

DIVISION DE ESTUDIOS DE POSGRADO
FACULTAD DE INGENIERÍA

01149

36

ANALISIS DE DATOS CON DISTRIBUCION
EXPONENCIAL

TESIS
PRESENTADA A LA DIVISIÓN DE ESTUDIOS DE
POSGRADO DE LA

FACULTAD DE INGENIERIA
DE LA
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

COMO REQUISITO PARA OBTENER
EL GRADO DE
MAESTRO EN INGENIERIA
(ESTRUCTURAS)

PRESENTA:

MARIO GUSTAVO | ORDAZ SCHROEDER

EJEMPLAR UNICO

MEXICO, D.F. 2002



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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 esta 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.

ANALISIS DE DATOS CON DISTRIBUCION EXPONENCIAL

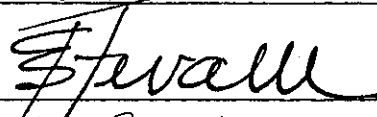
Créditos asignados a la tesis: 10 (Diez)

Aprobado por el jurado:

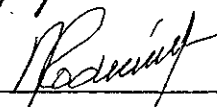
Presidente: DR. EMILIO ROSENBLUETH DEUTSCH



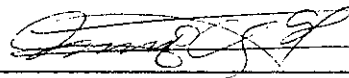
Vocal: DR. LUIS ESTEVA MARABOTO



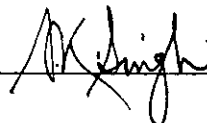
Secretario: ING. NEFTALI RODRIGUEZ CUEVAS



Suplente: DR. MARIO CHAVEZ GONZALEZ



Suplente: DR. SHRI KRISNA SINGH



TESIS CON
FALLA DE ORIGEN



UNIVERSIDAD NACIONAL
AUTONOMA

Profr. Emilio Rosenblueth Deutsch

Comunico a usted que a propuesta del Subjefe del área de INGENIERIA CIVIL ha sido designado como director de tesis del alumno MARIO GUSTAVO ORDAZ SCHROEDER para obtener el grado de M EN I EN ESTRUCTURAS.

Mucho he de agradecerle su comunicación, por escrito, de la aceptación a esta designación y el nombre de la tesis a desarrollár.

Atentamente,
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria a 18 de julio de 1985.
EL JEFE DE LA DIVISION

gu
R. Springall Galindo
DR. ROLANDO SPRINGALL GALINDO

**TESIS CON
FALLA DE ORIGEN**

E.5.1



86.6.30

Dr Gabriel Echávez
Jefe, División de Estudios de Posgrado
Facultad de Ingeniería
U N A M

En respuesta al oficio del Jefe de la División, fechado el 18 de julio de 1985, me complace manifestar a usted mi anuencia para dirigir la tesis de maestría del Ing Civil Mario Ordaz Schroeder.

El nombre de la tesis será "Análisis de datos con distribución exponencial".

Atentamente

Emilio Rosenblueth

TESIS CON
FALLA DE ORIGEN

INDICE

INTRODUCCION

1. POBLACIONES INDEPENDIENTES CON DISTRIBUCIONES ARBITRARIAS	1
2. DOS POBLACIONES INDEPENDIENTES CON DISTRIBUCIONES EXPONENCIALES	4
3. TRATAMIENTO DE ERRORES	12
4. SUBPOBLACIONES	24
5. EJEMPLOS NUMERICOS	30
CONCLUSIONES	48
AGRADECIMIENTOS	50
REFERENCIAS	51
FIGURAS	52
TABLAS	60

INTRODUCCION

En el presente trabajo se aborda el problema de valores atípicos en poblaciones estadísticas. Este problema ha sido tratado de diferentes maneras a lo largo de la historia. Una línea de pensamiento sostiene que deben rechazarse los valores sospechosos cuando difieran sustancialmente de los valores medios, mientras que otra forma de pensar arguye que a todo dato debe dársele igual peso sin importar qué tan lejos se encuentre de los valores "normales". Ambas concepciones tienen puntos débiles puesto que cada dato tiene siempre una probabilidad finita de ser falso, de haber sido malinterpretado o de contener equivocaciones o errores humanos.

El problema general de los valores atípicos puede ser planteado así: dada la densidad de probabilidades previa para la variable en cuestión, las probabilidades previas de que cada valor individual de la muestra sea válido, y los valores de una muestra, hallar la probabilidad posterior de que cada valor sea válido y por tanto la distribución posterior de la variable de interés.

Diremos que un valor es válido (o que pertenece a la población de valores válidos) cuando en efecto refleja el resultado de un experimento o la opinión de alguien sobre algún estado de cosas sin otros errores que no sean de medición u otros tipos de aproximación.

TESIS CON
FALLA DE ORIGEN

A este tipo de errores se les denominará errores aleatorios para diferenciarlos de los grandes errores o errores humanos que son los que hacen que un valor pertenezca a la población de valores no válidos o valores falsos.

El problema general de valores atípicos usualmente se complica al considerar la posibilidad de valores falsos, de errores aleatorios y de que los datos válidos puedan provenir de subpoblaciones con diferentes parámetros.

Este trabajo consta de 5 partes. En la primera se plantean las ecuaciones generales para valores de una variable aleatoria que puede pertenecer a una de dos poblaciones cuyos valores tienen distribuciones arbitrarias. En la segunda parte se particularizan las ideas de la primera a distribuciones exponenciales, las cuales tienen un amplio rango de aplicación práctica: funciones de confiabilidad de sistemas y componentes, procesos de espera tales como llamadas telefónicas llegando a una central o temblores que afectan a un sitio en particular, etc. A continuación se explora la posibilidad de que los valores de la muestra se encuentren afectados de errores aleatorios y de que en una muestra algunos valores no pertenezcan a la población de datos válidos y otros se encuentren afectados de errores aleatorios. Posteriormente se analiza el caso en que los valores válidos pueden pertenecer a una de varias subpoblaciones y finalmente se presenta la solución de ejemplos numéricos.

El enfoque del trabajo es bayesiano. La teoría general sobre procesamiento de información dudosa se encuentra desarrollada en la ref 1. En la ref 2 el análisis se particulariza al caso de distribuciones gaussianas, por lo que estos trabajos constituyen las bases del presente.

1. POBLACIONES INDEPENDIENTES CON DISTRIBUCIONES ARBITRARIAS

Presentamos aquí las expresiones generales que servirán de base para desarrollos posteriores. Estas expresiones se encuentran deducidas en la ref 1. La notación se ha simplificado.

Sea Y la población de valores verdaderos y Z la de valores falsos. Supondremos independencia entre ellas. Sean $p'_Y(\cdot)$ y $p'_Z(\cdot)$ sus respectivas funciones previas de densidad de probabilidad. Sea Q'_i la probabilidad previa de que el i-ésimo valor de una muestra $\{x\}$ de tamaño n pertenezca a Y, es decir, $Q'_i = P'(x_i \in Y)$ y $1-Q'_i = P'(x_i \in Z)$. El teorema de la probabilidad total proporciona la densidad posterior:

$$P''_Y(y) = \sum_{j=1}^{2^n} Q''_j P''_Y(y|\{x\}_{k_j}) \tag{1.1}$$

donde $\{x\}_{k_j} \subset \{x\}$ es alguna de las 2^n combinaciones de los n valores de la muestra, tomados de m en m, $0 \leq m \leq n$ y Q''_j es la probabilidad posterior de que todos los elementos de $\{x\}_{1_j}$ y sólo estos, sean verdaderos:

$$Q''_j = (1/D) p'_Y(\{x\}_{k_j}) p'_Z(\{x\}_{1_j}) \prod_{k=1}^m Q'_k \prod_{k=n-m+1}^n (1 - Q'_k) \tag{1.2}$$

donde $\{x\}_{1_j} = \{x\} - \{x\}_{k_j}$, $\{x\}_{1_j} \cap \{x\}_{k_j} = \phi$, $\phi =$ conjunto vacío, y la constante de normalización D es la suma de los 2^n términos que tienen la forma del numerador. Además, $p'_Y(\{x\}_{k_j})$ y $p'_Z(\{x\}_{1_j})$ son las densidades conjuntas previas de los conjuntos $\{x\}_{k_j}$ y $\{x\}_{1_j}$ respectivamente.

En la ec 1.1 podemos reemplazar las funciones de densidad con funciones de distribución acumulada.

Por ejemplo, si $n = 2$ y se reporta que se observaron x_1 y x_2 , de acuerdo con la ec 1.1 tenemos

$$p_Y''(x) = Q_1'' p_Y'(x) + Q_2'' p_Y''(x|x_1) + Q_3'' p_Y''(x|x_2) + Q_4'' p_Y''(x|x_1, x_2)$$

Q_1'' es la probabilidad posterior de que x_1 y x_2 sean falsos ($\{x\}_{k_1} = \phi$ y $\{x\}_{l_1} = \{x_1, x_2\}$); Q_2'' es la probabilidad posterior de que x_1 sea verdadero y x_2 sea falso ($\{x\}_{k_1} = \{x_1\}$, $\{x\}_{l_2} = \{x_2\}$) y Q_4'' es la probabilidad de que x_1 y x_2 sean verdaderos ($\{x\}_{k_4} = \{x_1, x_2\}$, $\{x\}_{l_4} = \phi$)

De acuerdo con la ec 1.2,

$$\begin{aligned} DQ_1'' &= (1-Q_1') (1 - Q_2') p_Z'(x_1, x_2) \\ &= (1 - Q_1') (1 - Q_2') p_Z''(x_1|x_2) p_Z'(x_2) \\ &= (1 - Q_1') (1 - Q_2') p_Z''(x_2|x_1) p_Z'(x_1) \end{aligned}$$

$$DQ_2'' = Q_1' (1 - Q_2') p_Y'(x_1) p_Z'(x_2)$$

$$DQ_3'' = Q_2' (1 - Q_1') p_Y'(x_2) p_Z'(x_1)$$

$$DQ_4'' = Q_1' Q_2' p_Y'(x_1, x_2) = Q_1' Q_2' p_Y''(x_1, x_2) p_Y'(x_2)$$

Igualando a 1 la suma de las Q'' puede calcularse el valor de D . La función posterior de densidad $p_Y''(\cdot)$ obtenida de una muestra de tamaño n es

