II.2.3. VARIABLES ALEATORIAS N-DIMENSIONALES

Hasta el momento, sólo se ha considerado una sola variable aleatoria; sin embargo, generalmente en un estudio de simulación se tienen simultáneamente n variables aleatorias X_1, X_2, \dots, X_n .

Si X_1, X_2, \dots, X_n son n variables aleatorias, entonces se denomina a (X_1, X_2, \dots, X_n) *variable aleatoria n-dimensional*.

Por simplificación se asumirá que $n=2$.

(X, Y) es una *variable aleatoria bidimensional discreta* si los valores posibles de (X, Y) son finitos o infinitos numerables. (X, Y) es una *variable aleatoria bidimensional continua* si (X, Y) puede tomar todos los valores en un conjunto no numerable del plano euclidiano.

Sea (X, Y) una variable bidimensional discreta. Sea $p(x_i, y_j) = P(X=x_i, Y=y_j)$ la función de probabilidad definida sobre todo (x_i, y_j) que cumple:

$$i) p(x_i, y_j) \geq 0 \quad \forall (x, y)$$

$$ii) \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i, y_j) = 1$$

Además,

$$p(x_i) = P(X=x_i) = \sum_{j=1}^{\infty} p(x_i, y_j)$$

representa la *distribución marginal de probabilidades* de X. Análogamente,

$$q(y_j) = P(Y=y_j) = \sum_{i=1}^{\infty} p(x_i, y_j)$$

representa la *distribución marginal de probabilidades* de Y.

Sea (X, Y) una variable aleatoria bidimensional continua que toma todos los valores en una región R del plano euclidiano. La *función de densidad de probabilidades conjuntas* f es una función que satisface las siguientes condiciones:

$$i) f(x, y) \geq 0 \quad \forall (x, y) \in R$$

$$ii) \int \int_R f(x, y) \, dx dy = 1$$

Se definen g y h como las *funciones de densidad de probabilidades marginales* de X y Y respectivamente, como sigue:

$$g(x) = \int_{-\infty}^{+\infty} f(x, y) dy \qquad h(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Sea (X, Y) una variable aleatoria bidimensional. La *función de distribución acumulativa* F está definida por:

$$F(x, y) = P(X \leq x, Y \leq y)$$

II.2.4. VARIABLES ALEATORIAS INDEPENDIENTES

Intuitivamente, X y Y son *variables aleatorias independientes* si el resultado de X no influye de ninguna manera en el resultado de Y .

Sea (X, Y) una variable aleatoria bidimensional discreta. Se dice que:

X y Y son independientes si y solo si $p(x_i, y_j) = p(x_i)p(y_j) \quad \forall i, j$

Sea (X, Y) una variable aleatoria bidimensional continua. Se dice que:

X y Y son independientes si y solo si $f(x, y) = g(x)h(y) \quad \forall x, y$

en donde f es la función de densidad de probabilidad conjunta y g y h son las funciones de densidad de probabilidad marginales de X y Y .

Es posible extender los conceptos presentados anteriormente para el caso bidimensional al caso n -dimensional.

II.2.5. PARAMETROS

La *media o valor esperado* de la variable aleatoria X_i ($i=1, 2, \dots, n$) es denotada por μ_i o $E(X_i)$ y se define por:

$$\mu = \begin{cases} \sum_{i=1}^{\infty} x_i p(x_i) & \text{Si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{Si } X \text{ es continua} \end{cases}$$

La media o valor esperado, como una medida de tendencia central, indica donde está situado el centro de la masa de la distribución de probabilidades de una variable aleatoria.

Sea c_i una constante. Entonces, se tienen las siguientes propiedades importantes para las medias:

1. $E(cX) = cE(X)$

2. $E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i E(X_i)$

La mediana $x_{0.5}$ de una variable aleatoria X_i se define como el menor valor de x tal que $F_X(x) \geq 0.5$. Si X_i es una variable aleatoria continua, entonces $F(x_{0.5}) = 0.5$. Cuando X_i toma valores extremos, la mediana puede resultar una mejor medida de la tendencia central que la media ya que los valores extremos pueden afectar considerablemente a la media (ver gráfica II.1).

La mediana tiene la siguiente propiedad: la suma de las desviaciones absolutas de las observaciones con respecto a la mediana es menor que la suma de las desviaciones absolutas con respecto a cualquier otro punto de la distribución.

$$\sum_i |x_i - x_{0.5}| = \text{Un M\u00ednimo}$$

La varianza de una variable aleatoria X_i es denotada por σ_i^2 o $\text{Var}(X_i)$ y se define por:

$$\sigma_i^2 = E[(X_i - \mu_i)^2] = E(X_i^2) - \mu_i^2$$

La varianza es la medida de dispersi\u00f3n de la variable aleatoria sobre la media (ver gr\u00e1fica II.2). Mientras m\u00e1s grande sea la varianza, m\u00e1s se alejaran de la media los valores.

La varianza tiene las siguientes propiedades:

1. $\text{Var}(X) \geq 0$

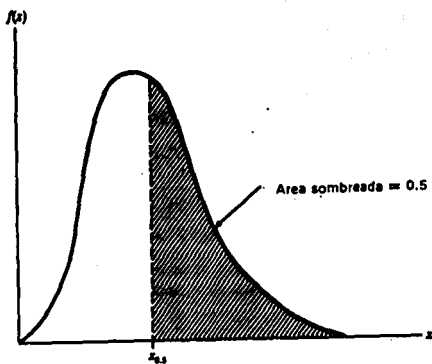
2. $\text{Var}(X+c) = \text{Var}(X)$

3. $\text{Var}(cX) = c^2 \text{Var}(X)$

4. $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$ si las X_i son independientes

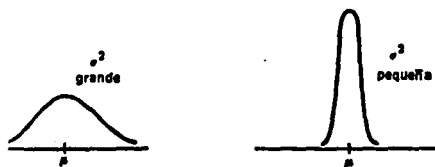
En las aplicaciones, resulta m\u00e1s conveniente tener una medida de variabilidad representada en las unidades originales, en vez de utilizar su cuadrado. Esto conduce a:

Grafica II.1.



La Mediana $x_{0.5}$ de una variable aleatoria continua

Grafica II.2.



Funciones de Densidad para variables aleatorias continuas

La *desviación estándar* de la variable aleatoria X_i que se define como:

$$\sigma_i = \sqrt{\sigma_i^2}.$$

Ahora, se considerarán las medidas de dependencia entre dos variables aleatorias.

La *covarianza* de X_i y X_j (donde $i=1,2,\dots,n$; $j=1,2,\dots,n$) que mide la dependencia (lineal) entre ambas variables y se denota por C_{ij} o $\text{Cov}(X_i, X_j)$, se define por:

$$C_{ij} = E\{(X_i - \mu_i)(X_j - \mu_j)\} = E(X_i X_j) - \mu_i \mu_j$$

(Nótese que las covarianzas son simétricas, i.e., $C_{ij} = C_{ji}$; y si $i=j$, entonces: $C_{ij} = C_{ii} = \sigma_i^2$).

Si $C_{ij} = 0$, entonces las variables aleatorias X_i y X_j no están correlacionadas. Es fácil mostrar que si X_i y X_j son independientes, entonces $C_{ij} = 0$. Sin embargo, lo contrario no siempre se cumple.

Si $C_{ij} > 0$, entonces X_i y X_j están *correlacionadas positivamente*. En este caso, $X_i > \mu_i$ y $X_j > \mu_j$ tienden a ocurrir simultáneamente, así como $X_i < \mu_i$ y $X_j < \mu_j$. Esto significa que si una de las variables es grande la otra también lo será, y viceversa. Por otro lado, si $C_{ij} < 0$ entonces X_i y X_j están *correlacionadas negativamente*. En este caso si una variable es grande, la otra será pequeña.

Si X_1, X_2, \dots, X_n son los resultados de una simulación, no solamente se requiere estimar la media μ_i y la varianza σ_i^2 para $i=1,2,\dots,n$ sino también la medida de la dependencia entre X_i y X_j para $i \neq j$. Sin embargo, no es recomendable utilizar C_{ij} ya que no es dimensionable, sino que es conveniente utilizar el *coeficiente de correlación* ρ_{ij} , definido por:

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad i=1,2,\dots,n \text{ y } j=1,2,\dots,n$$

como una medida de la dependencia lineal entre X_i y X_j . Debido a que el denominador en la ecuación anterior es positivo, es claro que ρ_{ij} tiene el mismo signo que C_{ij} . Más aún, es posible probar que: $-1 \leq \rho_{ij} \leq 1$ Vi, j (página 149 de "Probabilidad y Aplicaciones Estadísticas" de Paul L. Meyer). Si ρ_{ij} es cercano a +1, entonces X_i y X_j están altamente correlacionadas lineal y positivamente; de lo contrario, si ρ_{ij} se acerca a -1, entonces X_i y X_j se encuentran altamente correlacionadas lineal y

negativamente. Es necesario aclarar que existe una idea errónea sobre la interpretación del coeficiente de correlación: un valor de ρ próximo a cero, sólo indica la ausencia de una relación *lineal* entre X_i y X_j . No impide la posibilidad de una relación *no lineal*.

Ahora, se presenta una Definición muy importante: se dice que las variables X_1, X_2, \dots, X_n son i.i.d. si son independientes y tienen la misma Función de distribución de probabilidad (independientes e idénticamente distribuidas).

II.3. RESULTADOS DE LA SIMULACION Y PROCESOS ESTOCASTICOS

Debido a que la mayor parte de los modelos de simulación utilizan variables aleatorias de entrada, los resultados también son aleatorios y por lo tanto las conclusiones deben ser obtenidas cuidadosamente. Esta sección proporciona las bases para un análisis de resultados adecuado que se tratará a detalle en el siguiente capítulo.

Un *proceso estocástico* es una colección de variables aleatorias ordenadas en el tiempo y que tienen en común el espacio muestral. El conjunto de todos los valores que estas variables pueden tomar se denomina *espacio estado*. Si el conjunto de variables es de la forma X_1, X_2, \dots , entonces se tiene un *proceso estocástico discreto en el tiempo*; en cambio, si el conjunto de variables es tal que $\{X(t), \forall t \geq 0\}$, entonces se tiene un *proceso estocástico continuo en el tiempo*.

Para poder realizar inferencias del proceso estocástico obtenido de los resultados de una simulación, es necesario hacer ciertas suposiciones sobre el proceso estocástico que en la realidad pueden no ser completamente ciertas; sin embargo, sin estas suposiciones es imposible llevar a cabo un análisis estadístico de los resultados. Un ejemplo de esto es asumir que un proceso estocástico discreto en el tiempo X_1, X_2, \dots tiene una covarianza estacionaria, por lo que se caracteriza por:

$$\mu_i = \mu \quad \text{para } i=1,2,\dots \text{ y } -\infty < \mu < \infty$$

$$\sigma_i^2 = \sigma^2 \quad \text{para } i=1,2,\dots \text{ y } \sigma^2 < \infty$$

y $C_{i,i+j} = \text{Cov}(X_i, X_{i+j})$ es independiente de i para $j=1,2,\dots$

Entonces, en un proceso de covarianza estacionaria la media y la varianza son estacionarias a lo largo del tiempo y la covarianza entre dos observaciones X_i y X_{i+j} depende solamente de la separación j y no de los valores actuales i y $i+j$. (También es posible definir de manera análoga una covarianza estacionaria para un proceso estocástico continuo en el tiempo).

En un proceso de covarianza estacionaria, se denota la covarianza y la correlación entre X_i y X_{i+j} por C_j y ρ_j respectivamente, y se tiene que

$$\rho_j = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2} \quad \text{para } j=1,2,\dots$$

Si X_1, X_2, \dots es un proceso estocástico que empieza en el tiempo cero de la simulación no existirá covarianza estacionaria. Sin embargo, para algunas simulaciones X_{k+1}, X_{k+2}, \dots habrá una covarianza estacionaria aproximada si k es lo suficientemente grande (en donde k es la longitud del *período de calentamiento*, período en el que el sistema logra un estado estable). Este resultado es de suma importancia para un estudio de simulación y será analizado a fondo en el siguiente capítulo.

II.4. ESTIMACION DE MEDIAS, VARIANZAS Y CORRELACIONES

Sea X una variable aleatoria con una distribución de probabilidad que depende de un parámetro desconocido θ .

Sea X_1, X_2, \dots, X_n una muestra de X , si $g(X_1, X_2, \dots, X_n) = \hat{\theta}$ es una función de la muestra que va a ser utilizada para estimar a θ nos referimos a g (o $\hat{\theta}$) como *estimador* de θ .

Si $E(\hat{\theta}) = \theta$, entonces $\hat{\theta}$ será un *estimador insesgado* de θ . Se puede decir que *insesgado* significa que el valor promedio de la estimación estará próximo al valor verdadero del parámetro.

Supóngase que X_1, X_2, \dots, X_n son variables aleatorias i.i.d. (independientes e idénticamente distribuidas) con una media poblacional finita μ y una varianza poblacional finita σ^2 .

Entonces, la *media muestral*

$$\bar{X}(n) = \frac{\sum_{i=1}^n X_i}{n}$$

es un estimador insesgado de μ , es decir, $E[\bar{X}(n)] = \mu$. Similarmente, la varianza muestral:

$$S^2(n) = \frac{\sum_{i=1}^n [X_i - \bar{X}(n)]^2}{n-1}$$

será un estimador insesgado para σ^2 ya que $E[S^2(n)] = \sigma^2$.

(Nota: los estimadores $\bar{X}(n)$ y $S^2(n)$ se les suele denotar también por $\bar{\mu}$ y $\hat{\sigma}^2$ respectivamente.)

Resulta de poca utilidad trabajar solamente con $\bar{X}(n)$ como un estimador de μ ya que no existe una manera de determinar qué tan cercano se encuentra $\bar{X}(n)$ de μ . Debido a que $\bar{X}(n)$ es una variable aleatoria con varianza $\text{Var}[\bar{X}(n)]$, en algún experimento $\bar{X}(n)$ puede quedar cerca de μ , y en otro diferir considerablemente del mismo. Entonces, para proporcionar precisión a $\bar{X}(n)$ como un estimador puntual de μ se construye un intervalo de confianza para μ .

El primer paso para construir un intervalo de confianza consiste en estimar la varianza $\text{Var}[\bar{X}(n)]$:

$$\begin{aligned} \text{Var}[\bar{X}(n)] &= \text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \\ &= \frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n X_i \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \text{ debido a que los } X_i \text{'s son i.i.d.} \\ &= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

En general, es claro que mientras mayor sea n , más cercano estará $\bar{X}(n)$ de μ . Además, es posible obtener un estimador insesgado de $\text{Var}[\bar{X}(n)]$ al reemplazar σ^2 por $S^2(n)$:

$$\text{Var}[\bar{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^n [X_i - \bar{X}(n)]^2}{n(n-1)}$$

Sin embargo, la experiencia en simulación ha demostrado que los resultados obtenidos siempre están correlacionados, por lo que la discusión anterior sobre observaciones i.i.d. no es *directamente* aplicable en el análisis de resultados de simulación.

Para poder entender los peligros de utilizar las observaciones de una simulación como si fueran independientes, se utilizará el modelo de covarianza estacionaria de la sección anterior. En particular, supóngase que las variables aleatorias X_1, X_2, \dots, X_n provienen de un proceso estocástico de covarianza estacionaria. Se cumple que la media muestral $\bar{X}(n)$ es un estimador insesgado para μ ; sin embargo, la varianza muestral $S^2(n)$ ya no es un estimador insesgado de la varianza poblacional. De hecho, es posible mostrar lo siguiente:

$$E[S^2(n)] = \sigma^2 \left[1 - \frac{2 \sum_{j=1}^{n-1} (1-j/n) \rho_j}{n-1} \right]$$

Entonces, si $\rho_j > 0$ (correlacionado positivamente), $S^2(n)$ tendrá un sesgo negativo: $E[S^2(n)] < \sigma^2$. Es muy importante recalcar esto, ya que la mayoría de los lenguajes de simulación utilizan $S^2(n)$ para estimar la varianza y esto puede tener serias implicaciones en el análisis.

Considérese ahora el problema de estimar la varianza de la media muestral $\text{Var}[\bar{X}(n)]$ para un proceso de covarianza estacionaria. Es posible mostrar que:

$$\text{Var}[\bar{X}(n)] = \sigma^2 \frac{[1 + 2 \sum_{j=1}^{n-1} (1-j/n) \rho_j]}{n}$$

Entonces, si se estima $\text{Var}[\bar{X}(n)]$ a partir de $S^2(n)/n$ (que es correcto cuando las variables son i.i.d.), se tienen dos tipos de error:

- a) La desviación de $S^2(n)$ como un estimador de σ^2 .
- b) La negligencia de los términos de correlación de la ecuación anterior.

En algunas ocasiones resulta interesante estimar las ρ_j 's (o C_j 's) de los datos X_1, X_2, \dots, X_n . En este caso, ρ_j (para $j=1, 2, \dots, n-1$) puede ser estimado de la siguiente manera:

$$\hat{\rho}_j = \frac{\hat{C}_j}{S^2(n)}, \quad \hat{C}_j = \frac{\sum_{i=1}^{n-j} [X_i - \bar{X}(n)][X_{i+j} - \bar{X}(n)]}{n-j}$$

El inconveniente del estimador $\hat{\rho}_j$ es que es sesgado y tiene una varianza elevada para n pequeño; además, está correlacionado con otros estimadores de correlación, i.e., $\text{Cov}(\hat{\rho}_j, \hat{\rho}_k) \neq 0$. (En particular $\hat{\rho}_{n-1}$ será un estimador muy pobre para ρ_{n-1} ya que estará basado en el único producto: $[X_1 - \bar{X}(n)][X_n - \bar{X}(n)]$). En general, es difícil obtener "buenos" estimadores para las ρ_j 's al menos que n sea grande y j relativamente pequeño en comparación a n .

Es necesario notar que si las X_i 's son independientes no necesariamente los estimadores de correlación serán iguales a cero, ya que el estimador ρ_j es también una variable aleatoria.

Se ha observado que los resultados de una simulación están correlacionados por lo que no es posible aplicar las fórmulas clásicas de estadística basadas en observaciones i.i.d. Sin embargo, en el siguiente capítulo se estudiará la manera de agrupar los resultados en nuevas "observaciones" de tal manera que las fórmulas basadas en observaciones i.i.d. puedan ser aplicadas. Entonces, las fórmulas basadas en observaciones i.i.d. estudiadas en esta sección y en las dos siguientes, pueden ser aplicables indirectamente en el análisis de resultados de la simulación.

II.5. INTERVALOS DE CONFIANZA Y PRUEBAS DE HIPOTESIS PARA LA MEDIA

En esta sección se estudia la construcción de un intervalo de confianza para la media μ y la prueba de hipótesis acerca de que $\mu = \mu_0$.

Sea la función de distribución de la variable aleatoria normal $\phi(z)$ con media $\mu=0$ y varianza $\sigma_z^2=1$ (también conocida como variable aleatoria normal estándar), que está definida por:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-y^2/2} dy \quad \text{para } -\infty < z < \infty$$

donde la constante $1/\sqrt{2\pi}$ es un vector escalar necesario para que el área total sea igual a 1.

Teorema del Límite Central.

Sean X_1, X_2, \dots, X_n variables aleatorias independientes con $E(\hat{X}_i) = \mu_i$ y $V(\hat{X}_i) = \sigma_i$ para $i=1, 2, \dots$. Sea

$$\bar{X}(n) = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Luego, bajo ciertas condiciones generales (que no se indicarán explícitamente en esta tesis):

$$Z_n = [\bar{X}(n) - \mu] / \sqrt{\sigma^2/n} \quad (1)$$

tiene aproximadamente la Distribución Normal $N(0,1)$.

Es decir, si $F_n(s) = P(Z_n \leq s)$ es la Función de Distribución para Z_n , se tiene que:

$$\lim_{n \rightarrow \infty} F_n(s) = \phi(s)$$

Nótese que μ y σ^2/n son respectivamente la media y la varianza de $\bar{X}(n)$.

Entonces, por el Teorema del Límite Central se tiene que si n es "suficientemente grande", la variable aleatoria Z_n estará aproximadamente distribuida como una variable aleatoria normal estándar sin importar la distribución original de las X_i 's. O lo que es lo mismo: a medida que aumente el tamaño de la muestra n , la distribución de la media $\bar{X}(n)$ calculada a partir de una muestra aleatoria extraída prácticamente de cualquier población se acerca a la distribución normal estándar.

El problema del resultado anterior consiste que en la práctica la σ^2 es generalmente desconocida. Sin embargo, para n grande, la varianza muestral $S^2(n)$ converge hacia σ^2 . Entonces, para n suficientemente grande, un intervalo de confianza aproximado para μ al $100(1-\alpha)\%$ está dado al despejar (1) por:

$$\bar{X}(n) \pm z_{1-\alpha/2} \sqrt{\frac{S^2(n)}{n}} \quad (2)$$

donde $z_{1-\alpha/2}$ (para $0 < \alpha < 1$) es el punto crítico máximo $1-\alpha/2$ para una variable aleatoria normal estándar.

Para un conjunto dado de valores X_1, X_2, \dots, X_n , el mínimo valor del intervalo de confianza $l(n, \alpha) = \bar{X}(n) - z_{1-\alpha/2} \sqrt{S^2(n)/n}$ y el máximo valor $u(n, \alpha) = \bar{X}(n) + z_{1-\alpha/2} \sqrt{S^2(n)/n}$ solamente son números y el intervalo de confianza $[l(n, \alpha), u(n, \alpha)]$ puede o no contener a la media μ . Una vez obtenido el intervalo de confianza no se tiene nada probabilístico y la interpretación correcta sobre el mismo es la siguiente: Si se construye un gran número de intervalos de confianza independientes, al $100(1-\alpha)\%$ y basados en n observaciones, si n es suficientemente grande la proporción de estos intervalos de confianza que contendrán a μ será de $1-\alpha$. Esta proporción se conoce como la cobertura del intervalo de confianza.

El problema de construir un intervalo de confianza de la manera anterior consiste en que se desconoce que significa "n suficientemente grande". Si n es muy pequeña, la verdadera cobertura para un intervalo de confianza al $100(1-\alpha)\%$ será menor o igual a $1-\alpha$, y es por esta razón por la que se habla de un intervalo de confianza aproximado.

Debido a lo anterior, se construirá un intervalo de confianza alternativo. Si las X_i 's son variables aleatorias normales, la variable aleatoria $t_n = [\bar{X}(n) - \mu] / \sqrt{S^2(n)/n}$ tiene una distribución t con $n-1$ grados de libertad. Entonces, se puede construir un intervalo de confianza exacto al $100(1-\alpha)\%$ ($n \geq 2$) para μ de la siguiente manera:

$$\bar{X}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}} \quad (3)$$

donde $t_{1-1, 1-\alpha/2}$ se obtiene en las Tablas de la Distribución t en base a n y α .

La cantidad que se suma o resta a $\bar{X}(n)$ para construir un intervalo de confianza se denomina *longitud-media* del intervalo de confianza y es una medida de la precisión con que se conoce μ .

En la práctica, la distribución de las X_i 's en muy pocas ocasiones será normal y el intervalo de confianza será aproximado en términos de la cobertura. Debido a que $t_{n-1, 1-\alpha/2} > z_{1-\alpha/2}$, el intervalo de confianza obtenido en (3) será mayor que el intervalo calculado por (2) por lo que generalmente tendrá una cobertura más cercana a $1-\alpha$. Por esta razón se recomienda el uso de (3) para la construcción de un intervalo de confianza para μ . Nótese que $t_{n-1, 1-\alpha/2} \rightarrow z_{1-\alpha/2}$ cuando $n \rightarrow \infty$; en particular, $t_{40, 0.95}$ difiere de $z_{0.95}$ en menos del 3%.

Sea ν una medida de simetría definida por:

$$\nu = \frac{E\{(\bar{X}(n) - \mu)^3\}}{(\sigma^2)^{3/2}}$$

en donde ν es igual a cero en el caso de una distribución totalmente simétrica tal como la normal. La experiencia ha demostrado que mientras mayor sea ν , menor será la cobertura; por lo tanto, mientras más grande sea ν , mayor tendrá que ser n para que se obtenga una cobertura satisfactoria.

Supóngase que X_1, X_2, \dots, X_n están normalmente distribuidas (o se aproximan) y se desea probar la hipótesis nula H_0 de que $\mu = \mu_0$, donde μ_0 es un valor fijo e hipotético para μ . Intuitivamente se espera que si $|\bar{X}(n) - \mu_0|$ es grande, H_0 no será verdadero. Sin embargo, para desarrollar una prueba estadísticamente válida, se requiere de un estadístico (función de las X_i 's) cuya distribución se conozca cuando H_0 es verdadero. De acuerdo a lo visto anteriormente, si H_0 es verdadero, el estadístico $t_n = [\bar{X}(n) - \mu_0] / \sqrt{S^2(n)/n}$ tendrá una distribución t con $n-1$ grados de libertad. Entonces, la forma para la prueba de hipótesis para $\mu = \mu_0$ será:

$$\text{Si } |t_n| \begin{cases} > t_{n-1, 1-\alpha/2} \\ \leq t_{n-1, 1-\alpha/2} \end{cases} \quad \begin{array}{l} \text{rechazar } H_0 \\ \text{aceptar } H_0 \end{array} \quad (4)$$

Esta prueba es conocida como la prueba t .

El conjunto de números reales que corresponde al rechazo de H_0 , i.e., el conjunto de x tal que $|x| > t_{n-1, 1-\alpha/2}$, se denomina la *región crítica* de la prueba; y la probabilidad de que el estadístico t_n caiga en la región crítica dado que H_0 es verdadero, se denomina *nivel* de la prueba y es claro que es igual a α .

Cuando se lleva a cabo una prueba de hipótesis, se pueden producir dos tipos de errores. Si se rechaza H_0 cuando es verdadera, se tiene un *error de tipo I*. La probabilidad de que ocurra un error tipo I es igual al nivel α y por lo tanto está bajo el control del experimentador. Si se acepta H_0 cuando es falsa, se tiene un *error de tipo II* (denotado por β). Para un nivel fijo α y una muestra de tamaño n , la probabilidad de que ocurra un error de tipo II, depende de lo que actualmente es verdadero y se desconoce. Sea $\delta = 1 - \beta$ (denominada la potencia de la prueba) la probabilidad de rechazar H_0 cuando es falsa. Al aumentar el tamaño de la muestra n disminuirá σ/\sqrt{n} y, por consiguiente, también

disminuirá la extensión de la distribución de $\bar{X}(n)$. Esto permite una reducción de n y J , o una reducción aún mayor de β si α permanece constante.

Debido a que el poder de una prueba puede ser bajo y desconocido, se dice que se falla en rechazar a H_0 (en lugar de aceptar H_0) cuando el estadístico t_n no cae en la región crítica. (Cuando H_0 no es rechazada, no es posible saber con certeza si H_0 es falsa o verdadera, debido a que la prueba no es lo suficientemente poderosa para detectar diferencias entre H_0 y la realidad)

Es muy importante mencionar que existe una relación muy cercana entre el intervalo de confianza dado por (3) y la prueba de hipótesis dada por (4). De hecho, el rechazo de la hipótesis nula H_0 de que $\mu = \mu_0$ es equivalente a decir que μ_0 no se encuentra contenida en el intervalo de confianza para μ .

II.6. DISTRIBUCIONES DE PROBABILIDAD

En las secciones anteriores se describieron las propiedades de las variables aleatorias. En esta sección se describen distribuciones específicas de variables aleatorias que son importantes para el modelado de procesos aleatorios.

Si resulta posible recolectar datos sobre las variables aleatorias de interés del modelo, esta información puede ser utilizada para la simulación de acuerdo a las siguientes técnicas:

1. Los valores son utilizados directamente en la simulación. Por ejemplo, si los datos representan tiempos de servicio, se utilizará uno de estos cada vez que se requiera un tiempo de servicio en la simulación.
2. Los datos son utilizados para definir una función de distribución empírica. Si estos datos representan tiempos de servicio, se muestrearán de esta distribución cada vez que se requiera de un tiempo de servicio en la simulación.
3. Se utilizarán técnicas estándar de inferencia estadística para "ajustar" los datos a una distribución teórica y se realizarán pruebas para determinar la bondad del ajuste. Si cierta distribución teórica con ciertos valores para sus parámetros resulta un buen modelo para los datos de tiempos de servicio, se realizará un muestreo a partir de esta distribución cada vez que se requiera de un tiempo de servicio

en la simulación.

El inconveniente de la primer técnica consiste en que la simulación solamente puede reproducir lo que ha ocurrido históricamente y puede no tenerse la información suficiente para llevar a cabo el número de corridas deseadas. Con la segunda técnica se pueden efectuar las corridas que se deseen, ya que para datos continuos se puede escoger cualquier valor entre el mínimo y el máximo. Es preferible el uso de la técnica 2 sobre la 1; sin embargo, la primera se recomienda cuando se desea validar el modelo al comparar los resultados del modelo para un sistema existente con los resultados del sistema real.

Si es posible encontrar una distribución teórica que se ajuste a los datos de manera razonable de acuerdo a ciertas pruebas (a explicarse en la sección II.7), es preferible utilizar ésta en vez de la distribución empírica debido a que:

- Una distribución empírica puede presentar ciertas irregularidades, sobre todo si la información es escasa. Además, una distribución teórica "extiende" los datos y proporciona información sobre toda la distribución.
- Al utilizarse una distribución empírica no es posible generar valores fuera del rango de los datos observados. Esto es un inconveniente, ya que muchas medidas de rendimiento de los sistemas simulados dependen en gran parte de la probabilidad de que ocurra un evento "extremo"; por ejemplo, la generación de un tiempo de servicio muy alto. En cambio, una distribución teórica puede generar valores fuera del rango de los datos observados.
- Una distribución teórica es una forma compacta de representar a un conjunto de datos. Si se tienen n datos disponibles para una distribución continua, en muchos lenguajes de simulación será necesario almacenar en el computador $2n$ valores (los datos y las probabilidades acumuladas correspondientes); por lo que si el número de datos es excesivo, el uso de una distribución empírica consumirá mucha memoria.

A continuación se presentan aquellas distribuciones teóricas que han resultado de utilidad en el modelado de sistemas, así como sus propiedades básicas.

II.6.1. Parametrización de Distribuciones Continuas

Los parámetros de las funciones de distribución pueden ser clasificados en tres tipos: localización, escala y forma.

El *parámetro de localización* γ especifica el punto de localización en el eje de las x del rango de valores de la distribución; generalmente γ es el punto medio o el mínimo del rango de la distribución. Si γ cambia, su distribución asociada solamente se mueve hacia la derecha o izquierda sin alterarse.

El *parámetro de escala* β determina la unidad de medida de los valores de la distribución. Un cambio en β comprime o expande a la distribución asociada sin alterar su forma original.

El *parámetro de forma* α determina la forma básica de la distribución. Generalmente, un cambio en α altera las propiedades de la distribución (i.e. sesgo). Existen distribuciones (normal, exponencial, etc.) que no poseen este parámetro, mientras que otras poseen varios (la distribución beta tiene dos).

II.6.2. Distribuciones Continuas.

A continuación se presentan las distribuciones continuas más usadas en los modelos de simulación, incluyéndose las posibles aplicaciones teóricas para cada distribución.

UNIFORME $U(a,b)$

Aplicaciones posibles:

Utilizada como el "primer" modelo para una cantidad que se tiene idea que varía entre a y b pero que se desconoce el resto. La distribución $U(0,1)$ es muy importante en la generación de valores aleatorios de otras distribuciones.

Función de densidad (ver figura II.3):

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases}$$

Función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b < x \end{cases}$$

Parámetros:

$a, b \in \mathbb{R}$; a es un parámetro de localización y $b-a$ es un parámetro de escala.

Rango: $[a, b]$

Media: $\frac{a+b}{2}$

Varianza: $\frac{(b-a)^2}{12}$

Moda: No existe únicamente

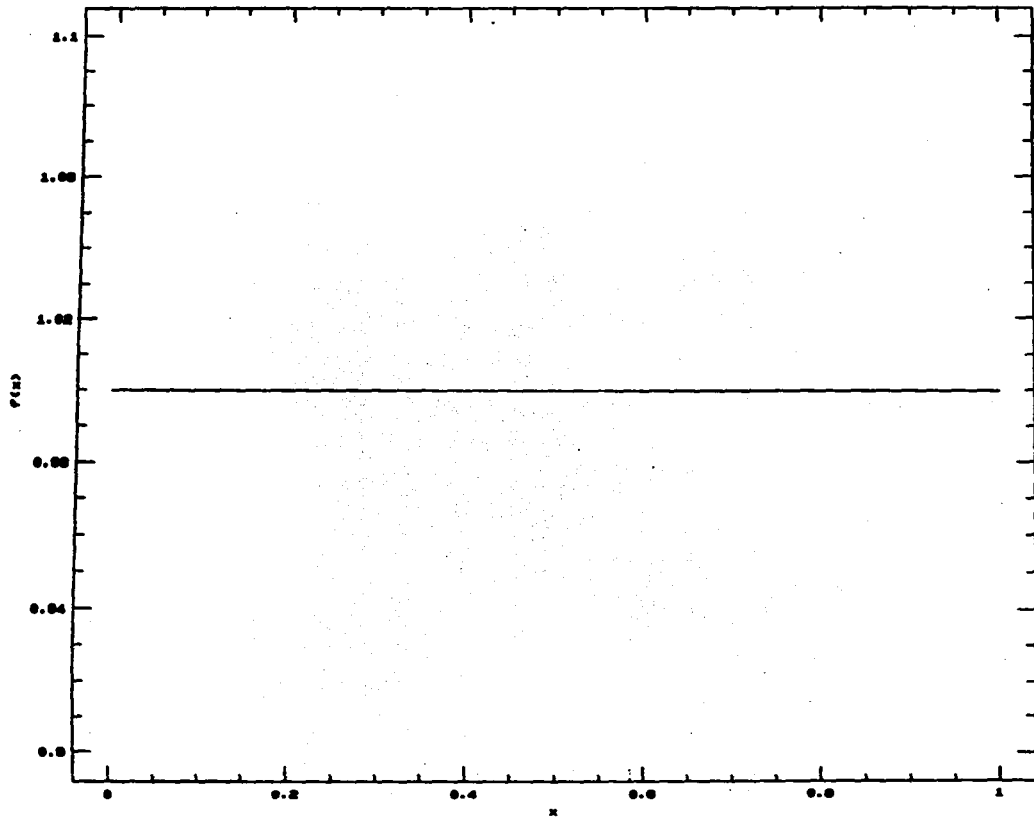
Comentarios:

- 1) La distribución $U(0, 1)$ es un caso especial de la distribución beta (cuando $\alpha_1 = \alpha_2 = 1$).
- 2) Si $X \sim U(0, 1)$ y $[x, \Delta x]$ es un subintervalo de $U(0, 1)$ con $\Delta x \geq 0$, se tiene que:

$$P(X \in [x, x+\Delta x]) = \int_x^{x+\Delta x} 1 dy = (x+\Delta x) - x = \Delta x$$

que justifica el nombre de "uniforme".

Graph 11.3
Uniform



EXPONENCIAL $\text{expo}(\beta)$

Esta distribución desempeña un papel importante en la descripción de una gran cantidad de fenómenos. Representa el intervalo de tiempo requerido para obtener un número específico de éxitos. Tiene la propiedad de "no tener memoria".

Aplicaciones posibles:

Tiempos de llegadas de los clientes.

Función de Densidad (ver gráfica II.4):

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{si } x \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{si } x \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

Parámetros:

Parámetro de escala $\beta > 0$.

Rango: $[0, \infty)$

Media: β

Varianza: β^2

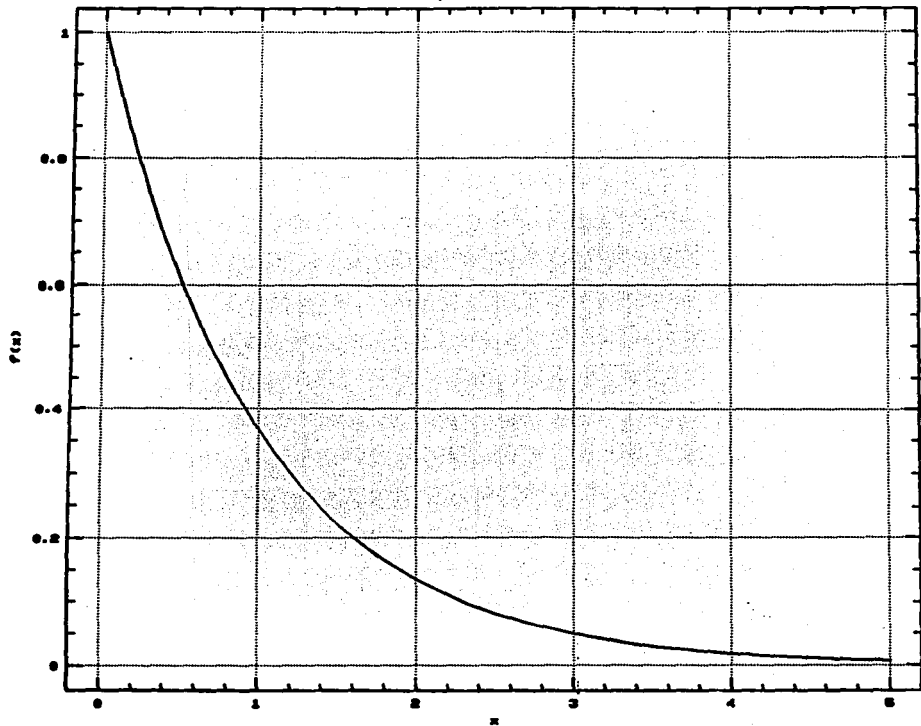
Moda: 0

Comentarios:

1) La distribución $\text{expo}(\beta)$ es un caso especial tanto de la distribución gamma como de la Weibull (con parámetro de forma $\alpha=1$ y de escala β en ambos casos).

2) Si X_1, X_2, \dots, X_m son variables aleatorias independientes con distribución $\text{expo}(\beta)$, entonces:

Grafica II.4
Esperimental (1.0)



$X_1 + X_2 + \dots + X_m \sim \text{gamma}(m, \beta)$, también llamada la distribución *m-Erlang*.

GAMMA $\text{gamma}(\alpha, \beta)$

Aplicaciones posibles:

Tiempo para realizar una tarea, por ejemplo: tiempo de servicio a un cliente, tiempo de reparación de una máquina.

Función de Densidad (ver gráfica II.5):

$$f(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

donde $\Gamma(\alpha)$ es la función *gamma* definida por: $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ \forall número real $z > 0$.

Función de Distribución:

Si α no es entero, entonces no existe una forma cerrada. Si α es entero, se tiene que:

$$F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \frac{(x/\beta)^j}{j!} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

Parámetros:

Parámetro de forma $\alpha > 0$, parámetro de escala $\beta > 0$.

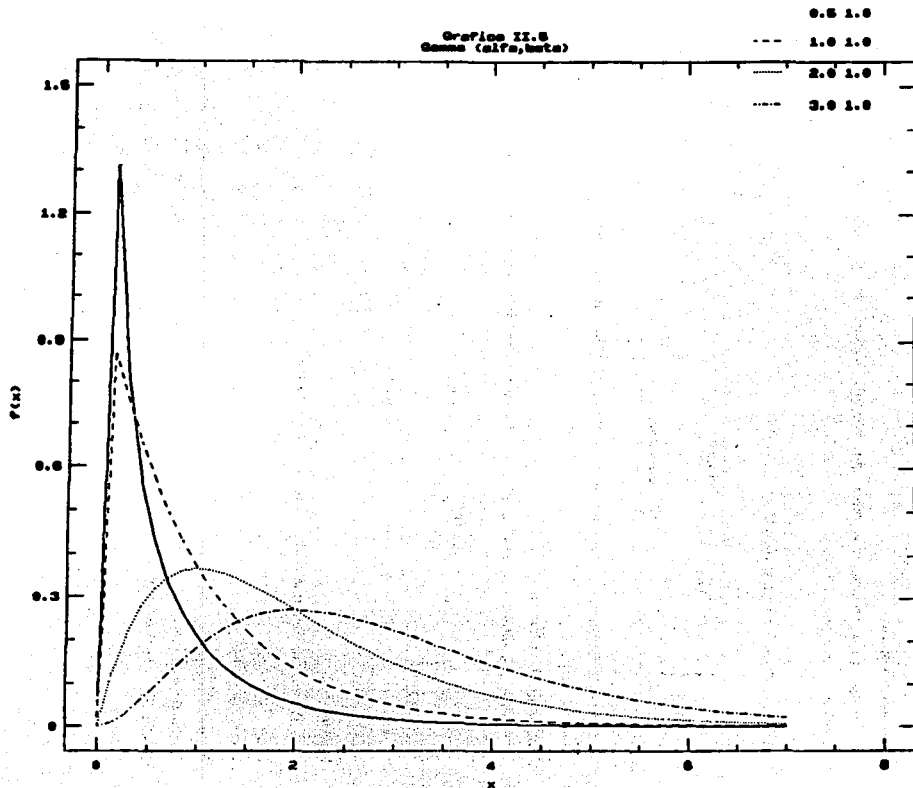
Rango: $(0, \infty)$

Media: $\alpha\beta$

Varianza: $\alpha\beta^2$

Moda: $\beta(\alpha - 1)$ si $\alpha \geq 1$, 0 si $\alpha < 1$

Grafico II.5
Gamma (alfa, beta)



Comentarios:

- 1) Las distribuciones $\text{expo}(\beta)$ y $\text{gamma}(1, \beta)$ son idénticas.
- 2) Para un número entero positivo m , la distribución $\text{gamma}(m, \beta)$ es denominada la distribución *m-Erlang*(β).
- 3) La distribución chi-cuadrada con k grados de libertad es igual a la distribución $\text{gamma}(k/2, 2)$.
- 4) Si X_1, X_2, \dots, X_m son variables aleatorias independientes con una distribución $\text{gamma}(\alpha_i, \beta)$, entonces $X_1 + X_2 + \dots + X_m \sim \text{gamma}(\alpha_1 + \alpha_2 + \dots + \alpha_m, \beta)$.
- 5) Si X_1 y X_2 son variables aleatorias independientes con $X_i \sim \text{gamma}(\alpha_i, \beta)$, entonces $X_1/(X_1 + X_2) \sim \text{beta}(\alpha_1, \alpha_2)$.
- 6) $X \sim \text{gamma}(\alpha, \beta)$ si y sólo si $Y=1/X$ tiene una distribución Pearson tipo V con parámetros de forma y de escala α y $1/\beta$, denotada por $\text{PT5}(\alpha, 1/\beta)$.
- 7)

$$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } \alpha < 1 \\ \frac{1}{\beta} & \text{si } \alpha = 1 \\ 0 & \text{si } \alpha > 1 \end{cases}$$

WEIBULL Weibull(α, β)

Aplicaciones posibles:

Tiempo para completar una tarea (las funciones de densidad toman formas similares a las funciones de densidad de la gamma), tiempo para que falle una pieza del equipo.

Función de Densidad (ver gráfica II.6):

$$f(x) = \begin{cases} \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

Parámetros:

Parámetro de forma $\alpha > 0$, parámetro de escala $\beta > 0$.

Rango: $(0, \infty)$

Media: $\frac{\beta}{\alpha} \Gamma\left(\frac{1}{\alpha}\right)$

Varianza: $\frac{\beta^2}{\alpha^2} \left\{ 2\Gamma\left(\frac{2}{\alpha}\right) - \left(\Gamma\left(\frac{1}{\alpha}\right)\right)^2 \right\}$

Moda: $\begin{cases} \beta \left(\frac{\alpha-1}{\alpha}\right)^{1/\alpha} & \text{si } \alpha \geq 1 \\ 0 & \text{si } \alpha < 1 \end{cases}$

donde Γ es la función gamma.

Comentarios:

- 1) Las distribuciones $\text{expo}(\beta)$ y $\text{Weibull}(1, \beta)$ son iguales.
- 2) $X \sim \text{Weibull}(\alpha, \beta) \Leftrightarrow X^\alpha \sim \text{expo}(\beta^\alpha)$.
- 3) Si $\alpha \rightarrow \infty$, la distribución Weibull se degenera en β . Entonces, para α grande las densidades de Weibull tienen un alto pico en la moda.

4)

$$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } \alpha < 1 \\ \frac{1}{\beta} & \text{si } \alpha = 1 \\ 0 & \text{si } \alpha > 1 \end{cases}$$

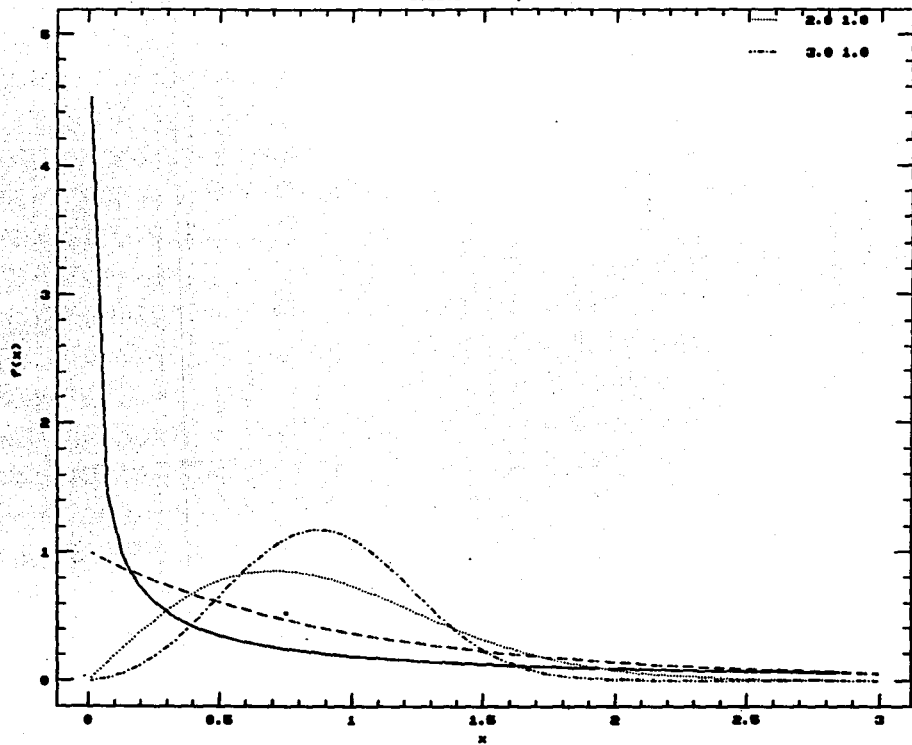
Graphics II.0
Unitball (alpha,beta)

0.5 1.0

--- 1.0 1.0

--- 2.0 1.0

--- 3.0 1.0



NORMAL $N(\mu, \sigma^2)$

La distribución Normal sirve como una aproximación excelente a una gran cantidad de distribuciones.

Aplicaciones posibles:

Errores de diferentes tipos, por ejemplo: en el punto de impacto de una bomba; cantidades que resultan de sumar grandes números u otras cantidades (debido al teorema del límite central).

Función de Densidad (ver gráfica II.7):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad \forall x \in \mathbb{R}$$

Función de Distribución:

No existe forma cerrada.

Parámetros:

Parámetro de localización $\mu \in (-\infty, \infty)$, parámetro de escala $\sigma > 0$

Rango: $(-\infty, \infty)$

Media: μ

Varianza: σ^2

Moda: μ

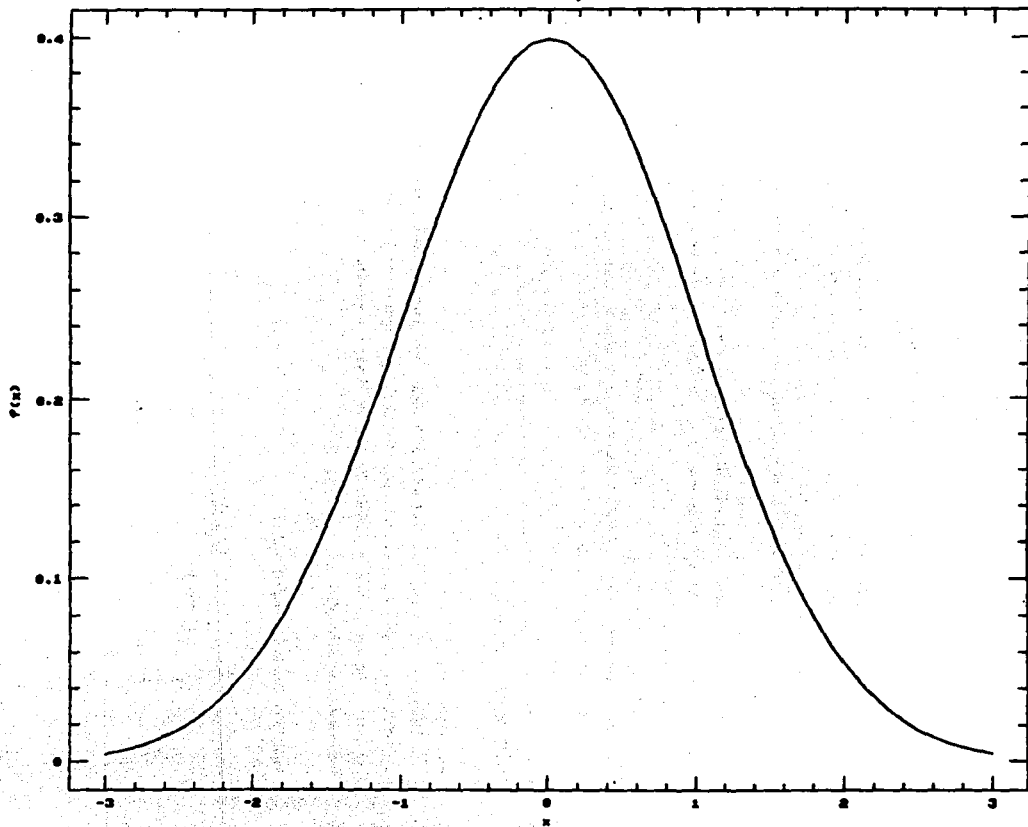
Comentarios:

1) Si dos variables aleatorias cuya distribución conjunta también es normal no están correlacionadas, entonces también son independientes.

2) Supóngase que la distribución conjunta de X_1, X_2, \dots, X_m es una normal multivariada, y sea $\mu_i = E(X_i)$ y $C_{ij} = \text{Cov}(X_i, X_j)$. Entonces, para cualquier conjunto de números reales a, b_1, b_2, \dots, b_m la variable aleatoria $a + b_1 X_1 + b_2 X_2 + \dots + b_m X_m$ tiene una distribución normal con media $\mu = a + \sum_{i=1}^m b_i \mu_i$ y varianza de

$$\sigma^2 = \sum_{i=1}^m \sum_{j=1}^m b_i b_j C_{ij}$$

Gráfico II.7
Normal (0,1)



Nótese que es necesario no asumir independencia en las X_i 's. Si las X_i 's son independientes, entonces:

$$\sigma^2 = \sum_{i=1}^m b_i^2 \text{Var}(X_i)$$

3) La distribución $N(0, 1)$ es comúnmente denominada distribución estándar o distribución normal unitaria.

4) Si X_1, X_2, \dots, X_k son variables aleatorias normales, estándares e independientes, entonces $X_1^2 + X_2^2 + \dots + X_k^2$ tiene una distribución chi-cuadrada con k grados de libertad (que también es equivalente a una distribución gamma($k/2, 2$)).

5) Si $X \sim N(\mu, \sigma^2)$, entonces e^X tiene una distribución lognormal con parámetros μ y σ , y se denota por $LN(\mu, \sigma^2)$.

6) Si $X \sim N(0, 1)$, si Y tiene una distribución chi-cuadrada con k grados de libertad y X y Y son independientes, entonces $X/\sqrt{Y/k}$ tiene una distribución t con k grados de libertad (generalmente denominada distribución t de Student).

7) Si la distribución normal es utilizada para representar una cantidad no negativa (por ejemplo: el tiempo), es necesario truncar su densidad en $x=0$.

8) A medida que $\sigma \rightarrow 0$, la distribución normal se degenera en μ .

LOGNORMAL $LN(\mu, \sigma^2)$

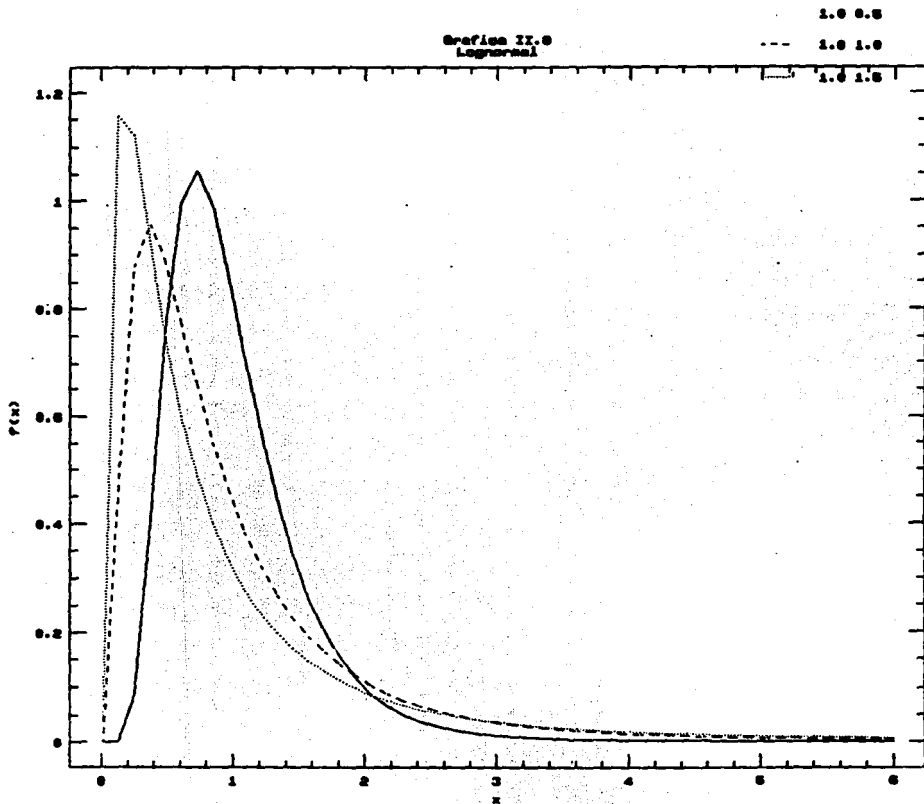
Aplicaciones posibles:

Tiempo de realizar una tarea (la función de densidad toma formas similares a la gamma(α, β) y a la Weibull(α, β) para $\alpha > 1$; sin embargo, posee un alto pico cerca de $x=0$ que generalmente es de utilidad); cantidades que son el producto de números elevados o de otras cantidades (en virtud al teorema del límite central).

Función de Densidad (ver gráfica II.8):

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \frac{-(\ln x - \mu)^2}{2\sigma^2} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

Graph II.0
Lognormal



Función de Distribución:

No existe forma cerrada.

Parámetros:

Parámetro de forma $\sigma > 0$, parámetro de escala $\mu \in (-\infty, \infty)$.

Rango: $(0, \infty)$

Media: $e^{\mu + \sigma^2/2}$

Varianza: $e^{2\mu + 2\sigma^2}(e^{\sigma^2} - 1)$

Moda: $e^{\mu - \sigma^2}$

Comentarios:

1) $X \sim LN(\mu, \sigma^2) \Leftrightarrow \ln X \sim N(\mu, \sigma^2)$. Entonces, si se tienen datos X_1, X_2, \dots, X_n y se piensa que son lognormales, los logaritmos de estos datos $\ln X_1, \ln X_2, \dots, \ln X_n$ pueden ser considerados como normalmente distribuidos para efectos de asumir una distribución hipotética, estimar parámetros y realizar pruebas de bondad y ajuste.

2) A medida que $\sigma \rightarrow 0$, la distribución lognormal se degenera en e^μ . Entonces, las funciones de densidad de la lognormal para σ pequeña tienen un pico muy pronunciado en la moda.

3) $\lim_{x \rightarrow 0} f(x) = 0$, sin importar el valor de los parámetros.

BETA $\text{beta}(\alpha_1, \alpha_2)$

Aplicaciones posibles:

Utilizada como un modelo tosco en el caso de ausencia de datos; distribución de una proporción aleatoria (Ejemplo: la proporción de piezas defectuosas en un embarque); tiempo para completar una tarea.

Función de Densidad (ver gráfica II.9):

$$f(x) = \begin{cases} \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{si } 0 < x < 1 \\ 0 & \text{otro caso} \end{cases}$$

donde $B(\alpha_1, \alpha_2)$ es la función beta definida por:

$$B(z_1, z_2) = \int_0^1 t^{z_1-1}(1-t)^{z_2-1} dt$$

✓ números reales $z_1 > 0$ y $z_2 > 0$.

Función de Distribución:

Normalmente, no existe forma cerrada. Si α_1 o α_2 es un entero positivo, es posible utilizar una expansión binomial para obtener $F(x)$, que será un polinomio en x . Por lo general, las potencias de x serán números positivos reales con un rango que variará desde 0 hasta $\alpha_1 + \alpha_2 - 1$.

Parámetros:

Parámetros de forma $\alpha_1 > 0$ y $\alpha_2 > 0$.

Rango: $[0, 1]$

Media: $\frac{\alpha_1}{\alpha_1 + \alpha_2}$

Varianza: $\frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)}$

Moda:

$$\frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2}$$

si $\alpha_1 > 1, \alpha_2 > 1$

0 y 1

si $\alpha_1 < 1, \alpha_2 < 1$

0

si $(\alpha_1 < 1, \alpha_2 \geq 1)$ o si $(\alpha_1 = 1, \alpha_2 > 1)$

1

si $(\alpha_1 \geq 1, \alpha_2 < 1)$ o si $(\alpha_1 > 1, \alpha_2 = 1)$

no existe únicamente

si $\alpha_1 = \alpha_2 = 1$

Comentarios:

1) Las distribuciones $U(0, 1)$ y $\text{beta}(1, 1)$ son iguales.

2) Si X_1 y X_2 son variables aleatorias independientes con $X_i \sim \text{gamma}(\alpha_i, \beta)$, entonces $X_1/(X_1+X_2) \sim \text{beta}(\alpha_1, \alpha_2)$.

3) Una variable aleatoria X en $[0, 1]$, por medio de la transformación $a+(b-a)X$ puede ser relocalizada y cambiado su tamaño para obtener una variable aleatoria beta en $[a, b]$ de la misma forma.

4) $X \sim \text{beta}(\alpha_1, \alpha_2) \Leftrightarrow 1-X \sim \text{beta}(\alpha_2, \alpha_1)$.

5) $X \sim \text{beta}(\alpha_1, \alpha_2) \Leftrightarrow Y=X/(1-X)$ tiene una distribución Pearson tipo VI con parámetros de forma α_1, α_2 y parámetro de escala 1, se denota por $\text{PT6}(\alpha_1, \alpha_2, 1)$.

6) La función de densidad $\text{beta}(1,2)$ es un triángulo izquierdo, y la $\text{beta}(2,1)$ es un triángulo derecho.

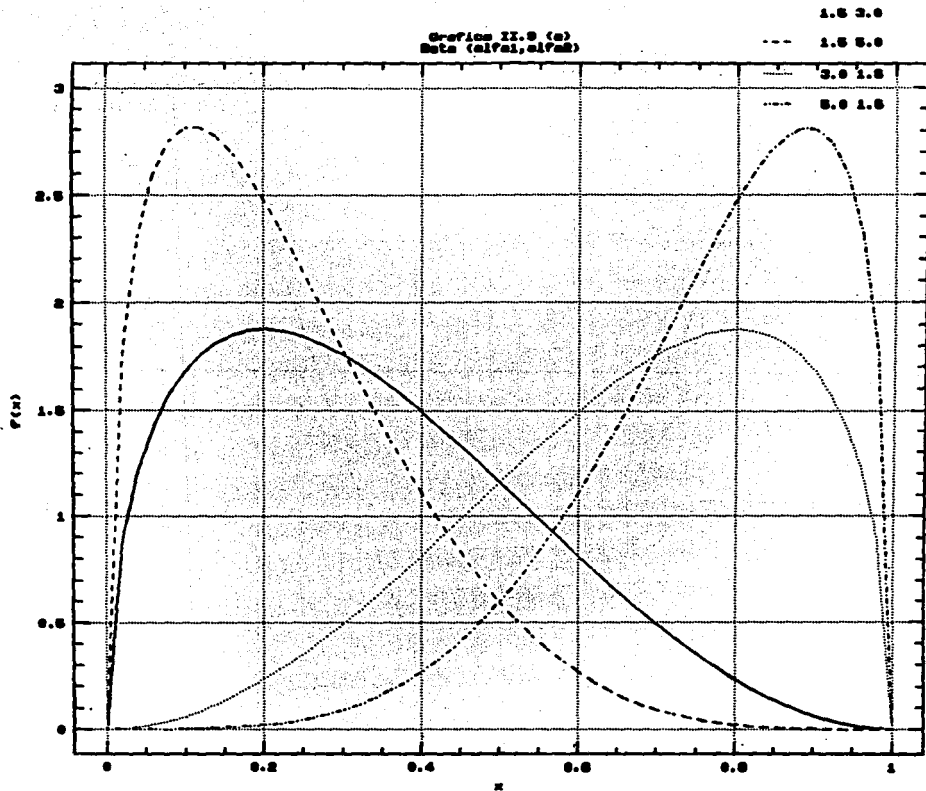
7)

$$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } \alpha_1 < 1 \\ \alpha_2 & \text{si } \alpha_1 = 1 \\ 0 & \text{si } \alpha_1 > 1 \end{cases}$$

$$\lim_{x \rightarrow 1} f(x) = \begin{cases} \infty & \text{si } \alpha_2 < 1 \\ \alpha_1 & \text{si } \alpha_2 = 1 \\ 0 & \text{si } \alpha_2 > 1 \end{cases}$$

8) La densidad es simétrica en $x=\frac{1}{2}$ si y sólo si $\alpha_1 = \alpha_2$. Además, la media y la moda son iguales si y sólo si $\alpha_1 = \alpha_2$.

Graphs II.9 (a)
Data (a1/a1, a1/a2)



Graph 11.0 (b)
Data (a1/a1, a1/a2)

1.0 1.0

--- 2.0 2.0

--- 3.0 3.0

--- 5.0 5.0

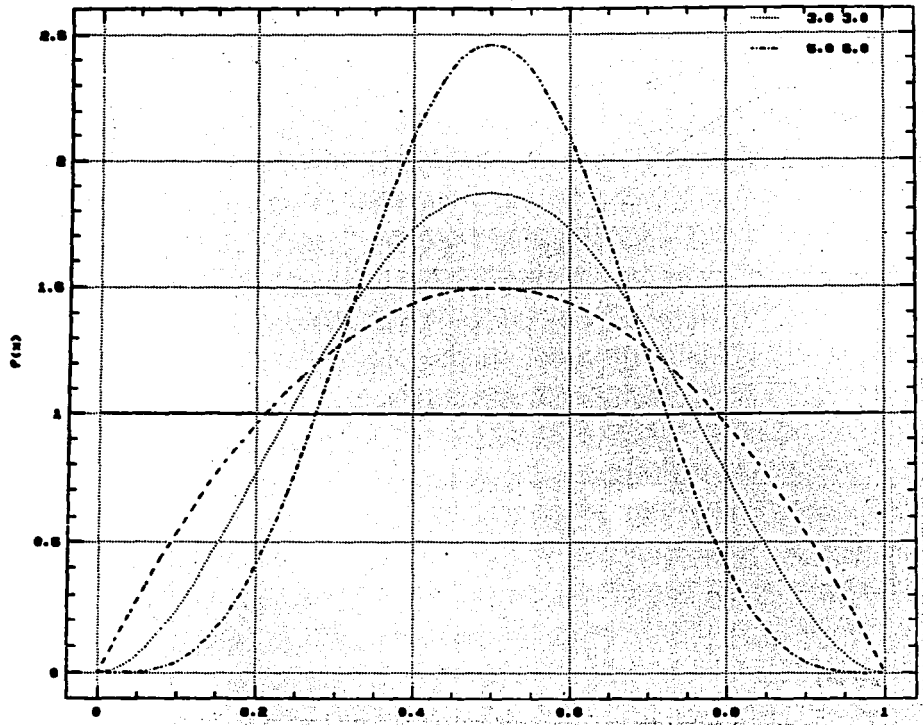
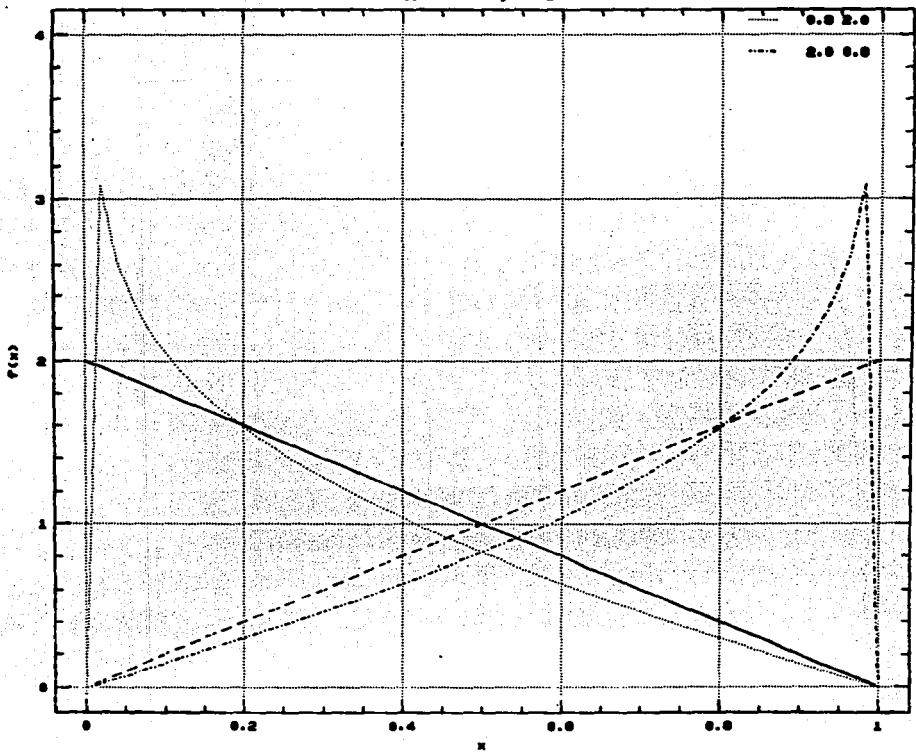
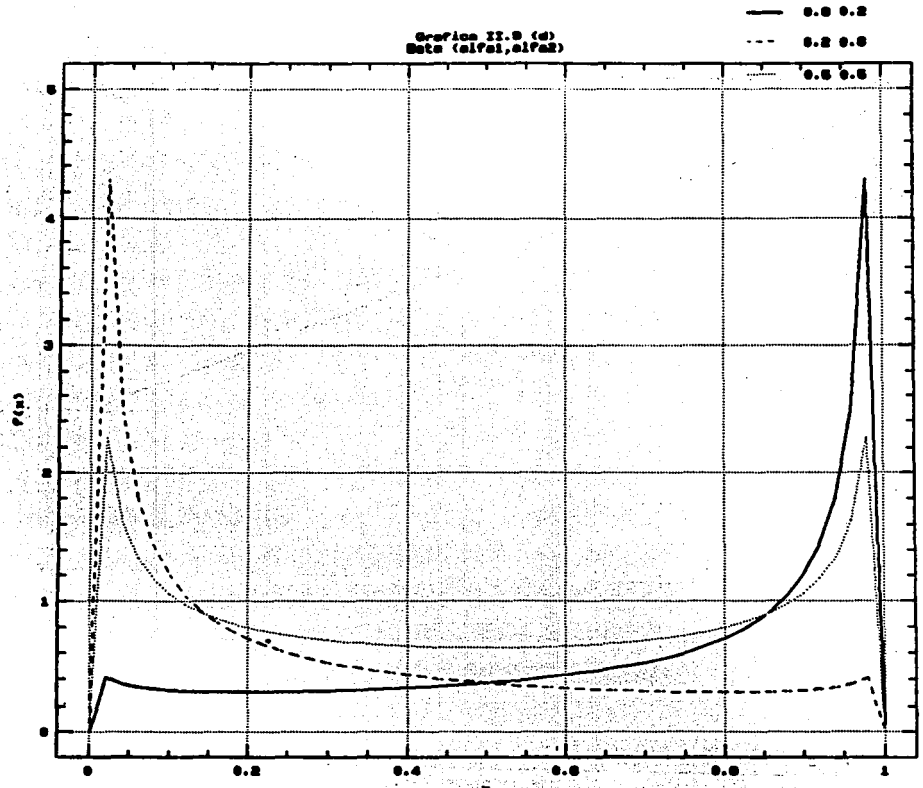


Gráfico 21.9 (a)
 Nota: (airal, airab)

- 1.0 2.0
- - - 2.0 1.0
- 0.0 2.0
- - - 2.0 0.0



Graph II.9 (d)
Beta (alpha, alpha)



PEARSON TIPO V $PT5(\alpha, \beta)$

Aplicaciones posibles:

Tiempo para realizar una tarea (la función de densidad toma formas similares a la lognormal, pero puede tener un pico más grande cuando se encuentra cercano a $x=0$).

Función de Densidad (ver gráfica II.10):

$$f(x) = \begin{cases} \frac{x^{-(\alpha+1)} e^{-\beta/x}}{\beta^{-\alpha} \Gamma(\alpha)} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 1 - F_G\left(\frac{1}{x}\right) & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

donde $F_G\left(\frac{1}{x}\right)$ es la función de distribución de una variable aleatoria $\text{gamma}(\alpha, 1/\beta)$.

Parámetros: De Forma $\alpha > 0$, de Escala $\beta > 0$

Rango: $[0, \infty)$

Media: $\frac{\beta}{\alpha-1}$ para $\alpha > 1$

Varianza: $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ para $\alpha > 2$

Moda: $\frac{\beta}{\alpha+1}$

Comentarios:

1) $X \sim PT5(\alpha, \beta)$ si y sólo si $Y=1/X \sim \text{gamma}(\alpha, 1/\beta)$. Entonces, la distribución Pearson tipo V suele ser denominada *distribución gamma invertida*.

2) Nótese que la media y la varianza únicamente existen para ciertos valores del parámetro de forma.

PEARSON TIPO VI $PTS(\alpha_1, \alpha_2, \beta)$

Aplicaciones posibles:

Tiempo para realizar una tarea.

Función de Densidad (ver gráfica II.11):

$$f(x) = \begin{cases} \frac{(x/\beta)^{\alpha_1-1}}{\beta B(\alpha_1, \alpha_2) [1+(x/\beta)]^{\alpha_1+\alpha_2}} & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(X) = \begin{cases} F_B\left(\frac{x}{\beta}\right) & \text{si } x > 0 \\ 0 & \text{otro caso} \end{cases}$$

donde $F_B(x)$ es la función de distribución de una variable aleatoria beta(α_1, α_2).