2. DOS POBLACIONES INDEPENDIENTES CON DISTRIBUCIONES EXPONENCIALES

Considérese que la variable correspondiente al proceso válido tiene distribución exponencial con parámetro Λ_y (la "intensidad" del proceso). Entonces, su función de densidad de probabilidad es

$$p_y(y) = \Lambda_y e^{-\Lambda_y y} \quad (2.1)$$

o bien,

$$Y \stackrel{d}{=} \text{EXP}(\Lambda_y) \quad (2.2)$$

El parámetro Λ_y no se conoce con precisión. Lo trataremos como una variable aleatoria, asignándole una distribución previa y calculando su distribución posterior. Como es costumbre en el análisis Bayesiano, de no haber razón en contra, suele asignarse a los parámetros una distribución conjugada natural*. La conjugada natural de la ec 2.1 es la distribución Gamma³. Su función de densidad de probabilidad puede ser escrita de la siguiente forma

$$p'_{\Lambda_y}(\lambda_y) = \frac{\lambda_y^{r'_y-1} t'^{r'_y} e^{-\lambda_y t'_y}}{\Gamma(r'_y)} \quad (2.3)$$

* Se dice que un parámetro de la función de distribución de una variable aleatoria tiene una distribución conjugada natural cuando la forma de esta última (y por tanto la de la distribución Bayesiana de la variable aleatoria) no cambia al incorporar información estadística; sólo sus parámetros cambian. Más adelante restringiremos el término "conjugada natural" a aquellas distribuciones de parámetros, que cumplan con la definición y que tengan ellas mismas el menor número posible de parámetros.

donde $\Gamma(\cdot)$ es la función Gamma y r' y t' son los parámetros de la distribución. De la ec 2.3 se sigue que

$$E'\Lambda_Y = r'_Y/t'_Y \quad (2.4)$$

$$\text{VAR}'\Lambda_Y = r'_Y/t'^2_Y \quad (2.5)$$

Escribiremos la ec 2.3 como sigue:

$$\Lambda_Y \stackrel{d'}{=} G(r'_Y, t'_Y)$$

En la misma forma, elegiremos para Z -la población de datos falsos- y para su parámetro las distribuciones previas

$$Z \stackrel{d}{=} \text{EXP}(\Lambda_Z)$$

$$\Lambda_Z \stackrel{d}{=} G(r'_Z, t'_Z)$$

La función Bayesiana de densidad de probabilidad de Y , incluyendo la incertidumbre en Λ_Y , es ⁴

$$p'_Y(y) = \frac{r'_Y t'^{r'_Y}_Y}{(t'_Y + y)^{r'_Y + 1}} \quad (2.6)$$

la cual denotaremos como

$$Y \stackrel{d'}{=} P(r'_Y, t'_Y)$$

Donde P representa la función de distribución de potencia (trasladada).

De manera análoga

$$Z \stackrel{d'}{=} P(r'_Z, t'_Z)$$

De la ec 2.6 se obtiene que

$$E'Y = t'_Y / (r'_Y - 1)$$

$$\text{VAR}'Y = \left(\frac{t'_Y}{r'_Y - 1} \right)^2 \frac{r'_Y}{r'_Y - 2}$$

Supongamos que de la población Y se extrae una muestra y_1, y_2, \dots, y_{n_Y} de tamaño n_Y . La aplicación del teorema de Bayes conduce a que

$$Y \stackrel{d''}{=} P(r''_Y, t''_Y) \quad (2.7)$$

con

$$t''_Y = t'_Y + \sum_i y_i \quad (2.8)$$

$$r''_Y = r'_Y + n_Y \quad (2.9)$$

y expresiones similares para Z.

Reemplazando primas con dobles primas obtenemos las esperanzas y varianzas posteriores de Λ_Y y Λ_Z .

Como ocurre con otras conjugadas naturales en estadística Bayesiana, los

parámetros de la distribución previa pueden ser interpretados heurísticamente como resultados de experimentos ficticios. Entonces, puede considerarse que la esperanza y varianza previas de Λ_y y Y son resultado de una muestra imaginaria de tamaño r'_y , cuya suma de valores vale t'_y . Posteriormente, tomamos una muestra real de tamaño n_y , y los valores de t''_y y r''_y resultan de agrupar los experimentos real y ficticio. Lo mismo se aplica para Z . Sin embargo, r'_y y r'_z -y por tanto r''_y y r''_z - no tienen que ser enteros.

En contraste con las distribuciones Gaussianas tratadas en la ref 2, el parámetro que se considera incierto no es la esperanza de la población. Si $Y \stackrel{d}{=} \text{EXP}(\Lambda_y)$, entonces $EY = 1/\Lambda_y$.

Empleando las ecs 2.6 y 2.8 es posible evaluar las distribuciones Bayesianas posteriores, así como las credibilidades posteriores, sustituyendo en las ecs 1.1-1.2.

Considérese el siguiente ejemplo:

$$t'_y = 6; r'_y = 7; \text{ entonces } E'Y = 1; \sigma'_y = 1.1832$$

$$t'_z = 12; r'_z = 4; \text{ entonces } E'Z = 4; \sigma'_z = 5.66$$

De la ec 2.6,

$$\begin{aligned} p'_y(y) &= \frac{8 \times 7^7}{(y+7)^8} \\ &= \left(\frac{6.1167}{y+6}\right)^8 \end{aligned}$$

Supongamos que se observan tres valores: $y_1 = 1$, $y_2 = 3$, $y_3 = 4, 8, 20$.

Si sólo y_1 perteneciera a la población Y tendríamos

$$r_y'' = 7 + 1 = 8$$

$$t_y'' = 6 + 1 = 7$$

Entonces,

$$\begin{aligned} p_y''(y|y_1) &= \frac{8 \times 7^8}{(y+7)^9} \\ &= \left(\frac{7.1046}{y+7} \right)^9 \end{aligned}$$

Si y_1 y y_2 pertenecieran a Y ,

$$\begin{aligned} p_y''(y|y_1, y_2) &= \frac{9 \times 10^9}{(y+10)^{10}} \\ &= \left(\frac{9.8952}{y+10} \right)^{10} \end{aligned}$$

En forma similar pueden determinarse el resto de las distribuciones posteriores y las credibilidades correspondientes. Las funciones de densidad de interés son combinaciones lineales de funciones de densidad como las presentadas anteriormente. Las funciones de distribución acumulada son combinaciones lineales de funciones de la forma

$$P_y''(y) = 1 - \left(\frac{t_y''}{y + t_y''} \right)^{r_y''}$$

Las distribuciones acumuladas resultantes se muestran en papel semilogarítmico en las figs 2.1-2.3 para $Q'_i = 0.6, 0.8, 0.9$. Las credibilidades posteriores se muestran en la tabla 2.1

A partir de las figs 2.1-2.3, es claro que la importancia de las credibilidades previas aumenta conforme los valores observados difieren más y más de la esperanza previa. Compárense las diferencias entre las tres curvas para $y_3 = 4$ y $y_3 = 20$.

Como con las distribuciones Gaussianas, es deseable averiguar si cada dato debe ser tratado como sospechoso, en virtud de que el número de distribuciones que deben combinarse linealmente es 2^n , donde n es el número de datos. Para explorar esta cuestión, se repitieron los cálculos con los datos anteriores pero con $Q'_1 = Q'_2 = 1$ $Q'_3 = 0.6, 0.8, 0.9$. Las distribuciones resultantes se presentan en las figs 2.4-2.6 y se comparan con las curvas del análisis Bayesiano convencional ($Q = 0$ o 1). Las credibilidades posteriores correspondientes se proporcionan en la tabla 2.2. La comparación indica que con $Q'_i > 0.8$ no hay diferencias apreciables entre asignar $Q'_i = 1$ a todos los datos excepto a los atípicos y asignar a cada dato su credibilidad previa real. Aun con credibilidades previas tan bajas como $Q'_i = 0.6$, las diferencias entre valores de Y asociados a una probabilidad de excedencia dada, son menores de 5% y sólo son apreciables para valores pequeños de Y . Se concluye que con fines prácticos basta con tratar a los datos atípicos como sospechosos, es decir, con $Q'_i < 1$.

Nótese que para el cálculo de la distribución posterior de Y también se puede proceder como se indica a continuación:

Sea A_j el evento consistente en observar que de la muestra $\{y\}$, de tamaño N , el conjunto $\{y\}_{k_j}$ es verdadero y el $\{y\}_{1_j}$ es falso. El conjunto $\{y\}_{k_j}$ está formado por n_{y_j} elementos, y el $\{y\}_{1_j}$ por n_{z_j} ; $n_{y_j} + n_{z_j} = N$, $j = 1, 2, \dots, 2^N$; $\{y\}_{k_j} \cap \{y\}_{1_j} = \phi$, $\{y\}_{k_j} \cup \{y\}_{1_j} = \{y\}$. La función de verosimilitud del evento A_j , dados Λ_y y Λ_z , vale

$$\begin{aligned} l_{A_j} | \Lambda_y, \Lambda_z &= l_{\{y\}_{k_j}} | \Lambda_y \cdot l_{\{y\}_{1_j}} | \Lambda_z \\ &= \lambda_{y_j}^{n_{y_j}} e^{-\lambda_{y_j} \Sigma y_{k_j}} \lambda_{z_j}^{n_{z_j}} e^{-\lambda_{z_j} \Sigma y_{1_j}} \end{aligned} \quad (2.10)$$

donde Σy_{k_j} es la suma de los elementos de $\{y\}_{k_j}$ y Σy_{1_j} es la suma de los elementos de $\{y\}_{1_j}$.

La verosimilitud del evento A , consistente en la observación de $\{y\}$, es:

$$l_A | \Lambda_y, \Lambda_z = \sum_{j=1}^{2^N} Q'_j \lambda_{y_j}^{n_{y_j}} e^{-\lambda_{y_j} \Sigma y_{k_j}} \lambda_{z_j}^{n_{z_j}} e^{-\lambda_{z_j} \Sigma y_{1_j}} \quad (2.11)$$

donde Q'_j es la probabilidad previa de la ocurrencia del evento A_j . La verosimilitud posterior conjunta de Λ_y y Λ_z vale

$$l''_{\Lambda_y, \Lambda_z} = l'_A | \Lambda_y, \Lambda_z \cdot l'_{\Lambda_y, \Lambda_z}$$

por lo que

$$l''_Y = \int_0^\infty \int_0^\infty l'_A | \Lambda_y, \Lambda_z \cdot l''_{\Lambda_y, \Lambda_z} d\Lambda_y d\Lambda_z$$

Efectuando las operaciones correspondientes se llega a que

$$p_Y''(y) = \frac{1}{D} \sum_{j=1}^{2^N} Q_j' \frac{\Gamma(r_{z_j}'') \Gamma(r_{y_j}'')}{t_{z_j}''^{r_{z_j}''} t_{y_j}''^{r_{y_j}''}} \frac{r_{y_j}''^{r_{y_j}''}}{(y + t_{y_j}'')^{r_{y_j}'' + 1}} \quad (2.12)$$

donde

$$D = \sum_{j=1}^{2^N} Q_j' \frac{\Gamma(r_{z_j}'') \Gamma(r_{y_j}'')}{t_{z_j}''^{r_{z_j}''} t_{y_j}''^{r_{y_j}''}} \quad (2.13)$$

con

$$r_{y_j}'' = r_y' + n_{y_j}; \quad t_{y_j}'' = t_y' + \sum y_{k_j} \quad (2.14)$$

$$r_{z_j}'' = r_z' + n_{z_j}; \quad t_{z_j}'' = t_z' + \sum y_{1_j} \quad (2.15)$$

Es claro entonces que

$$Q_j'' = \frac{1}{D} \frac{\Gamma(r_{z_j}'') \Gamma(r_{y_j}'')}{t_{z_j}''^{r_{z_j}''} t_{y_j}''^{r_{y_j}''}} Q_j' \quad (2.16)$$

3. TRATAMIENTO DE ERRORES

3.1 Sólo errores aleatorios

Sea $Y = UW$, donde U representa la población de valores verdaderos y W es un coeficiente de error experimental o de observación. Supongamos que $U \stackrel{d}{=} \text{EXP}(\Lambda_u)$ y $\Lambda_u \stackrel{d}{=} G(r'_u, t'_u)$. En general, el coeficiente W deberá ser tratado como una variable aleatoria, por lo que se hace necesario asignarle una distribución de probabilidad. Dependiendo del fenómeno que se intente modelar, W puede tener diferentes distribuciones, y por tanto, conducir a diferentes distribuciones para Y ; la mayoría de estas resultan de manejo analítico complejo. Sin embargo, se ha observado⁵ que en numerosos problemas de toma de decisiones, el uso de concentraciones de probabilidad que satisfagan los primeros momentos de la variable en cuestión conduce a resultados que son poco sensibles a momentos de orden superior, con un considerable ahorro en la labor de cálculo. Por estas razones y para dar mayor generalidad, elegiremos para W una distribución de concentraciones como la que se muestra en la fig 3.1. De esta figura es claro que

$$P_o + P_a + P_b = 1 \quad (3.1)$$

$$w_o P_o + w_1 P_a + w_2 P_b = EW \quad (3.2)$$

Los valores de P_o , P_a , P_b , w_o , w_1 y w_2 deben asignarse de manera que reflejen lo más fielmente posible nuestro conocimiento de la variable W .

En particular, será razonable suponer para w_0 un valor igual a la esperanza previa de W ; w_1 y w_2 deberán ser tales que sea improbable que $W < w_1$ y $W > w_2$. Si los valores de w_0 , w_1 y w_2 son asignados con estos criterios, los valores de P_a y P_b pueden calcularse a partir de P_0 con las ecs 3.1 y 3.2. Asignaremos a este parámetro la siguiente densidad de probabilidad:

$$p'_{p_0}(p_0) = P'_{00} \delta(p_0) + P'_{01} \delta(p_0 - 1)$$

donde $\delta(\cdot)$ es la función Delta de Dirac. El primero de los sumandos refleja la posibilidad de que W tenga una distribución casi plana; el segundo, la posibilidad de que los valores de W estén fuertemente concentrados alrededor de w_0 , para el cual, de no existir errores sistemáticos (como los de calibración), será razonable suponer un valor de 1.

En estas condiciones,

$$E'W = P'_{01} w_0 + P'_{00} (P_1 w_1 + P_2 w_2) \quad (3.3)$$

$$\text{VAR}'W = P'_{00} [P_1 w_1^2 + P_2 w_2^2] + P'_{01} w_0^2 - (E'W)^2 \quad (3.4)$$

con

$$P_1 = P(W = w_1 | P_0 = 0)$$

$$P_2 = P(W = w_2 | P_0 = 0)$$

$$P_1 + P_2 = 1 \quad (3.5)$$

$$P'_{01} + P'_{00} = 1 \quad (3.6)$$

Si w_0 , w_1 y w_2 se asignan de manera que las distribuciones resultantes cubran el rango de posibles varianzas de W , los valores de P'_{01} , P'_{00} , P_1 y P_2 pueden determinarse con las ecs 3.3-3.6.

En este punto no deseamos conocer la distribución posterior de Y sino la de U , pero no podemos proceder directamente porque desconocemos los valores observados de esta variable.

Sean y_i , $i = 1, 2 \dots n$, los valores observados. Para cada uno de ellos,

$$y_i = u_i/w_i$$

por lo que

$$u_i = y_i w_i$$

Analizaremos primero el caso $P_0 = 0$ que se presenta con probabilidad P'_{00} . La variable W puede tomar los valores w_1 y w_2 con probabilidades P_1 y P_2 respectivamente.

Si se observa un conjunto de n valores de Y , los respectivos valores de W serán algún subconjunto del conjunto de todas las posibles combinaciones con repetición de dos valores, w_1 y w_2 tomados de n en n . A cada uno de estos subconjuntos le corresponderá una probabilidad, dada por el producto de las probabilidades individuales de cada uno de los valores de W . Por ejemplo, el subconjunto $C_{i_1 i_2 i_3 \dots i_n}$, $i_j = 1, 2$, estará formado por los valores $\{w_{i_1}, w_{i_2}, \dots, w_{i_n}\}^*$ y tendrá una probabi-

* Como i_j puede valer sólo 1, 2, un posible conjunto C podría ser $C_{12112 \dots}$, formado por $\{w_1, w_2, w_1, w_1, w_2 \dots\}$

lidad asociada que vale $P_{i_1} P_{i_2} \dots P_{i_n}$. En general, existen 2^n subconjuntos posibles. A cada uno de ellos le corresponde un conjunto de valores de U dados $U = \{y_1/w_{i_1}, y_2/w_{i_2}, \dots, y_n/w_{i_n}\}$ por lo que, de ser esta la combinación de valores de W que se presentaron, la distribución posterior de Λ_u tendría los siguientes parámetros:

$$r''_u = r'_u + n \quad (3.7)$$

$$t''_u = t'_u + y_1/w_{i_1} + y_2/w_{i_2} + \dots + y_n/w_{i_n} \quad (3.8)$$

Como cualquiera de los 2^n subconjuntos C podría ocurrir, la distribución posterior de Λ_u resulta ser una combinación lineal de 2^n distribuciones Gamma, ponderadas de acuerdo con la probabilidad de ocurrencia de cada subconjunto C específico. Si j designa cualquier subconjunto C_{i_1, i_2, \dots, i_n} , diremos que

$$S_j = y_1/w_{i_1} + y_2/w_{i_2} + \dots + y_n/w_{i_n} \quad (3.9)$$

$$P_j = P_{i_1} P_{i_2} \dots P_{i_n} \quad (3.10)$$

$$t''_{u_j} = t'_u + S_j \quad (3.11)$$

Ahora, si $E''\Lambda_{u_k} = E''(\Lambda_u | P_o = k)$ y $\text{VAR}''\Lambda_{u_k} = \text{VAR}''(\Lambda_u | P_o = k)$; $k = 0, 1$, entonces

$$E''\Lambda_{u_0} = \sum_{j=1}^{2^n} P_j \frac{r''_u}{t''_{u_j}}$$

$$\text{VAR}'' \Lambda_{u_0} = \sum_{j=1}^{2^n} \mathbb{P}_j \frac{r_u''(r_u'' + 1)}{t_{y_j}''^2} - (E'' \Lambda_{u_0})^2$$

Denotaremos $p''_{\Lambda_{u_0}}(\cdot | P_0 = k)$ como $P''_{\Lambda_{u_k}}(\cdot)$; $k = 0, 1$. Entonces,

$$p''_{\Lambda_{u_0}}(\lambda_u) = \sum_{j=1}^{2^n} \mathbb{P}_j \frac{\lambda_u^{r_u''-1} t_{u_j}''^{r_u''} e^{-\lambda_u t_{u_j}''}}{\Gamma(r_u'')}$$

De lo anterior resulta que la distribución posterior de U será una combinación lineal de 2^n distribuciones $P(\cdot, \cdot)$. Sea $p''_{u_k}(\cdot) = p''_{\Lambda_{u_k}}(\cdot | P_0 = k)$, $k = 0, 1$. Entonces,

$$p''_{u_0}(u) = \sum_{j=1}^{2^n} \mathbb{P}_j \frac{r_u'' t_{u_j}''^{r_u''}}{(u + t_{u_j}'')^{r_u''+1}} \quad (3.12)$$

$$E'' U_0 = \sum_{j=1}^{2^n} \mathbb{P}_j \frac{t_{u_j}''}{(r_u'' - 1)} \quad (3.13)$$

$$\text{VAR}'' U_0 = \sum_{j=1}^{2^n} \mathbb{P}_j \frac{2 t_{u_j}''^2}{(r_u''-1)(r_u''-2)} - (E'' U_0)^2 \quad (3.14)$$

Ahora, si $P_0 = 1$, lo cual ocurre con probabilidad P'_{01} , tendremos que $W = w_0$ en forma determinística y $u_i = y_i / w_0$, $i = 1, 2, \dots, n$. En este caso,

$$r_u'' = r_u' + n \text{ y } t_u'' = \sum_{i=1}^n y_i / w_0 = t_{u_0}''$$

$$p''_{\lambda_{u_i}}(\lambda_u) = \frac{\lambda_u^{r_u''-1} t_{u_0}''^{r_u''} e^{-\lambda_u t_{u_0}''}}{\Gamma(r_u'')}$$

$$p''_{u_1}(u) = \frac{r_u'' t_{u_0}''^{r_u''}}{(t_{u_0}'' + u)^{r_u''+1}} \quad (3.16)$$

con

$$E''U_1 = \frac{t''_{u_0}}{r''_{u-1}}, r''_u > 1 \quad (3.17)$$

$$\text{VAR}''U = \left(\frac{t''_{u_0}}{r''_{u-1}} \right) \frac{r''_u}{r''_{u-2}}, r''_u > 2 \quad (3.18)$$

El teorema de Bayes permite actualizar P'_{00} y P'_{01} :

$$P''_{00} = \frac{P'_{00} L_{Y_0}(y_1, y_2, \dots, y_n)}{H} \quad (3.19)$$

$$P''_{01} = \frac{P'_{01} L_{Y_1}(y_1, y_2, \dots, y_n)}{H} \quad (3.20)$$

$$H = P'_{00} L_{Y_0}(y_1, y_2, \dots, y_n) + P'_{01} L_{Y_1}(y_1, y_2, \dots, y_n) \quad (3.21)$$

Sabemos que

$$p''_{Y_0}(y) = \frac{P_1}{w_1} p''_{U_0}(y/w_1) + \frac{P_2}{w_2} p''_{U_0}(y/w_2) \quad (3.22)$$