Parámetros:

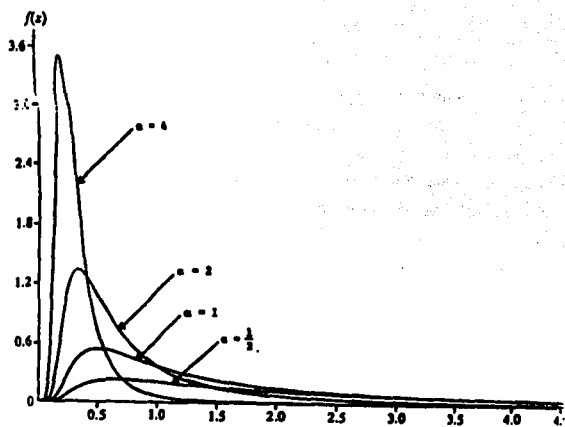
Parámetros de forma $\alpha_1 > 0$ y $\alpha_2 > 0$, parámetro de escala $\beta > 0$.

Rango: $(0, \infty)$

Media: $\frac{\beta \alpha_1}{\alpha_2 - 1}$ para $\alpha_2 > 1$

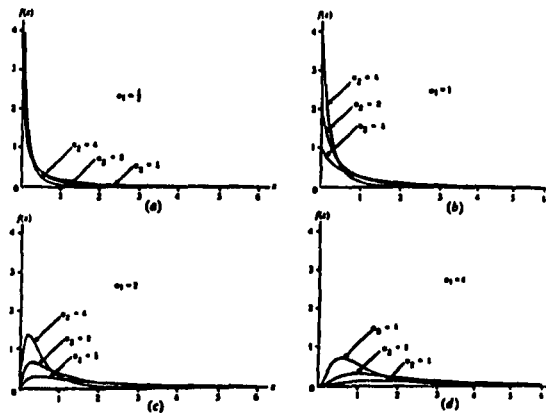
Varianza: $\frac{\beta^2 \alpha_1 (\alpha_1 + \alpha_2 - 1)}{(\alpha_2 - 1)^2 (\alpha_2 - 2)}$ para $\alpha_2 > 2$

Grafica II.10



$PTV(\alpha, 1)$

Grafica II.11



$PTVI(\alpha_1, \alpha_2, 1)$

Moda:

$$\begin{cases} \frac{\beta(\alpha_1 - 1)}{\alpha_2 + 1} & \text{si } \alpha_1 \geq 1 \\ 0 & \text{otro caso} \end{cases}$$

Comentarios:

- 1) $X \sim \text{PT6}(\alpha_1, \alpha_2, 1)$ si y sólo si $Y = X/(1+X) \sim \text{beta}(\alpha_1, \alpha_2)$.
- 2) Si X_1 y X_2 son variables aleatorias independientes con $X_i \sim \text{gamma}(\alpha_i, \beta)$, entonces $Y = X_1/X_2 \sim \text{PT6}(\alpha_1, \alpha_2, \beta)$.
- 3) Nótese que la media y la varianza únicamente existen para ciertos valores del parámetro de forma α_2 .

TRIANGULAR $\text{triang}(a, b, c)$

Aplicaciones posibles:

Utilizado como un modelo aproximado en la ausencia de datos.

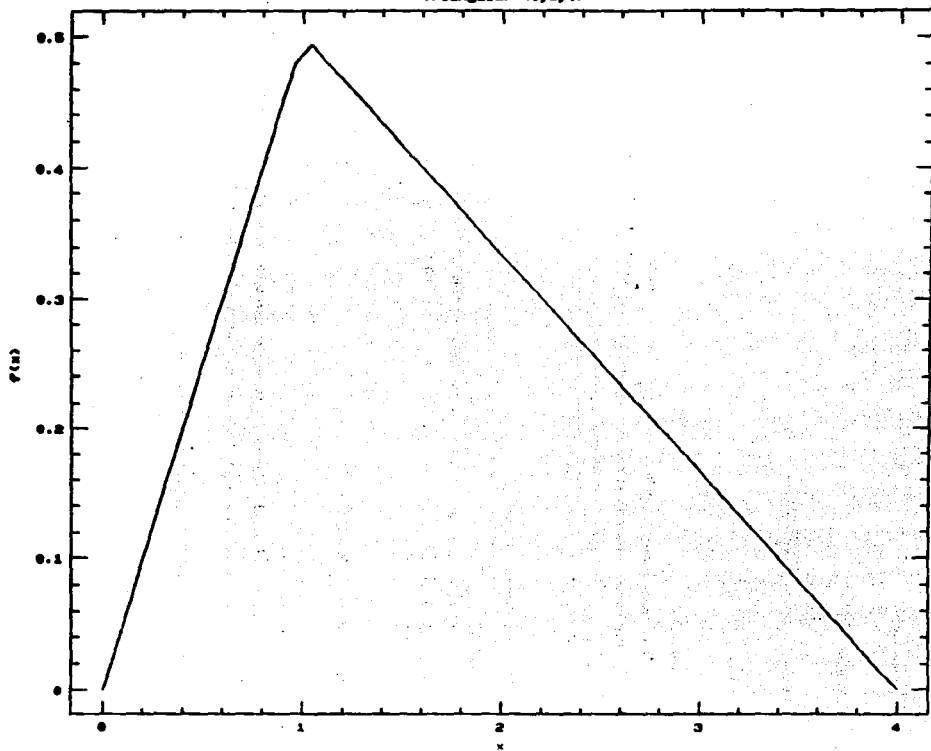
Función de Densidad (ver gráfica II.12):

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 1 & \text{si } b < x \end{cases}$$

Graph II.12
Triangular (0,1,4)



Parámetros:

a, b y c son números reales con $a < c < b$. a es un parámetro de localización, $b - a$ es un parámetro de escala. c es un parámetro de forma.

Rango: [a, b]

$$\text{Media: } \frac{a+b+c}{3}$$

$$\text{Varianza: } \frac{a^2+b^2+c^2-ab-ac-bc}{18}$$

Moda: c

Comentarios:

Los casos límites cuando $c = b$ y $c = a$ son denominados *distribución triangular derecha* y *triangular izquierda* respectivamente. Para $a=0$ y $b=1$ las distribuciones triangular derecha e izquierda son casos especiales de la distribución beta.

II.6.3 Distribuciones Discretas.

A continuación se presentan las distribuciones discretas más usuales en los modelos de simulación incluyendo las aplicaciones teóricas posibles para cada distribución.

BERNOULLI Bernoulli(p)

Aplicaciones posibles:

Suceso aleatorio en el que pueden obtenerse solamente dos resultados. Utilizada para generar otras variables aleatorias discretas como binomial, geométrica y binomial negativa.

Probabilidad (ver gráfica II.13):

$$p(x) = \begin{cases} 1 - p & \text{si } x = 0 \\ p & \text{si } x = 1 \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{cases}$$

Parámetro: $p \in (0, 1)$

Rango: $\{0, 1\}$

Media: p

Varianza: $p(1-p)$

Moda:

$$\begin{cases} 0 & \text{si } p < \frac{1}{2} \\ 0 \text{ y } 1 & \text{si } p = \frac{1}{2} \\ 1 & \text{si } p > \frac{1}{2} \end{cases}$$

Comentarios:

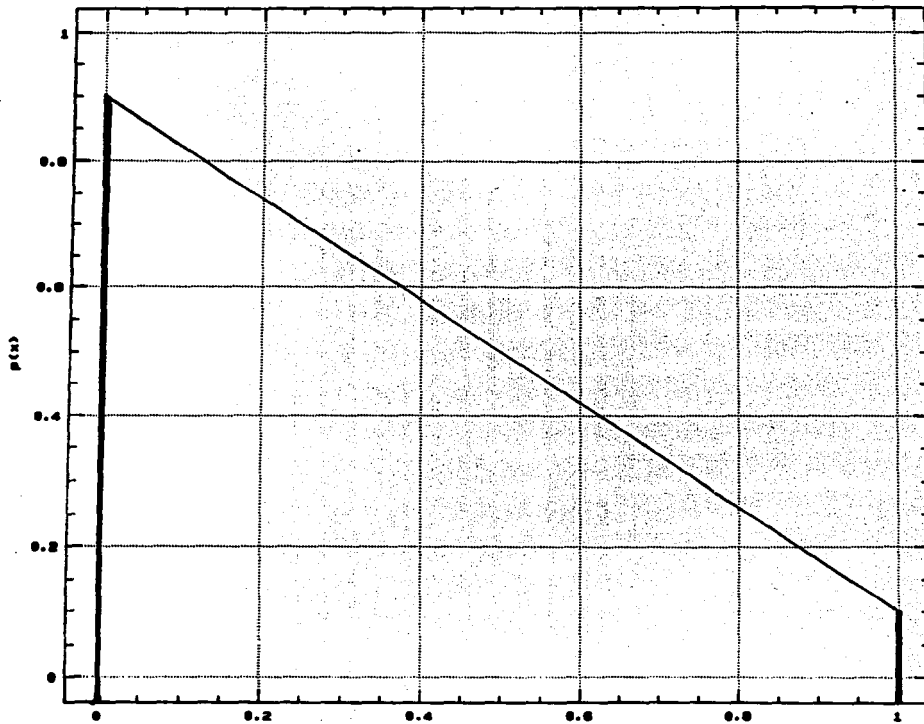
1) Una variable aleatoria X con distribución bernoulli(p) puede ser considerada como un experimento que "falla" o "tiene éxito". Si la probabilidad de éxito es p , y sea $X=0$ si el experimento falla y $X=1$ si tiene éxito, entonces $X \sim \text{bernoulli}(p)$. Este experimento denominado *prueba de Bernoulli*, proporciona una manera conveniente para relacionar otras distribuciones discretas con la distribución bernoulli.

2) Si t es un entero positivo y X_1, X_2, \dots, X_t son variables aleatorias independientes bernoulli(p), $X_1 + X_2 + \dots + X_t$ tiene una distribución binomial con parámetros p y t . Entonces, una variable aleatoria binomial puede ser considerada como el número de éxitos obtenidos al realizar de manera independiente un número fijo de pruebas bernoulli.

3) Supóngase que se realizan pruebas bernoulli independientes con probabilidad de éxito p . Entonces, el número de fallas *antes* de observar la primer prueba tiene una distribución geométrica con parámetro p . Para un entero positivo s , el número de fallas antes de observar la s -ésima prueba tiene una distribución binomial negativa con parámetros s y p .

4) La distribución bernoulli(p) es un caso especial de la distribución binomial (con $t=1$ y el mismo valor para p).

Grafica II.13
Bernoulli (0.1)



Nota: $p(x)$ se indica por las barras vert.

UNIFORME DISCRETA $DU(i, j)$

Aplicaciones posibles:

Suceso aleatorio en el que pueden ocurrir diversos resultados con la misma probabilidad. Utilizado como "primer" modelo para una cantidad que se conoce que varía entre los enteros i y j , y sobre los cual no se posee información adicional.

Probabilidad (ver gráfica H.14):

$$p(x) = \begin{cases} \frac{1}{j-i+1} & \text{si } x \in \{i, i+1, \dots, j\} \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < i \\ \frac{\lfloor x \rfloor - i + 1}{j - i + 1} & \text{si } i \leq x \leq j \\ 1 & \text{si } j < x \end{cases}$$

donde $\lfloor x \rfloor$ representa el mayor entero $\leq x$.

Parámetros:

i, j números enteros tal que $i \leq j$; i es un parámetro de localización, $j-i$ es un parámetro de escala.

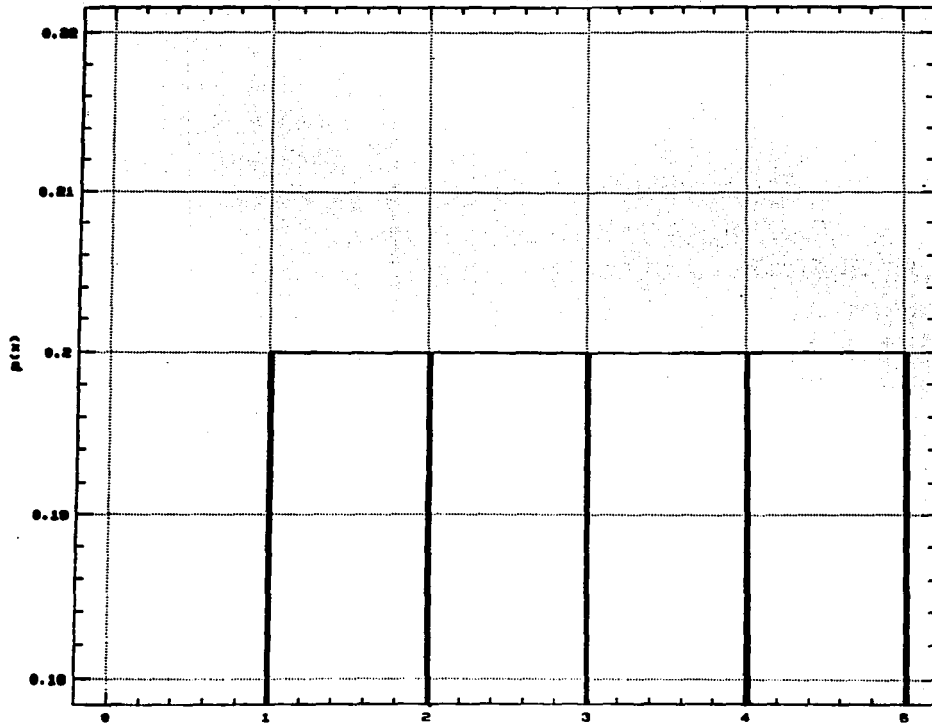
Rango: $\{i, i+1, \dots, j\}$

Media: $\frac{i+j}{2}$

Varianza: $\frac{(j-i+1)^2 - 1}{12}$

Moda: No existe únicamente

Grafico II.14
Uniforme Discreta (1,6)



Nota: $p(x)$ se indica por las barras vert.

Comentarios:

Las distribuciones $DU(0,1)$ y $Bernoulli(\frac{1}{2})$ son iguales.

BINOMIAL $\text{bin}(t, p)$

Sea A un suceso en un Experimento. Supongamos que la probabilidad que ocurra A es p . Se define la variable aleatoria X como el número de veces que ocurrió el suceso A en n intentos. X será la variable aleatoria binomial con los parámetros n y p , sus valores posibles serán: $0, 1, 2, \dots, n$.

Aplicaciones posibles:

Número de éxitos en t pruebas Bernoulli independientes con probabilidad de éxito p en cada prueba; número de artículos defectuosos en un lote de tamaño t ; número de artículos en un lote (por ejemplo grupo de personas) de tamaño aleatorio; número de artículos solicitados un inventario.

Probabilidad (ver gráfica II.15):

$$p(x) = \begin{cases} \binom{t}{x} p^x (1-p)^{t-x} & \text{si } x \in \{0, 1, \dots, t\} \\ 0 & \text{otro caso} \end{cases}$$

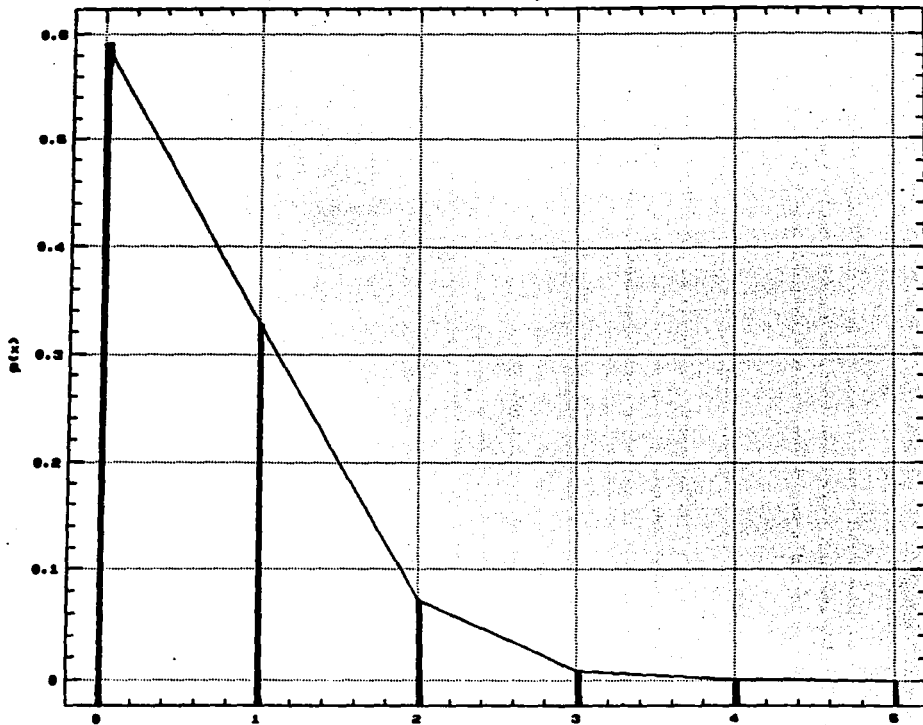
donde $\binom{t}{x}$ es el *coeficiente binomial* definido por

$$\binom{t}{x} = \frac{t!}{x!(t-x)!}$$

Función de Distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{\lfloor x \rfloor} \binom{t}{i} p^i (1-p)^{t-i} & \text{si } 0 \leq x \leq t \\ 1 & \text{si } t < x \end{cases}$$

Gráfico II.18 (a)
Binomial (8, 0, 0.1)



Nota: $p(x)$ se indica por las barras vert.

Gráfico II.18 (b)
Bimomial (10.0,0.1)

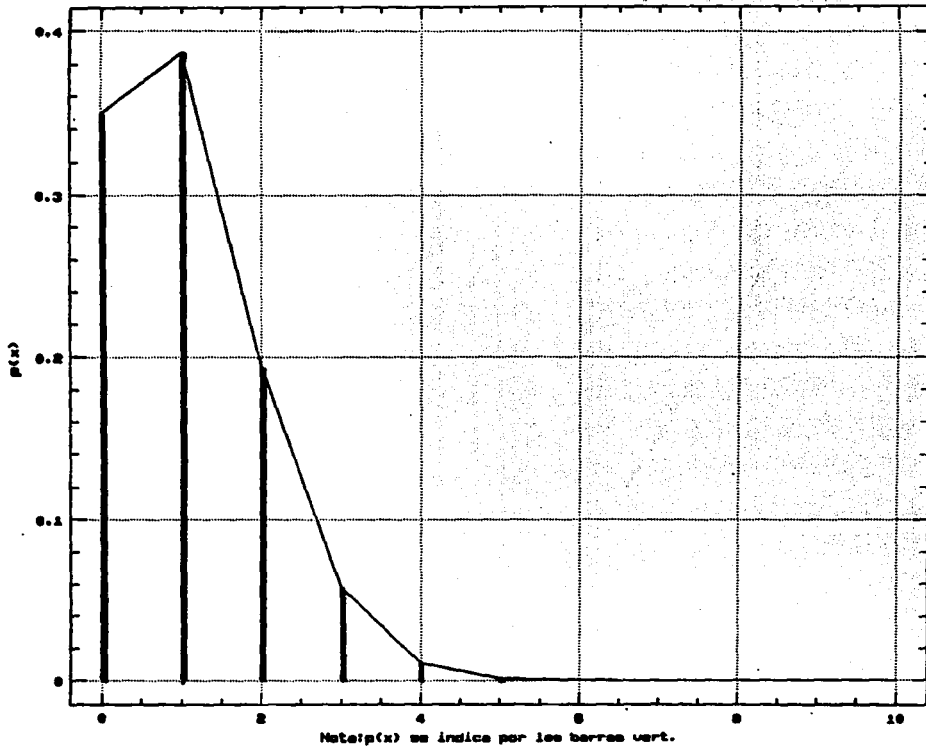
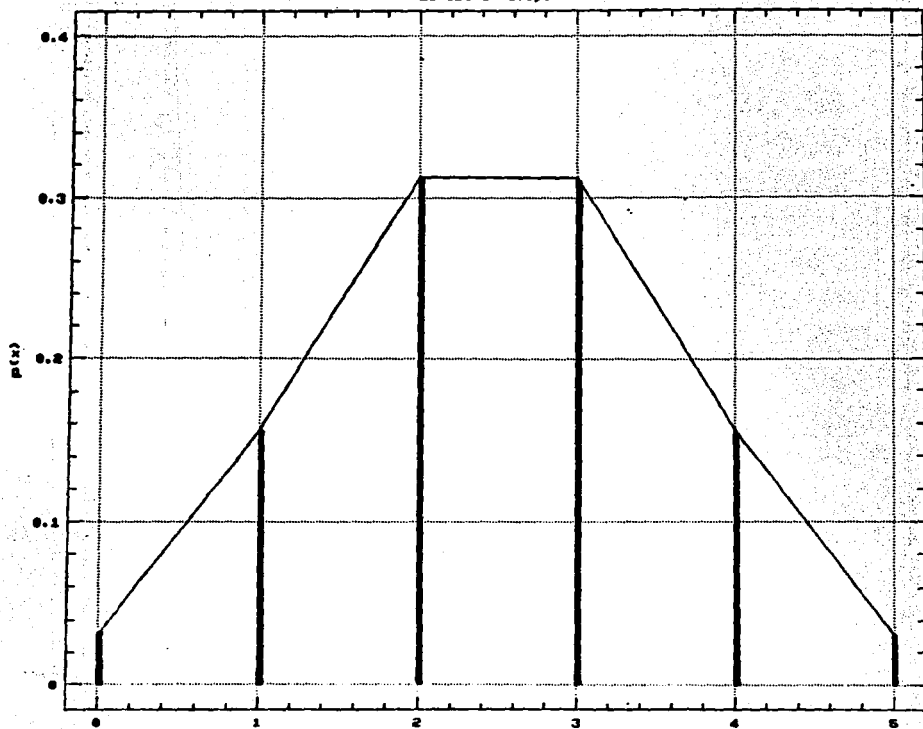
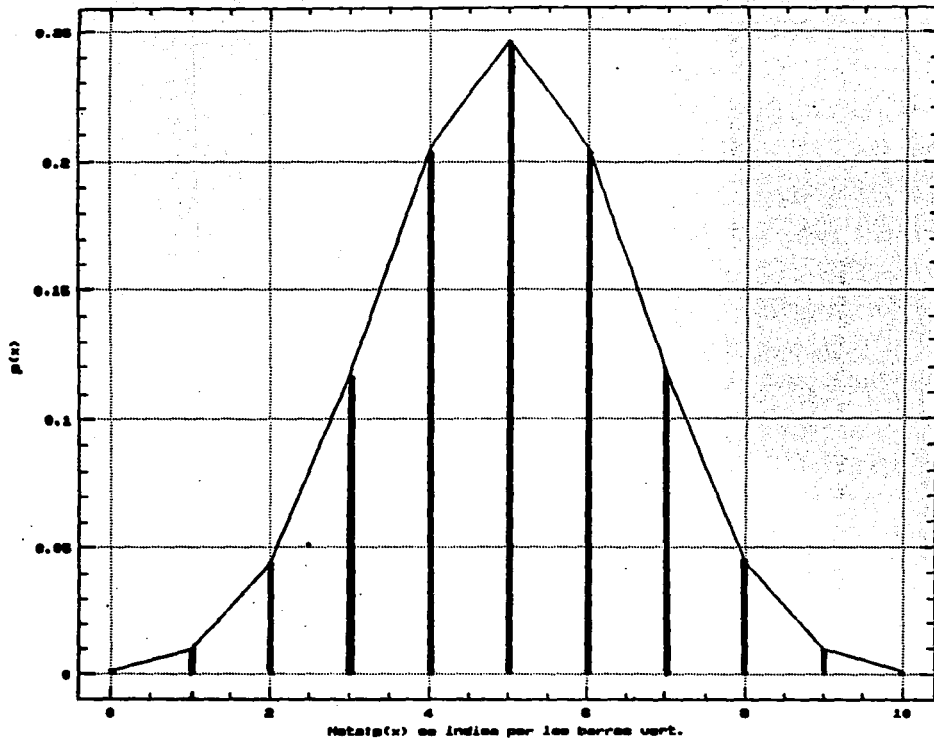


Grafico II.18 (a)
Binomial (0.5, 0.5)



Nota: $p(x)$ se indica por las barras vert.

Grafico II.18 (d)
Binomial (10, 0,5)



Parámetros:

t es un número entero positivo, $p \in (0,1)$.

Rango: $\{0,1,\dots,t\}$

Media: tp

Varianza: $tp(1-p)$

Moda:

$p(t+1)-1$ y $p(t+1)$	si $p(t+1)$ es un entero
$\lfloor p(t+1) \rfloor$	otro caso

Comentarios:

1) Si Y_1, Y_2, \dots, Y_t son variables aleatorias independientes Bernoulli(p), entonces $Y_1 + Y_2 + \dots + Y_t \sim \text{bin}(t, p)$.

2) Si X_1, X_2, \dots, X_m son variables aleatorias independientes y $X_i \sim \text{bin}(t_i, p)$, entonces $X_1 + X_2 + \dots + X_m \sim \text{bin}(t_1, t_2, \dots, t_m, p)$.

3) La distribución $\text{bin}(t, p)$ es simétrica si y sólo si $p = \frac{1}{2}$.

4) $X \sim \text{bin}(t, p) \Leftrightarrow t-X \sim \text{bin}(t, 1-p)$.

5) Las distribuciones $\text{bin}(1, p)$ y $\text{bernoulli}(p)$ son iguales.

GEOMETRICA geom(p)

Aplicaciones posibles:

Número de fallas antes de que ocurra el primer éxito en una secuencia de pruebas independientes bernoulli con probabilidad de éxito p para cada prueba; número de artículos inspeccionados antes de que sea encontrado el primer artículo defectuoso; número de artículos en un lote de tamaño aleatorio; número de artículos solicitados de un inventario.

Probabilidad (ver gráfica II.16):

$$p(x) = \begin{cases} p(1-p)^x & \text{si } x \in \{0,1,\dots\} \\ 0 & \text{otro caso} \end{cases}$$

Grafico II.10 (a)
 Geometria (0.25)

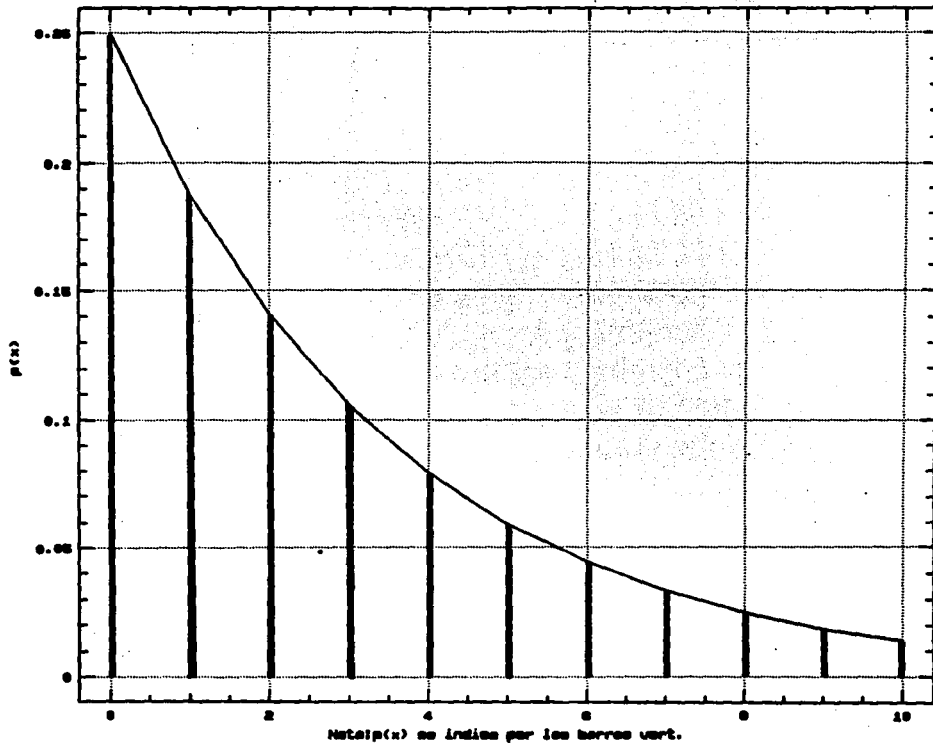
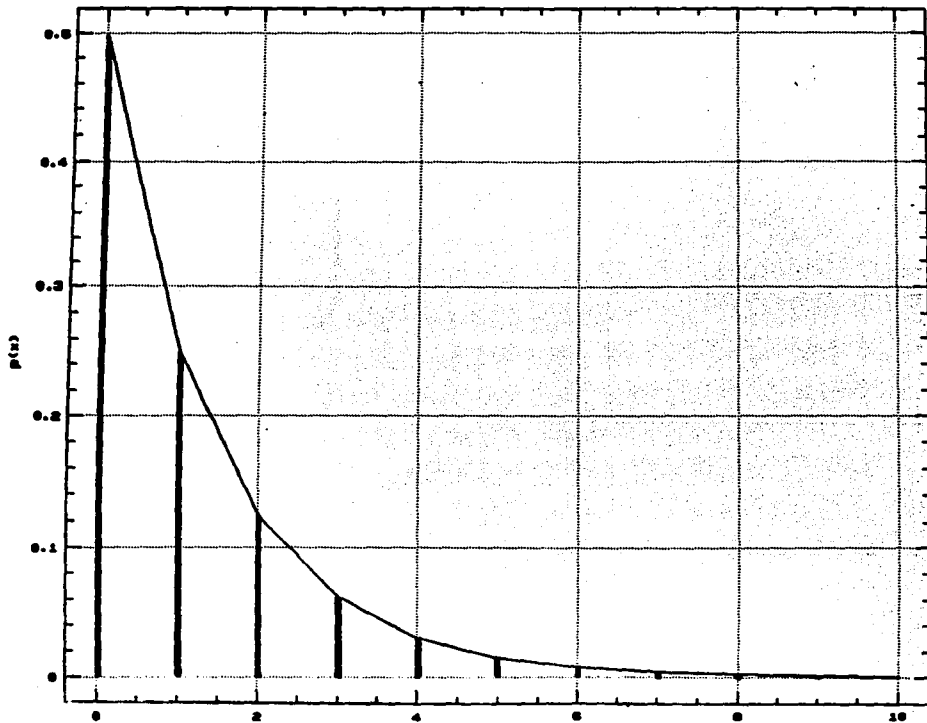


Gráfico II.18 (b)
Geometria (0.6)



Nota: $p(x)$ em índices por las barras vert.

Función de Distribución:

$$F(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor + 1} & \text{si } x \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

Parámetros:

$$p \in (0, 1)$$

Rango: $\{0, 1, \dots\}$

Media: $\frac{1-p}{p}$

Varianza: $\frac{1-p}{p^2}$

Moda: 0

Comentarios:

1) Si Y_1, Y_2, \dots es una secuencia de variables aleatorias independientes bernoulli(p) y $X = \min\{i: Y_i = 1\} - 1$, entonces $X \sim \text{geom}(p)$.

2) Si X_1, X_2, \dots, X_s son variables aleatorias independientes $\text{geom}(p)$, entonces $X_1 + X_2 + \dots + X_s$ tiene una distribución binomial negativa con parámetros s y p .

3) La distribución $\text{geom}(p)$ es un caso especial de la distribución binomial negativa (con $s=1$ y para el mismo valor de p).

BINOMIAL NEGATIVA $\text{binneg}(s, p)$

Aplicaciones posibles:

Número de fallas antes de que ocurra el s -ésimo suceso en una secuencia de pruebas independientes bernoulli con probabilidad de éxito p en cada prueba; número de artículos correctos antes de que se encuentre el s -ésimo artículo defectuoso; número de artículos en un lote de tamaño aleatorio; número de artículos solicitados de un almacén.

Probabilidad (ver gráfica II.17):

$$p(x) = \begin{cases} \binom{s+x-1}{x} p^x (1-p)^s & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{s+i-1}{i} p^i (1-p)^s & \text{si } x \geq 0 \\ 0 & \text{otro caso} \end{cases}$$

Parámetros:

s es un entero positivo, $p \in (0, 1)$.

Rango: $\{0, 1, \dots\}$

Media: $\frac{s(1-p)}{p}$

Varianza: $\frac{s(1-p)}{p^2}$

Moda:

Sea $y = [s(1-p) - 1]/p$; entonces

$$\text{Moda} = \begin{cases} y \wedge y+1 & \text{si } y \text{ es un entero} \\ \lfloor y \rfloor + 1 & \text{otro caso} \end{cases}$$

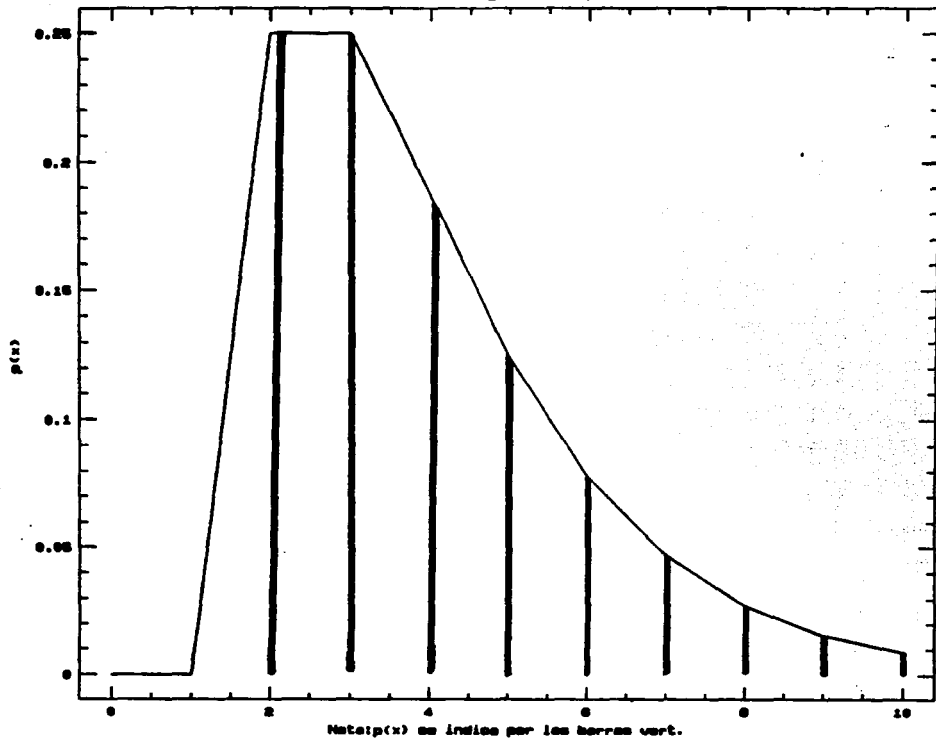
Comentarios:

1) Si Y_1, Y_2, \dots, Y_s son variables aleatorias independientes $\text{geom}(p)$, entonces $Y_1 + Y_2 + \dots + Y_s \sim \text{binneg}(s, p)$.

2) Si Y_1, Y_2, \dots es una secuencia de variables aleatorias independientes $\text{bernoulli}(p)$ y $X = \min\{i: \sum_{j=1}^i Y_j = s\} - s$, entonces $X \sim \text{binneg}(s, p)$.

3) Si X_1, X_2, \dots, X_m son variables aleatorias independientes y $X_i \sim \text{binneg}(s_i, p)$, entonces

Gráfico II.17
Barras Negativas (E.O.O.S)



$X_1 + X_2 + \dots + X_m \sim \text{binneg}(s_1 + s_2 + \dots + s_m, p)$.

4) Las distribuciones $\text{binneg}(1, p)$ y $\text{geom}(p)$ son iguales.

POISSON Poisson(λ)

Aplicaciones posibles:

Número de eventos que ocurren en un intervalo de tiempo cuando los eventos ocurren de manera constante; número de artículos en un lote de tamaño aleatorio; número de artículos solicitados en un inventario.