Por tanto,

$$L_{Y_0}(y_1, \dots, y_n) = \prod_{j=1}^n \left[\sum_{i=1}^2 \frac{P_i}{w_i} p''_{U_0}(y_i/w_i) \right] \quad (3.23)$$

También,

$$p''_{Y_1}(y) = \frac{1}{w_0} p''_{U_1}(y/w_0), \quad (3.24)$$

por lo que

$$L_{Y_1}(y_1, y_2, \dots, y_n) = \prod_{j=1}^n \frac{1}{w_0} p''_{Y_1}(y_j/w_0) \quad (3.25)$$

Finalmente,

$$p''_U(u) = P''_{00} p''_{U_0}(u) + P''_{01} p''_{U_1}(u) \quad (3.26)$$

por lo cual,

$$E''U = P''_{00} E''U_0 + P''_{01} E''U_1 \quad (3.27)$$

$$\text{VAR}''U = P''_{00} E''U_0^2 + P''_{01} E''U_1^2 - (E''U)^2 \quad (3.28)$$

siendo

$$E''U_k^2 = \text{VAR}''U_k + (E''U_k)^2; k = 0,1$$

De la ec 3.3,

$$E''W = P''_{01} w_0 + P''_{00} (P_1 w_1 + P_2 w_2) \quad (3.29)$$

De 3.4,

$$\text{VAR}''W = P''_{00} (P_1 w_1^2 + P_2 w_2^2) + P''_{01} w_0^2 - (E''W)^2 \quad (3.30)$$

De la independencia entre U y W se tiene

$$E''Y = E''U E''W \quad (3.31)$$

$$\text{VAR}''Y = [\text{VAR}''U + (E''U)^2] [\text{VAR}''W + (E''W)^2] - (E''Y)^2 \quad (3.32)$$

Puede observarse en la ec 3.12 que la evaluación de $p''_{U_0}(\cdot)$ requiere de gran labor numérica aun para n moderado.

Se ha encontrado que se obtienen resultados suficientemente precisos si se emplea la siguiente aproximación:

$$p''_{U_0}(u) = \sum_{j=1}^{2^n} p_j \frac{r''_u t''_{u_j} r''_u}{(t''_{u_j} + u)^{r''_u + 1}} \sim \frac{R'' T'' R''}{(T'' + u)^{R'' + 1}} \quad (3.33)$$

donde R'' y T'' son tales que el miembro derecho tenga iguales primero y segundo momentos que el izquierdo, es decir,

$$\frac{T''}{R''-1} = E''U_0 \quad (3.34)$$

$$\left(\frac{T''}{R''-1}\right)^2 \frac{R''}{R''-2} = \text{VAR}''U_0 \quad (3.35)$$

con $E''U_0$ y $\text{VAR}''U_0$ dadas por las ecs 3.13 y 3.14

Esta aproximación es válida siempre que

$$\sum_{k=1}^m p_k'' \frac{r_k''}{(r_k''-2)} \approx \frac{R''}{(R''-2)} ; \sum_{k=1}^m p_k'' = 1$$

Esto ocurre siempre que la varianza previa de W no sea excesivamente grande.

Por ejemplo, sea $p_{U_0}''(\cdot) = 0.3 P(3,4) + 0.15 P(4,5) + .35 P(4.5,6) + 0.2 P(5,7)$

$$E''U_0 = 0.3 \times 412 + 0.15 \times 513 + 0.35 \times 613.5 + 0.2 \times 7/4 = 1.800$$

$$\begin{aligned} \text{VAR}''U_0 &= 0.3 \frac{2 \times 4^2}{2 \times 1} + 0.15 \frac{2 \times 5^2}{3 \times 2} + 0.35 \frac{2 \times 6^2}{3.5 \times 2.5} + 0.2 \frac{2 \times 7^2}{4 \times 3} = 1.8 \\ &= 7.3233 \end{aligned}$$

Entonces, $T''/(R''-1) = 1.8$ y $(T''/(R''-1))^2 (R''/R''-2) = 7.3233$, de donde $R'' = 3.587$ y $T'' = 4.6565$.

Es decir,

$$p_{U_0}''(\cdot) \approx P(3.587, 4.6565)$$

Tenemos que

$$\sum_{k=1}^4 p_k (r_k''/r_k'' - 2) = 0.3 \times \frac{3}{1} + 0.15 \frac{4}{2} + 0.35 \frac{4.5}{2.5} + 0.2 \frac{5}{3}$$

$$= 2.1633$$

$$\frac{R}{R-2} = \frac{3.587}{1.587} = 2.2602$$

La relación entre $\sum p_k (r_k''/r_k'' - 2)$ y $R/R'' - 2$ es cercana a 1, por lo que la aproximación es válida.

En la tabla 3.1 se presentan los valores exactos y aproximados de p_{u_0}'' para diferentes valores de U_0 .

La aproximación también puede emplearse, en forma reiterada, para la evaluación de p_{y_0}'' , en virtud de las siguientes relaciones:

Sea

$$p_{U_0}''(\cdot) \approx P(R'', T'')$$

De la ec 3.22,

$$p_{y_0}''(y) = \frac{P_1}{w_1} p_{U_0}''(y/w_1) + \frac{P_2}{w_2} p_{U_0}''(y/w_2) \quad (3.36)$$

Pero si $p_{U_0}''(\cdot)$ es de la forma $P(\cdot, \cdot)$, entonces,

$$\frac{1}{w_1} p_{U_0}''(y/w_1) = \frac{R''(w_1 T'')^{R''}}{(y + w_1 T'')^{R''}}$$

es decir, una distribución P con parámetros R'' y $w_1 T''$. Por tanto,

$$p_{y_0}''(\cdot) = P_1 P(R'', w_1 T'') + P_2 P(R'', w_2 T'') \quad (3.37)$$

Como $P_1 + P_2 = 1$, es posible aplicar la ec 3.33

Con los mismos datos anteriores, supongamos que $w_1 = 0.8$ y $w_2 = 1.2$ con $P_1 = 0.4$, $P_2 = 0.6$:

$$p''_{y_0}(y) = \frac{0.4}{0.8} p''_{U_0}(y/0.8) + \frac{0.6}{1.2} p''_{U_0}(y/1.2)$$

$$w_1 T = 3.725; w_2 T = 5.588$$

$$p''_{y_0}(\cdot) \approx 0.4 P(3.587, 3.725) + 0.6 P(3.587, 5.588)$$

Aplicando 3.33,

$$p''_{y_0}(\cdot) \approx P(3.454, 4.593)$$

En la tabla 3.2 se presentan los valores de $p''_{y_0}(\cdot)$ para distintas Y , empleando la ecuación exacta y la aproximación reiterada.

3.2 Errores aleatorios y equivocaciones
(mínusc. como en 3.1)

3.2 VALORES DUDOSOS

Sean nuevamente $Y = UW$, U valores verdaderos y W el coeficiente de error. Y representa la población de valores válidos. Designaremos como Z a la población de valores falsos y supondremos independencia entre todas las variables.

Sea $\{Q_i\}$ el vector de las probabilidades previas de que una muestra independiente, $\{y_i\}$ pertenezca a la población de valores válidos.

Al igual que en el inciso 3.1, estamos interesados en calcular la distribución posterior de U , aunque ahora tomaremos en cuenta el hecho de que

algunas de las observaciones pueden ser falsas.

La distribución posterior de U puede calcularse a partir de la ec 1.1:

$$p_U''(u) = \sum_j Q_j'' p_U''(u | \{y\}_{k_j}) \quad (3.36)$$

Si $U = \text{EXP}(\Lambda_u^d)$, $\Lambda_Y^d = G(r_u', t_u')$ y W tiene la distribución indicada en el inciso anterior, entonces $p_U''(u | \{y\}_{k_j})$ puede determinarse con la ec 3.26.

Para el cálculo de $Q_{\{j\}}''$ puede hacerse uso de la ec 1.2:

$$Q_j'' = \frac{p_Y''(\{y\}_{k_j}) p''(\{y\}_{1_j}) \prod_{k=1}^m Q_k' \prod_{k=n-m+1}^n (1-Q_k')}{D} \quad (3.37)$$

Nótese que debemos emplear la distribución posterior de Y y no la de U puesto que no se observan los valores de esta variable. La distribución posterior de Y suponiendo que todos los valores observados son válidos, es la siguiente

$$p_Y''(y) = P_{00}'' p_{Y_0}''(y) + P_{01}'' p_{Y_1}''(y) \quad (3.38)$$

con $p_{Y_0}''(\cdot)$ y $p_{Y_1}''(\cdot)$ dadas por las ecs 3.22 y 3.24 respectivamente.

Con $p_U''(u)$ y $p_Y''(y)$ dadas por las ecs 3.36 y 3.38, el procedimiento numérico es similar al del inciso 2.

Nótese que es posible calcular los valores de Q'' en forma directa con la ec 2.16, siempre que se encuentren los valores de r''_{z_j} , r''_{y_j} , t''_{y_j} y t''_{z_j} equivalentes, con la aplicación reiterada de la ec 3.33.

En este caso, no es necesario considerar que si un valor pertenece a la población de datos falsos puede también contener errores. Será suficiente con considerar que es falso.

TESIS CON
FALLA DE ORIGEN

4. SUBPOBLACIONES

Sea Y nuevamente una variable aleatoria tal que $Y \stackrel{d}{=} \text{EXP}(\Lambda)$ y $\Lambda \stackrel{d}{=} G(r', t')$. Llamaremos "la población" al conjunto de valores que puede tomar Y . Supongamos que la población se encuentra dividida en un número finito, N , de subpoblaciones. Cada subpoblación será un subconjunto de valores de Y , y no existen intersecciones entre estos subconjuntos.

A su vez, la j -ésima subpoblación tendrá su propio parámetro, Λ_j , $j = 1, 2, \dots, N$, desconocido. Estableceremos que

$$\Lambda_j = \xi_j \Lambda \quad (4.1)$$

Donde ξ_j , $j = 1, 2, \dots, N$ es un conjunto de coeficientes constantes que por ser desconocidos trataremos como variables aleatorias. Las supondremos independientes de Λ e independientes entre sí, con las condiciones de que $\xi_j > 0$ para toda j y que $\sum_j a_j \xi_j = a$ con $a = \text{CTE}$, $a_j > 0$, $j = 1, \dots, N$. La última condición puede reflejar cualquiera de los siguientes casos: El parámetro de la población es la suma de los parámetros de las subpoblaciones, en cuyo caso ξ_j puede asimilar a a_j y $\sum_j \xi_j = 1$; o bien el parámetro de la población es el promedio ponderado de los parámetros Λ_j de las subpoblaciones, y debe satisfacerse que $\sum_j a_j = a$.