Probabilidad (ver gráfica II.18):

$$p(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{otro caso} \end{cases}$$

Función de Distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{si } 0 \leq x \end{cases}$$

Parámetros:

$$\lambda > 0$$

Rango: $\{0, 1, \dots\}$

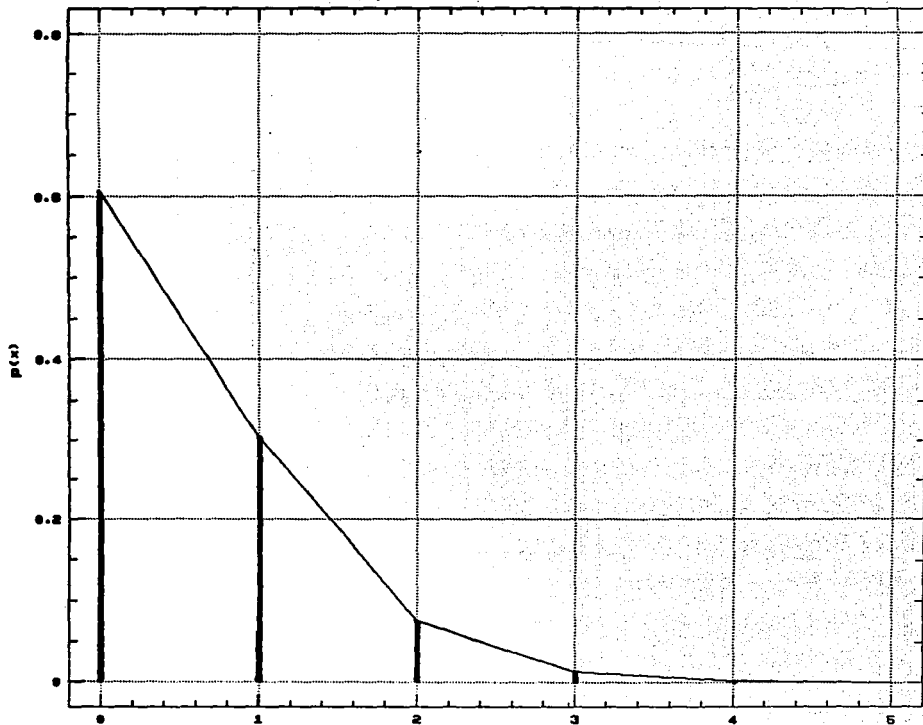
Media: λ

Varianza: λ

Moda:

$$\begin{cases} \lambda - 1 \text{ y } \lambda & \text{si } \lambda \text{ es un entero} \\ \lfloor \lambda \rfloor & \text{otro caso} \end{cases}$$

Gráfico II.18 (a)
Polinômio (G.B.)



Nota: p(x) se indica por las barras vert.

Grafico II.18 (b)
Pulsos (1.0)

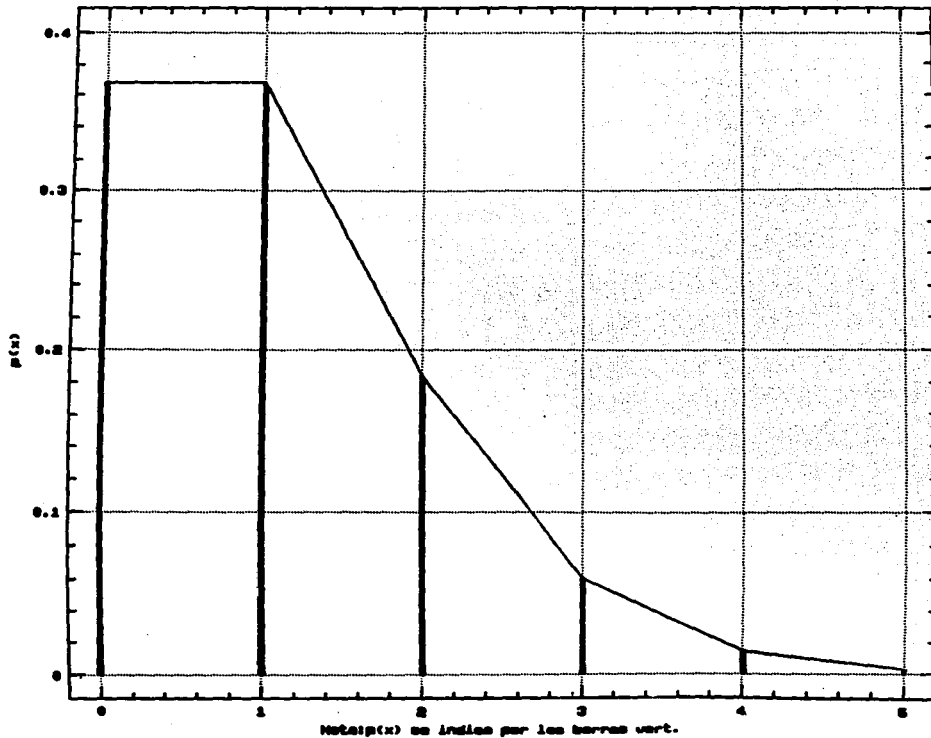


Gráfico II.18 (a)
Poisson (2.0)

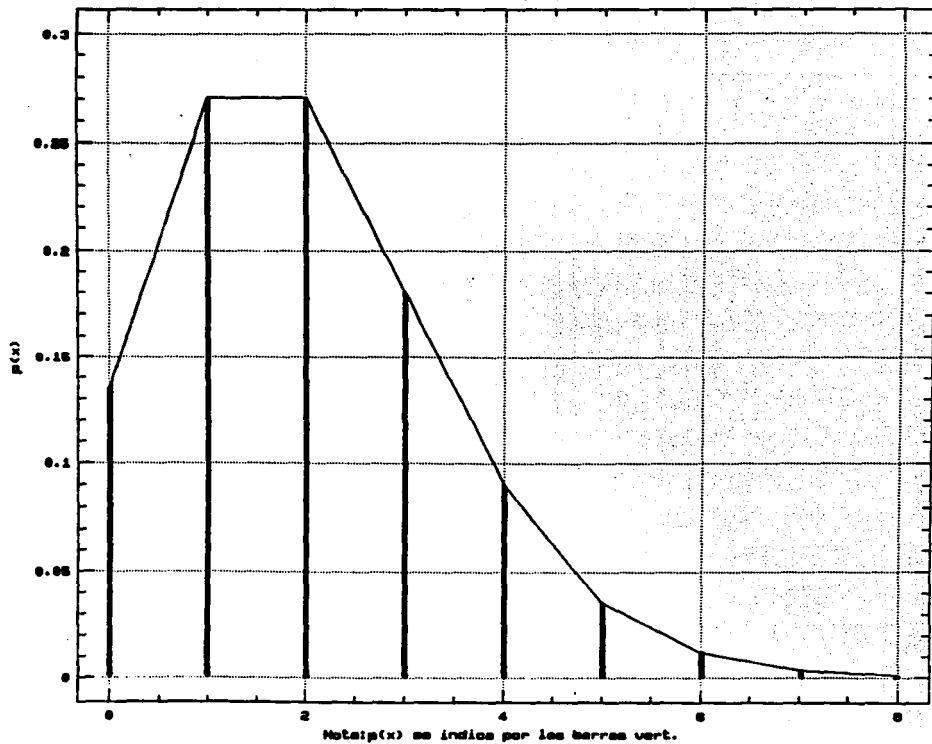
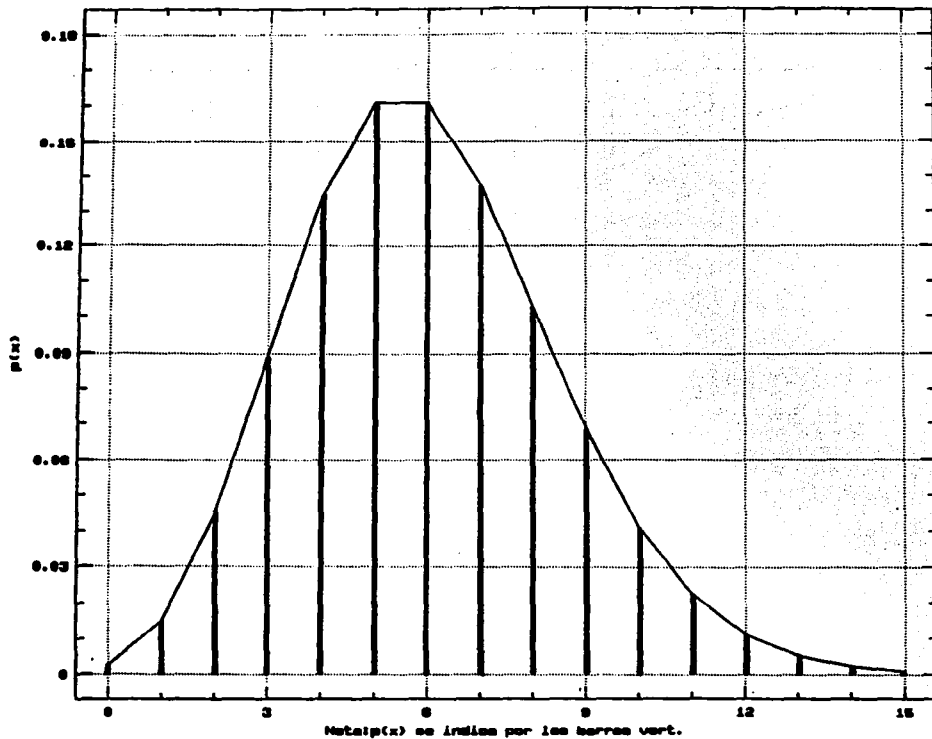


Gráfico II.18 (d)
Poisson (6.8)



Comentarios:

1) Sea Y_1, Y_2, \dots una secuencia de variables aleatorias i.i.d. y sea $X = \max\{i: \sum_{j=1}^i Y_j \leq 1\}$. Entonces la distribución de las Y_i 's es exponencial $\Leftrightarrow X \sim \text{Poisson}(\lambda)$. Además, si $X' = \max\{i: \sum_{j=1}^i Y_j \leq \lambda\}$, entonces las Y_i 's son $\text{expo}(1) \Leftrightarrow X' \sim \text{Poisson}(\lambda)$.

2) Si X_1, X_2, \dots, X_m son variables aleatorias independientes y $X_i \sim \text{Poisson}(\lambda_i)$, entonces $X_1 + X_2 + \dots + X_m \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_m)$.

II.6.4 Distribuciones Empíricas

Para variables aleatorias continuas, la definición de la distribución empírica depende si se conocen las observaciones X_1, X_2, \dots, X_n o si se posee solamente el número de X_i 's que caen en un determinado intervalo. Si se tienen los datos originales, se puede definir una función lineal continua F al ordenar las X_i 's en orden ascendente. Si $X_{(i)}$ es el i -ésimo menor valor de las X_j 's, se tiene que: $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Entonces, F está dada por:

$$F(x) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{i-1}{n-1} + \frac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & \text{si } X_{(i)} \leq x \leq X_{(i+1)} \text{ para } i=1, 2, \dots, n-1 \\ 1 & \text{si } X_{(n)} \leq x \end{cases}$$

$F(x)$ crece más rápido en aquellos rangos para los que las X_i 's se encuentran más densamente distribuidas. Para cada i se tiene que: $F(X_{(i)}) = (i-1)/(n-1)$, que para n grande será aproximadamente la proporción de X_j 's que son menores que $X_{(i)}$, siendo ésta la manera en que se espera que se comporte una función de distribución continua. Sin embargo, esta distribución presenta la desventaja de que las variables aleatorias generadas no pueden ser menores que $X_{(1)}$, ni mayores que $X_{(n)}$; además, la media de $F(X)$ no es igual a la media muestral $\bar{X}(n)$ obtenida a partir de los datos $X_{(i)}$'s.

Por otro lado, si los datos se encuentran previamente agrupados es necesario utilizar un enfoque diferente. Supóngase que n variables aleatorias X_i 's se encuentran agrupadas en k intervalos

adyacentes: $[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$, de tal manera que el intervalo j -ésimo contiene n_j observaciones y donde $n_1 + n_2 + \dots + n_k = n$ (generalmente los intervalos tienen la misma medida pero no es necesario asumir esto). Una función de distribución empírica puede ser especificada de la siguiente manera: $G(a_0) = 0$ y $G(a_j) = (n_1 + n_2 + \dots + n_j) / n$ para $j = 1, 2, \dots, k$. Entonces, interpolando linealmente entre las a_j 's se define:

$$G(x) = \begin{cases} 0 & \text{si } x < a_0 \\ G(a_{j-1}) + \frac{x - a_{j-1}}{a_j - a_{j-1}} [G(a_j) - G(a_{j-1})] & \text{si } a_{j-1} \leq x < a_j \text{ para } j = 1, 2, \dots, k \\ 1 & \text{si } a_k \leq x \end{cases}$$

En donde $G(a_j)$ es la proporción de X_i 's menores que a_j , y $G(x)$ tiene un crecimiento más rápido sobre los rangos de x en donde el número de observaciones es mayor. Sin embargo, los valores aleatorios generados por esta distribución también están acotados por a_0 y a_k . Esto representa un problema ya que la generación de valores elevados puede ocasionar un impacto significativo en una corrida de simulación. Para remediar esta desventaja, se propone anexar en el lado derecho una distribución exponencial que permitirá que se generen valores superiores a $X_{(n)}$ [para una explicación más detallada consultar Law (1991, p. 353)].

En el caso de datos discretos es muy sencillo definir una distribución empírica. Si se encuentran disponibles los datos originales X_1, X_2, \dots, X_n , para cada valor posible x se define una función de masa empírica $p(x)$ como la proporción de X_i 's que son iguales a x . En el caso de datos discretos agrupados, se define una función de masa de tal manera que la suma de los $p(x)$ sobre todos los valores posibles de x en un intervalo, sea igual a la proporción de X_i 's dentro del intervalo. La forma en que los $p(x)$ son asignados sobre los posibles valores de x en el intervalo es arbitraria.

II.7. ELECCION DE DISTRIBUCIONES DE PROBABILIDAD

El primer paso para la elección de una distribución de probabilidad específica consiste en agrupar los datos y graficarlos en un histograma, para así poder determinar visualmente las familias de distribución (exponencial, normal, etc.) que pueden resultar adecuadas de acuerdo a la forma que

presentan los datos. Si los datos no presentan ninguna forma de distribución aparente, es posible agruparlos de manera diferente y volver a graficarlos.

En algunas ocasiones en que se busca una distribución adecuada, es posible utilizar el conocimiento a priori acerca de la manera en que una variable aleatoria funciona dentro del sistema. Esto se basa en conocimientos teóricos previos y no se requiere de datos.

Asimismo, la elección del rango de una distribución puede determinar si ésta resulta adecuada o no. Por ejemplo, los tiempos de servicio no deben ser generados a partir de una distribución normal debido a que una variable aleatoria generada por una distribución normal puede resultar negativa y los tiempos de servicio nunca serán negativos. Por otro lado, la proporción de partes defectuosas no debe tener una distribución gamma ya que la proporción de defectos debe variar entre 0 y 1, mientras que las variables aleatorias generadas por la distribución gamma no tienen límite superior.

La información a priori debe de ser utilizada si se encuentra disponible; sin embargo, se recomienda respaldar la distribución elegida por medio de datos y pruebas estadísticas de ajuste de la distribución.

Existen diversas pruebas empíricas (comparaciones de frecuencias, gráficas de probabilidad, etc.) que permiten determinar si un conjunto de datos se ajusta a determinada distribución. Sin embargo, es necesario realizar pruebas de bondad y de ajuste que son pruebas de hipótesis estadísticas que permiten determinar si las observaciones X_1, X_2, \dots, X_n son muestras independientes de una distribución particular con función de distribución F . Es decir, una prueba de bondad y de ajuste puede ser utilizada para probar la siguiente hipótesis nula:

$H_0 =$ Las X_i 's son variables aleatorias i.i.d. con función de distribución F .

Es necesario aclarar, que el fallar en rechazar H_0 no debe ser interpretado como "aceptar H_0 como verdadero". Generalmente, estas pruebas no resultan muy poderosas para una muestra de tamaño pequeño o mediano; es decir, no son muy sensibles a pequeños desacuerdos entre los datos y la distribución a ajustar. Entonces, deben ser consideradas como una aproximación en la detección de diferencias grandes. Por otro lado, si el tamaño de la muestra n es grande por lo general las pruebas rechazarán a H_0 . Esto se debe a que H_0 casi nunca resulta completamente verdadera, por lo que para n grande se detectará cualquier pequeña diferencia entre los datos y la distribución. Esto representa una

desventaja en estas pruebas, ya que generalmente en simulación es suficiente tener una distribución que sea "casi" correcta.

Existen básicamente dos pruebas de bondad y ajuste: Chi cuadrada y Kolmogorov-Smirnov. La mayoría de los programas estadísticos incluyen un módulo de aplicación de estas pruebas a un conjunto de datos dado. En este caso particular, se aplicó el paquete estadístico StatGraf que realiza ambas pruebas.

II.7.1. Pruebas Chi-Cuadrada.

Una prueba de bondad y ajuste puede ser considerada como la comparación formal entre un histograma o una gráfica lineal con una función de densidad o de masa específica.

Para llevar a cabo esta prueba, tanto en el caso continuo como en el discreto, es necesario dividir la distribución a ajustar en k intervalos adyacentes $[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$, en donde es posible que $a_0 = -\infty$ y $a_k = \infty$.

Sea N_j = número de X_i 's en el j -ésimo intervalo $[a_{j-1}, a_j)$ para $j=1, 2, \dots, k$. (Nótese que $\sum_{j=1}^k N_j = n$).

Se calcula la proporción esperada p_j de X_i 's que caerían en el j -ésimo intervalo si se muestreara la distribución a ajustar.

Para el caso continuo:

$$p_j = \int_{a_{j-1}}^{a_j} \hat{f}(x) dx$$

donde $\hat{f}(x)$ es la función de densidad de la distribución a ajustar.

En caso de datos discretos:

$$p_j = \sum_{a_{j-1} \leq X_i < a_j} \hat{p}(X_i)$$

donde \hat{p} es la función de masa de la distribución a ajustar.

Finalmente, la prueba estadística será:

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}$$

Debido a que np_j es el número esperado de las nX_i 's que caerían en el intervalo j -ésimo si H_0 fuera verdadero, se espera que χ^2 sea pequeño si el ajuste es bueno. Entonces, se rechaza H_0 si χ^2 es elevado. La precisión de la forma de la prueba depende de si los parámetros de la distribución fueron estimados a partir de los datos.

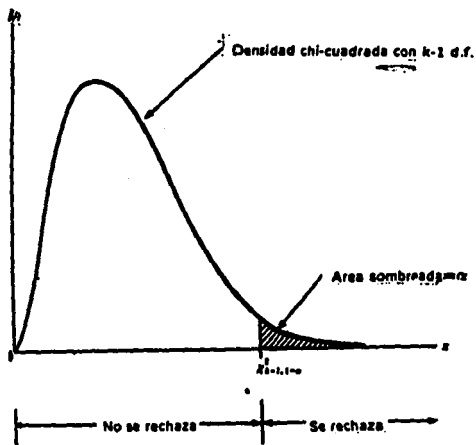
En primer lugar, supóngase que todos los parámetros de la distribución a ajustar son conocidos, es decir, se especificó la distribución sin que se hayan utilizado los datos. El hacer esta suposición no parece ser muy práctico; sin embargo, existen dos aplicaciones en simulación:

1. En la prueba del proceso de Poisson (no considerada en este trabajo) se analiza si los tiempos de llegada pueden ser considerados como variables aleatorias i.i.d. $U(0, T)$, donde T es una constante independiente de los datos.
2. En las pruebas empíricas de generadores de números aleatorios se prueba para una distribución $U(0,1)$ su independencia.

Si H_0 es verdadera, a medida que $n \rightarrow \infty$ χ^2 converge a una distribución chi-cuadrada con $k-1$ grados de libertad, que es equivalente a la distribución gamma $((k-1)/2, 2)$. Entonces, para n elevado y en una prueba con un nivel aproximado α , se rechaza H_0 si $\chi^2 > \chi_{k-1, 1-\alpha}^2$ (ver gráfica II.19), donde $\chi_{k-1, 1-\alpha}^2$ es el punto crítico superior $1-\alpha$ para una distribución chi-cuadrada con $k-1$ grados de libertad (para consultar los valores para $\chi_{k-1, 1-\alpha}^2$ ver Law, pag. 739). Nótese que la prueba chi-cuadrada solamente es válida a medida que $n \rightarrow \infty$.

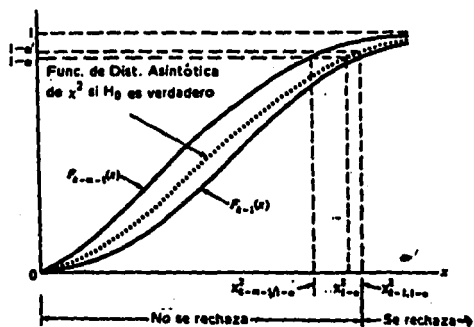
Ahora, supóngase que para poder especificar la distribución a ajustar resulta necesario calcular m parámetros ($m \geq 1$) a partir de los datos recolectados. Si H_0 es verdadera, a medida que $n \rightarrow \infty$ la función de distribución de χ^2 converge a una función de distribución que yace entre las funciones de

Grafica II.19



Prueba chi-cuadrada cuando se conocen todos los parámetros

Grafica II.20



Prueba chi-cuadrada cuando m parámetros son estimados

distribución chi-cuadrada con $k-1$ y $k-m-1$ grados de libertad. (ver gráfica II.20, donde F_{k-1} y F_{k-m-1} representan las funciones de distribución de las distribuciones chi-cuadrada con $k-1$ y $k-m-1$ grados de libertad respectivamente, y la función de distribución punteada es aquella a la que se aproxima la función de distribución χ^2 a medida que $n \rightarrow \infty$). Si se asume que $\chi_{1-\alpha}^2$ es el punto crítico superior $1-\alpha$ de la distribución asintótica de χ^2 , entonces:

$$\chi_{k-m-1, 1-\alpha}^2 \leq \chi_{1-\alpha}^2 \leq \chi_{k-1, 1-\alpha}^2$$

como se muestra en la gráfica II.20. Desafortunadamente, por lo general se desconoce el valor de $\chi_{1-\alpha}^2$. Es obvio que H_0 debe ser rechazado si $\chi^2 > \chi_{k-1, 1-\alpha}^2$ y H_0 no debe ser rechazado si $\chi^2 < \chi_{k-m-1, 1-\alpha}^2$. Sin embargo, se presenta una situación ambigua cuando ocurre que:

$$\chi_{k-m-1, 1-\alpha}^2 \leq \chi^2 \leq \chi_{k-1, 1-\alpha}^2$$

(ver gráfica II.20).

Se recomienda rechazar H_0 solamente si $\chi^2 > \chi_{k-1, 1-\alpha}^2$ debido a que esto es un enfoque tradicional, es decir, la probabilidad actual α' de cometer un error Tipo I [rechazar H_0 cuando es verdadero] es por lo menos tan pequeña como la probabilidad α . Sin embargo, esta opción implica pérdida en el poder de la prueba (probabilidad de rechazar una H_0 falsa). Usualmente, m no es mayor que 2 y si k es lo suficientemente larga, la diferencia entre $\chi_{k-m-1, 1-\alpha}^2$ y $\chi_{k-1, 1-\alpha}^2$ no resultará muy grande. Entonces, H_0 se rechaza si y sólo si $\chi^2 > \chi_{k-1, 1-\alpha}^2$, (al igual que el caso en el que se conocen todos los parámetros).

El mayor problema que se presenta en una prueba chi-cuadrada es la elección adecuada del número y tamaño de los intervalos. No existe una regla que garantice resultados adecuados en cuanto al poder y a la validez (que el nivel real de la prueba se encuentre cerca del nivel deseado α) para todos los tipos de distribuciones hipotéticas y los diferentes tamaños de muestras. Sin embargo, existen algunos lineamientos que resulta conveniente tomar en cuenta:

1) Se elimina parte de la ambigüedad en la selección de intervalos si estos son elegidos de tal manera que: $p_1 = p_2 = \dots = p_k$ (denominado *enfoque equiprobable*). En el caso continuo, esto puede presentar inconvenientes para algunas distribuciones ya que es necesario invertir la distribución a ajustar. Para el caso de las distribuciones discretas solamente será posible establecer las p_i 's

aproximadamente iguales.

2) Sea $a = \min_{1 \leq j \leq k} np_j$ y sea $y(5)$ el número de np_j 's menor que 5. Entonces, en base a investigaciones teóricas y empíricas, la prueba chi-cuadrada será aproximadamente válida si $k \geq 3$ y $a \geq 3y(5)/k$. En el caso de intervalos equiprobables, las condiciones son satisfechas si $k \geq 3$ y $np_j \geq 5 \forall j$.

Existen ocasiones en las que debido a la elección de los intervalos, se puede llegar a conclusiones totalmente diferentes a partir de los mismos datos. A pesar de lo anterior, esta prueba continúa siendo ampliamente utilizada ya que puede ser aplicada a cualquier distribución hipotética.

Ejemplos Prácticos de la Prueba χ^2 :

a) Se tira un dado 300 veces y se observan las siguientes frecuencias:

$$X_1 = 57$$

$$X_2 = 43$$

$$X_3 = 59$$

$$X_4 = 55$$

$$X_5 = 63$$

$$X_6 = 23$$

Probar la Hipótesis H_0 : el dado no está cargado, con un nivel de significancia de .05.

Debido a que la prueba χ^2 se define como:

$$\chi^2_{k-1} = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j}$$

$$\chi^2_5 = \sum_{j=1}^6 \frac{(N_j - n/6)^2}{n/6}$$

Bajo la Hipótesis H_0 , las frecuencias esperadas np_j deben ser iguales a $300 \times \frac{1}{6} = 50$. Entonces:

$$\chi^2_5 = \frac{(57-50)^2}{50} + \frac{(43-50)^2}{50} + \frac{(59-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(63-50)^2}{50} + \frac{(23-50)^2}{50}$$

$$\chi^2_5 = 22.04$$

Debido a que de acuerdo a las Tablas $\chi^2_{5,.95} = 11.07$ se rechaza la Hipótesis H_0 al nivel 0.05.

b) Los datos de la siguiente Tabla representan el número de accidentes automovilísticos observados en un lapso de 100 días en un crucero.

Número de Accidentes	0	1	2	3	4	≥ 5
Frecuencia	16	24	22	21	11	6

Si se supone una distribución Poisson con $\lambda=2$, se tendrían las siguientes probabilidades esperadas:

X	0	1	2	3	4	≥ 5
p_j	.1353	.2707	.2707	.1804	.0902	.0527
np_j	13.50	27.10	27.10	18.04	9.02	5.30

La prueba estadística será:

$$\chi^2_5 = \frac{(16-13.5)^2}{13.5} + \frac{(24-27.1)^2}{27.1} + \frac{(22-27.1)^2}{27.1} + \frac{(21-18)^2}{18} + \frac{(11-9)^2}{9} + \frac{(6-5.3)^2}{5.3}$$

$$\chi^2_5 = 2.81$$

Debido a que de acuerdo a las Tablas $\chi^2_{5,.95} = 11.07$, se acepta la Hipótesis nula de una distribución Poisson, con media 2. Si el parámetro (media) de la distribución Poisson no hubiera sido dado, podría haber sido estimado a partir de los datos. Sin embargo, el Estadístico chi-cuadrada hubiera tenido un grado de libertad menos, es decir, $\chi^2_{4,.95}$.

II.7.2. Pruebas Kolmogorov-Smirnov

Como se observó, la prueba chi-cuadrada puede ser considerada como la comparación formal entre los histogramas o gráficas lineales de los datos con las funciones de probabilidad o de densidad de la distribución hipotética a ajustar. En el caso continuo, esta prueba presenta dificultades en la elección de los intervalos a usar.

La prueba Kolmogorov-Smirnov (K-S) compara una función de *distribución empírica* con la función de *distribución* \hat{F} de la distribución hipotética. La prueba K-S no requiere de la agrupación de los datos por lo que no se pierde información y se elimina el problema de la selección de intervalos. Además, esta prueba es válida para cualquier tamaño de muestra n , mientras que la prueba chi-cuadrada solamente es válida en el sentido asintótico (n grande).

Sin embargo, la prueba K-S también presenta sus inconvenientes:

1. En el caso de datos discretos, los valores críticos requeridos deben ser calculados a partir de un conjunto de fórmulas muy complejo. Se estima que la computadora únicamente puede realizar estos cálculos cuando $n \leq 50$.

2. La forma *original* de la prueba K-S solamente es válida si la distribución es continua y se conocen todos los parámetros de la distribución hipotética, es decir, los parámetros no pueden ser estimados a partir de los datos. Sin embargo, en la práctica, la prueba K-S ha sido directamente aplicada a distribuciones continuas y discretas con parámetros estimados, lo que produce una prueba conservadora (la probabilidad de que ocurra un error Tipo I es menor que la especificada), aunque en algunos casos estas pruebas no están basadas en fundamentos teóricos.

A pesar de que no siempre existe un fundamento teórico, para efectos de esta tesis se utilizará la Tabla II.22 para la Prueba Kolmogorov-Smirnoff aplicable a distribuciones continuas únicamente.

Para definir el estadístico K-S, es necesario definir a partir de los datos X_1, X_2, \dots, X_n la función de distribución empírica $F_n(x)$ como:

$$F_n(x) = \frac{\text{número de } X_i\text{'s } \leq x}{n} \quad x \in \mathfrak{R}$$

Entonces, $F_n(x)$ es una función escalera continua por la derecha tal que $F_n(X_{(i)}) = i/n$ para $i=1, 2, \dots, n$. Si $\hat{F}(x)$ es la función de distribución hipotética a ajustar, una evaluación natural de la bondad del ajuste es algún tipo de medida de la proximidad entre F_n y \hat{F} . El estadístico de prueba D_n es la distancia (vertical) más larga entre $F_n(x)$ y $\hat{F}(x) \forall x$, y se define formalmente por:

$$D_n = \sup_x \{ | F_n(x) - \hat{F}(x) | \}$$

[El "sup" de un conjunto de números A es el menor valor que es mayor o igual que todos los miembros del conjunto A. El "sup" se utiliza en lugar del "máximo" ya que en ciertos casos el máximo no se encuentra bien definido. Por ejemplo, si $A = (0,1)$, no existe un máximo pero el "sup" será 1.]

D_n se obtiene al calcular:

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - \hat{F}(X_{(i)}) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ \hat{F}(X_{(i)}) - \frac{i-1}{n} \right\}$$

y sea:

$$D_n = \max \{ D_n^+, D_n^- \}$$

En esta prueba se rechaza la Hipótesis nula H_0 , si $D_n > k$, donde k es determinada en la Tabla II.22.

A continuación se verá un caso práctico, para entender el funcionamiento de la prueba:

Los Tiempos de llegada de un Tren entre las 8:00 y 8:15 A.M. son registrados como: 8:01, 8:03, 8:07, 8:04, 8:11 y 8:14. Se va a probar la Hipótesis nula H_0 : La Distribución de Tiempos de llegada del Tren es Uniforme (0,15) con un nivel de significancia de 0.05.

Entonces:

$$\hat{F}(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{15} & 0 \leq x \leq 15 \\ 1 & x > 15 \end{cases}$$

(ver Gráfica II.21).

La Función de Distribución obtenida a partir de los datos tiene saltos de tamaño $\frac{1}{6}$ ($n=6$) en los minutos 1, 3, 4, 7, 11 y 14. Tenemos que:

$$D_6 = \max \{ D_6^+, D_6^- \}$$

Tabla II.22 Valores críticos para la Prueba Kolmogorov-Smirnov

Tamaño de la Muestra (n)	Nivel de Significancia				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.829
4	.494	.525	.564	.624	.734
5	.446	.474	.510	.563	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.409	.486
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.391
17	.250	.266	.286	.318	.380
18	.244	.259	.278	.309	.370
19	.237	.252	.272	.301	.361
20	.231	.246	.264	.294	.352
25	.21	.22	.24	.264	.32
30	.19	.20	.22	.242	.29
35	.18	.19	.21	.23	.27
Fórmula Asintótica	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Gráfica 11.21

Funciones
Acumulativas
de Distribución

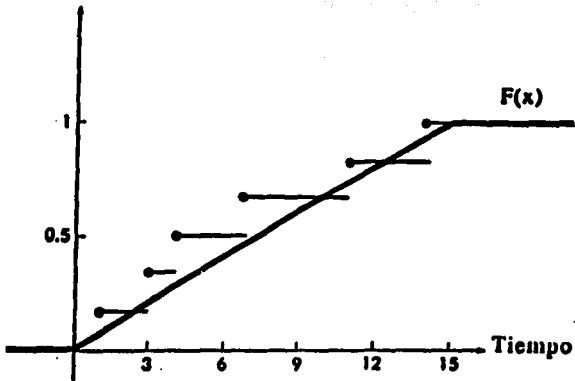
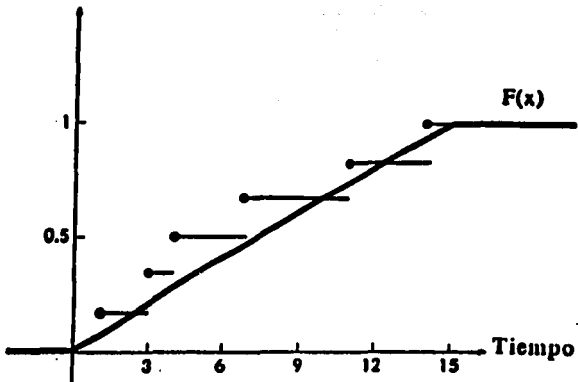


Tabla II.22 Valores críticos para la Prueba Kolmogorov-Smirnov

Tamaño de la Muestra (n)	Nivel de Significancia				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.829
4	.494	.525	.564	.624	.734
5	.448	.474	.510	.563	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.409	.486
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.391
17	.250	.266	.286	.318	.380
18	.244	.259	.278	.309	.370
19	.237	.252	.272	.301	.361
20	.231	.246	.264	.294	.352
25	.21	.22	.24	.264	.32
30	.19	.20	.22	.242	.29
35	.18	.19	.21	.23	.27
Fórmula Asintótica	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Gráfica 11.21

**Funciones
Acumulativas
de Distribución**



En los puntos de salto de la Gráfica, los valores D_6 están dados por la siguiente Tabla, donde $D_n^+ [F_n(x)]$ proporciona el valor de $F_n(x)$ después de efectuarse el salto y $D_n^- [F_n(x)]$ proporciona el valor de $F_n(x)$ antes de efectuarse el salto:

X	D_n^+ $F_n(x+) - \hat{F}(x)$	D_n^- $\hat{F}(x) - F_n(x-)$
1	$\frac{1}{6} - \frac{1}{15} = \frac{3}{30}$	$\frac{1}{15} - 0 = \frac{1}{15}$
3	$\frac{2}{6} - \frac{3}{15} = \frac{4}{30}$	$\frac{3}{15} - \frac{1}{6} = \frac{1}{30}$
4	$\frac{3}{6} - \frac{4}{15} = \frac{7}{30}$ *	$\frac{4}{15} - \frac{2}{6} = -\frac{2}{30}$
7	$\frac{4}{6} - \frac{7}{15} = \frac{6}{30}$	$\frac{7}{15} - \frac{3}{6} = -\frac{1}{30}$
11	$\frac{5}{6} - \frac{11}{15} = \frac{3}{30}$	$\frac{11}{15} - \frac{4}{6} = \frac{2}{30}$
14	$1 - \frac{14}{15} = \frac{2}{30}$	$\frac{14}{15} - \frac{5}{6} = \frac{3}{30}$

Entonces,

$$D_6 = \frac{7}{30}$$

Debido a que $k = D_{6,.90} = .47$ (de acuerdo a la Tabla II.22) se acepta la Hipótesis nula H_0 .

El cálculo directo de D_n^+ y D_n^- requiere que los datos sean ordenados para así obtener las $X_{(i)}$'s. Para valores moderados de n , el proceso de ordenación puede ser realizado rápidamente por medio de métodos sencillos; sin embargo, para n elevada aún los métodos sofisticados resultan lentos.

CAPITULO III

METODOLOGIA

III. METODOLOGIA

Como se indicó anteriormente, la metodología aquí descrita está enfocada a generar modelos de simulación orientados a satisfacer necesidades en el diseño y rediseño de la distribución de plantas de manufactura de nuestro país.

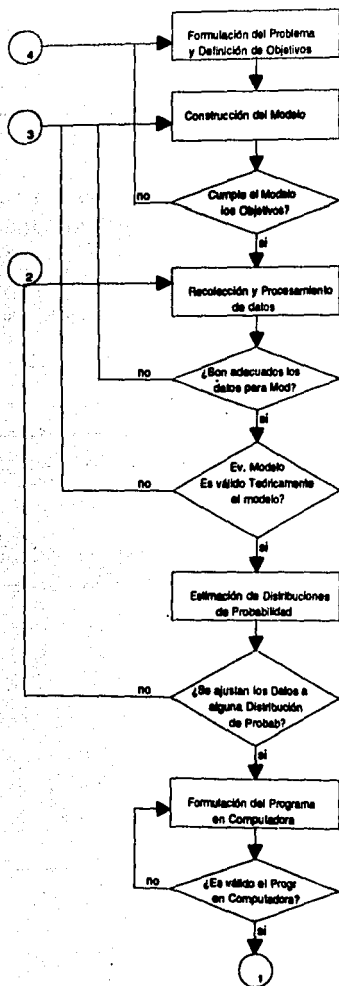
Lo más importante para la creación de un modelo de simulación exitoso es el compromiso existente entre el modelador y el cliente para trabajar en equipo. El modelador debe aclarar que él es sólo un miembro del equipo y que el cliente es una parte muy importante dentro del mismo, especialmente en lo que respecta a la definición de los objetivos y a las pruebas de validez del modelo. Un aspecto relevante del trabajo de un modelador es mantener una total comunicación con el cliente (usuario).

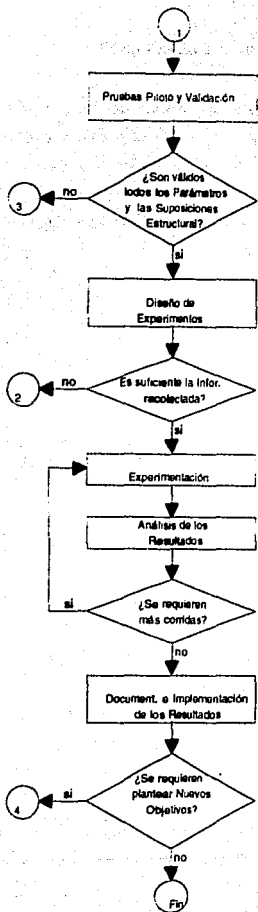
La gerencia de planta debe respaldar totalmente el proyecto. Además del tiempo y esfuerzo del modelador, un modelo requiere de los conocimientos del personal de ingeniería industrial, de los supervisores y de los mismos trabajadores de la planta. Ellos proporcionan los supuestos necesarios sobre el funcionamiento del sistema real o futuro y validan el modelo.

Para llevar a cabo un experimento de simulación se sugiere la siguiente metodología:

1. Formulación del problema y definición de objetivos.
2. Construcción del modelo.
3. Recolección y procesamiento de datos.
4. Evaluación del modelo y estimación de las Distribuciones de Probabilidad.
5. Formulación de un programa para la computadora.
6. Pruebas piloto y validación.
7. Diseño de experimentos.
8. Experimentación.
9. Análisis de los resultados.
10. Documentación e implementación de los resultados.

DIAGRAMA DE FLUJO DE LA SIMULACION





III.1. FORMULACION DEL PROBLEMA Y PLANEACION DEL ESTUDIO

El primer objetivo del modelador debe ser aprender lo más posible acerca del funcionamiento del sistema real o futuro. Este esfuerzo requiere de un diálogo continuo entre el modelador y el personal de la planta encargado del proyecto de simulación.

Como en otras áreas de la investigación científica, el estudio de simulación debe comenzar con la formulación del problema en la que es muy importante definir:

a) Objetivos por alcanzar. Pueden ser de la forma:

- Preguntas por contestar
- Hipótesis por probar
- Efectos por estimar

b) Conjunto de criterios que evalúan el grado de satisfacción al que deba sujetarse el experimento a fin de que los objetivos sean cumplidos.

Es necesario observar que la formulación del problema es un proceso secuencial que evoluciona con el tiempo; mientras mejor se conozca el problema, los objetivos y criterios de evaluación serán más precisos y confiables. Es probable que a lo largo del tiempo, el problema tenga que ser analizado desde diversos puntos de vista y, por consiguiente, modificado. Siempre deben tomarse en cuenta el tiempo estimado del proyecto, los costos, los recursos de cómputo, etc. Una vez realizadas las primeras corridas es conveniente volver a analizar la formulación del problema.

III.2. CONSTRUCCION DEL MODELO

Es necesario entender claramente las estructuras y reglas de operación del sistema real para así extraer la esencia del mismo.

Se requiere definir las variables del modelo y sus relaciones. Como ya se mencionó en el capítulo I, las variables son de tres tipos:

a) *Variables endógenas o de salida.*

Se determinan al establecerse los objetivos del estudio.

b) *Variables exógenas o de entrada.*

Definen el sistema. Es muy importante determinar el número apropiado de variables exógenas ya que la existencia de pocas variables puede ocasionar modelos inválidos, en cambio, el exceso de ellas puede imposibilitar la simulación en computadora y complicar innecesariamente el modelo.

c) *Variables de estado.*

Llevar la información sobre el estado del sistema de un periodo a otro.

Sólo aquellas variables que influyan en la toma de decisiones deben ser incluidas. Al efectuarse las primeras corridas, es posible determinar las variables exógenas que influyen considerablemente en el sistema para ser analizadas con más detalle, ignorándose aquellas variables cuya influencia es casi nula.

La construcción del modelo puede enfrentarse a los siguientes problemas:

- Imposibilidad de cuantificar ciertas variables.
- El número de variables exceda la capacidad de los recursos de cómputo.
- Se desconozcan ciertas variables exógenas significativas al sistema.
- Se ignoren algunas relaciones entre las variables.
- Sea imposible definir en términos del modelo las relaciones entre las variables.

La información recolectada puede ocasionar que el modelo sufra modificaciones debido a:

- No exista información sobre ciertas variables y se tengan que asumir ciertos supuestos sobre el modelo.
- De acuerdo a la naturaleza de la información existente, la estructura del modelo deba ser modificada.

III.3. RECOLECCION Y PROCESAMIENTO DE DATOS

Este paso es uno de los más importantes ya que la información recolectada se utiliza en la definición detallada del modelo de la siguiente manera:

- Estimación de las distribuciones de probabilidad de las variables.
- Generación de nuevas hipótesis que afectan el modelo. Como ya se mencionó anteriormente, la construcción del modelo es recursiva y la obtención de información puede mostrar la necesidad de rediseñar el modelo.
- Validación del modelo.

El proceso de recolección y procesamiento de datos se desglosa en:

- a) Recolección y almacenamiento
- b) Validación
- c) Conversión
- d) Manipulación

III.3.1 Recolección y almacenamiento

La recolección de datos es uno de los procesos más importantes en un estudio de simulación; sin embargo, por lo general es sumamente costosa y laboriosa pues comprende la identificación, revisión, edición y codificación de la información.

Por datos se entiende tiempos de proceso de las partes, tiempos de transporte y espera del mismo, tiempos de descompostura y reparación de maquinaria, etc. El modelo define por sí mismo los datos que requiere, pero la falta de disponibilidad de los mismos puede llegar a restringir el modelo. Por otro lado, los datos de "salida" del sistema también son necesarios para la validación de los resultados arrojados por el modelo.

Las fuentes de información más comunes son:

- Estudios de tiempos y movimientos
- Registros históricos (incluyen archivos de contabilidad, sistemas, etc)
- Información proporcionada por los proveedores
- La mejor estimación del ingeniero de manufactura
- La mejor estimación del modelador

En México, la mayoría de las empresas no poseen la información requerida para llevar a cabo un estudio de simulación. A pesar de que muchas plantas de manufactura generan una gran cantidad de información, ésta es desechada casi inmediatamente, se encuentra en un formato incorrecto o simplemente contiene datos equivocados. Frecuentemente, los tiempos de proceso son reportados como un promedio sobre un período largo de tiempo (o sobre diversos productos), mientras que la simulación requiere de tiempos de proceso individuales. Es necesario crear conciencia sobre la importancia del conjunto de valores que toman en una muestra las variables principales, para así poder conocer su comportamiento aleatorio real y no trabajar únicamente con los promedios.

Si se realiza un estudio de simulación como consultor externo, es conveniente efectuar un análisis previo para determinar el tipo, cantidad y calidad de información que posee la empresa. En caso de que ésta no se encuentre completa, se recomienda solicitar al cliente la información deseada y programar las siguientes fases del estudio una vez terminada la recopilación de la información faltante.

Si son diseñados adecuadamente, los estudios de tiempos y movimientos proporcionarán información bastante exacta. Es conveniente observar que los estudios de tiempos y movimientos no siempre son realistas ya que durante el estudio el obrero se encuentra bajo condiciones que no son las normales. De disponerse de suficiente tiempo se puede solicitar al obrero o supervisor que durante un determinado periodo proporcione la información requerida, como:

- Parte procesada.
- Tamaño del lote.
- Tiempo real de proceso (sin incluir descansos, tiempos de descompostura, reparación y preparación, etc.).

para así respaldar el estudio de tiempos y movimientos.

Sin embargo, para poder llevar a cabo lo anterior probablemente sea necesario capacitar al personal, lo cual no siempre es factible.

Los registros históricos (tales como reportes de producción y de paros de las máquinas) deben de ser analizados cuidadosamente para verificar su exactitud e importancia. Por ejemplo, no se debe asumir que el tiempo en que una máquina estuvo desocupada y que empezó con un paro por falla, se debió en su totalidad al tiempo empleado en repararla. Probablemente, en el momento en que ocurrió la falla, el calendario de producción fue modificado y la máquina no fue requerida sino hasta la siguiente semana, por lo que la reparación fue pospuesta. Entonces, el tiempo de reparación no será igual al tiempo en que la máquina permaneció bloqueada. Debido a que la mayoría de los registros históricos se encuentran incompletos, la información debe de ser utilizada cuidadosamente y sólo después de haber discutido la situación con el personal que recolectó la información.

En el caso de equipo nuevo, el proveedor puede proporcionar cierta información la cual resulta útil inicialmente; pero, posteriormente debe ser modificada de acuerdo a la experiencia y conocimientos adquiridos por el personal de la planta y el modelador.

Cuando no se dispone de información, los estimados proporcionados por el personal de la planta capacitado pueden ser utilizados hasta que se obtenga más información. Asimismo, el modelador puede tener experiencia suficiente para "estimar" los parámetros y variables de un sistema. Por ejemplo, en el caso del tiempo transcurrido entre fallas, es razonable asumir una distribución exponencial si el único estimado que se posee es la media del tiempo entre las fallas. De igual manera, si la información disponible es limitada, el modelador puede asumir una distribución uniforme o triangular.

Para almacenar la información se recomienda el uso de hojas electrónicas de cálculo que permiten ordenar, formatear, etc. las veces que se requiera. Además, pueden ser convertidas a archivos planos y utilizados en otros paquetes de cómputo.

III.3.2 Validación

Debido a la falta de conocimientos sobre los estudios de simulación, generalmente la información es proporcionada subjetivamente por el personal de la planta que no muestra compromiso ni interés por el estudio. Es conveniente que el personal firme de haber brindado la información como tal. Hay que hacer notar que el hecho de solicitar el acuerdo por escrito sobre la información proporcionada, propicia que ésta vuelva a ser revisada.

A causa de la falta de capacitación del personal de planta en México, se recomienda que la información a validar sea presentada al personal en un orden muy semejante al que ellos están acostumbrados a manejar, para así evitar confusiones y pérdida de tiempo.

Por otro lado, el personal debe firmar un documento en el que notifique que se ha entregado toda la información necesaria para llevar a cabo el estudio de simulación. En el caso de entregas posteriores debe notificarse la fecha y el contenido.

Es necesario realizar una junta con los usuarios en la que se planteen los supuestos que se van a utilizar para el modelo. El usuario debe de estar de acuerdo con los supuestos y firmar la validación. A medida que el estudio vaya progresando, surgirá la necesidad de asumir nuevos supuestos por lo que se requerirá realizar juntas periódicas con los usuarios para la validación de los nuevos supuestos.

III.3.3. Conversión

En muchas ocasiones la información que se recolectó y validó no se encuentra en el formato adecuado para el estudio, por lo que será necesario convertirla a un nuevo formato.

III.3.4 Manipulación

Puede resultar necesario el realizar con los datos operaciones aritméticas y lógicas, así como clasificar, cotejar, ordenar, intercalar, etc.

En este paso resulta evidente la ventaja del uso de hoja electrónica de cálculo.

En algunos casos, los valores exactos de los datos de entrada tienen poco efecto en los resultados de la simulación. La sensibilidad de los resultados que arroje el estudio ante cambios en los datos de entrada puede ser evaluada al correr varias veces el proceso con diferentes valores de los datos de entrada y analizar los resultados (diseño de experimentos); de esta manera, se usa la simulación para determinar el tiempo y dinero requeridos para refinar la información inicial.

III.4. EVALUACION DEL MODELO Y ESTIMACION DE LAS DISTRIBUCIONES DE PROBABILIDAD

Este paso sólo representa la primer etapa en la prueba de un modelo de simulación . Esta evaluación es anterior a las corridas en computadora por lo que el interés se centra en las pruebas de validez de los datos de entrada.

Durante esta etapa es conveniente formular las siguientes preguntas:

- ¿ Se incluyen variables exógenas no pertinentes, i.e., que contribuyan muy poco en la predicción del comportamiento de las variables endógenas del sistema ?
- ¿ Se omitieron variables exógenas que pueden llegar a afectar el comportamiento de las variables endógenas ?
- ¿ Se formularon de manera correcta las relaciones entre las variables endógenas y exógenas del sistema ?

Estas preguntas son evaluadas con el personal de la planta que se encuentra íntimamente relacionado con el funcionamiento del sistema (por ejemplo: operadores de las máquinas, ingenieros industriales, etc.). Así mismo, los modeladores deben interactuar constantemente con las personas encargadas de la toma de decisiones para así considerar cambios y mejoras al modelo. De esta manera cuando se entreguen los resultados al cliente, se evitará que se detecten características del sistema que no habían sido considerados por el modelo.

Es recomendable realizar evaluaciones periódicas con el cliente. El número adecuado de estas evaluaciones dependerá de la magnitud del proyecto y de la cantidad de cambios que vayan surgiendo. Es necesario que estas evaluaciones se efectúen con diferentes niveles de la administración, para que tanto las gerencias como los niveles inferiores tengan conocimiento de lo que está ocurriendo y se mantenga un alto nivel de compromiso por parte de la organización.

Por otro lado, una vez que se recolectaron y validaron los datos, es necesario encontrar una Distribución de Probabilidad a la que se ajusten, para así poder trabajar estadísticamente con el modelo.

Existen diversas técnicas para encontrar las Distribuciones de Probabilidad de los datos. Por motivos prácticos se recomienda la siguiente:

1. Almacenar los datos a analizar en una hoja electrónica de cálculo. De esta manera es posible cambiar el formato, depurar y ordenar la información cuantas veces sea necesario.
2. Trasladar la información al paquete estadístico que se utilice (por ejemplo el paquete StatGraf posee una utilidad que permite importar directamente hojas de cálculo de Lotus).
3. Probar diferentes Distribuciones de Probabilidad y aplicarles las pruebas de χ^2 y de Kolmogorov-Smirnov. (Ver capítulo II). Generalmente el Paquete Estadístico se encarga de realizar estas pruebas. Para efectos de este trabajo se utilizó el Paquete Estadístico StatGraf que realiza tanto la prueba χ^2 como la prueba de Kolmogorov-Smirnov. Además para realizar estas pruebas, calcula la media y la varianza a partir de los datos utilizando estimadores de máxima verosimilitud.
4. Graficar los histogramas de los datos y las Distribuciones de Probabilidad.
5. De acuerdo a la gráfica que más se aproxime y a los grados de significancia (que se acerquen a 1) de las pruebas, se elige la Distribución de Probabilidad más adecuada.

La teoría parece sencilla, sin embargo en la práctica se pueden presentar los siguientes problemas:

- No exista información suficiente para llevar a cabo las pruebas. En este caso lo más conveniente es utilizar una Distribución Triangular de acuerdo a criterio del analista.
- Los datos se encuentran muy dispersos por lo que ninguna Distribución de Probabilidad se ajusta. En este caso es necesario depurar los datos dispersos y volver a buscar una Distribución que se ajuste. La dificultad es que no se tiene una regla para depurar los datos y se desconoce si al efectuarse la depuración se llegue a obtener una Distribución. En muchas ocasiones las gráficas pueden ayudar a determinar los datos que requieren ser depurados.
- En ocasiones a pesar de que la información se encuentra dentro de un rango muy pequeño, no se tiene una Distribución de Probabilidad satisfactoria. Probablemente se pueda apreciar gráficamente la forma de la Distribución y puede resultar que tenga la forma de una parte de una Distribución de

Probabilidad (por ejemplo, que tenga la forma de la mitad de una Normal); en este caso, se recomienda utilizar la Distribución de Probabilidad a la que se ajusta en parte y acotarla por programa. Es lógico pensar que los grados de significancia se encontrarán lejos de 1, pero lo importante es que los datos se pueden aproximar a una parte de alguna Distribución. Por lo general, lo anterior se aplica cuando el estudio está restringido por el tiempo y por la falta de más información.

- Finalmente, si ninguna de las opciones anteriores resulta conveniente es posible utilizar una distribución empírica con los datos muestreados.

Recientemente, ha surgido nuevos paquetes estadísticos que permiten realizar con mayor agilidad las pruebas de bondad y ajuste, tal es el caso de UNIFIT-II. Sin embargo, el análisis de estos paquetes es motivo de otro trabajo.

III.5. FORMULACION DE UN PROGRAMA PARA LA COMPUTADORA

Actualmente, existen paquetes de simulación que permiten trasladar fácilmente el modelo a computadora. Sin embargo, estos paquetes están orientados a problemas específicos y no siempre son lo suficientemente flexibles para satisfacer el nivel de detalle acorde a los objetivos. Por otro lado, la formulación de un programa en computadora permitirá modelar todo el detalle que se requerirá para un problema específico.

La formulación de un programa requiere de las siguientes actividades:

1. Diagrama de flujo.

El objetivo de este diagrama es bosquejar la secuencia lógica de los eventos que realizará la computadora, al generar los tiempos planeados para las variables endógenas y de estado del modelo.

2. Lenguaje de la computadora.

Una vez terminado el diagrama de flujo se procede a escribir el código para la computadora que se utilizará para las corridas del experimento.

3. Verificación y búsqueda de errores del programa.

Se chequea que el programa funcione tal como se espera.

La industria del software enfocada a simulación está empezando a dirigir sus esfuerzos hacia la técnica de programación orientada a objetos que sin duda impactará fuertemente en las características del nuevo software que se produzca en el futuro.

Para este trabajo en particular se utilizó el paquete de simulación SimScript que cuenta con las siguientes características:

- i) Trabaja con 14 funciones de distribución para generar los parámetros del modelo.
- ii) Permite visualizar en pantalla las funciones de densidad de probabilidad y de distribución acumulada.
- iii) El usuario puede definir sus propias distribuciones al crear una tabla de datos, asignarle un nombre y utilizarla como distribución empírica.
- iv) Los números aleatorios son seleccionados por medio de un mecanismo denominado Arreglo de Números Aleatorios.

En la función de distribución se especifica el Arreglo a ser utilizado. Cada Arreglo genera una secuencia de números aleatorios estadísticamente independientes entre sí (la prueba es motivo de otro trabajo).

El arreglo requiere de un valor de inicio (entero y positivo) denominado *semilla*. SimScript proporciona por default 10 Arreglos de Números Aleatorios, siendo iniciado cada uno de estos por un valor semilla. Sin embargo, el usuario puede ampliar hasta a 100 el número de Arreglos al especificar semillas adicionales. Cada semilla nueva es agregada al final de la lista, por lo que el Arreglo correspondiente a la nueva semilla es colocado automáticamente en el siguiente número disponible; esto significa que los Arreglos deben ser proporcionados en forma secuencial.

v) SimScript permite utilizar Arreglos de Números negativos que producen las Variables Antitéticas: para cualquier muestra de números aleatorios R , $1-R$ es la muestra Antitética (Ver la sección III.9.2. sobre técnicas de reducción de varianza).

Esto resulta de utilidad en el análisis de sensibilidad. Por ejemplo, supóngase que un modelo se corre una vez, para la segunda corrida se cambia el arreglo de números aleatorios R por sus negativos $1-R$. Si los resultados de la segunda corrida son similares a los de la primera, significa que el modelo no es sensible a la selección de números aleatorios y viceversa.

III.6. PRUEBAS PILOTO Y VALIDACION

Teóricamente, es necesario verificar por medio de pruebas piloto que exista una correspondencia exacta entre el modelo de simulación y el sistema real. Es necesario validar: datos de entrada, elementos del modelo, subsistemas y puntos de interfase.

Sin embargo, en la práctica, el proceso de validar un modelo de simulación es sumamente difícil ya que incluye aspectos de tipo práctico, teórico, estadístico e inclusive filosófico. Aparentemente, sólo dos preguntas se consideran apropiadas para validar un modelo:

- ¿ Qué tan bien coinciden los valores simulados de las variables endógenas con datos históricos del sistema real (si es que se conocen) ?

- ¿ Qué tan exactas son las predicciones del comportamiento del sistema real hechas por el modelo de simulación para periodos de tiempo futuros ?

Una buena costumbre en la validación de un estudio de simulación consiste en tratar de modelar el sistema real en el estado en el que se encuentra y con datos determinísticos. De esta manera, la validación se lleva a cabo al verificar si los resultados arrojados por el modelo simulado son semejantes a los que el sistema arroja. Una vez validado el sistema actual, se puede proceder al análisis de sistemas alternativos.

III.7. DISEÑO DE EXPERIMENTOS

El Diseño de Experimentos permite elegir una configuración particular para el modelo a simular, de tal manera que se pueda obtener la información deseada con el menor número de corridas de simulación. Además, es posible determinar los parámetros y suposiciones estructurales que tienen mayor efecto sobre las medidas de rendimiento del sistema.

En la terminología del Diseño de Experimentos se tienen los siguientes conceptos:

1. **Factores.** Son los parámetros de entrada y las suposiciones estructurales.

(se incluyen las variables exógenas).

2. **Respuestas.** Son los resultados obtenidos.

(se incluyen las variables endógenas).

La decisión sobre qué parámetros de entrada y suposiciones estructurales deben considerarse como aspectos fijos del modelo y cuáles como aspectos experimentales, depende de los objetivos del estudio y no de la forma del modelo.

Los Factores pueden ser clasificados en:

a) *Cuantitativos.* Son aquellos que asumen valores numéricos.

b) *Cualitativos.* Representan las suposiciones estructurales que no son cuantificables fácilmente (i.e. distribuciones de probabilidad).

III.7.1. Diseños Factoriales 2^k

Si un modelo tiene solamente un factor, el Diseño de Experimentos es muy simple: se corre la simulación para diferentes valores (o niveles) del factor y se determina cómo influye éste sobre las respuestas del modelo. En el caso de factores cuantitativos, es conveniente el uso de una gráfica de respuesta que sea función de los niveles del factor.

Ahora, supóngase que se tienen k factores ($k \geq 2$) y se desea obtener una idea inicial de la manera en que los factores afectan a las respuestas, así como determinar el grado de interacción de los factores entre sí. Un método para medir el efecto de un factor particular, consiste en fijar los niveles de los otros $k-1$ factores y realizar varias simulaciones para diferentes niveles del factor de interés y observar cómo reacciona la respuesta a cambios en el factor. Sin embargo, esta técnica resulta muy ineficiente por el gran número de corridas que requiere, además no es posible medir la interacción entre los factores.

Entonces, es conveniente utilizar la técnica de Diseño Factorial 2^k . En esta técnica se eligen únicamente dos niveles para cada factor y se realizan corridas de simulación para cada una de las 2^k combinaciones posibles de factores-niveles, denominados *puntos de diseño*. Se asigna un signo "-" a un nivel del factor y un signo "+" al otro nivel del factor (por lo general, para factores cuantitativos se asigna el signo "-" al menor nivel). No existe una regla general para la elección de los niveles y se espera que el analista tenga una idea sobre los valores que deberían tomar estos.

Para mayor claridad, esta técnica puede ser representada en forma de tabla, por ejemplo se tiene la *matriz de diseño* para $k=3$:

Punto de Diseño	Factor 1	Factor 2	Factor 3	Respuesta
1	-	-	-	R_1
2	+	-	-	R_2
3	-	+	-	R_3
4	+	+	-	R_4
5	-	-	+	R_5
6	+	-	+	R_6
7	-	+	+	R_7
8	+	+	+	R_8

Esta matriz de diseño facilita el cálculo de los efectos de los factores sobre el modelo y de las interacciones entre ellos.

El *efecto principal* del factor j mide el cambio promedio en la respuesta al variar el factor j del nivel "-" al nivel "+", mientras que los demás factores permanecen fijos. Este promedio es calculado sobre todas las combinaciones de todos los niveles de los otros factores.

Por ejemplo, el efecto principal para el factor 1 será:

$$e_1 = ((R_2 - R_1) + (R_4 - R_3) + (R_6 - R_5) + (R_8 - R_7)) / 4$$

y una manera más fácil de expresar lo anterior:

$$e_1 = (-R_1 - R_3 - R_5 - R_7 + R_2 + R_4 + R_6 + R_8) / 4$$

en donde los signos son los mismos que los signos del factor 1 en la matriz de diseño. Entonces, para calcular cualquier e_j (efecto para el factor j) simplemente se aplican los signos del "factor j " a las R_i 's correspondientes, se suman y se dividen entre 2^{k-1} .

En la práctica, es posible que el efecto del factor j_1 dependa del nivel del factor j_2 , en cuyo caso se dice que j_1 y j_2 interactúan entre sí. Esta interacción se mide por el efecto de interacción de dos factores, $e_{j_1 j_2}$, y se define como la mitad de la diferencia entre el efecto promedio del factor j_1 cuando el factor j_2 se encuentra en el nivel "+" (y todos los otros factores son constantes), y el efecto promedio del factor j_1 cuando j_2 se encuentra en el nivel "-". $e_{j_1 j_2}$ también se denomina la interacción $j_1 \times j_2$.

Por ejemplo:

$$e_{12} = \frac{1}{2} \left[\frac{(R_4 - R_3) + (R_6 - R_7)}{2} - \frac{(R_2 - R_1) + (R_5 - R_8)}{2} \right]$$

Sin embargo, es más fácil expresar lo anterior de la siguiente manera:

$$e_{12} = \frac{R_1 - R_2 + R_3 - R_4 - R_5 + R_6 - R_7 + R_8}{4}$$

en donde los signos son calculados al multiplicar los signos de las columnas 1 y 2 y aplicarlos a las R_i 's correspondientes. Esta fórmula es válida para interacciones de n -factores.

Debido a que los R_i son variables aleatorias, los efectos resultantes también lo son. Para determinar si un efecto es "real" o si se origina por fluctuaciones aleatorias, es necesario estimar su varianza. Una técnica muy común es realizar n corridas y obtener n valores independientes para los efectos, posteriormente se calculan los intervalos de confianza para cada uno de los efectos por medio de la fórmula:

$$\bar{X}(n) \pm t_{n-1, \alpha/2} \sqrt{\frac{S^2(n)}{n}}$$

Si para un efecto particular el intervalo de confianza no contiene cero, se concluye que el efecto es real; de lo contrario no se posee evidencia estadística para mostrar que se encuentra presente. Como siempre, n grande reduce la amplitud del intervalo.

Obsérvese que para reducir a la mitad la amplitud del intervalo es necesario incrementar 3 veces el tamaño muestral (número de corridas).

Los Diseños Factoriales presentan el problema de que si existen interacciones fuertes, no es posible utilizar los efectos principales por sí solos para interpolar o extrapolar los valores de respuesta para otros niveles.

III.7.2. Tratando con muchos Factores

Si el número de factores k es muy grande, el Diseño Factorial 2^k se vuelve muy difícil de manejar por lo que resulta necesario utilizar otras técnicas. En esta sección se describen diversos procedimientos a aplicar en el caso de que el número de factores sea muy elevado; por supuesto, siempre se tiene la esperanza de que después de unas cuantas pruebas se puedan ignorar los factores no clave y solamente prestar atención a aquellos que influyen notoriamente en los resultados.

a) Diseños Factoriales Fraccionados 2^{k-p}

Este diseño proporciona un método para estimar solamente los efectos principales y posiblemente los dos primeros niveles de interacciones. Se elige un subconjunto (de tamaño 2^{k-p}) de los 2^k puntos de diseño originales. Debido a que solamente se corren $1/2^p$ de las 2^k combinaciones existentes, si $p=1$ se denomina "fracción media", si $p=3$ "fracción octava", etc. Desde el punto de vista de rendimiento del computador se espera que p sea grande; sin embargo, si p es grande se perderá información.

La elección de las 2^{k-p} combinaciones a partir de las 2^k no es tarea fácil. A continuación se presenta una "receta de cocina" que en algunas ocasiones puede ser utilizada eficazmente en la selección de las combinaciones. Sin embargo, antes de continuar es necesario aclarar algunos conceptos.

Confusión en los Diseños Factoriales Fraccionados

Significa que para diferentes efectos se tendrá la misma expresión algebraica. Por ejemplo, en la Fracción media 2^{4-1} , la fórmula para el efecto principal e_4 es idéntica a la fórmula del efecto triple -interacción e_{123} (ver Tabla III.1). Es decir, el efecto principal del factor 4 se *confunde* con el efecto triple -interacción para los factores 1, 2 y 3.

Lo que en realidad significa es que la fórmula común para e_4 y e_{123} es un estimador incesgado para $E(e_4) + E(e_{123})$. Ahora, si se asume que $e_{123}=0$ (esto es posible ya que mientras mayor sea el número de factores en la interacción, más pequeño será su valor), entonces $E(e_4)=e_4$. Sin embargo, esto no siempre resulta tan sencillo; por ejemplo, si las fórmulas de e_{12} y e_{34} son idénticas, resulta difícil determinar que interacción debe ser igualada a cero.

Para cuantificar la severidad de la confusión se tiene el concepto de *resolución* que garantiza que dos efectos no sean confundidos si la suma de sus "caminos" (números de las interacciones) es estrictamente menor que la resolución. Por ejemplo, si el diseño tiene resolución IV, los efectos principales no son confundidos con las doble-interacciones ($1+2<4$); sin embargo, las doble-interacciones pueden ser confundidas entre sí ($2+2=4$). Mientras mayor sea la resolución, la confusión será menor.

Una vez identificado el número de factores k y la resolución deseada, se sigue el siguiente procedimiento:

- i) Se identifica p y los 2^{k-p} renglones que serán utilizados.
- ii) Se escribe la Matriz Factorial para los factores 1,2,...,k-p, y las p columnas restantes se calculan al multiplicar algunas de las primeras k-p columnas de acuerdo a ciertas reglas establecidas por la Tabla III.1.
- iii) Se calcula el efecto principal para el factor j al aplicar los signos de la columna j a los números correspondientes de la columna de respuesta R_j 's, se suman y se dividen entre 2^{k-p-1} . Las interacciones son calculadas de manera similar: se multiplican los signos de las columnas de los factores afectados, se aplican a los números correspondientes de la columna de respuesta R_j 's, se suman y se dividen entre 2^{k-p-1} .

TABLA III.1

Reglas para construcción de Dígitos Factoriales Fraccionados

k	Resolución		
	III	IV	V
3	2_{III}^{3-1} 3 = ± 12		
4		2_{IV}^{4-1} 4 = ± 123	
5	2_{III}^{5-2} 4 = ± 12 5 = ± 13		2_V^{5-1} 5 = ± 1234
6	2_{III}^{6-3} 4 = ± 12 5 = ± 13 6 = ± 23	2_{IV}^{6-2} 5 = ± 123 6 = ± 234	
7	2_{III}^{7-4} 4 = ± 12 5 = ± 13 6 = ± 23 7 = ± 123	2_{IV}^{7-3} 5 = ± 123 6 = ± 234 7 = ± 134	
8		2_{IV}^{8-4} 5 = ± 234 6 = ± 134 7 = ± 123 8 = ± 124	2_V^{8-2} 7 = ± 1234 8 = ± 1256

9

 2_{III}^{9-5}

5 = ± 123

6 = ± 234

7 = ± 134

8 = ± 124

9 = ± 1234

 2_{IV}^{9-1}

6 = ± 2345

7 = ± 1345

8 = ± 1245

9 = ± 1235

10

 2_{III}^{10-6}

5 = ± 123

6 = ± 234

7 = ± 134

8 = ± 124

9 = ± 1234

(10) = ± 12

 2_{IV}^{10-5}

6 = ± 1234

7 = ± 1235

8 = ± 1245

9 = ± 1345

(10) = ± 2345

 2_V^{10-3}

8 = ± 1237

9 = ± 2345

(10) = ± 1346

11

 2_{III}^{11-7}

5 = ± 123

6 = ± 234

7 = ± 134

8 = ± 124

9 = ± 1234

(10) = ± 12

(11) = ± 13

 2_{IV}^{11-6}

6 = ± 123

7 = ± 234

8 = ± 345

9 = ± 134

(10) = ± 145

(11) = ± 245

 2_V^{11-4}

8 = ± 1237

9 = ± 2345

(10) = ± 1346

(11) = ± 1234567

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

A pesar de sus limitaciones, el Diseño Factorial Fraccionado puede proporcionar información útil sin que se requiera de tantos cálculos computacionales.

b) Estrategias de Reducción de Factores

Existen ocasiones en que el número de factores es tan elevado que ni siquiera el uso de un Diseño Factorial Fraccionado es adecuado; en este caso es necesario utilizar un *diseño supersaturado*.

La construcción de un diseño supersaturado proviene de la observación de que cada columna de una Matriz de Diseño 2^k posee la mitad de signos "+" y la mitad de signos "-". Es posible construir Matrices de Diseño con esta propiedad de *balanceo* (se asignan signos de forma aleatoria a las columnas con la condición de que se utilice la misma cantidad de signos) y un número adecuado de puntos de diseño, entonces, los efectos serán calculados a partir de los signos de esta matriz. El inconveniente de los diseños supersaturados consiste en que los efectos resultantes también son aleatorios. Debido a lo anterior, se han desarrollado *diseños supersaturados sistemáticos* en los cuales los signos son asignados de tal manera que se minimizan las confusiones resultantes; sin embargo, estos diseños solamente se encuentran disponibles para ciertas combinaciones de k y un número determinado de puntos de diseño; pero, el estudio de lo anterior no es motivo de este trabajo.