En virtud de que los valores de Y tienen distribución exponencial, su parámetro Λ puede interpretarse como la intensidad de un proceso de Poisson, de manera que para obtener información sobre su valor puede

procederse de dos formas: medir directamente el "tiempo" entre la ocurrencia de dos eventos sucesivos o contar cuántos de estos eventos ocurren en un "tiempo" determinado. Seguiremos el segundo camino por presentar ventajas de sencillez analítica.

Sea n_j el número de eventos observados durante el tiempo t_j en la j -ésima subpoblación. La probabilidad de que ocurra el evento conjunto de la observación en las N subpoblaciones es⁶

$$P_{\epsilon|\underline{\xi},\Lambda} = \left\{ \prod_j \frac{(\xi_j \Lambda t_j)^{n_j}}{n_j!} \right\} \exp\{-\sum_j \xi_j \Lambda t_j\} \quad (4.2)$$

es decir, el producto de las probabilidades de que se presente el evento correspondiente a cada subpoblación. En esta expresión, $\underline{\xi}$ es el vector de ξ_j , $j = 1, \dots, N$. Además, la cantidad $\sum_j \xi_j t_j$ es constante dada la independencia entre Λ y ξ_j , e igual a $\sum_j t_j$ por las condiciones impuestas a $\underline{\xi}$.

Así, la verosimilitud del evento, dado $\underline{\xi}$ vale

$$L_{\epsilon|\underline{\xi}} = \left[\prod_j (\xi_j \Lambda)^{n_j} \right] e^{-\Lambda \sum_j \xi_j t_j}$$

Si hacemos $\eta_j = \xi_j t_j / t$ con $t = \sum_j t_j$, $\sum_j \eta_j = 1$, puede verse que el exponente de e se convierte en $-\Lambda t$, y como $\eta_j \propto \xi_j$,

$$L_{\epsilon|\underline{\eta}} = \prod_j \eta_j^{n_j} \quad (4.3)$$

La forma de esta verosimilitud sugiere asignar a $\underline{\eta}$ una distribución conjunta de Dirichlet⁷ como distribución previa, por ser una conjugada natural de $L_{\epsilon|\underline{\eta}}$. Diremos entonces que

$$p_{\underline{\eta}}'(\eta_1, \eta_2, \dots, \eta_N) = \Gamma(m') \prod_j \frac{\eta_j^{m'_j - 1}}{\Gamma(m'_j)} \quad (4.4)$$

donde m'_j , $j = 1, \dots, N$ son los parámetros previos de la distribución y $m' = \sum_j m'_j$. Tendremos que

$$\begin{aligned} \underline{\eta}'' &= L_{\epsilon|\eta} \underline{\eta}' \\ \underline{\eta}' &= \prod_j \eta_j^{m'_j - 1} \\ \underline{\eta}'' &= \prod_j \eta_j^{n_j} \eta_j^{m'_j - 1} = \prod_j \eta_j^{m''_j - 1} \end{aligned} \quad (4.5)$$

donde $m''_j = m'_j + n_j$. De la ec 4.5 se concluye que

$$p_{\underline{\eta}}''(\eta_1, \eta_2, \dots, \eta_N) = \Gamma(m'') \prod_j \frac{\eta_j^{m''_j - 1}}{\Gamma(m''_j)} \quad (4.6)$$

con $m'' = \sum_j m''_j$. Es posible demostrar⁷ que la distribución marginal de cada una de las η_j es:

$$p_j''(\eta_j) = \frac{1}{B} \eta_j^{m''_j - 1} (1 - \eta_j)^{m'' - m''_j - 1} \quad (4.7)$$

$$B = \frac{\Gamma(m''_j) \Gamma(m'' - m''_j)}{\Gamma(m'')} \quad (4.8)$$

Es decir, una distribución Beta. Por tanto,

$$E''\eta_j = m''_j/m'' \quad (4.9)$$

$$c''^2\eta_j = \frac{m'' - m''_j}{m''_j (m'' + 1)} \quad (4.10)$$

Como $\eta_j = \xi_j t_j / t$,

$$\begin{aligned} E''\xi_j &= E''\eta_j t''/t''_j \\ &= (m''_j/m'') (t''/t''_j) \end{aligned} \quad (4.11)$$

con $t'' = t' + \sum_j t_j$ y $t''_j = t'_j + t_j$. En esta expresión $\sum_j t'_j = t'$ para la condición $\sum_j \eta_j = 1$. Además,

$$c''^2\xi_j = c''^2\eta_j \quad (4.12)$$

De la ec 4.1 y la independencia entre Λ y ξ_j ,

$$\begin{aligned} E''\Lambda_j &= E''\xi_j E''\Lambda \\ c''^2\Lambda_j &= c''^2\xi_j + c''^2\xi_j c''^2\Lambda + c''^2\Lambda \end{aligned}$$

Donde $E''\Lambda$ y $c''^2\Lambda$ se calculan como en el inciso 1:

$$E''\Lambda = r''/t'' \quad (4.13)$$

$$r'' = r' + \sum_j r_j \quad (4.14)$$

$$c''^2\Lambda = 1/r'' \quad (4.15)$$

Por tanto,

$$\begin{aligned} E''\Lambda_j &= (m''_j/m'') (t''/t''_j) (r''/t'') \\ &= (m''_j/m'') (r''/t''_j) \end{aligned} \quad (4.16)$$

$$c''^2\Lambda_j = \frac{m'' - m''_j}{m''_j (m'' + 1)} \left(1 + \frac{1}{r''}\right) + \frac{1}{r''} \quad (4.17)$$

En numerosos problemas de interés práctico procede tomar $E'\eta_j$ proporcional a la importancia relativa de la subpoblación correspondiente, aunque esta condición no es indispensable. Si se hace esta suposición, t'_j y m'_j también deberán ser proporcionales al tamaño de la subpoblación, como se desprende de las condiciones $\sum_j m'_j = m'$ y $\sum_j t'_j = t'$. En todo caso, $m'_j, t'_j \propto E'\eta_j, j = 1, \dots, N$.

Nótese que para tener definidas las distribuciones previa y posterior de \underline{n} es necesario contar con un valor de m' . Este parámetro crece con el grado de correlación entre las η_j y puede estimarse considerando que la relación entre Λ_i y $\Lambda_j, i \neq j$, es tal que puede probarse que

$$\rho'_{ij} = \frac{c'^2\Lambda}{c' \Lambda_i c' \Lambda_j} \quad (4.18)$$

donde ρ'_{ij} es el coeficiente de correlación entre Λ_i y Λ_j . De la ec 4.17, si se sustituyen las dobles primas con primas, es posible determinar la relación de m' con ρ'_{ij} . En particular, si todas las subpoblaciones son iguales, se llega a que

$$m' = \frac{(N - 1) (1/c'^2\Lambda + 1)}{1/\rho'_{ij} - 1} - 1 \quad (4.19)$$

Nótese que si todas las subpoblaciones tienen el mismo tamaño, es decir, $E'\xi_i = E'\xi_j = 1/N$, $i, j = 1, 2, \dots, N$, la distribución de Dirichlet obliga a que todos los coeficientes de correlación entre parejas de Λ_j sean iguales. Nótese también que no todas las posibles combinaciones de coeficientes de correlación entre parejas de Λ_j son admisibles con el esquema de Dirichlet cuando las subpoblaciones tienen diferentes tamaños. En este caso, si se sustituyen las ecs 4.13 y 4.14 en la 4.15 -cambiando dobles primas por primas-, es posible calcular $\rho'_{i,j}$ para una m' dada y para una pareja Λ_i, Λ_j . Si se efectúa este cálculo para todas las parejas Λ_i y Λ_j y para diferentes valores de m' , podrá obtenerse la combinación admisible de coeficientes de correlación que mejor refleje el conocimiento previo que el decisor tiene sobre las relaciones entre las subpoblaciones. De no existir una combinación adecuada, no será posible utilizar la distribución de Dirichlet como conjugada natural. La ref 8 presenta un tratamiento alternativo que puede usarse en estos casos y la ref 6 proporciona criterios para estimar m' cuando se procesa información sísmica.

Desde este enfoque, en general, las Λ_j no tendrán distribución Gamma. Si $r'' > m'' > m''_j + 1$, la densidad de probabilidad de Λ_j será una combinación lineal de distribuciones Gamma con la misma t'' pero diferentes r'' . Sólo en el caso en que $m'' = r''$ y $m''_j = r'' - 1$ Λ_j tendrá distribución $G(r'' - 1, t'')$. Sin embargo, pueden calcularse tantos momentos de Λ_j como se desee, y siempre podrá calcularse numéricamente la densidad de probabilidad. También puede sustituirse la densidad de Λ_j por concentraciones⁵ sin cometer gran error.

5. EJEMPLOS NUMERICOS

A continuación se presentan ejemplos numéricos que ilustran la teoría que se desarrolló en los incisos anteriores.

El ejemplo 1 está dedicado al procesamiento de datos que pueden no pertenecer a la población de datos válidos; en particular, se presenta el uso de las ecs 2.12-2.16. Los resultados se comparan con los que se obtendrían del análisis Bayesiano convencional.

En el ejemplo 2 se presenta el tratamiento de datos cuyos valores reales se encuentran afectados por errores. Los resultados son comparados con los que se tendrían si se ignora la componente de error. En este ejemplo se considera que todos los datos son válidos.

En el ejemplo 3 se ilustra un caso similar al anterior, pero en el que cada dato tiene una probabilidad diferente de cero de ser falso.

Los ejemplos 2 y 3 ilustran el uso de la ec 3.33.

El ejemplo 4 trata el caso de información proveniente de subpoblaciones parcialmente correlacionadas cuando las subpoblaciones son de igual tamaño. El ejemplo 5 trata el caso más general de subpoblaciones de diferente tamaño. En el ejemplo 4 se comparan los resultados con los que se hubieran obtenido de haber ignorado la correlación entre subpoblaciones. En el ejemplo 5 se presenta el cálculo de la distribución posterior de los valores de una subpoblación, señalando las simplificaciones que pueden hacerse para disminuir la complejidad analítica y comparando los resultados con los de un cálculo exacto.