Otro método de reducción de factores consiste en agrupar a los mismos de alguna manera (grupo de tiempos de servicio, grupo de tiempos de inspección, etc). Cada grupo se considera un factor, y los efectos serán calculados de acuerdo a cambios en los niveles de cada grupo. Algunos grupos aparecerán como importantes mientras que otros no lo harán; entonces, será necesario un proceso recursivo en el que se reagrupen los factores de diferentes maneras hasta que sea posible determinar los factores más importantes. Este método reduce el número de factores y el tiempo de proceso; sin embargo, debido a su naturaleza recursiva es imposible determinar con anticipación el número de corridas necesarias. Tampoco es obvia la manera de agrupar los factores; de hecho, un grupo puede contener factores importantes y aparecer como si no los tuviera (los factores pueden estar en direcciones opuestas y cancelarse).

Finalmente, se tiene un método que es aplicable en simulación dinámica: se oscilan los parámetros de entrada en el transcurso de la simulación. La oscilación de cada parámetro debe realizarse con una frecuencia diferente. Posteriormente, se examina el proceso de salida para observar que frecuencias oscilatorias en los parámetros de entrada pueden ser detectadas en la salida. Un factor

importante producirá una oscilación notoria, mientras que la oscilación de un factor poco importante no será aparente. Se recomienda el uso de gráficas para una eficaz detección de las oscilaciones.

III.8. EXPERIMENTACION

Consiste en la ejecución del modelo de simulación para la obtención de resultados.

La simulación funciona igual que en la realidad y si se toma una única muestra del sistema la información no resultará representativa. Es necesario correr el modelo el mayor número de veces que sea posible. Sin embargo, por limitaciones de recursos de cómputo y/o de tiempo, no siempre es posible realizar el número adecuado de corridas.

III.9. ANALISIS DE LOS RESULTADOS

III.9.1. Análisis de los datos de salida para un solo sistema

En muchos estudios de simulación se utiliza una gran cantidad de tiempo en la elaboración del modelo y programación del mismo, pero los resultados no son analizados adecuadamente. De hecho, se realiza una única corrida y se trata a los resultados obtenidos como las características únicas del sistema. Como se mencionó anteriormente, los resultados también son variables aleatorias y como tales pueden diferir considerablemente de la realidad.

Una simulación es un experimento muestral estadístico realizado por medio de una computadora. Si se busca que los resultados de una simulación tengan significado, es necesario utilizar técnicas estadísticas apropiadas para el análisis de experimentos de simulación. Las técnicas estadísticas clásicas basadas en observaciones i.i.d. no son aplicables directamente. De hecho, actualmente existen problemas de análisis de resultados para los que no existe una solución completamente aceptada, y los métodos existentes generalmente son difíciles de aplicar.

A continuación se describe la naturaleza aleatoria de los resultados de simulación.

Sea Y_1, Y_2, \dots un proceso estocástico de salida para una sola corrida de simulación.

Las Y_i 's son variables aleatorias que no son independientes ni idénticamente distribuidas. Entonces, la mayor parte de las fórmulas mencionadas en el capítulo anterior no pueden ser aplicadas directamente.

Sean $y_{11}, y_{12}, \dots, y_{1m}$ un conjunto de las variables aleatorias Y_1, Y_2, \dots, Y_m resultantes de una corrida de simulación de longitud m al usarse los números aleatorios u_{11}, u_{12}, \dots (el i -ésimo número aleatorio utilizado en la j -ésima corrida se denota por u_{ji}). Si se corre otra simulación con un conjunto diferente de números aleatorios u_{21}, u_{22}, \dots , se obtendrá un subconjunto diferente: $y_{21}, y_{22}, \dots, y_{2m}$ de las variables aleatorias Y_1, Y_2, \dots, Y_m .

Supóngase que se realizan n corridas independientes de longitud m cada una, y sean las observaciones resultantes:

$$\begin{array}{l} y_{11} \dots y_{1i} \dots y_{1m} \\ y_{21} \dots y_{2i} \dots y_{2m} \\ \vdots \quad \quad \quad \vdots \\ y_{n1} \dots y_{ni} \dots y_{nm} \end{array}$$

Es claro que las observaciones de una corrida particular (renglón) no son i.i.d. Sin embargo, nótese que $y_{1i}, y_{2i}, \dots, y_{ni}$ (de la i -ésima columna) son observaciones i.i.d. de la variable aleatoria Y_i , para $i=1, 2, \dots, m$. Esta independencia a lo largo de las corridas es la clave para los métodos de análisis de resultados descritos a continuación. Generalmente hablando, el objetivo de un análisis de resultados es utilizar las observaciones y_{ji} ($i=1, 2, \dots, m; j=1, 2, \dots, n$) para realizar inferencias acerca de las variables aleatorias Y_1, Y_2, \dots, Y_m .

El objetivo de esta sección es presentar los métodos estadísticos de análisis de resultados más sencillos que son relativamente fáciles de implementar y que se ha mostrado que son útiles en la práctica.

a) *Comportamiento Transitorio o de Estado-Lugar en un Proceso Estocástico*

Sea Y_1, Y_2, \dots un proceso estocástico de salida y sea $F_i(y) = P(Y_i \leq y)$ para $i=1, 2, \dots$ la Distribución Transitoria en el tiempo discreto i .

Si para todo y , $F_i \rightarrow F(y)$ cuando $i \rightarrow \infty$, entonces $F(y)$ es la *distribución estado-lugar* del proceso de salida Y_1, Y_2, \dots . En un sentido estricto, $F(y)$ solamente se obtiene cuando $i \rightarrow \infty$. Sin embargo, en la práctica existe un índice de tiempo finito $k+1$ a partir del cual se tiene aproximadamente el estado-lugar. Esto no significa que las variables aleatorias Y_{k+1}, Y_{k+2}, \dots , tomarán el mismo valor, sino que tienen aproximadamente la misma distribución.

b) *Tipos de Simulación*

Las simulaciones pueden ser de dos tipos:

i) *Terminal.*

Es aquella en la que existe un evento natural E que especifique la longitud de la corrida. El evento E es determinado antes de que se realicen las corridas y el momento en que ocurra será una variable aleatoria dentro de la simulación.

Por ejemplo, si se desea simular un banco que permanece abierto de las 9:00 a.m. a las 5:00 p.m. y se quiere estimar la calidad de servicio al cliente desde las 9 de la mañana hasta que haya sido atendido el último cliente que entró antes de las 5 de la tarde. Entonces, el evento E va a requerir que hayan transcurrido por lo menos 8 horas de simulación y que el sistema se encuentre vacío (es decir que no queden clientes en espera de ser atendidos). El tiempo en que ocurra el evento E es una variable aleatoria ya que se desconoce el momento en que el último cliente termine de ser atendido.

ii) *No Terminal.*

Es aquella simulación en la que no existe un evento natural E que especifique la longitud de la corrida.

Se dice que en una simulación no terminal, una medida del desempeño del sistema será un *parámetro de estado-lugar*, si es una característica de la distribución estado-lugar $F(Y)$ para algún proceso estocástico de salida Y_1, Y_2, \dots . Por ejemplo: $E(Y)$.

Sea Y_1, Y_2, \dots un proceso estocástico que no tiene distribución de estado-lugar y supóngase que se divide el tiempo en intervalos de la misma longitud denominados ciclos. Sea Y_i^c una variable aleatoria definida en el i -ésimo ciclo y asuma que Y_1^c, Y_2^c, \dots son comparables. Supóngase que el proceso Y_1^c, Y_2^c, \dots tiene una distribución de estado-lugar F^c y que $Y^c \sim F^c$. Entonces, una medida del rendimiento será un *parámetro cíclico de estado-lugar* si es una característica de Y^c .

En el caso de que en una simulación no terminal el proceso estocástico Y_1, Y_2, \dots no tenga una distribución estado-lugar ni una distribución cíclica estado-lugar, se requerirá de una cantidad fija de datos que describa la manera en la que los parámetros de entrada cambian a lo largo del tiempo. Esto proporcionará un evento terminal E y las técnicas de análisis para simulaciones terminales serán aplicables.

c) Análisis Estadístico para Simulaciones Terminales

Supóngase que se realizan n corridas independientes de una simulación terminal y que cada corrida es terminada por el evento E e iniciada con las "mismas" condiciones iniciales. Además, por simplificación, asúmase que solamente se tiene una medida del desempeño del sistema.

Sea X_j una variable aleatoria definida en la j -ésima corrida y se asume que las X_j son comparables entre sí; entonces, las X_j 's son variables aleatorias i.i.d. (por ejemplo, las X_j 's pueden ser el tiempo promedio de proceso, el número promedio de partes en almacén, etc)

Media

Sea $\bar{X}(n) = \sum_{i=1}^n \frac{X_i}{n}$ un estimador insesgado para la media μ

y la varianza muestral $S^2 = \sum_{i=1}^n \frac{[X_i - \bar{X}(n)]^2}{n-1}$

Entonces, un intervalo de confianza para μ al $100(1-\alpha)$ por ciento se calculará de la siguiente manera:

$$\bar{X}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}}$$

Se dice que este intervalo de confianza está basado en el *procedimiento de tamaño-muestral-fijo*.

Además, el valor promedio de la longitud-media del intervalo de confianza está dado por:

$$\pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}}$$

Al dividirse el promedio de longitud-media entre $\bar{X}(n)$ se obtiene una medida de precisión del intervalo de confianza.

La desventaja del procedimiento de tamaño-muestral-fijo consiste en que no se tiene control sobre la longitud-media del intervalo (o sobre la precisión de $X(n)$) ya que depende directamente de la $\text{Var}(X)$. A continuación se enunciarán diversas técnicas para poder estimar μ con un error específico.

Si el estimado \bar{X} es tal que:

$$|\bar{X} - \mu| = \beta \quad (\beta > 0)$$

se dice que \bar{X} tiene un error absoluto de β .

Si el estimado \bar{X} es tal que:

$$|\bar{X} - \mu| / |\mu| = \gamma$$

se dice que \bar{X} tiene un error relativo de γ o que el porcentaje de error en \bar{X} es de $100\gamma\%$.

1. Método.

Se construye un intervalo de confianza para μ basado en n corridas. Si se asume que $S^2(n)$ no varía al aumentar n , se requiere calcular el número necesario de corridas, $n_a^*(\beta)$, para que se obtenga un error absoluto β .

$$n_a^*(\beta) = \min \left\{ i \geq n: t_{i-1, 1-\alpha/2} \sqrt{\frac{s^2(n)}{i}} \leq \beta \right\}$$

es posible calcular $n_{\alpha}^*(\beta)$ iterativamente, al aumentar i en 1 hasta que se cumpla con:

$$t_{i-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{i}} \leq \beta$$

Si $n_{\alpha}^*(\beta) > n$ y se realizan $n_{\alpha}^*(\beta) - n$ corridas adicionales; entonces, el estimado \bar{X} basado en el total de corridas tendrá un error absoluto de β . Esta exactitud va a depender de la aproximación de $S^2(n)$ a $\text{Var}(X)$.

2. Método

Se realizan n corridas y se calculan $\bar{X}(n)$ y $S^2(n)$. Si se asume que estos estimados no variarán notablemente al aumentar n , el número de corridas necesario, $n_r^*(\gamma)$, para obtener un error relativo de γ será de

$$n_r^*(\gamma) = \left\{ i \geq n: \frac{t_{i-1, 1-\alpha/2} \sqrt{S^2(n)/i}}{\bar{X}(n)} \leq \gamma \right\}$$

donde $\gamma' = \gamma(\gamma+1)$ es el error relativo ajustado que se requiere para calcular el error relativo actual y.

Si $n_r^*(\gamma) > n$ y se realizan $n_r^*(\gamma) - n$ corridas adicionales; entonces, el estimado $\bar{X}(n)$ calculado sobre todas las corridas tendrá un error relativo de aproximadamente γ .

Este método presenta algunos problemas: Si $n_r^*(\gamma)$ es mayor que el número de corridas requerido, se habrán realizado corridas innecesarias. Por otro lado, si $n_r^*(\gamma)$ es muy pequeño, un estimado basado en él no será tan preciso como se piensa.

3. Método (Secuencial).

En este método las corridas se van agregando una a la vez.

Sea n_0 el número inicial de corridas y sea :

$$\delta(n, \alpha) = t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}} \text{ la longitud-media del intervalo de confianza.}$$

Sean X_1, X_2, \dots variables aleatorias i.i.d. provenientes de la simulación. Entonces, el algoritmo utilizado para este método será:

a) Realizar n_0 corridas e igualar $n = n_0$.

b) Calcular $\bar{X}(n)$, $S^2(n)$ y $\delta(n, \alpha)$ a partir de X_1, X_2, \dots ,

c) Si $\delta(n, \alpha)/|\bar{X}(n)| \leq \gamma'$, entonces $\bar{X}(n)$ será el estimador puntual de μ y se:

$$I(\alpha, \gamma) = [\bar{X}(n) - \delta(n, \alpha), \bar{X}(n) + \delta(n, \alpha)]$$

el intervalo de confianza para μ con un $100(1-\alpha)\%$ de grados de precisión.

Por otro lado, si $\delta(n, \alpha)/|\bar{X}(n)| > \gamma'$, se calcula $n = n + 1$, se realiza una corrida adicional y se regresa al paso b).

Para que el proceso secuencial funcione correctamente, es decir, que la cobertura del intervalo se acerque a $1-\alpha$ es necesario establecer $\gamma \leq 0.15$ y $n_0 \geq 10$.

Recomendaciones para el uso de los tres métodos:

- Para una prueba piloto es conveniente usar el procedimiento de tamaño-muestral-fijo. Sin embargo, si las X_j 's resultan altamente no normales y n es pequeño, la cobertura del intervalo será menor a lo deseado.

A partir de la prueba piloto es posible obtener un estimado de $n_{\alpha}^*(\beta)$ para un error absoluto β y un estimado $n_r^*(\gamma)$ para un error relativo γ .

- Si se desea un error relativo $\gamma \leq 0.15$ se recomienda el uso del procedimiento secuencial con $n_0 \geq 10$. Sin embargo, si se desea $\gamma > 0.15$ es conveniente aplicar sucesivamente el procedimiento de tamaño - muestral-fijo. Es decir, se estima $n_r^*(\gamma)$, se realizan $(n_r^*(\gamma)-n)/2$ simulaciones adicionales y se utiliza el tamaño-muestral-fijo para construir un intervalo de confianza basado en las $[n+n_r^*(\gamma)]/2$ corridas existentes. Si el error relativo es mayor que γ' , entonces $n_r^*(\gamma)$ puede ser reestimado en base al nuevo estimado de la variancia, y así sucesivamente hasta que el error relativo sea menor a γ' .

- Para construir un intervalo de confianza con un error absoluto pequeño, es conveniente aplicar sucesivamente el procedimiento de tamaño-muestral-fijo.

En todos los casos, se debe realizar un mínimo de 3 a 5 corridas. Si no es posible realizar esto, es mejor no llevar a cabo el estudio de simulación.

Existen ocasiones en las que si solamente se analizan las medias de los datos se puede obtener información errónea. Entonces, es necesario estimar otras medidas del desempeño del sistema.

Proporción

El estimador puntual de una proporción se estima de la siguiente manera:

$$E(X) = E\left(\sum_{i=1}^n \frac{I_i(n_1, n_2)}{n}\right)$$

$$\text{en donde } I_i(n_1, n_2) = \begin{cases} 1 & \text{si } n_1 < D_i < n_2 \\ 0 & \text{en otro caso} \end{cases}$$

Probabilidad

Sea la probabilidad $p = P(X \in B)$, la probabilidad de que X pertenezca a un subconjunto de los números reales B . Sea S el número de X_j 's que pertenece a B . Entonces, S tiene una distribución binomial con parámetros n y p , y un estimador insesgado de la probabilidad p estará dado por:

$$\hat{p} = \frac{S}{n}$$

Q-ésimo Cuantil X_q

Si $X(1), X(2), \dots$, son los estadísticos de orden correspondientes a las X_j 's de n corridas independientes; entonces, un estimador puntual de X_q será:

$$\hat{x}(q) = \begin{cases} x(nq) & \text{si } nq \text{ es entero} \\ 0 & \text{otro caso} \end{cases}$$

NOTA: En simulaciones terminales es muy importante la manera en que se eligen las condiciones iniciales del sistema.

d) Análisis Estadístico para Parámetros Estado-Lugar

Supóngase que Y es una variable aleatoria de estado-lugar con función de distribución F , y sea ϕ un parámetro de la distribución estado-lugar al ser una característica de Y tal como $E(Y)$, $P(Y \leq y)$, etc.

En la práctica es difícil estimar ϕ ya que la función de distribución Y_i ($i=1,2,\dots$) es diferente de F , debido a que no es posible elegir las condiciones iniciales de tal manera que Y_i sea representativa del "Comportamiento estado-lugar".

La técnica más común para resolver el problema de las condiciones iniciales es el *calentamiento del modelo o borrado de datos iniciales*. El objetivo de esta técnica es borrar algunas observaciones del inicio de la corrida y utilizar la información restante para estimar los parámetros de estado-lugar.

Por ejemplo, dadas las m observaciones Y_1, Y_2, \dots, Y_m , se sugiere utilizar la siguiente fórmula para estimar la media muestral:

$$\bar{Y}(m,l) = \sum_{i=l+1}^m \frac{Y_i}{m-l}$$

La elección de l y m no resulta sencilla. Si l y m son pequeños, entonces $E[\bar{Y}(m,l)]$ estará muy alejado de ν . Por otro lado, si l es más grande de lo requerido, $\bar{Y}(m,l)$ tendrá una varianza larga e innecesaria.

La técnica más simple para determinar l es el procedimiento gráfico de Welch en el que se determina un índice l tal que: $E(Y_i) \rightarrow \nu$ para $i > l$. En este algoritmo se busca el momento en que $E(Y_i)$ se estabiliza en ν . Este proceso consta de los siguientes pasos:

1) Se realizan $n \geq 5$ corridas de longitud m . Sea Y_{ji} la i -ésima observación de la j -ésima corrida ($j=1,2,\dots,n$; $i=1,2,\dots,m$).

2) Se calcula el proceso promedio:

$$\bar{Y}_i = \sum_{j=1}^n \frac{Y_{ji}}{n} \quad i=1,2,\dots,m$$

en donde resulta que el proceso promedio tiene la misma media que el proceso original, pero solamente $1/n$ de su varianza.

3) Para eliminar las oscilaciones de alta frecuencia en $\bar{Y}_1, \bar{Y}_2, \dots$ ocasionadas por las condiciones iniciales y dejar únicamente las oscilaciones de baja frecuencia que ocurren naturalmente a lo largo del tiempo, se define el promedio móvil $\bar{Y}_i(w)$ (donde $w \leq [m/2]$) de la siguiente manera:

$$\bar{Y}_i(w) = \begin{cases} \frac{\sum_{s=i-w}^i \bar{Y}_{i+s}}{2w+1} & \text{si } i=w+1, \dots, m-w \\ \frac{\sum_{s=-(i-1)}^{i-1} \bar{Y}_{i+s}}{2i-1} & \text{si } i=1, \dots, w \end{cases}$$

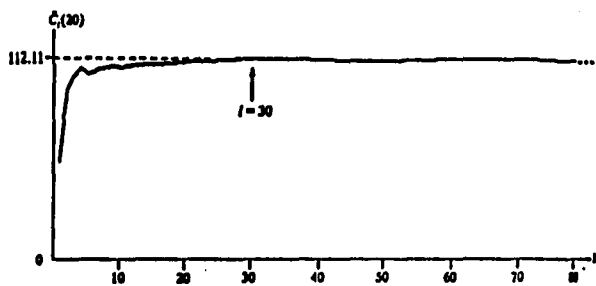
4) Grafique $\bar{Y}_i(w)$ para $i=1,2,\dots,m-w$ y sea l el valor de i a partir del cual $\bar{Y}_1(w), \bar{Y}_2(w), \dots$ converge gráficamente.

5) En el caso de que no exista una gráfica adecuada para las w existentes, es necesario realizar entre 5 y 10 corridas adicionales y repetir todo el procedimiento (Ver gráfica III.2).

Para la elección de n , m y l se recomienda lo siguiente:

- Iniciar con n entre 5 y 10 corridas y m tan grande como sea posible. En particular, m debe ser mucho mayor que el valor pronosticado subjetivamente para l y debe permitir que los eventos poco frecuentes ocurran un número razonable de ocasiones durante la simulación.

Grafica III.2



Promedio móvil con $w=20$

- Graficar $\hat{Y}_i(w)$ para diferentes valores de w , y escoger el menor valor w a partir del cual la gráfica se empieza a pulir. Utilice esta gráfica para determinar el valor de l .

Si el valor w es pequeño, la gráfica de $\hat{Y}_i(w)$ será escabrosa. En caso contrario, las observaciones Y_i estarán sobreagregadas y $E(Y_i)$ no será estimado correctamente.

- Se sugiere estimar l grande (aún cuando la varianza aumente) para así garantizar que el estimado puntual ν tenga un sesgo pequeño.

Una vez calculado el índice de calentamiento l , es necesario calcular $\nu = E(Y)$. Existen varias técnicas pero la más común es la *Técnica de Corrida/Borrado* debido a su buen rendimiento y facilidad estadística. Esta técnica consiste en lo siguiente:

Se realizan n' corridas cada una con longitud m' , y sea Y_{ji} la i -ésima observación de la j -ésima corrida, entonces

$$X_j = \sum_{i=1}^{m'} \frac{Y_{ji}}{m'} \quad j=1,2,\dots,n'$$

Debido a que las X_j 's son variables aleatorias i.i.d. con $E(X_j) = \nu$. Entonces $\bar{X}(n')$ es un estimador puntual insesgado para ν y para un intervalo de confianza al $100(1-\alpha)\%$ se tiene que:

$$\bar{X}(n') \pm t_{n'-1, 1-\alpha/2} \sqrt{\frac{S^2(n')}{n'}}$$

En donde $\bar{X}(n')$ y $S^2(n')$ se calculan de acuerdo a los métodos estadísticos tradicionales.

Para que se obtenga una mejor precisión estadística se recomienda que las corridas para calcular l sean independientes de las corridas para calcular $\bar{X}(n')$ y $S^2(n')$.

Como ya se mencionó anteriormente, la media no siempre proporciona una medida adecuada del desempeño del sistema por lo que resulta necesario considerar otros parámetros de estado-lugar:

Probabilidad Estado-Lugar, $p=P(Y \in B)$

La probabilidad p es un caso especial de la estimación de la media ν . Sea la variable aleatoria estado-lugar definida por:

$$Z = \begin{cases} 1 & \text{si } Y \in B \\ 0 & \text{otro caso} \end{cases}$$

Es posible demostrar que estimar p es equivalente a estimar la media estado-lugar de $E(Z)$. Por lo tanto, se puede aplicar la técnica de corrida/borrado al proceso de salida Z_i para obtener un estimador puntual y un intervalo de confianza para $E(Z)$.

Nótese que el período de calentamiento del proceso Z_1, Z_2, \dots puede diferir del período del proceso original Y_1, Y_2, \dots .

Q-ésimo cuantil, y_q

Sea y_q el valor de y tal que:

$$P(Y \leq y_q) = q \quad \text{donde } Y \text{ es una variable aleatoria de estado-lugar}$$

No existe un método estándar para calcular los cuantiles.

NOTA: La estimación de un parámetro cíclico de estado-lugar es un caso especial de la estimación de un parámetro de estado-lugar. Es decir, todas las técnicas estudiadas se aplican a las variables aleatorias cíclicas Y_i^c en lugar de a las originales Y_i .

e) Múltiples Medidas del Desempeño del Sistema

En las secciones anteriores se presentaron procedimientos para la construcción de intervalos de confianza para una sola medida del desempeño del sistema. Sin embargo, en el mundo real es común que se requieran simultáneamente diferentes medidas. Supóngase que I_s es un intervalo de confianza al $100(1-\alpha_s)\%$ para la medida del desempeño del sistema μ_s (donde $s=1,2,\dots,k$ y k es el número de medidas requeridas) y que todas las μ_s 's pertenecen al mismo tipo de simulación (terminal o no

terminal). Entonces, la probabilidad de que los k intervalos de confianza contengan simultáneamente a las medias verdaderas, debe satisfacer:

$$P(\mu_s \in I_s, \forall s=1,2,\dots,k) \geq 1 - \sum_{s=1}^k \alpha_s$$

Este resultado tiene serias implicaciones en un estudio de simulación. Por ejemplo, si se construyen 10 intervalos de confianza al 90% (i.e. $\alpha_s=0.1$) para 10 diferentes medidas del funcionamiento, la probabilidad de que cada uno de los intervalos de confianza contenga su verdadera media será mayor o igual a cero. Esto significa que si se tiene un gran número de medidas del funcionamiento, la única solución consiste en construir los intervalos de confianza usuales y tener siempre en cuenta que uno o más de estos intervalos no contendrá su verdadera media.

Una solución práctica al problema mencionado anteriormente y cuando k es relativamente "pequeña" consiste en:

i) Determinar el nivel de confianza total α asociado con los k intervalos de confianza con un valor de al menos $100(1-\alpha)\%$.

ii) Escoger las α 's de tal manera que $\sum_{s=1}^k \alpha_s = \alpha$.

Nótese que las α 's no tiene que ser iguales. Entonces, las α 's correspondientes a las medidas del desempeño más importantes pueden ser las más pequeñas.

Así, se pueden construir 10 intervalos de confianza al 99% y tener un intervalo de confianza total con un valor de al menos 90%. La dificultad de esta solución consiste en que los intervalos de confianza serán más largos en el método del tamaño muestral fijo, o se requerirá de más datos en un proceso secuencial.

f) Gráficas de Tiempo para Variables Clave

En esta sección se ha estudiado como construir estimadores puntuales e intervalos de confianza para diversas medidas del funcionamiento. A pesar de que estas medidas son muy útiles, existen ocasiones en que se necesita un mejor indicador que muestre como varía el sistema durante el tiempo (esto es particularmente cierto cuando las características del sistema cambian como una función del tiempo). El graficar una o más variables clave durante la simulación permite observar fácilmente como varía el sistema a lo largo del tiempo. Por otro lado, la animación proporciona una comprensión del

sistema a corto término.

III.9.2. Técnicas de reducción de varianza

Debido a que un estudio de simulación puede consumir una gran cantidad de tiempo de computadora y de almacenamiento, un análisis estadístico apropiado (que necesite de una gran cantidad de corridas) puede resultar muy costoso. En algunas ocasiones, el costo del análisis estadístico puede resultar tan elevado que ocasione que la precisión de los resultados sea muy baja. Entonces, el analista debe de utilizar todos los medios que estén a su alcance para incrementar la exactitud de los resultados.

Si existe alguna forma para reducir la varianza de una variable aleatoria de salida sin que la media sea afectada, se puede obtener una mayor precisión, i.e., menores intervalos de confianza para el mismo número de corridas de simulación o, alternativamente, obtener la precisión deseada con un menor número de corridas. Esta sección explicará brevemente algunas de las técnicas de reducción de varianza (VRT por sus siglas en inglés: Variance Reduction Technique).

La técnica de reducción de varianza a aplicar dependerá del modelo, siendo imposible conocer de antemano el grado de reducción de varianza que se obtendrá. Sin embargo, es posible realizar corridas preliminares para analizar la diferencia entre aplicar las VRT's o no hacerlo. Es conveniente observar que el uso de VRT's generalmente implica un incremento en el costo de los recursos de cómputo, por lo que será necesario evaluar la conveniencia del decremento en la eficiencia computacional a cambio de un incremento en la exactitud estadística. Por lo general, el uso de todos los VRT's requiere de un esfuerzo adicional por parte del analista, y esto debe de ser considerado.

Existen diferentes técnicas de reducción de varianza:

a) *Números Aleatorios Comunes*

Denominada CRN por sus siglas en inglés: Common Random Numbers.

Esta técnica compara diferentes configuraciones "bajo condiciones experimentales similares", para

así garantizar que las diferencias observadas son ocasionadas por el uso de configuraciones alternas y no debido a las "condiciones experimentales" del modelo.

Las "condiciones experimentales similares" son obtenidas al utilizarse durante la simulación de las configuraciones alternas los mismos números aleatorios (CRN).

Idealmente, un número aleatorio específico que sea utilizado en una configuración para un propósito particular, debe de ser usado para exactamente el mismo propósito en las otras configuraciones. Por ejemplo, si un número aleatorio U_k es utilizado en la primera configuración para generar un tiempo específico de servicio; entonces, en la segunda configuración debe de ser usado para generar el mismo tiempo de servicio.

En particular, no es suficiente iniciar la simulación de todas las configuraciones con la misma semilla o arreglo de números aleatorios.

La sincronización de números aleatorios entre diferentes configuraciones depende de la estructura del modelo y sus parámetros, así como de los métodos usados en la generación de variables aleatorias. Durante la programación es posible utilizar algunos "trucos" que ayudan a mantener la sincronización:

1. Dedicar un arreglo de números aleatorios a una variable de entrada específica.

La mayoría de los lenguajes de simulación permiten el uso de diferentes arreglos de números aleatorios. Sin embargo, el número de estos arreglos puede no resultar suficiente en el caso de simulaciones grandes. Más aún, debido a que generalmente los arreglos de números aleatorios son segmentos adyacentes de un único generador de números aleatorios, en el caso de simulaciones grandes será necesario cuidar que los números aleatorios no se sobrepongan unos con otros. Una solución a lo anterior consiste en saltar algunas asignaciones de arreglos; por ejemplo: utilizar el arreglo 1 y la primera mitad del arreglo 2 para los tiempos de llegadas, el arreglo 6 y la primera mitad del arreglo 7 para los tiempos de servicio, etc.

2. Desperdiciar ciertos números aleatorios en ciertos puntos de la simulación.

3. En la simulación de líneas de espera, se puede generar el número aleatorio del tiempo de servicio de un cliente en el momento en que éste llega, y almacenarse como un atributo del cliente.

Aún con los trucos de programación, puede resultar casi imposible el obtener una sincronización completa de las variables aleatorias del modelo. Además, es necesario tomar en cuenta que el tiempo de programación, el tiempo de computadora y los costos de almacenamiento pueden no valer la reducción de la varianza. Debido a lo anterior, se recomienda sincronizar únicamente las variables clave y generar el resto de manera independiente.

b) Variables Antitéticas

Denominada AV por sus siglas en inglés: Antithetic Variates.

Se aplica a la simulación de un sola configuración del modelo.

El objetivo de AV es generar parejas de corridas, de tal manera que una observación pequeña sea anulada por su pareja que será una observación grande proveniente de otra corrida. Entonces, si se utiliza el promedio de las dos observaciones éste tenderá a acercarse a la media μ .

En su forma más simple, AV trata de inducir la correlación negativa al utilizar un par números aleatorios complementarios. Es decir, si U_k es un número aleatorio específico utilizado para un propósito particular durante la primer corrida, entonces en la segunda corrida se usará $1-U_k$ para el mismo fin.

Un requisito indispensable para que AV funcione correctamente consiste en que la respuesta generada por un número aleatorio en particular sea monótona en cualquier dirección. Puede darse el caso de que la varianza aumente en lugar de disminuir: cuando se tenga que la respuesta es grande para U_k pequeño, pequeña para U_k cercano a 0.5 y otra vez grande para U_k grande.

Al utilizarse esta técnica, es necesario que el analista proporcione evidencia de que AV funciona correctamente, ya sea argumentando propiedades físicas o por experimentación inicial.

c) Variables de Control

Denominada CV por sus siglas en inglés: Control Variates.

Sea X una variable aleatoria de salida y se quiere estimar $\mu = E(X)$. Supóngase que Y es otra variable aleatoria que se supone correlacionada con X ; además, se conoce el valor de $\nu = E(Y)$.

Si X y Y están correlacionadas positivamente, se corre una simulación y se observa que $Y > \nu$; entonces se puede sospechar que X está por encima del valor de μ por lo que se puede ajustar X hacia abajo en una determinada cantidad. Por otro lado, si resulta que $Y < \nu$ se puede asumir que $X < \mu$ y se debe ajustar X hacia arriba. De esta manera, se aprovecha el conocimiento de la media de Y para ajustar a X hacia su media μ , reduciendo así la variabilidad de X con respecto a μ de una corrida a otra. Y se denomina *variable de control* de X ya que es utilizada para ajustar X o para "controlarla" parcialmente.

Una variable de control Y apropiada tiene que estar altamente correlacionada con la variable X (para que el ajuste sea más amplio) y debe tener una varianza pequeña. La búsqueda de estas variables de control se realiza por medio del análisis de la estructura del modelo o por la experimentación inicial. Existen tres fuentes de variables de control:

- Internas

Son las variables aleatorias de entrada o simples funciones de ellas. Generalmente sus medias son conocidas y un simple análisis puede mostrar el grado de correlación que tienen con respecto a X . Más aún, estas variables son necesarias para llevar a cabo la simulación por lo que no agregan ningún costo adicional al estudio.

- Externas

En algunas ocasiones es imposible calcular analíticamente a $\mu = E(X)$. Sin embargo, existe la probabilidad de que si se altera el modelo asumiendo varios supuestos, se pueda calcular la media ν de la variable aleatoria de salida Y del modelo simplificado. A pesar de que no se desea utilizar el modelo simplificado ya que se puede alterar la validez del modelo, Y puede servir como variable de control para X . Si se simulan ambas configuraciones por medio de CRN (técnica de números aleatorios comunes) y X y Y están correlacionadas positivamente, entonces se puede utilizar a Y como variable de control de X . A diferencia de las variables de control internas, este enfoque tiene un costo ya que se

requiere de una segunda simulación para determinar si la variable de control es adecuada. Esto significa que la correlación entre X y Y debe ser muy elevada para que valga la pena utilizar a Y como variable de control de X.

III.10. DOCUMENTACION E IMPLEMENTACION DE LOS RESULTADOS

Es necesario elaborar un reporte explicando detalladamente los resultados obtenidos durante la simulación, para que así el personal involucrado confíe en el modelo y el estudio de simulación y decida tomar decisiones de acuerdo a los resultados obtenidos.

Cabe recalcar que un estudio de simulación no se debe considerar completo y terminado sino hasta que los resultados hayan sido implementados: ¿De qué sirve realizar un estudio de simulación y no ponerlo en práctica?

CAPITULO IV

CASO PRACTICO

IV. CASO PRACTICO

Con el objetivo de aplicar la metodología propuesta para la industria manufacturera nacional, se buscó una empresa mexicana que se encontrará en proceso de cambio o expansión de su planta y que necesitará de un estudio de simulación para poder realizar las modificaciones necesarias de manera adecuada.

Se contactó una empresa cuya distribución de planta era obsoleta y que estaba en proceso de rediseño de una nueva. Se llegó a un acuerdo en el que la empresa se comprometía a proporcionar toda la información necesaria y a asignar a un líder de proyecto que fuera el contacto entre la empresa e INDUCE quien por su parte se comprometió a entregar resultados en un plazo de dos meses. El nombre de dicha empresa no será mencionado por no convenir a sus intereses.

IV.1. FORMULACION DEL PROBLEMA Y PLANEACION DEL ESTUDIO

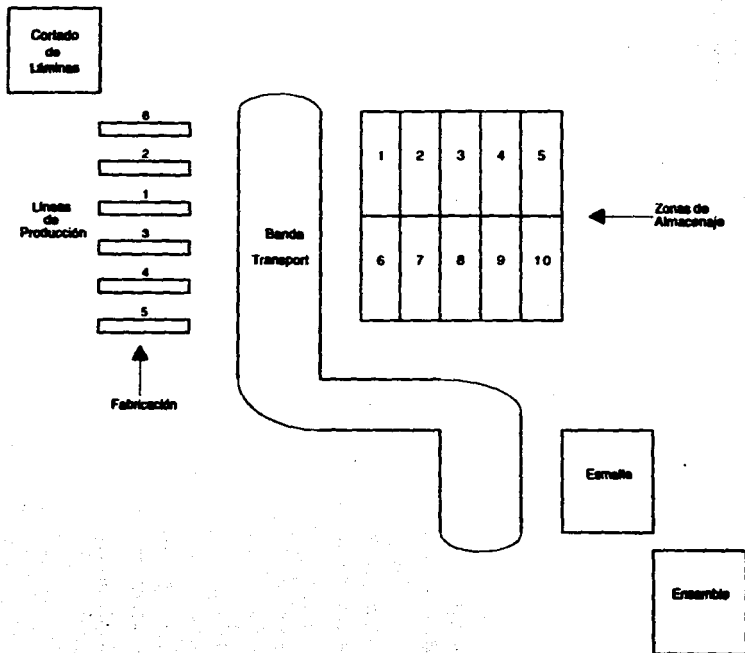
Por tratarse de un proyecto piloto se convino en limitarlo a una duración de dos meses. Fue necesario encontrar una sección de la distribución de la planta en la que la empresa tuviera necesidad de rediseño y que el estudio de simulación no requiriera más del tiempo establecido.

La planta de la empresa contactada se encuentra dividida en tres áreas: fabricación, esmalte y ensamble (en la figura IV.1 se muestra un esquema de la distribución de la planta).

A grandes rasgos la planta funciona de la siguiente manera: al área de producción llega la materia prima en forma de láminas y salen piezas terminadas. Un porcentaje de estas piezas es colgado en el transportador aéreo y otro porcentaje es almacenado. Actualmente, el porcentaje es variable pero se espera que en un futuro se cuelgue el 20% y se almacene el 80% del lote producido (de acuerdo a los ciclos de producción que son de 5.09 días). Las partes almacenadas son colgadas en el transportador aéreo de acuerdo a la demanda. Estas piezas son lavadas y esmaltadas, vuelven a ser almacenadas (en otra bodega correspondiente al área de esmalte) y, finalmente, son solicitadas para el área de ensamble.

GRAFICA IV.1.

DISTRIBUCION DE LA PLANTA



Por tratarse de un proyecto limitado a una duración de 2 meses, se decidió simular únicamente el área de fabricación. Anteriormente esta área no estaba organizada, sino que las piezas se procesaban de acuerdo a la demanda del momento a la disponibilidad de las máquinas, por lo que existía mucho inventario en proceso almacenado en forma de canastillas. La empresa decidió establecer 6 líneas de producción, por limitaciones del proyecto se acordó analizar a fondo únicamente las líneas 2 y 5 (las más próximas a ser implementadas) y superficialmente el resto de las líneas.

De acuerdo a lo anterior y a las necesidades prioritarias de la empresa, se establecieron los siguientes objetivos:

- a) Determinación de requerimientos de equipo mediante la detección de posibles cuellos de botella en el plan de producción y bandas transportadoras, así como por tiempos ociosos de máquinas.
- b) Verificación de que la distribución propuesta cumple con los objetivos del plan de producción.
- c) Dimensionamiento de los almacenes que reciben productos de las líneas 2 y 5, mediante el análisis de capacidad.
- d) Determinación del número de montacargas necesario.

Para facilitar el estudio, en las líneas 1, 3, 4 y 6 se utilizaron únicamente los tiempos estándar y promedios.

IV.2. CONSTRUCCION DEL MODELO

El modelo construido abarcó desde el inicio de producción de las partes en las 6 líneas (siendo más específico para las líneas 2 y 5) hasta que las partes son entregadas a la lavadora por medio de la banda transportadora. Se incluye el proceso de producción, operación de colgado, movimientos del montacargas de las líneas al almacén y de éste a la zona de colgado, y almacenamiento de las partes en canastillas. El proceso llevado a cabo en la lavadora y los procesos posteriores son ignorados por este modelo, así como los procesos anteriores a las líneas de producción (ver gráfica IV.1).

Debido a lo anterior, se manejan los siguientes supuestos:

- a) El abasto de materia prima a las líneas de producción es ilimitado.
- b) La lavadora tiene capacidad para recibir todo lo que llega por la banda transportadora.

Las variables establecidas para este modelo fueron:

Exógenas

- Flujo de materiales
- Tiempos de proceso de cada parte en cada máquina a lo largo de la línea
- Tiempos de descompostura y reparación de las máquinas
- Tiempos de preparación de las máquinas
- Tiempos de descompostura y reparación de los montacargas
- Tiempos de descompostura y reparación del transportador aéreo

Dentro de las variables exógenas se consideraron determinísticamente:

- Ciclo de producción
- Número de partes por canastilla
- Velocidad del transportador aéreo
- Forma de colgado de las partes en el transportador aéreo
- Horario de producción
- Zonas de embarque
- Asignación de espacios físicos en el almacén
- Número de montacargas

Endógenas

- Número de canastillas en el almacén
- Retraso en la producción de las partes (verificación de que se cumple el plan de producción)
- Número de piezas en espera de ser colgadas

A pesar de que determinar el número de montacargas era un objetivo inicial, por limitaciones de tiempo del proyecto no fué posible realizar corridas con diferentes números de montacargas.

De Estado

Básicamente son definidas por el programa en computadora. Se utilizaron las siguientes variables de estado:

- Capacidad del almacén
- Retraso de la producción
- Localización de los montacargas
- Estado del transportador aéreo
- Lotes en espera de ser procesados

Debido a que el número de variables resultaba excesivo para el modelo, se decidió utilizar un modelo simplificado en el que las líneas de producción se reducían a una sola máquina. Es decir, se consideró el trabajo de todas las máquinas de la línea como una sola máquina. Esto fue posible gracias a que las máquinas están conectadas entre sí por medio de una banda transportadora que puede ser ignorada por el modelo (ya que su velocidad depende de la máquina más lenta).

Se realizó una simulación previa con un modelo muy sencillo que únicamente representaba a la línea de producción y así se determinó el tiempo de proceso de cada parte a través de la línea. Se observó para todas las partes que el tiempo de proceso de la parte a través de la línea era muy similar al tiempo de proceso en la máquina más lenta. Los tiempos de proceso de cada parte a través de la línea fueron capturados en un archivo, procesados en el paquete estadístico Statgraf y así se obtuvo la distribución de probabilidad del tiempo de proceso de cada parte a través de la línea. Estas distribuciones de probabilidad se encuentran en el Anexo A.

Asimismo, fue necesario concentrar los paros de la línea provocados por la descompostura de alguna máquina en un solo factor. Para lograr esto, en un modelo independiente se simularon los paros de la línea por descomposturas de alguna de las máquinas y los paros fueron concentrados en lo que sería el paro de la línea en su totalidad. Estos datos fueron guardados en un archivo, procesados en Statgraf y así se obtuvo la distribución de probabilidad de los tiempos de descompostura de toda la línea.

Debido a que los tiempos de reparación de las máquinas de una línea eran muy similares se concentraron en un solo archivo, se procesaron en Statgraf y se obtuvo la distribución de probabilidad del tiempo que tarda la línea en ser reparada. Las distribuciones de probabilidad de los tiempos de descompostura y reparación de las líneas y de cada máquina se encuentran en el Anexo B.

El tiempo de preparación de toda la línea se calculó al simular el tiempo de preparación de cada máquina y guardar en un archivo el máximo tiempo resultante. Estos tiempos fueron procesados en Statgraf y se obtuvieron las distribuciones de probabilidad correspondientes. Cabe hacer notar que las distribuciones de probabilidad asignadas a cada parte, dependen del número de máquinas que requiera una parte para ser procesada. Las distribuciones de probabilidad de los tiempos de preparación de la línea calculados de acuerdo al número de máquinas que se utilicen se encuentran en el Anexo C.

Una vez que se simplificó el modelo y se establecieron sus aspectos generales, se mostrarán los detalles específicos del mismo. A continuación se presentan las reglas utilizadas para definir las variables endógenas del modelo.

IV.2.1. Flujo de Materiales

Cabe recordar que por limitaciones del proyecto, se analizaron detenidamente las líneas 2 y 5 y de manera superficial las líneas restantes. Se consideraron 29 partes (numeradas del 1 al 31 y excluyéndose los números 2 y 5).

Por simplificación, todas las partes de cada una de las líneas 1, 3, 4 y 6 se concentraron en una sola parte: 1, 3, 4 y 6 respectivamente. El número total de partes producidas fué la suma de todas las partes de la línea; así mismo, el tiempo de proceso y número de partes por canastilla consistió en el promedio de los datos de todas las partes de la línea.

Las piezas de la línea 2 están numeradas del 7 al 17, y de la línea 5 del 18 al 31. Esta numeración se utilizó para simplificar la programación del modelo en computadora y por así convenir a los intereses de confidencialidad de la empresa.

Se consideraron 10 ubicaciones de almacén para las partes. A continuación se muestra la asignación de las partes en cada almacén.

UBICACION	PARTES ALMACENADAS
Línea 2	
1	7, 8, 9, 10 Y 11
2	12, 13, 14 Y 15
3	16 Y 17
Línea 5	
4	18, 19, 20 y 21
5	22 y 27
6	23, 24, 25 y 26
7	28
8	29
9	30 y 31
Líneas restantes (1, 3, 4 y 6)	
10	1, 3, 4 y 6

Cuando una parte sale de la línea de producción y existe demanda en la lavadora, la pieza es colgada inmediatamente en el transportador aéreo; de lo contrario, es mandada al almacén. Por otro lado, si existe demanda de la pieza en la lavadora y la línea de producción no la puede satisfacer en ese momento (ya que está produciendo otra parte), la pieza es sacada del almacén y colgada en la banda aérea.

Los montacargas son utilizados para transportar las partes de las líneas de producción al almacén, y de éste a la banda transportadora. Cabe aclarar que por simplificación del modelo, las partes correspondientes a las líneas 1, 3, 4 y 6 son transportadas "mágicamente" sin utilizar el montacargas.

IV.2.2. Ciclos de Producción

La nueva distribución de la planta contempla que se produzca un lote de cada parte cada 5.09 días. Según datos proporcionados por la empresa el tamaño del lote es suficiente para satisfacer la demanda de la lavadora; sin embargo, no se toma en cuenta la posibilidad de que se presente demanda imprevista.

El modelo genera un lote de cada parte en el tiempo cero de la simulación. Existe un orden de producción y cada lote espera su turno para ser procesado. El segundo lote de cada parte es generado una vez que se terminó de producir el primero y debe esperar su turno. Si el primer lote se terminó de producir antes del día 5.09, entonces el segundo lote se programa para que empiece en el día 5.09 o después si la línea no está desocupada en ese momento. Si el primer lote se terminó de producir después del día 5.09, que indicará un retraso en la producción, el segundo lote será programado inmediatamente o después si la línea no se encuentra desocupada.

A medida que cada pieza sale de la línea de producción, se verifica si existe demanda en la lavadora (ver sección de demanda en la lavadora). En caso afirmativo la pieza es colgada en el transportador aéreo (ver sección de colgado de piezas); de lo contrario, la pieza es "almacenada en canastilla" en espera de ser transportada al almacén por medio del montacargas.

IV.2.3. Generación de la Demanda en la Lavadora

Debido a que los ciclos de producción son de 5.09 días, es decir, la producción de cada lote es suficiente para abastecer durante 5.09 días a la lavadora; para efectos del modelo la lavadora debe demandar diariamente $1/5.09$ del tamaño del lote de cada pieza. Entonces, la primer demanda de cada pieza se genera en el instante en que se empieza a producir el primer lote correspondiente. Las demandas posteriores se generan a partir de la primera cada 24 horas. Así se garantiza que si en la simulación todavía no se ha producido por primera vez la pieza, no se tenga demanda sobre algo que no existe y el modelo no funcione correctamente.

Una vez generada la demanda se verifica si existen piezas recién salidas de la línea de producción y "almacenadas" en canastillas junto a ésta. En caso afirmativo, las piezas son colgadas inmediatamente en el transportador aéreo; de lo contrario, las partes son sacadas del almacén por medio del montacargas y colgadas en la banda transportadora en la zona de embarque correspondiente. En el caso de las piezas de las líneas 1, 3, 4 y 6 el montacargas no es utilizado.

IV.2.4. Descompostura de las Líneas de Producción

Como se mencionó anteriormente, el tiempo entre descomposturas y tiempos de reparación de cada línea fueron calculados en simulaciones previas.

Al iniciar la simulación el modelo calcula el tiempo en que se producirá la primer descompostura de cada línea. Al llegar la simulación al tiempo de descompostura, se calcula el tiempo de reparación y durante el transcurso de éste se suspende toda actividad en la línea. Una vez concluido el tiempo de reparación, se calcula el tiempo en que ocurrirá la siguiente descompostura. Este ciclo se repite a lo largo de toda la corrida de simulación.

IV.2.5. Movimiento de la Banda Transportadora

Debido a que el proceso de colgado de las piezas en las perchas de la banda depende del número y tamaño de las piezas que ya se encuentran colgadas en la percha, fué necesario simular el movimiento de la banda transportadora.

Se consideró que la velocidad de la banda es de 6.8 mts/min (es decir, las perchas cambian de posición cada 0.117 minutos). El movimiento de la banda y la carga por percha son considerados únicamente desde la primer zona de colgado (posición 8 en el almacén) hasta la última zona de colgado al final de la línea 5 (posición 5). Esto permitió concluir que a lo largo de este tramo de la banda existen 88 perchas. Para llevar a cabo la actividad de colgado en este tramo (explicada posteriormente) se asignaron 8 posiciones de colgado. (ver gráfica IV.2).

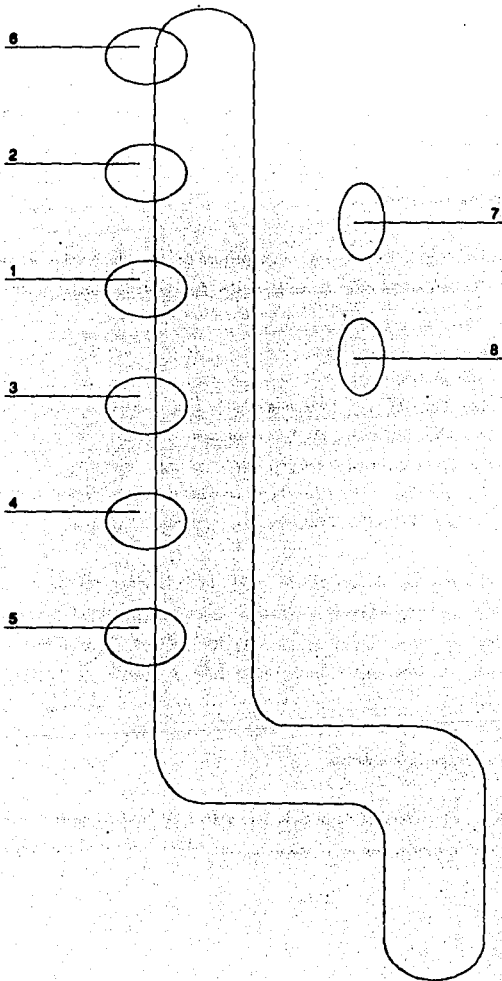
Para efectos de este estudio se asumió que el movimiento de la banda es constante y que ésta nunca se para. Esto se debió a que una vez pasado el período de aprendizaje de los obreros, el efecto del paro de la banda es despreciable para el resto de las variables. Además, no se poseía suficiente información acerca de la banda y la que se tenía correspondía al período de aprendizaje (no representativo para este estudio).

IV.2.6. Colgado en la Banda Transportadora Aérea

En el modelo se definieron 8 posiciones de colgado (ver gráfica IV.2). Aunque en la realidad el colgado se realiza a lo largo de la banda, esto no afecta al modelo debido a que el efecto final sobre las

GRAFICA IV.2.

UBICACIONES DEL COLGADO



piezas en espera de ser colgadas es el mismo si se considera que hay 9 colgadores en cada posición de colgado.

Para colgar una pieza en una determinada zona de colgado se consideró lo siguiente:

1. Las líneas de producción tienen asignadas las posiciones de colgado que llevan el mismo número que la línea (ver gráfica IV.2).
2. Todas las piezas de la línea 2 que se encuentran en almacén son colgadas en la posición 7.
3. Todas las piezas de las líneas 5, 1, 3, 4 y 6 que se encuentran almacenadas son colgadas en la posición 8.
4. Debido a que la determinación del número de colgadores no es un objetivo de este estudio, se decidió este número fuera lo suficientemente grande para no afectar al resto de las variables del modelo. Entonces, se asignaron 9 colgadores en cada zona de colgado, y así el modelo funciona como si se tuvieran colgadores a lo largo de toda la banda.
5. El tiempo de colgado de las piezas depende de su tamaño:

TAMAÑO	TIEMPO DE COLGADO EN MINUTOS
Chica	0.112
Mediana	0.170
Grande	0.240
Extragrande	0.240

Por simplificación, las partes de las líneas 1, 3, 4 y 6 se consideraron chicas.

6. Para el acomodo de las piezas en las perchas a cada pieza se le asignó un tipo que se clasifica en:

TIPO DE PIEZA	DESCRIPCION
20	Pieza de 20"
30	Pieza de 30"
40	Piezas que no son ni de 20" ni de 30"

Los tipos se asignaron de la siguiente manera:

TIPO	PARTE
40	1 Corresponde a todas las partes de la línea 1
40	3 Corresponde a todas las partes de la línea 3
40	4 Corresponde a todas las partes de la línea 4
40	6 Corresponde a todas las partes de la línea 6
Partes de la línea 2	
20	7
30	8
20	9
20	10
30	11
20	12
30	13
20	14
30	15
40	16
40	17
Partes de la línea 5	
20	18
20	19
30	20
30	21
20	22
20	23
20	24
30	25
30	26
30	27
40	28
40	29
40	30
40	31

7. El acomodo de las piezas en las perchas tiene las siguientes reglas:

- a) A cada percha se le asigna el tipo de la primer pieza que es colgada en ella.
- b) Si la percha es de tipo 20, solamente se le pueden colgar partes del tipo 20 o 40 (la percha seguirá siendo de tipo 20). Si la percha es de tipo 30, sólo se le pueden colgar partes de tipo 30 o 40 (la percha seguirá siendo de tipo 30). Si la percha es de tipo 40, se le pueden colgar partes de 40, 20 o 30. Ver figuras IV.3 y IV.4).
- c) Si una percha es de tipo 40 y se le cuelga una pieza de tipo 20, la percha cambia a tipo 20.
- d) Si una percha es de tipo 40 y se le cuelga una pieza de tipo 30, la percha cambia a tipo 30.
- e) Si se trata de colgar una parte en una percha y la parte no se ajusta al tipo de la percha, es necesario esperar 0.117 minutos a que llegue la nueva percha para volver a intentar colgar la parte.

En las gráficas IV.3 y IV.4 se puede apreciar claramente las zonas de colgado y los arreglos por percha de las piezas.

IV.2.7. Turnos de Producción

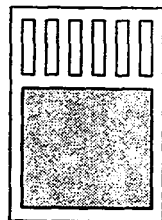
En el modelo se consideró que las líneas de producción suspenden labores de las 22:00 a las 6:00. Sin embargo, la banda transportadora, la lavadora, los almacenes y montacargas operan las 24 horas del día.

IV.2.8. Asignación de Espacios en el Almacén

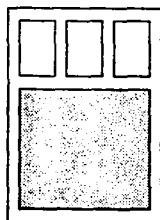
Como ya se mencionó anteriormente, cada parte tiene asignada una ubicación en el almacén. El espacio físico que utiliza cada parte dentro de su ubicación depende del número de partes que caben en una canastilla. A continuación se muestra una tabla con el número de partes que caben por canastilla de acuerdo al número de parte:

ARREGLOS POR PERCHA

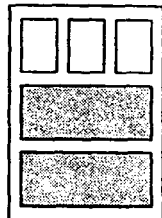
SOLO PIEZAS DE 20"



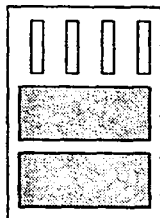
EXTRAGrandes - Chicas



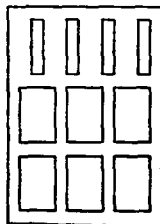
EXTRAGrandes-Mediana



Grandes-Mediana

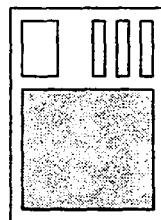


Grandes-Chicas

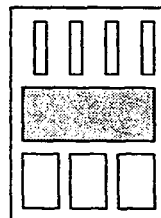


Mediana - Chicas

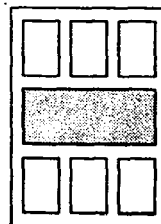
SOLO PIEZAS DE 30"



EXGrandes-Mediana-Chicas



Grandes-Mediana-Chicas



Grandes - Mediana

GRAFICA IV.4.

ARREGLOS POR PERCHA

NUMERO DE PIEZAS POR TIPO

SOLO PIEZAS DE 20"

EXGde	Gde	Med	Chica
1	0	0	6
1	0	3	0
0	2	3	0
0	2	0	4
0	0	6	4

SOLO PIEZAS DE 30"

EXGde	Gde	Med	Chica
1	0	1	3
0	1	3	4
0	1	6	0

PARTE	PIEZAS/ CANASTILLA
1	498*
3	218*
4	233*
6	3,000*

* Estos datos son un promedio de todas las partes de la línea correspondiente

Línea 2

7	190
8	90
9	190
10	190
11	90
12	145
13	110
14	145
15	110
16	130
17	130

Línea 5

18	230
19	230
20	150
21	150
22	750
23	250
24	250
25	160
26	160
27	450
28	750
29	600
30	1,500
31	1,500

Durante la simulación se considera el número de canastillas por ubicación de almacén para calcular la máxima ocupación del mismo.

IV.2.9. Operación y Descompostura de los Montacargas

Para efectos de este estudio, los montacargas solamente se utilizaron para el transporte de las partes provenientes de las líneas 2 y 5. Las partes de las líneas 1, 3, 4 y 6 son transportadas "mágicamente" por el modelo.

Las variables consideradas para los montacargas de la línea 2 son:

- Tiempo que tarda el montacargas de ir de la línea 2 al área de almacenes de las partes de la línea 2.
- Tiempo que tarda el montacargas en moverse del área de almacenes de las partes de la línea 2 a la zona de colgado 7 (correspondiente a las partes de la línea 2).
- Tiempo que tarda el montacargas en transportarse de la zona 7 a la línea 2.
- Tiempo de carga y descarga.
- Tiempo entre descompostura de los montacargas (ver Anexo D).
- Tiempo de reparación de los montacargas (ver Anexo D).

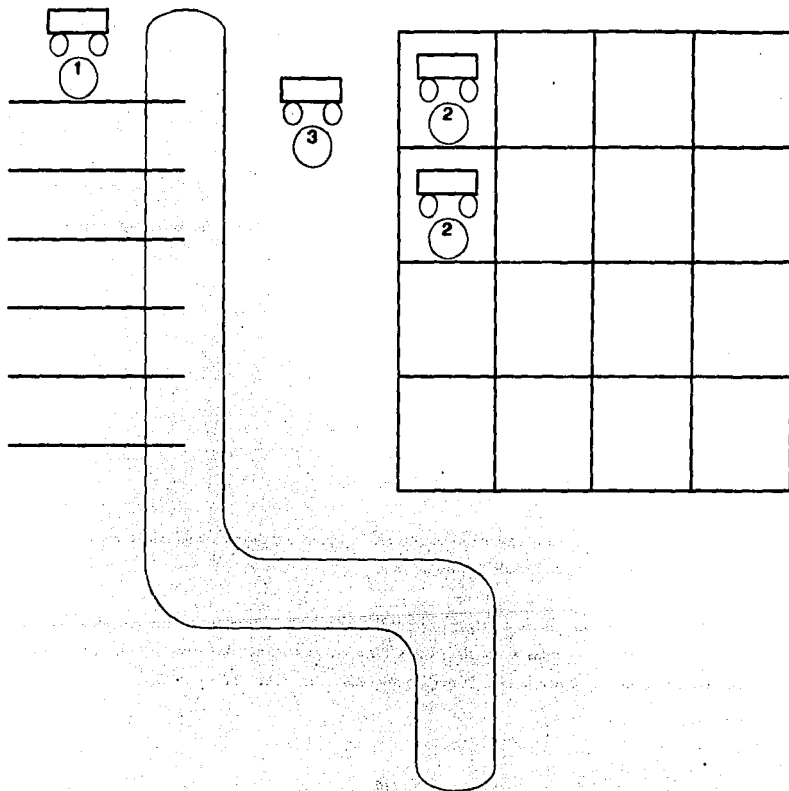
En el caso de los montacargas de la línea 5 se hicieron consideraciones análogas, excepto que en lugar de utilizarse la zona de colgado 7 se utilizó la 8.

Modo de operación del montacargas (independientemente de la línea que se trate): se identificará la posición 1 del montacargas en la zona de la línea correspondiente, la posición 2 en la zona de almacenes y la posición 3 será la zona de colgado asignada. (Ver gráficas IV.5 y IV.6).

GRAFICA IV.5.

LINEA 2

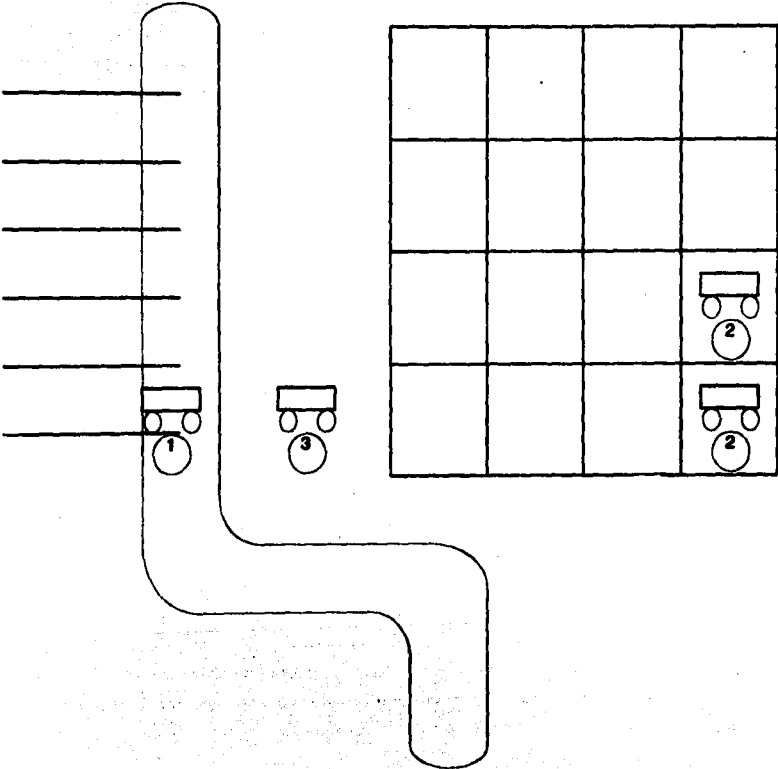
Posiciones del Montacargas



GRAFICA IV.6.

LINEA 5

Posiciones del Montacargas



El montacargas funciona de la siguiente manera:

1. Al inicio de la simulación se encuentra en la posición 1.
2. El máximo número de canastillas transportadas por un montacargas es 3. No se consideró capacidad de carga en kilogramos.
3. La velocidad del montacargas es de 15km/hr.
4. Cada montacargas proporciona servicio exclusivo a una línea (2 o 5). Transporta el material de la posición 1 a la 2 y de la 2 a la 3, y se mueve vacío de la posición 2 a la 3 y 2 a la 1 para recoger más carga.
5. El tiempo de carga y descarga es de 0.5 minutos.

Una vez que la simulación está en proceso, el montacargas se mueve de una posición a otra. En el momento en que llega a una determinada posición, investiga en que posición se encuentra y de acuerdo a esto toma decisiones. A continuación se describen las acciones a tomar dependiendo de la posición en la que se encuentre (ver gráficas IV.5 y IV.6):

Posición 1

El montacargas recoge canastillas de las partes que estén junto a la línea y las lleva a la posición 2 (almacén). Empieza recogiendo el número de parte que tenga más piezas.

El tiempo que tarda el montacargas en moverse de la posición 1 a la 2, para la línea 2 es de 0.2780 minutos y para la línea 5 de 0.1936 minutos.

Si no existen piezas por transportar en la posición 1, verifica si es requerido en la posición 2 en cuyo caso viaja vacío en esa dirección. En caso de no ser necesitado en la posición 2, espera 0.5 minutos y vuelve a revisar si existen piezas por transportar.

Posición 2

Esta posición se encuentra en el almacén correspondiente a la línea. En este lugar, el montacargas recoge canastillas con el número de parte que tenga piezas en existencia y que sea demandado por la lavadora. Recoge prioritariamente el número de parte que tenga más piezas a ser transportadas.

Las partes son llevadas a la posición 3 (zona de colgado). El tiempo que tarda el montacargas en ir de la posición 2 (almacén) a la posición 3 es de 0.1586 minutos para la línea 2 y de 0.1528 para la línea 5.

Si el montacargas no encuentra piezas por transportar en la posición 2, verifica si es requerido en la posición 1. En caso afirmativo, el montacargas se dirige a la posición 1; de lo contrario, espera 0.5 minutos y vuelve a revisar si existen piezas por transportar.

Posición 3

Al encontrarse en esta posición, el montacargas verifica si es requerido en la posición 2 en cuyo caso se transporta hacia allá. Si no es solicitado en 2, verifica si es requerido en 1; en caso contrario, espera 0.5 minutos y vuelve a revisar si es necesitado.

El tiempo que tarda el montacargas de la posición 3 a la 1 es de 0.0580 minutos para la línea 2 y de 0.0408 para la línea 5..

IV.3. RECOLECCION Y PROCESAMIENTO DE DATOS

IV.3.1. Recolección y Almacenamiento

La obtención de los datos para llevar a cabo el estudio de simulación no fué tarea sencilla ya que la empresa no estaba preparada con la información necesaria para llevar a cabo un estudio de este tipo. Afortunadamente, debido a que se encontraban en proceso de rediseño de la distribución de la planta se contaba con información determinística de gran utilidad concentrada en reportes muy sencillos de entender y manejar. Esta información consistía en ciclo de producción, tamaño de los lotes, número de piezas por canastilla, plano de la nueva distribución de la planta, máquinas que conforman una línea, etc.

Sin embargo, no se contaba con datos con los cuales se pudieran obtener las distribuciones de probabilidad correspondientes, ya que el personal de la empresa no tiene una idea clara de la diferencia entre un estudio determinístico y uno aleatorio. En el caso de los tiempos de proceso generalmente se poseían únicamente promedios; en muy pocos casos se contaba con estudios de tiempos y movimientos ya que la política de la empresa es realizar estos estudios, obtener promedios y desechar la información. También se contaba con los reportes diarios de la producción de los obreros; sin embargo, estos reportes son muy poco confiables ya que los obreros los llenan de manera incorrecta y no especifican los motivos

de un paro.

Para la obtención de los tiempos de descompostura y reparación de las máquinas y montacargas se revisaron registros históricos con ayuda del personal encargado del mantenimiento; así fué posible determinar los tiempos reales de descompostura y reparación. El problema consistió en que mucha de la información estaba basada en la memoria del personal.

Los tiempos de preparación de cada máquina fueron proporcionados como un estimado (mínimo, media y máximo) por ingeniería industrial.

A pesar de que con la recolección de datos se obtuvo mucha información sobre el sistema real y los objetivos, no se logró obtener una visión completa del modelo. Por ejemplo, hasta que se hubieron realizado las corridas de producción y se detectó un retraso considerable en la línea 5, la empresa informó que existían partes que se producían simultáneamente.

Cabe hacer notar que la comunicación entre el personal de sistemas informáticos y el de la planta es mínima por lo que la información proporcionada por el departamento de sistemas (básicamente las hojas de proceso) era totalmente obsoleta para el estudio.

IV.3.2. Validación

La validación resultó un paso muy importante en el estudio ya que en el momento en que se solicitaba al personal que validara la información firmándola, muchas veces la volvían a revisar y surgían modificaciones importantes.

IV.3.3. Conversión y Manipulación

La hoja electrónica de cálculo Lotus 1-2-3 fue de gran utilidad para el almacenamiento, conversión y manipulación de los datos. Permitió realizar operaciones aritméticas, de orden, de depuración, etc. Además, fue posible convertir fácilmente el archivo de Lotus 1-2-3 a un archivo de Statgraf para así poder obtener las distribuciones de probabilidad a partir de los datos proporcionados.

IV.4. EVALUACION DEL MODELO Y ESTIMACION DE LAS DISTRIBUCIONES DE PROBABILIDAD

Desgraciadamente por limitaciones del proyecto y falta de tiempo del personal de la planta, las evaluaciones realizadas conjuntamente fueron muy informales por lo que durante la programación del modelo se asumieron muchos supuestos sin haber consultado previamente al personal de la empresa. En el momento de entregar resultados se vió que algunos de estos supuestos habían sido incorrectos.

Una vez recolectados los datos fue necesario encontrar su distribución de probabilidad. Esto se llevó a cabo por medio del paquete estadístico Statgraf.

Por la naturaleza de la información recolectada se tenían muchos datos dispersos, por lo que en ocasiones resultó necesario depurar la información varias veces hasta poder encontrar una distribución de probabilidad adecuada. Así mismo, fue necesario acotar por abajo por programa algunas de las distribuciones de probabilidad; ya que, por ejemplo, los datos de los tiempos de proceso y reparación se ajustaban a una distribución particular, sin embargo estos tiempos no podían ser inferiores a un determinado tiempo por lo que resultó necesario acotar por programa. Por otro lado, algunas veces la información obtenida resultó tan escasa que fue necesario asumir distribuciones triangulares.

En ocasiones no se tenía información sobre los tiempos de proceso de alguna parte, por lo que se asumía su distribución de probabilidad de acuerdo a (en orden de prioridad):

- Una pieza similar con una operación similar
- La misma pieza en una operación diferente
- Una pieza similar en una operación diferente

IV.5. FORMULACION DE UN PROGRAMA PARA LA COMPUTADORA

Inicialmente se trató de trasladar el modelo a la computadora por medio del paquete de simulación SIMFACTORY II.5 el cual es muy sencillo de utilizar; sin embargo, no resultó adecuado para efectos del modelo ya que no podía simular correctamente la demanda de la lavadora (diariamente un 20% del lote), la manera de colgar las piezas en la banda aérea, entre otras cosas. Además, su capacidad resultaba limitada para el número de partes que maneja el modelo y no era posible determinar el periodo de calentamiento del sistema.

Entonces, el modelo fue desarrollado en el lenguaje de simulación SIMSCRIPT II.5 que permitía establecer todo el nivel de detalle requerido por el modelo. (Ver sección III.5).

IV.6. PRUEBAS PILOTO Y VALIDACION

Una vez que el programa funcionó adecuadamente, se realizaron las pruebas piloto para detectar errores del modelo. En las primeras corridas se observó que algo funcionaba mal ya que los almacenes se vaciaban durante la noche. Debido a que no fue posible detectar fácilmente el error en el modelo, se decidió animar el modelo en la computadora y así se detectó la falla. En este caso se demostró ampliamente la utilidad del uso de la animación.

Como el modelo simulado no existía en la realidad no fue posible validarlo contra éste. Sin embargo, se observó que el modelo predecía con bastante exactitud el comportamiento del modelo real.

IV.7. DISEÑO DE EXPERIMENTOS

Los factores que intervinieron en el modelo son:

- Tiempo de proceso de cada parte
- Tiempo de preparación de cada parte en la línea
- Tiempo de descompostura de la línea
- Tiempo de reparación de la línea
- Cantidad producida
- Cantidad demandada
- Velocidad del transportador aéreo
- Capacidad del transportador aéreo
- Velocidad de los montacargas
- Capacidad de los montacargas
- Número de montacargas
- Tiempo de descompostura de los montacargas
- Tiempo de reparación de los montacargas
- Número de piezas por canastilla

Las respuestas obtenidas son:

- Número de canastillas en almacenes
- Retraso de la producción de las partes
- Piezas en espera de ser colgadas

Por limitaciones del proyecto no se realizó el diseño de experimentos, es decir, no se analizó qué factores tienen un mayor impacto sobre los resultados.