EJEMPLO 1

Una empresa investiga las propiedades de duración de cierto componente electrónico cuya vida se supone de distribución exponencial con parámetro desconocido. El analista, de acuerdo con su experiencia en componentes similares, decide asignar al parámetro una distribución Gamma con $r'_y = 7$ y $t'_y = 6$ años; es decir, $\Lambda'_y \stackrel{d'}{=} G(7, 6)$. Por tratarse de un artefacto sumamente costoso sólo podrán ser ensayados tres especímenes, y los resultados serán reportados por un laboratorio externo al cual el analista asigna una credibilidad previa de 0.9 para cualquier dato reportado. Asimismo, el analista considera que de haber datos falsos, el parámetro Λ'_z tendría distribución Gamma con parámetros $r'_z = 4$ y $t'_z = 12$ años. El laboratorio externo reporta $y_1 = 8$; $y_2 = 3$, $y_3 = 8$; y_i en años. Se pide la distribución posterior de la vida del componente.

$$N = 3; 2^N = 8$$

$$\text{-- para } j = 1 \quad \{y\}_{k1} = \phi, \{y\}_{\ell_1} = \{1, 3, 8\}; Q'_{\{1\}} = 0.1^3$$

$$n_{y_1} = 0$$

$$n_{z_1} = 3$$

$$r''_{y_1} = 7 + 0 = 7$$

$$r''_{z_1} = 3 + 4 = 7$$

$$t''_{y_1} = 6 + 0 = 6$$

$$t''_{z_1} = 12 + 12 = 24$$

$$DQ''_1 = \frac{\Gamma(7) \Gamma(7)}{24^7 6^7} 0.1^3 = 4.0376 \times 10^{-13}$$

TESIS CON
FALLA DE ORIGEN

$$j = 2 \quad \{y\}_{k2} = \{1\}; \{y\}_{\ell2} = \{3, 8\}; Q'_2 = 0.9 \times 0.1^2 = 9 \times 10^{-3}$$

$$n_{y2} = 1 \quad n_{z2} = 2$$

$$r''_{y2} = 8 \quad r''_{z2} = 6$$

$$t''_{y2} = 7 \quad t''_{z2} = 23$$

$$DQ''_2 = \frac{\Gamma(8) \Gamma(6)}{7^8 23^6} \times 9 \times 10^{-3} = 6.3783 \times 10^{-12}$$

$$j = 3 \quad \{y\}_{k3} = \{3\}; \{y\}_{\ell3} = \{1, 8\}; Q'_3 = 9 \times 10^{-3}$$

$$DQ''_3 = \frac{\Gamma(8) \Gamma(6)}{9^8 21^6} \times 9 \times 10^{-3} = 1.4743 \times 10^{-12}$$

$$j = 4 \quad \{y\}_{k4} = \{8\}; \{y\}_{\ell4} = \{1, 3\}; Q'_4 = 9 \times 10^{-3}$$

$$DQ''_4 = \frac{\Gamma(8) \Gamma(6)}{14^8 16^6} \times 9 \times 10^{-3} = 2.1984 \times 10^{-13}$$

$$j = 5 \quad \{y\}_{k5} = \{1, 3\}; \{y\}_{\ell5} = \{8\}; Q'_5 = 0.9^2 \times 0.1 = 8.1 \times 10^{-2}$$

$$DQ''_5 = \frac{\Gamma(9) \Gamma(5)}{10^9 20^5} \times 8.1 \times 10^{-2} = 2.4494 \times 10^{-11}$$

$$j = 6 \quad \{y\}_{k6} = \{1, 8\}; \{y\}_{\ell6} = \{3\}; Q'_6 = 8.1 \times 10^{-2}$$

$$DQ''_6 = \frac{\Gamma(9) \Gamma(5)}{15^9 15^5} \times 8.1 \times 10^{-2} = 2.6850 \times 10^{-12}$$

$$j = 7 \quad \{y\}_{k7} = \{3, 8\}; \{y\}_{\ell7} = \{1\}; Q'_7 = 8.1 \times 10^{-2}$$

$$DQ''_7 = \frac{\Gamma(9) \Gamma(5)}{19^9 13^5} \times 8.1 \times 10^{-2} = 1.7802 \times 10^{-12}$$

TESIS CON
FALLA DE ORIGEN

$$j = 8 \quad \{y\}_{k8} = \{1, 3, 8\}; \{y\}_{\lambda 8} = ; Q'_8 = 0.9^3$$

$$DQ''_8 = \frac{\Gamma(10)}{18^{10}} \frac{\Gamma(4)}{12^4} \times 0.9^3 = 2.1438 \times 10^{-22}$$

$$D = 5.8873 \times 10^{-11}$$

Entonces,

$$p''_Y(\cdot) = 0.00686 P(7, 6) + 0.10834 P(8, 7) + 0.02504 P(8, 9) + \\ + 0.00373 P(8, 14) + 0.41605 P(9, 10) + 0.04561 P(9, 15) + \\ + 0.03024 P(9, 17) + 0.36414 P(10, 18)$$

por lo que $E''Y = 1.55$ años

$$\text{VAR}''Y = 3.43 \text{ años}^2$$

$$c''Y = 1.19$$

Si se hubiese considerado que $Q'_i = 1$, $i = 1, 2, 3$; es decir, análisis Bayesiano convencional, se tendría:

$$E''Y = 2 \text{ años}$$

$$\text{VAR}''Y = 5 \text{ años}^2$$

$$c''Y = 1.12$$

EJEMPLO 2

En una situación similar a la del ejemplo anterior, el analista considera que si bien todos los datos reportados serán válidos, es posible que el aparato que simula grandes duraciones en la prueba acelerada de vida proporcione lecturas equivocadas. Se consulta al experto del laboratorio, quien considera que la máquina no produce errores sistemáticos; asevera también que el coeficiente de variación de los errores es 0.155 y que aunque esta estimación fuera equivocada, el coeficiente de variación difícilmente podría exceder a 0.2. El analista considera que $r'_u = 5$ y $t'_u = 6$ años. El laboratorio ensaya dos especímenes y reporta $y_1 = 1$, $y_2 = 3$, ambos en años. Se pide la distribución posterior de la vida del componente.

Con los datos proporcionados por el experto, el analista procede de la siguiente forma: como no existen errores sistemáticos, $E'W = 1$ y $w_0 = 1$. Si la estimación del coeficiente de variación fuera equivocada, se sabe que casi seguramente no sería mayor que 0.2. En estas condiciones, y como $E'W = 1$, el analista hace $w_1 = 0.8$ y $w_2 = 1.2$. De $c'W = 0.155$ concluye que $VAR'W = 0.024$ y con estos valores, y la aplicación de las ecs 3.3-3.6, obtiene que $P'_{00} = 0.6$, $P'_{01} = 0.4$, $P_1 = 0.5$, $P_2 = 0.5$.

a) Si $P_0 = 0$

$$\begin{aligned}
 j = 1 & \quad C_{11} = \{0.8, 0.8\}; \quad U_{11} = \{1/0.8, 3/0.8\} = \{1.25, 3.75\} \\
 j = 2 & \quad C_{12} = \{0.8, 1.2\}; \quad U_{12} = \{1.25, 2.5\} \\
 j = 3 & \quad C_{21} = \{1.2, 0.8\}; \quad U_{21} = \{0.833, 3.75\} \\
 j = 4 & \quad C_{22} = \{1.2, 1.2\}; \quad U_{22} = \{0.833, 2.5\}
 \end{aligned}$$

$$P_1 = P_2 = P_3 = P_4 = 0.25; r_u'' = 5 + 2 = 7$$

$$S_1 = 1.25 + 3.75 = 5 \quad ; \quad t_{u1}'' = 5 + 6 = 11$$

$$S_2 = 1.25 + 2.5 = 6.25; \quad t_{u2}'' = 6.25 + 6 = 12.25$$

$$S_3 = 4.5833 \quad ; \quad t_{u3}'' = 10.5833$$

$$S_4 = 3.3333 \quad ; \quad t_{u4}'' = 9.3333$$

De 3.12,

$$p_{u_0}''(\cdot) = 0.25 P(7, 11) + 0.25 P(7, 12.25) + 0.25 P(7, 10.5833) + \\ + 0.25 P(7, 9.3333)$$

$$E''U_0 = 1.7986\frac{1}{2}$$

$$E''U_0^2 = 7.8363$$

$$\text{VAR}''U_0 = 4.6014$$

De acuerdo con la ec 3.33

$$p_{u_0}''(\cdot) \approx P(6.735, 10.315)$$

en virtud de que $10.315/5.735 = 1.7986 = E''U_0$ y

$$(10.315/5.735)^2 \times 6.735/4.735 = 4.6014 = \text{VAR}''U_0$$

Nótese que $\sum_k P_k \frac{r_k''}{(r_k'' - 2)} = 1.4$ y $6.735/4.735 = 1.4224$;

$1.4224/1.4 = 1.02 \approx 1$. Por tanto, la aproximación es adecuada.

b) Si $P_0 = 1$, $r_u'' = 7$, $t_u'' = 6 + 4 = 10$

$$p_{u_1}''(\cdot) = P(7, 10)$$

$$E''U_1 = 1.6667$$

$$E''U_1^2 = 6.6667$$

$$\text{VAR}''U_1 = 3.8889$$

c) Cálculo de L_{Y_0}

De acuerdo con las ecs 3.36 y 3.37,

$$p_{Y_0}''(y) = \frac{P_1}{w_1} p_{u_0}''(y/w_1) + \frac{P_2}{w_2} p_{u_0}''(y/w_2)$$

$$p_{Y_0}''(\cdot) \approx P_1 P(6.735, 0.8 \times 10.315) + P_2 P(6.735, 1.2 \times 10.315)$$

$$p_{Y_0}''(\cdot) \approx 0.5 P(6.735, 8.252) + 0.5 P(6.735, 12.378)$$

$$p_{Y_0}''(\cdot) \approx P(5.8514, 8.7257)$$

y de acuerdo con la ec 3.23,

$$L_{Y_0} = \prod_{j=1}^2 p_{Y_0}''(y_j)$$

$$L_{Y_0} = (0.5 \times 0.3369 + 0.5 \times 0.2983) (0.5 \times 0.0741 + 0.5 \times 0.1015)$$

$$L_{Y_0} = 0.0279$$

d) Cálculo de L_{Y_1}

De la ec 3.24, y como $w_0 = 1$,

$$p_{Y_1}''(y) = p_{u_1}''(y)$$

$$p_{Y_1}''(\cdot) = P(7, 10)$$

y de acuerdo con la ec 3.25,

$$\begin{aligned} L_{Y_1} &= \prod_{j=1}^2 p_{Y_1}''(y_j) \\ &= 0.3266 \times 0.0858 \\ &= 0.0280 \end{aligned}$$

e) Cálculo de P''_{00} y P''_{01}

De las ecs 3.19-3.21

$$P''_{00} = 0.5991$$

$$P''_{01} = 0.4009$$

f) Distribución final de U

$$\begin{aligned} E''U &= P''_{00} E''U_0 + P''_{01} E''U_1 \\ &= 0.5991 \times 1.7986 + 0.4009 \times 1.6667 \\ &= 1.746 \text{ años} \end{aligned}$$

$$\begin{aligned} E''U^2 &= P''_{00} E''U_0^2 + P''_{01} E''U_1^2 \\ &= 0.5991 \times 7.8363 + 0.4009 \times 6.6667 \\ &= 7.367 \text{ años}^2 \end{aligned}$$

$$\text{VAR}''U = 7.367 - 1.746^2$$

$$= 4.318 \text{ años}^2$$

$$c''U = 1.190$$

$$p''_u(u) = P''_{00} p''_{u_0}(u) + P''_{01} p''_{u_1}(u)$$

$$p''_u(\cdot) \approx 0.5991 P(6.375, 10.315) + 0.4009 P(7, 10)$$

Aplicando la ec 3.33,

$$p''_u(\cdot) \approx P(6.7899, 10.1074)$$

Si se hubiese despreciado la posibilidad de tener errores, se habría
llegado a que

$$E^U = 1.667 \text{ años}$$

$$\text{VAR}^U = 3.889 \text{ años}^2$$

$$c^U = 1.183$$

TESIS CON
FALLA DE ORIGEN

EJEMPLO 3

En una situación similar a la del ejemplo 2, el analista estima que cada dato tiene una probabilidad de 0.9 de ser verdadero. El resto de los valores numéricos permanece inalterado, y se pide la distribución posterior de U , considerando que $r'_z = 4$ y $t'_z = 12$ años.

a) Los dos valores son verdaderos

Del ejemplo 2,

$$p''_u(\cdot) \approx P(6.7899, 10.1074)$$

$$p''_y(\cdot) \approx 0.5991 P(5.8514, 8.7257) + 0.4009 P(7, 10)$$

Aplicando la ec 3.33,

$$p''_y(\cdot) \approx P(6.1739, 9.0321)$$

Como los dos valores son verdaderos, $p''_z(\cdot) = P(4, 12)$

b) Sólo y_1 es verdadero

$$r''_u = 5 + 1 = 6$$

$$c_1 = \{0.8\}; U_1 = \{1.25\}; P_1 = 0.5; S_1 = 1.25; t''_{u_1} = .6 + 1.25$$

$$c_2 = \{1.2\}; U_2 = \{0.833\}; P_2 = 0.5; S_2 = 0.833; t''_{u_2} = 6.833$$

$$p''_{u_0}(\cdot) = 0.5 P(6, 7.25) + 0.5 P(6, 6.833)$$

$$p''_{u_0}(\cdot) \approx P(5.9820, 7.0162)$$

ESTA TESIS NO SALE
DE LA BIBLIOTECA

$$\begin{aligned}
 p''_{y_0}(\cdot) &\approx 0.5 P(5.9820, 0.8 \times 7.0162) + 0.5 P(5.9820, 1.2 \times 7.0162) \\
 &\approx 0.5 P(5.9820, 5.613) + 0.5 P(5.9820, 8.4194) \\
 &\approx P(5.3204, 6.0844)
 \end{aligned}$$

$$p''_{u_1}(\cdot) = P(6, 7)$$

$$p''_{y_1}(\cdot) = P(6, 7)$$

$$L_{y_0} = p''_{y_0}(1) = 0.3342$$

$$L_{y_1} = p''_{y_1}(1) = 0.3366$$

$$P''_{00} = 0.5983; P''_{01} = 0.4017$$

$$\begin{aligned}
 p''_u(\cdot) &\approx 0.5983 P(5.9820, 7.0162) + 0.4017 P(6, 7) \\
 &\approx P(5.989, 7.0094)
 \end{aligned}$$

$$\begin{aligned}
 p''_u(\cdot) &\approx 0.5983 P(5.3204, 6.0844) + 0.4017 P(6, 7) \\
 &\approx P(5.5616, 6.4089)
 \end{aligned}$$

$$p''_z(\cdot) = P(5, 13)$$

c) Sólo y_2 es verdadero

$$\begin{aligned}
 p''_{u_0}(\cdot) &= 0.5 P(6, 9.75) + 0.5 P(6, 8.5) \\
 &\approx P(5.9083, 8.9577)
 \end{aligned}$$

$$\begin{aligned}
 p''_{y_0}(\cdot) &\approx 0.5 P(5.9083, 0.8 \times 8.9577) + 0.5 P(5.9083, 1.2 \times 8.9577) \\
 &\approx P(5.2669, 7.7871)
 \end{aligned}$$

$$p''_{u1}(\cdot) = P(6, 9)$$

$$p''_{y1}(\cdot) = P(6, 9)$$

$$L_{y0} = p''_{y0}(3) = 0.0882$$

$$L_{y1} = p''_{y1}(3) = 0.0890$$

$$P''_{00} = 0.5979; P''_{01} = 0.4021$$

$$p''_u(\cdot) = P(5.9432, 8.9716)$$

$$p''_y(\cdot) = P(5.4887, 8.1930)$$

$$p''_z(\cdot) = P(5, 15)$$

d) y_1 y y_2 son falsos

$$p''_u(\cdot) = P(5, 6)$$

$$p''_y(\cdot) = 0.5 P(5, 0.8 \times 6) + 0.5 P(5, 1.2 \times 6) \\ \approx P(4.5862, 5.3793)$$

$$p''_z(\cdot) = P(6, 16)$$

e) Cálculo de Q''_j con la ec 2.16

$$DQ''_1 = \frac{\Gamma(4) \Gamma(6.1739)}{12^4 \cdot 9.0321^{6.1739}} \times 0.9^2 = 4.7661 \times 10^{-8}$$

$$DQ''_2 = \frac{\Gamma(5) \Gamma(5.5616)}{13^5 \cdot 6.4089^{5.5616}} \times 0.09 = 1.0961 \times 10^{-8}$$

$$DQ_3'' = \frac{\Gamma(5) \Gamma(5.4887)}{15^5 8.193^{5.4887}} \cdot 0.09 = 1.417 \times 10^{-9}$$

$$DQ_4'' = \frac{\Gamma(6) \Gamma(4.5862)}{16^6 5.3793^{4.5862}} \times 0.01 = 4.1087 \times 10^{-10}$$

Por tanto, $D = 6.0457 \times 10^{-8}$ y

$$Q_1'' = 0.7883$$

$$Q_2'' = 0.1813$$

$$Q_3'' = 0.0234$$

$$Q_4'' = 0.0068$$

Además,

$$P_u''(\cdot) \approx 0.7883 P(6.7899, 10.1074) + 0.1813 P(5.989, 7.0094) + \\ + 0.0234 P(5.9432, 8.9716) + 0.0068 P(5, 6)$$

Por lo que, estando U en años,

$$E''U = 1.684$$

$$VAR''U = -4.103$$

$$c''U = 1.203$$

De haber despreciado errores y equivocaciones, se tendría

$$E''U = 1.667$$

$$VAR''U = 3.889$$

$$c''U = 1.183$$

TESIS CON
FALLA DE ORIGEN

De haber despreciado equivocaciones pero no errores,

$$E^{\text{U}} = 1.746$$

$$\text{VAR}^{\text{U}} = 4.318$$

$$c^{\text{U}} = 1.190$$

De haber despreciado errores pero no equivocaciones,

$$E^{\text{U}} = 1.602$$

$$\text{VAR}^{\text{U}} = 3.690$$

$$c^{\text{U}} = 1.199$$

TESIS CON
FALLA DE ORIGEN

TESIS CON
FALLA DE ORIGEN

EJEMPLO 4

Se desea afinar la estimación que se tiene sobre el parámetro β de una cierta región sísmica, donde se supone que la tasa de excedencia, λ , es de la forma $\lambda = \alpha(e^{-\beta M_0} - e^{-\beta M_u})$, siendo M_0 el umbral de magnitudes sísmicas, arriba del cual se tienen datos completos, M_u la magnitud máxima posible y α un parámetro conocido. Sea $\tilde{M}_i = M_i - M_0$, donde M_i es la magnitud del i -ésimo temblor ocurrido durante el intervalo de observación. En la ref 6 se demuestra que \tilde{M} tiene aproximadamente distribución exponencial con parámetro β^* . La región de interés se encuentra dividida en dos regiones, de áreas sensiblemente iguales y cuya similaridad tectónica hace pensar al experto que el coeficiente de correlación entre tasas de excedencia es 0.6. Además, el experto estima que $E'\beta = 2$ y $c'\beta = 0.2$. Se dispone de los registros de magnitudes que se muestran en la tabla 5.1 para ambas regiones, con $M_0 = 4$.

Se asignará a β una distribución Gamma, por lo que $r' = 25$, $t' = 12.5$. Aquí y en lo que sigue, todos los tiempos están en años. Como las regiones son iguales, $E'\eta_1 = E'\eta_2 = 0.5$, $t'_1 = t'_2 = 6.25$.

De la tabla 5.1, $\eta_1 = 12$, $\eta_2 = 12$, $t_1 = 5.76$, $t_2 = 4.93$, $t = t_1 + t_2 = 10.690$. Por tanto, $t'' = t' + t = 12.5 + 10.60 = 23.190$
 $r'' = 25 + 24 = 49$; $t''_1 = 12.010$, $t''_2 = 11.180$.

De la ec 4.13, $E''\beta = 49/23.190 = 2.113$; de la ec. 4.15, $c''^2\beta = 1/49 = 0.0204$.

* La distribución se trunca en $M = M_u - M_0$, pero si se trata con magnitudes no muy cercanas a la máxima posible, la aproximación es satisfactoria. Esto ocurre, por ejemplo, en el diseño de edificios medianos.

De la ec 4.19, $m' = 38$. De la igualdad entre subregiones, $m'_1 = m'_2 = 19$. $m'' = 38 + 24 = 62$; $m''_1 = m''_2 = 19 + 12 = 31$. Entonces, de la ec 4.16, $E'\beta_1 = (31/62) (49/12.010) = 2.040$, $E''\beta_2 = 2.191$; $c''^2\beta_1 = (1/63) (1 + 1/49) + 1/49 = 0.0366$ y $c''^2\beta_2 = 0.0366$.

Si se hubiese despreciado la correlacion, se habrian obtenido las mismas esperanzas pero $c''^2\beta_1 = c''^2\beta_2 = 0.0612$.

EJEMPLO 5

En una situación similar a la del problema 4, se considera que la región 1 contribuye con 70% de los sismos, por ser de mayor tamaño. Nuevamente el experto estima que $E'\beta = 2$, $c'\beta = 0.2$ y $\rho_{12} = 0.6$. Los registros de magnitudes se muestran en la tabla 5.2.

También en esta ocasión $\beta \stackrel{d}{=} G(25, 12.5)$. A partir de las ecs 4.17 y 4.18, se llega a que $m'_1 = 17.9$, $m'_2 = 0.7 \times 17.9 = 12.53$, $m' = 5.37$. $t'_1 = 0.7 \times 12.5 = 8.75$, $t'_2 = 3.75$.