La cantidad producida y demandada, la velocidad y capacidad del transportador aéreo, la velocidad y capacidad de los montacargas y el número de piezas por canastilla son factores considerados fijos por la empresa.

Los tiempos de proceso de las partes son muy parecidos entre sí. Entonces, resulta que la variación de estos factores sería mínima y no afectaría considerablemente a las respuestas.

Durante las pruebas piloto se variaron los valores de tiempos de descompostura y reparación de la línea y montacargas, así como el número de montacargas; sin embargo, no se observaron cambios notables en las respuestas.

De lo anterior se concluyó que los factores que sería adecuado analizar por medio de Diseño de Experimentos es el tiempo de preparación de la línea para cada parte. Estos tiempos fueron proporcionados a nivel muy general por ingeniería industrial, y se consideró que estos podrían disminuir a medida que los obreros adquirieran práctica en el montaje de los nuevos troqueles.

IV.8. EXPERIMENTACION

Las corridas de producción se llevaron a cabo en INDUCE. Se dejaron varias computadoras trabajando durante las noches y a la mañana siguiente se recogían los resultados guardados en archivos (ya que las corridas duraban unas 10 horas y algunas de las computadoras carecían de no break).

IV.9. ANALISIS DE RESULTADOS

Debido a que el proceso de producción de la empresa es continuo no existe un evento terminal E que determine el fin de la simulación. Para calcular el período de calentamiento del sistema se realizaron 8 corridas iniciales con una duración de 28 días. Este período se calculó en base al número de canastillas en almacén ya que el retraso de la producción era acumulado a lo largo del tiempo y el número de piezas en espera de ser colgadas está limitado por programa (máximo 4) y no es representativo. A los resultados obtenidos en estas corridas previas se les aplicó la fórmula de Welch y se determinó gráficamente el período de calentamiento en 7 días (ver gráfica IV.7).

Las corridas de producción se realizaron para 28 días ya que así existe período de calentamiento, se simulan varios ciclos de producción (5 completos) y hay tiempo para que ocurran varias veces los eventos imprevistos: descomposturas de las líneas y montacargas.

Se realizaron 8 corridas definitivas y por programa se eliminaron los resultados de los primeros 7 días. Posteriormente se calcularon los intervalos de confianza para el número de canastillas en almacén, el número de piezas en espera de ser colgadas y el retraso de la producción.

Debido a que la varianza resultó pequeña no hubo necesidad de aplicar las técnicas de reducción de varianza mencionadas en el capítulo III. En el caso de los máximos retrasos de la producción en la línea 5 la varianza es elevada; esto se debe a que al tratarse de máximo únicamente se cuenta con 8 datos (uno por cada corrida). Además, como ya se mencionó anteriormente, los resultados arrojados para la línea 5 no son correctos ya que no se simuló la producción simultánea de algunas partes.

IV.10. DOCUMENTACION E IMPLEMENTACION DE LOS RESULTADOS

En esta fase se llevó a cabo la documentación de los resultados y la entrega de los mismos a la empresa seleccionada.

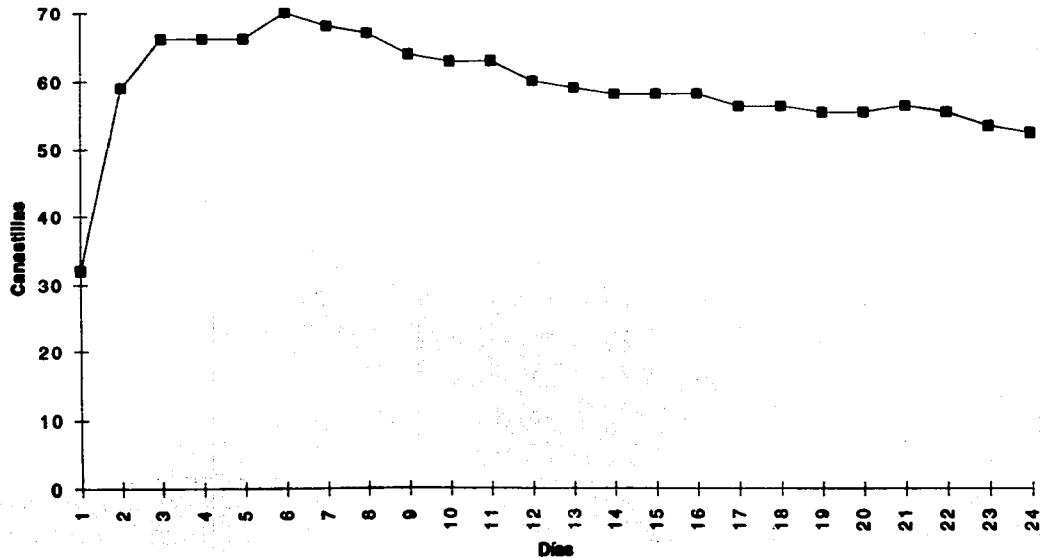
Se entregó un reporte completo a la empresa. Los resultados presentados fueron los siguientes:

RESULTADOS
FORMULA DE WELCH
NUMERO DE CANAETILLAS EN ALMACEN
(A FINAL DEL DIA)

DIAS	PROMEDIO	FORMULA
		k=4
1	32	32
2	70	59
3	75	56
4	83	66
5	72	66
6	64	70
7	68	58
8	67	67
9	62	64
10	62	63
11	56	63
12	61	50
13	60	59
14	63	58
15	60	56
16	46	58
17	57	56
18	55	56
19	59	55
20	60	55
21	44	56
22	56	55
23	58	53
24	58	52
25	56	
26	44	
27	45	
29	51	

Gráfica IV.7

Análisis Tiempo Calentamiento



IV.10.1. Retraso de la producción.

Se observó que la línea 2 (partes 7 a la 17) no sufre retrasos considerables. Únicamente la parte 17 sufrió un pequeño retraso durante 5 ciclos de producción. Esto puede ocurrir en la realidad pero con 1 o 2 horas extras se regulariza el plan de producción para esta línea.

La simulación arrojó que la línea 5 sufre retrasos considerables en el plan de producción. Sin embargo, esto se debió a que la empresa no notificó que existían parejas de partes que se producían simultáneamente, resultando que la información obtenida para esta línea no sea representativa.

Con respecto a las líneas 1, 3, 4 y 6 como se trataba de promedios muy generales los resultados arrojaron retrasos considerables en el plan de producción. Sin embargo, estos resultados no deben ser considerados ya que estas líneas únicamente fueron simuladas a nivel muy general.

IV.10.2. Número de canastillas en el almacén.

Los máximo valores del número de canastillas en el almacén son los representativos para este estudio. Únicamente son importantes los valores para los almacenes correspondientes a la línea 2 (ubicaciones: 1, 2 y 3) ya que los almacenes de la línea 5 siempre están vacíos debido al retraso que tiene esta línea en el plan de producción. Así mismo, el almacén 10 correspondiente a las líneas 1, 3 4 y 6 no es representativo por simularse estas líneas a nivel muy general.

IV.10.3. Número de partes en espera de ser colgadas.

Los valores medios son los representativos para la determinación de un cuello de botella en la banda transportadora. Debido a una restricción técnica de la herramienta de simulación utilizada, el máximo número de piezas en espera de ser colgadas fué de 4; sin embargo, se observa que el promedio más grande de piezas en espera de ser colgadas es de 0.68 lo que significa que no existen cuellos de botella en la banda.

Desgraciadamente para que estos resultados puedan ser implementados se requiere volver a realizar corridas con supuestos adicionales que quedó de proporcionar la empresa; sin embargo, hasta el momento de terminar esta tesis la empresa no había proporcionado la información por encontrarse en proceso de cambios organizacionales.

**RESULTADOS
RETRASOS DE LAS PARTES**

PARTE	CORRIDA 1			CORRIDA 2		
	MIN	FROM	MAX	MIN	FROM	MAX
1	0.00	8.47	11.41	0.00	8.47	11.41
2	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	12.10	12.10	0.00	12.10	12.10
4	0.00	4.29	7.09	0.00	4.22	7.02
5	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.05	0.05	0.00	0.00	0.00
18	0.00	2.47	3.92	0.00	2.42	3.90
19	0.00	2.71	4.12	0.00	2.77	4.24
20	0.00	3.52	4.90	0.00	3.58	5.09
21	0.00	2.92	5.79	0.00	4.39	5.90
22	0.00	3.24	5.96	0.00	3.19	6.23
23	0.00	3.62	6.15	0.00	3.46	6.71
24	0.00	4.02	6.80	0.00	3.99	7.02
25	0.00	4.17	6.89	0.00	4.09	7.11
26	0.00	4.44	7.07	0.00	4.30	7.28
27	0.00	3.75	5.31	0.00	3.59	5.06
28	0.00	3.26	6.34	0.00	3.13	6.05
29	0.00	4.19	7.18	0.00	4.06	7.04
30	0.00	4.88	7.82	0.00	4.37	7.36
31	0.00	5.10	8.05	0.00	4.96	7.96

**RESULTADOS
RETRASOS DE LAS PARTES**

PARTE	CORRIDA 3			CORRIDA 4		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	8.47	11.41	0.00	8.47	11.41
2	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	12.10	12.10	0.00	12.10	12.10
4	0.00	4.24	7.01	0.00	4.23	7.15
5	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	1.41	2.78	0.00	2.41	3.50
19	0.00	1.59	2.90	0.00	2.64	4.24
20	0.00	2.42	3.76	0.00	3.46	4.99
21	0.00	3.08	4.29	0.00	4.33	5.88
22	0.00	3.42	4.79	0.00	3.14	6.22
23	0.00	3.54	4.90	0.00	3.45	6.68
24	0.00	2.96	5.64	0.00	3.97	6.97
25	0.00	3.06	5.75	0.00	4.08	7.10
26	0.00	3.34	5.88	0.00	4.30	7.27
27	0.00	4.07	6.78	0.00	3.56	6.96
28	0.00	3.69	4.99	0.00	3.07	5.95
29	0.00	3.15	5.95	0.00	4.04	6.89
30	0.00	3.57	6.32	0.00	4.31	7.13
31	0.00	4.04	6.91	0.00	4.93	7.77

**RESULTADOS
RETRASOS DE LAS PARTES**

PARTE	CORRIDA 5			CORRIDA 6		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	8.48	11.41	0.00	8.47	11.41
2	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	12.10	12.10	0.00	12.10	12.10
4	0.00	4.60	7.15	0.00	4.60	7.15
5	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00
18	0.00	2.41	3.86	0.00	2.86	2.86
19	0.00	2.65	3.98	0.00	1.78	3.25
20	0.00	3.49	4.86	0.00	2.60	4.01
21	0.00	4.34	5.74	0.00	3.40	4.83
22	0.00	3.29	6.12	0.00	3.78	5.20
23	0.00	3.51	6.23	0.00	4.23	5.54
24	0.00	4.01	6.87	0.00	3.13	5.92
25	0.00	4.12	6.97	0.00	3.29	6.12
26	0.00	4.34	7.21	0.00	3.67	6.27
27	0.00	3.64	5.07	0.00	4.38	7.00
28	0.00	3.16	6.05	0.00	3.89	5.30
29	0.00	4.08	7.03	0.00	3.26	6.10
30	0.00	4.38	7.34	0.00	3.85	6.73
31	0.00	4.99	7.94	0.00	4.08	6.97

**RESULTADOS
RETRASOS DE LAS PARTES**

PARTE	CORRIDA 7			CORRIDA 8		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	8.47	11.41	0.00	8.47	11.41
2	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	12.10	12.10	0.00	12.10	12.10
4	0.00	4.24	6.95	0.00	4.29	7.09
5	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.05	0.05
18	0.00	2.46	3.96	0.00	2.47	3.92
19	0.00	2.69	4.30	0.00	2.71	4.12
20	0.00	3.49	5.00	0.00	3.52	4.90
21	0.00	2.89	5.82	0.00	2.92	5.79
22	0.00	3.39	6.68	0.00	3.24	5.96
23	0.00	3.66	6.78	0.00	3.62	6.15
24	0.00	4.08	7.13	0.00	4.02	6.80
25	0.00	4.22	7.23	0.00	4.17	6.89
26	0.00	3.11	4.79	0.00	4.44	7.07
27	0.00	3.74	5.30	0.00	3.75	5.31
28	0.00	3.23	6.33	0.00	3.26	6.34
29	0.00	4.14	7.19	0.00	4.19	7.18
30	0.00	4.54	7.82	0.00	4.88	7.82
31	0.00	5.01	8.06	0.00	5.10	8.05

**RESULTADOS
RETRASOS DE LAS PARTES**

PARTE	PROMEDIO			MAXIMO		
	MEDIA	VAR.	INTERVALO CONF.(95%)	MEDIA	VAR.	INTERVALO CONF.(95%)
1	8.47	0.00	(08.47, 08.47)	11.41	0.00	(11.41, 11.41)
2	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
3	12.10	0.00	(12.10, 12.10)	12.10	0.00	(12.10, 12.10)
4	4.34	0.03	(04.20, 04.47)	7.08	0.01	(07.01, 07.14)
5	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
6	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
7	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
8	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
9	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
10	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
11	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
12	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
13	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
14	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
15	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
16	0.00	0.00	(00.00, 00.00)	0.00	0.00	(00.00, 00.00)
17	0.01	0.00	(00.00, 00.03)	0.01	0.00	(00.00, 00.03)
18	2.36	0.17	(02.02, 02.71)	3.54	0.25	(03.21, 04.06)
19	2.44	0.22	(02.05, 02.84)	3.89	0.28	(03.45, 04.33)
20	3.26	0.22	(02.87, 03.65)	4.69	0.26	(04.27, 05.11)
21	3.53	0.49	(02.95, 04.12)	5.50	0.36	(05.00, 06.01)
22	3.34	0.04	(03.17, 03.51)	5.90	0.37	(05.39, 06.40)
23	3.64	0.06	(03.43, 03.85)	6.16	0.40	(05.63, 06.69)
24	3.77	0.20	(03.40, 04.15)	6.64	0.30	(06.19, 07.10)
25	3.90	0.20	(03.52, 04.28)	6.76	0.28	(06.32, 07.20)
26	3.99	0.29	(03.54, 04.44)	6.61	0.80	(05.86, 07.36)
27	3.81	0.08	(03.58, 04.05)	5.60	0.66	(04.92, 06.28)
28	3.34	0.09	(03.09, 03.58)	5.92	0.26	(05.49, 06.34)
29	3.89	0.18	(03.53, 04.24)	6.82	0.25	(06.40, 07.24)
30	4.35	0.21	(03.97, 04.73)	7.29	0.30	(06.83, 07.75)
31	4.78	0.20	(04.40, 05.15)	7.71	0.24	(07.30, 08.12)

RESULTADOS
NUMERO DE CANASTILLAS EN EL ALMACEN

ALMACEN	CORRIDA 1			CORRIDA 2		
	MIN	PROM	MAX	MIN	PROM	MAX
1	1	22	49	1	21	49
2	1	19	47	1	19	47
3	0	7	20	0	7	20
4	0	6	23	0	7	23
5	0	1	5	0	1	5
6	0	3	11	0	3	10
7	0	1	3	0	1	3
8	0	1	5	0	1	3
9	0	1	4	0	1	4
10	0	2	3	0	1	3

ALMACEN	CORRIDA 3			CORRIDA 4		
	MIN	PROM	MAX	MIN	PROM	MAX
1	1	22	49	1	21	49
2	1	19	47	1	19	47
3	0	7	20	0	7	20
4	0	8	25	0	7	24
5	0	1	5	0	1	5
6	0	3	11	0	3	10
7	0	1	3	0	1	3
8	0	1	3	0	1	3
9	0	1	4	0	1	4
10	0	1	3	0	1	3

RESULTADOS
NUMERO DE CANASTILLAS EN EL ALMACEN

CORRIDA 5

ALMACEN	MIN	PRGM	MAX
1	1	21	49
2	1	19	47
3	0	7	20
4	0	7	23
5	0	1	5
6	0	3	10
7	0	1	3
8	0	1	3
9	0	1	4
10	0	2	3

CORRIDA 6

MIN	PRGM	MAX
1	21	49
1	19	46
0	7	20
0	7	23
0	1	5
0	3	11
0	1	3
0	1	3
0	1	4
0	2	3

CORRIDA 7

ALMACEN	MIN	PRGM	MAX
1	1	22	48
2	1	19	46
3	0	8	20
4	0	7	23
5	0	1	5
6	0	3	11
7	0	1	3
8	0	1	3
9	0	1	4
10	0	1	3

CORRIDA 8

MIN	PRGM	MAX
1	22	49
1	19	47
0	7	20
0	6	23
0	1	5
0	3	11
0	1	3
0	1	3
0	1	4
0	2	3

RESULTADOS
NÚMERO DE CANASTILLAS EN EL ALMACEN

ALMACEN	PROMEDIO DE CANASTILLAS			MÁXIMO DE CANASTILLAS		
	MEDIA	VAR.	INTERVALO CONF.(95%)	MEDIA	VAR.	INTERVALO CONF.(95%)
1	21.54	0.03	(21.40, 21.69)	48.88	0.13	(48.58, 49.17)
2	18.91	0.03	(18.75, 19.06)	46.75	0.21	(46.36, 47.14)
3	7.27	0.03	(07.14, 07.41)	20.00	0.00	(20.00, 20.00)
4	6.80	0.18	(06.45, 07.16)	23.38	0.55	(22.75, 24.00)
5	1.30	0.02	(01.18, 01.42)	5.00	0.00	(05.00, 05.00)
6	2.86	0.01	(02.77, 02.96)	10.63	0.27	(10.19, 11.06)
7	0.92	0.00	(00.90, 00.93)	3.00	0.00	(03.00, 03.00)
8	0.90	0.00	(00.89, 00.91)	3.00	0.00	(03.00, 03.00)
9	1.11	0.00	(01.07, 01.16)	4.00	0.00	(04.00, 04.00)
10	1.48	0.01	(01.40, 01.56)	3.00	0.00	(03.00, 03.00)

RESULTADOS
PIEZAS EN ESPERA DE SER COLGADAS

POSICION	CORRIDA 1			CORRIDA 2		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.18	3.00	0.00	0.16	3.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.01	3.00	0.00	0.00	3.00
5	0.00	0.50	3.00	0.00	0.47	3.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.68	4.00	0.00	0.67	4.00
8	0.00	0.12	4.00	0.00	0.12	4.00

POSICION	CORRIDA 3			CORRIDA 4		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.14	3.00	0.00	0.15	3.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	3.00	0.00	0.00	3.00
5	0.00	0.47	3.00	0.00	0.44	3.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.67	4.00	0.00	0.67	4.00
8	0.00	0.13	4.00	0.00	0.11	4.00

RESULTADOS
PIEZAS EN ESPERA DE SER COLGADAS

POSICION	CORRIDA 5			CORRIDA 6		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.15	3.00	0.00	0.19	3.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.01	3.00	0.00	0.00	3.00
5	0.00	0.48	3.00	0.00	0.39	3.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.67	4.00	0.00	0.68	4.00
8	0.00	0.12	4.00	0.00	0.13	4.00

POSICION	CORRIDA 7			CORRIDA 8		
	MIN	PROM	MAX	MIN	PROM	MAX
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.16	3.00	0.00	0.18	3.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	3.00	0.00	0.01	3.00
5	0.00	0.49	3.00	0.00	0.50	3.00
6	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.71	4.00	0.00	0.68	4.00
8	0.00	0.12	4.00	0.00	0.12	4.00

RESULTADOS
PIEZAS EN ESPERA DE SER COLGADAS
(PROMEDIO)

POSICION	MEDIA	VARIANZA	INTERVALO CONF. (95%)
1	0.0000	0.000000	(0.00, 0.00)
2	0.1646	0.000342	(0.15, 0.18)
3	0.0000	0.000000	(0.00, 0.00)
4	0.0036	0.000002	(0.00, 0.00)
5	0.4684	0.001258	(0.44, 0.55)
6	0.0000	0.000000	(0.00, 0.00)
7	0.6794	0.000147	(0.67, 0.69)
8	0.1193	0.000038	(0.11, 0.12)

CONCLUSIONES

CONCLUSIONES

Desde hace un tiempo, la simulación es utilizada ampliamente en Estados Unidos y otros países industrializados y ya existen diversas metodologías para la realización de estudios de simulación. Sin embargo, la aportación de esta tesis consiste en proporcionar una metodología de simulación totalmente aplicable a las empresas de manufactura mexicanas.

Generalmente es necesario utilizar la simulación para diseñar o modificar la distribución de una planta, ya que la experimentación en la realidad puede no poder llevarse a cabo o resultar incosteable. Una empresa que utilice simulación puede predecir los efectos que ocasionará un cambio efectuado en determinada área sobre la totalidad de la planta, sin que tenga necesidad de experimentar en la realidad gastando dinero y pudiendo llegar hasta a detener el proceso de la planta durante un tiempo. Por otro lado, aquellas empresas que no hacen uso de la simulación tienen que experimentar forzosamente con la realidad (invirtiendo tiempo y dinero) o deben "adivinar teóricamente" el "mejor" diseño y trabajar en el futuro con éste. Esto significa que aquellas empresas que utilizan la simulación tienen una reducción en sus costos y una mejor planeación, siendo más productivas y competitivas en el mercado.

Debido a la firma del Tratado de Libre Comercio y a que el uso de la simulación es cada vez más frecuente en muchas empresas a nivel mundial, es necesario que las empresas mexicanas apliquen la simulación para así mantener un nivel de competitividad adecuado frente al mercado internacional.

Al ser la simulación un concepto nuevo en México es conveniente que las empresas busquen apoyo en las Universidades que siempre tienen responsabilidad social para con el progreso del país.

Uno de los principales retos a los que se enfrenta la simulación en México es el factor humano. Es conveniente aclarar que las aportaciones enunciadas a continuación están basadas en el perfil típico de la empresa mexicana, con su muy característica forma de trabajar; pero, claro está, existen excepciones.

En primer lugar, el concepto de simulación es poco conocido en las empresas mexicanas por lo que resulta necesario explicar claramente sus ventajas y desventajas. Una vez que se adquiere una noción básica del concepto de simulación pueden surgir dos interpretaciones erróneas: la simulación es complicada y al no entender su funcionamiento no satisface mis necesidades; o, por otro lado, la simulación es "mágica" y va a resolver todos mis problemas. A pesar de que el último enfoque es muy optimista y la persona cooperará ampliamente con el proyecto, es necesario especificar claramente

desde un principio los límites del proyecto para que posteriormente no haya desilusiones al no cumplir la simulación con las expectativas y que por lo mismo nunca lleguen a ser implementados los resultados obtenidos.

Como todo estudio, lo más importante es que el personal interesado defina claramente los Objetivos y conozca sus limitaciones. En general, una vez que el usuario se ha convencido de las ventajas de la simulación desea que el estudio abarque *todo* el sistema, por lo que es deber del modelador ayudarlo a establecer los objetivos prioritarios y explicarle con claridad las limitaciones. Es conveniente aclarar que los problemas adicionales que el usuario desea resolver, pueden ser objeto de otro estudio de simulación independiente.

Una vez definidos los objetivos, se requiere conocer a fondo el sistema a simular. Esto no resulta tarea fácil, ya que la mayor parte del personal de planta de las empresas mexicanas no tiene la capacidad de explicar con claridad el funcionamiento del sistema en forma parcial o total. El personal puede conocer el manejo del sistema a la perfección; pero, probablemente no posea los conocimientos para explicarlo claramente.

A pesar de que para un modelador externo no resulta sencillo el comprender el sistema en su totalidad, se llegó a la conclusión de que es mejor que el estudio de simulación sea realizado por personal externo en lugar de personal interno (al menos que se justifique la existencia de un departamento de simulación y éste se encuentre totalmente capacitado). Esto se debe a que se requieren conocimientos estadísticos y de programación; así mismo, se debe contar con equipo de cómputo suficiente (en ocasiones resulta necesario dejar varias máquinas en funcionamiento durante toda la noche calendarizando su uso) y con paquetería adecuada de simulación y estadística. Por otro lado, debido a la forma en que se dan las relaciones interpersonales en la mayoría de las plantas de manufactura en México, es probable que un modelador externo pueda recopilar la información con mayor facilidad que el personal interno ya que puede presentarse el sentimiento de "no me vayas a quitar mi puesto si te proporciono información". Una vez que la simulación ha sido ampliamente aceptada dentro de la empresa, es posible formar un departamento de simulación contando inicialmente con el apoyo y conocimientos del modelador externo en forma de cursos y asesorías.

En el caso que se contrate a modeladores externos, es necesario elegir dentro del personal de la planta a un líder de proyecto que sea el contacto entre la empresa y los consultores. Es muy importante mantener un contacto continuo con el líder para que así no se pierda el interés y los resultados tengan credibilidad y sean implementados. Es recomendable que este líder de proyecto tenga nociones básicas de estadística, para que así la comunicación sea óptima y los objetivos claramente definidos.

Debido a que la mayoría de las empresas mexicanas no toman en cuenta el aspecto estadístico de la información (por lo general únicamente conservan promedios de los datos), es conveniente realizar un análisis previo al estudio de simulación para determinar el tipo de información que posee la empresa en particular. Si los datos existentes no son suficientes para proporcionar aleatoriedad al modelo, se recomienda postergar el estudio durante un tiempo para que durante este período el personal interno recabe la información. Es necesario capacitar al personal en la recolección de los datos. El lapso de recolección depende de la frecuencia con la que se presenten todos los eventos a ser utilizados por el modelo. Durante el período de captura de datos, es necesario verificar periódicamente que la información se recopile de manera adecuada. En caso negativo, se requiere determinar la falla y reforzar la capacitación al personal. Es muy importante recalcar que la información proporcionada es la que dará lugar a la aleatoriedad en los datos y así la simulación emulará correctamente a la realidad. En el caso de que el modelador externo recolecte la información, es necesario que el personal la valide y la firme aceptándola (al pedir una firma, el personal revisa nuevamente la información a validar). Resulta de utilidad que la información a ser presentada para su validación lleve el mismo orden que utiliza normalmente el personal de la planta, para así evitar confusiones y que la validación sea efectuada de forma óptima.

Uno de los factores más importantes para el éxito de un estudio de simulación, es el compromiso del personal de la empresa con la veracidad y exactitud de la información proporcionada o recolectada. No solamente se deben validar los datos y los supuestos del sistema, sino que se debe garantizar que no exista información adicional que no haya sido proporcionada. En el caso en que se tengan datos faltantes es necesario especificarlo, así como la fecha de su entrega futura. De esta manera el modelador se cubre contra posibles cambios o adiciones.

Si la empresa tiene un control informático del proceso de manufactura y se desea obtener información a través del mismo, es necesario verificar que exista concordancia entre la realidad y lo reportado por el departamento de sistemas. Así mismo, si la empresa tiene procedimientos escritos es necesario verificar si en la práctica estos procedimientos se siguen o se realiza otra cosa. Es decir, es importante checar si se va a modelar la parte teórica de la empresa (de acuerdo a los procedimientos escritos y al control informático) o de acuerdo a lo que se lleva a cabo en la realidad.

Es importante recalcar que muchos estudios de simulación pueden no proporcionar los resultados deseados ya que la elaboración del modelo y las corridas de producción consumen la mayor parte del tiempo asignado al proyecto, y no se toma en cuenta el análisis estadístico de los resultados.

ANEXOS

ANEXO A

DISTRIBUCIONES DE PROBABILIDAD DE TIEMPOS DE PROCESO SINTETIZADOS PARA LA LINEA

A partir de los tiempos de proceso de una parte en cada operación, se calcularon las distribuciones de probabilidad de los tiempos de proceso de la parte a través de toda la línea.

LINEA 2

PARTE: 7

Lognormal

Media: 0.1863

Desviación Est: 0.0149

Límite Inferior: 0.1640

PARTE: 8

Lognormal

Media: 0.1750

Desviación Est: 0.0114

Límite Inferior: 0.1640

PARTE: 9

Lognormal

Media: 0.1750

Desviación Est: 0.0114

Límite Inferior: 0.1640

PARTE: 10

Triangular

Mínimo: 0.1689

Moda: 0.1690

Máximo: 0.2235

PARTE: 11

Triangular

Mínimo: 0.1689

Moda: 0.1690

Máximo: 0.2452

PARTE: 12

Constante:

0.1920

PARTE: 13

Constante:

0.1920

PARTE: 14
Constante: 0.1920

PARTE: 15
Triangular
Mínimo: 0.1649
Moda: 0.1650
Máximo: 0.2551

PARTE: 16
Constante: 0.1920

PARTE: 17
Constante: 0.1920

LINEA 5

PARTE: 18
Constante: 0.2399

PARTE: 19
Weibull
Alfa: 4.3223
Beta: 0.2543
Limite Inferior: 0.0700

PARTE: 20
Lognormal
Media: 0.2552
Desviación Est: 0.0054
Limite Inferior: 0.2420

PARTE: 21
Triangular
Mínimo: 0.2601
Moda: 0.2602
Máximo: 0.2757

PARTE: 22
Triangular
Mínimo: 0.2295
Moda: 0.2300
Máximo: 0.2756

PARTE: 23
Lognormal
Media: 0.2750
Desviación Est: 0.0296
Limite Inferior: 0.2350

PARTE: 24
Lognormal
Media: 0.2600
Desviación Est: 0.0333
Límite Inferior: 0.2300

PARTE: 25
Lognormal
Media: 0.2700
Desviación Est: 0.0320
Límite Inferior: 0.2350

PARTE: 26
Lognormal
Media: 0.2600
Desviación Est: 0.0350
Límite Inferior: 0.2400

PARTE: 27
Triangular
Mínimo: 0.2299
Moda: 0.2300
Máximo: 0.2833

PARTE: 28
Lognormal
Media: 0.2442
Desviación Est: 0.0236
Límite Inferior: 0.1950

PARTE: 29
Lognormal
Media: 0.2383
Desviación Est: 0.0171
Límite Inferior: 0.2050

PARTE: 30
Triangular
Mínimo: 0.0951
Moda: 0.0973
Máximo: 0.1125

PARTE: 31
Triangular
Mínimo: 0.0931
Moda: 0.0982
Máximo: 0.1478

ANEXO B

DISTRIBUCIONES DE PROBABILIDAD DE TIEMPOS DE DESCOMPOSTURA Y REPARACION DE LAS LINEAS

A partir de la información disponible sobre descomposturas de las prensas, se calcularon las distribuciones de probabilidad de los tiempos de descompostura de cada máquina. Posteriormente, utilizando estas distribuciones, se calcularon los tiempos de descompostura de las líneas.

TIEMPOS DE DESCOMPOSTURA DE LA LINEA 2

Weibull

Alfa: 1.11

Beta: 2,685.09

TIEMPOS DE DESCOMPOSTURA DE LA LINEA 5

Weibull

Alfa: 0.92

Beta: 1,217.71

Por otro lado, debido a que los tiempos de reparación para todas las prensas eran muy similares, se juntaron todos los datos y se estimó una distribución de probabilidad que es válida para todas las prensas, así como para ambas líneas.

TIEMPO DE REPARACION DE LAS LINEAS

Lognormal

Media: 0.5800

Desviación Est: 0.6000

Límite Inferior: 0.1600

ANEXO C

DISTRIBUCIONES DE PROBABILIDAD DE TIEMPOS DE PREPARACION

Algunas de las piezas solamente son procesadas en una prensa, mientras que otras son procesadas hasta en 4 máquinas. Esto significa que el tiempo de preparación de una pieza va a depender del número de prensas en la que sea procesada, y va a ser equivalente al tiempo de la prensa que más tardó en ser preparada.

Los tiempos de la máquina que más tardó en ser preparada fueron utilizados para calcular las distribuciones de probabilidad de los tiempos de preparación de una pieza de acuerdo al número de prensas en la que es operada:

1 PRENSA

Triangular

Mínimo: 45.00

Moda: 60.00

Máximo: 75.00

2 PRENSAS

Gamma

Alfa: 155.86

Beta: 4.72

3 PRENSAS

Normal

Media: 65.00

Desviación Est: 4.72

Límite Inferior: 45.00

4 PRENSAS

Normal

Media: 66.24

Desviación Est: 4.15

Límite Inferior: 45.00

ANEXO D

DISTRIBUCIONES DE PROBABILIDAD DE DESCOMPOSTURA Y REPARACION DE LOS MONTACARGAS

Actualmente, únicamente se tiene información de los tiempos de descompostura y reparación de dos tipos de montacargas. Por otro lado, se desconoce el tipo de montacargas que se asignarán permanentemente a las líneas de producción. Entonces, las distribuciones de probabilidad de los tiempos de descompostura y reparación se calcularon a partir de toda la información disponible.

TIEMPO DE DESCOMPOSTURA

Weibull

Alfa: 1.02

Beta: 8,173.34

TIEMPO DE REPARACION

Weibull

Alfa: 0.80

Beta: 1.69

BIBLIOGRAFIA

BIBLIOGRAFIA

LIBROS:

1. Averill M. Law y David Kelton, Simulation Modeling and Analysis, segunda edición, Estados Unidos: McGraw-Hill, 1991.
2. Varios, Técnicas de Simulación en Computadoras, séptima edición, México: Limusa, 1988.
3. Alan B. Priestker, Introduction to Simulation and SLAM II, segunda edición, Estados Unidos: Systems Publishing Corporation, 1984.
4. Hugh J. Watson, Computer Simulation in Business, Estados Unidos: John Wiley & Sons Editors, 1981.
5. Byron, J.T. Morgan, Elements of Simulation, Great Britain: Chapman and Hall Ltd, 1984.
6. Paul L. Meyer, Probabilidad y Aplicaciones Estadísticas, segunda edición, México: Fondo Educativo Interamericano, 1973.
7. Jagdish S. Rustagi, Introduction to Statistical Methods, volumen 1, Estados Unidos: Rowman & Allanheld Publishers, 1984.
8. Ya-Lun Chou, Análisis Estadístico, segunda edición, México: McGraw-Hill/ Interamericana de México, S.A. de C.V., 1992.
9. SimFactory II.5, Reference Manual and Users' Guide, Estados Unidos: CACI Products Company, 1991

REVISTAS:

1. Averill M. Law y Michael G. McCormac, "Pitfalls to avoid in the Simulation of Manufacturing Systems", Industrial Engineering, (Estados Unidos, Mayo de 1989).

2. Averill M. Law y Michael G. McComas, "Secrets of Successful Simulation Studies", Industrial Engineering, (Estados Unidos, Mayo de 1990).
3. Averill M. Law, "Introduction to Simulation: a Powerful Tool for analyzing Complex Manufacturing Systems", Industrial Engineering, (Estados Unidos, Mayo de 1986).
4. John S. Carson, "Convincing Users of Model's Validity is Challenging Aspect of Modeler's Job", Industrial Engineering, (Estados Unidos, Mayo de 1986).
5. David Kelton, "Statistical Analysis Methods enhance Usefulness and Reliability of Simulation Models", Industrial Engineering, (Estados Unidos, Septiembre de 1986).
6. Héctor F. Sierra, "Redes de Información y Cooperación", Expansión, (México, D.F., Septiembre 18 de 1991, Vol. XXIII, No. 574).
7. Victor Manuel Ornelas, "Por una Nueva Estrategia Competitiva", Expansión, (México, D.F., Enero 8 de 1992, Edición especial fuera de serie).
8. Alfredo Acle Tomasini, "Perspectivas Macro, Desafíos Micro", Expansión, (México, D.F., Enero 8 de 1992, Edición especial fuera de serie).