De la tabla 5.2, $n_1 = 11$, $t_1 = 5.91$, $n_2 = 5$, $t_2 = 1.52$. Entonces, $r'' = 25 + 11 + 5 = 41$; $t = 7.43$, $t'' = 12.5 + 7.43 = 19.93$; $E''\beta = 41/19.93 = 2.057$, $c''\beta = 1/41$, $c''\beta = 0.156$ $m'' = 12.53 + 11 = 23.53$, $m''_1 = 5.37 + 5 = 10.37$, $m''_2 = 33.9$. $t''_1 = 8.75 + 5.91 = 14.66$, $t''_2 = 3.75 + 1.52 = 5.27$. Entonces, a partir de la ec 4.16, $E''\beta_1 = (23.53/33.9) (41/14.66) = 1.941$, $E''\beta_2 = (10.37/33.9) (41/5.27) = 2.38$.

De la ec 4.1, $c''\eta_1 = 0.0126$, $c''\eta_2 = 0.0650$. Sustituyendo en 4.17, $c''\beta_1 = 0.0373$, $c''\beta_2 = 0.0910$.

Nótese que

$$\begin{aligned} \beta_1 &= \xi_1 \beta \\ &= (t''/t''_1) \eta_1 \beta \\ &= 1.359 \eta_1 \beta \end{aligned}$$

con $\beta \stackrel{d}{=} G(41, 19.93)$ y $\eta_1 \stackrel{d}{=} Be(33.9, 23.53)$ donde $Be(\cdot, \cdot)$ denota la

distribución Beta. En estas condiciones, el cálculo de la densidad de probabilidad de β_1 requiere de integración numérica. La densidad correspondiente se presenta en la fig 5.1 y es aproximadamente igual a una distribución $G(26.81, 13.81)$. Entonces, la variable $\tilde{M} \stackrel{d}{=} \text{EXP}(\beta)$ tendrá a posteriori una distribución aproximadamente $P(26.81, 13.81)$ con $E\tilde{M} = 13.81/(26.81 - 1) = 0.535$, $\text{VAR}\tilde{M} = 0.309$, $c\tilde{M} = 1.040$.

TESIS CON
FALLA DE ORIGEN

CONCLUSIONES

Se ha hecho ver que cuando una muestra estadística contiene valores atípicos puede errarse gravemente tanto si se desprecia a estos valores como si se les da el mismo peso que a los que no son atípicos. La forma racional de proceder consiste en reconocer que cada valor tiene una probabilidad finita de ser falso. Se ha mostrado que para cada problema puede existir un rango en que se comete poco error al asignar a un valor validez total, otro rango en que el valor puede ignorarse y otro rango más en que es indispensable tomar en cuenta la probabilidad de que el valor sea válido.

Las figuras correspondientes a valores observados muy lejanos a la esperanza previa muestran que en las colas de la distribución puede errarse en la estimación de la confiabilidad por más de un orden de magnitud. Se recomienda ampliamente el uso de este enfoque en estos casos. Nótese que aunque el rango de valores para el que es necesario tomar en cuenta la credibilidad sólo puede determinarse a posteriori, las soluciones de Bayes clásico y la del presente enfoque coincidirían si el efecto de la incertidumbre en las observaciones no fuera importante.

Se destaca el hecho de que cuando se recurre al análisis Bayesiano, el efecto del tamaño de la muestra se refleja en la dispersión de la distribución de los parámetros inciertos y, a través de ésta, en la distribución de la variable de interés.

Se han examinado varias situaciones. Primero, un dato puede pertenecer a la población de datos válidos o a la de datos falsos. Se asignan probabilidades previas a cada situación. En cualquiera de los dos casos el dato tendrá distribución exponencial con parámetro desconocido. A estos parámetros se asigna una distribución previa de probabilidades gamma, que constituye una conjugada natural. La probabilidad de que cada valor sea válido se actualiza a la luz de los datos observados, lo cual permite actualizar la distribución de valores válidos.

En una segunda situación, los valores válidos pueden estar distorsionados por errores aleatorios. El error se considera como un factor multiplicativo aleatorio, con una distribución suficientemente rica. La distribución de valores válidos es actualizado también.

Finalmente se ha analizado el caso de subpoblaciones, en donde las correlaciones entre miembros de una subpoblación son mayores que entre miembros de diferentes subpoblaciones.

Se han desarrollado métodos aproximados para reducir el trabajo numérico. Se ha mostrado que en muchos casos las aproximaciones son suficientemente precisas.

Los planteamientos teóricos se ilustran con aplicaciones a confiabilidad de componentes sistemas y a procesos de Poisson, puesto que

los tiempos entre eventos tienen distribución exponencial. Aunque, como se muestra en la ref 2, las variables con distribución exponencial pueden ser transformadas en gaussianas y ser sometidas a la teoría desarrollada en ese trabajo, existen ventajas conceptuales, analíticas y numéricas que justifican el tratar con la distribución exponencial directamente.

En el presente trabajo se ha partido de la suposición de que hay razones suficientes para pensar que la variable de interés tiene distribución exponencial. Esto, en general, no es algo evidente, por lo que se llama la atención sobre las posibles implicaciones que acarrearía una inadecuada selección de distribuciones subyacentes. En general, la elección de distribuciones previas no tiene un impacto tan grande, sobre todo en problemas de decisión en que sólo interesan los primeros momentos de la variable.

Aunque se considera que la estadística Bayesiana constituye la mejor herramienta para enfrentar problemas como los que se han discutido, es necesario hacer mención de que la herramienta no puede extrapolar los límites de información contenida en los datos, y los resultados deben ser interpretados en esta inteligencia.

AGRADECIMIENTOS

Agradezco al Dr. Emilio Rosenblueth la orientación y apoyo recibidos durante la elaboración de este trabajo, sin las cuales no hubiera sido posible su realización. La mayor parte del trabajo fue elaborado durante una visita del autor a la Universidad de Waterloo, por lo que agradezco al Dr. Niels C. Lind su apoyo.

REFERENCIAS

1. Rosenblueth, E. "On the processing of doubtful information. Part 1: General Theory" Risk, Structural Engineering and Human Error M Grigorio (Ed), University of Waterloo Press, Waterloo, Ontario, Canadá (1984).
2. Rosenblueth, E y Ferregut, C "On the processing of doubtful information. Part 2: Gaussian Distributions" Risk, Structural Engineering and Human Error M Grigoriu (Ed), University of Waterloo Press, Waterloo, Ontario, Canada (1984)
3. Raiffa, H y Schlaifer, R Applied Statistical Decision Theory Harvard University Press, Cambridge, Mass (1961)
4. Benjamin J R y Cornell, C A Probability, Statistics and Decisions for Civil Engineers, McGraw-Hill Book Co, Nueva York (1970)
5. Rosenblueth, E "Point estimates for probability" Appl Math Modelling, 5, 329, (1981)
6. Rosenblueth, E "Use of statistical data in assessing local seismicity" aceptado para su publicación en Earthq Engnrg Strct. Dyn
7. Koroliu, V S Manual de la Teoría de Probabilidades y Estadística Matemática, MIR, Moscú (1981)
8. Rosenblueth, E y Ordaz, M "Use of seismic data from similar regions" sometido para su publicación en Earthq Engnrg Strct. Dyn

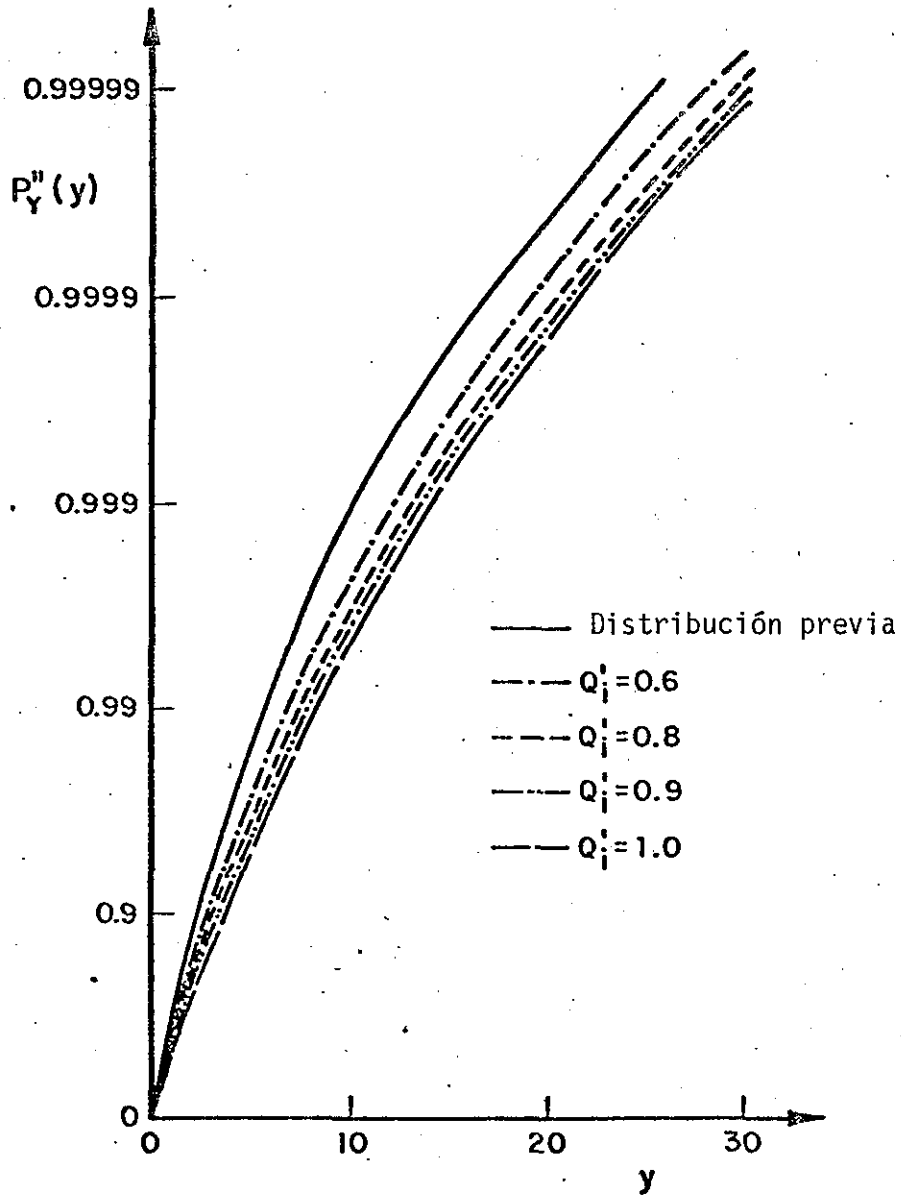


Fig 2.1 Densidades acumuladas $y_3=4$

TESIS CON
FALLA DE ORIGEN

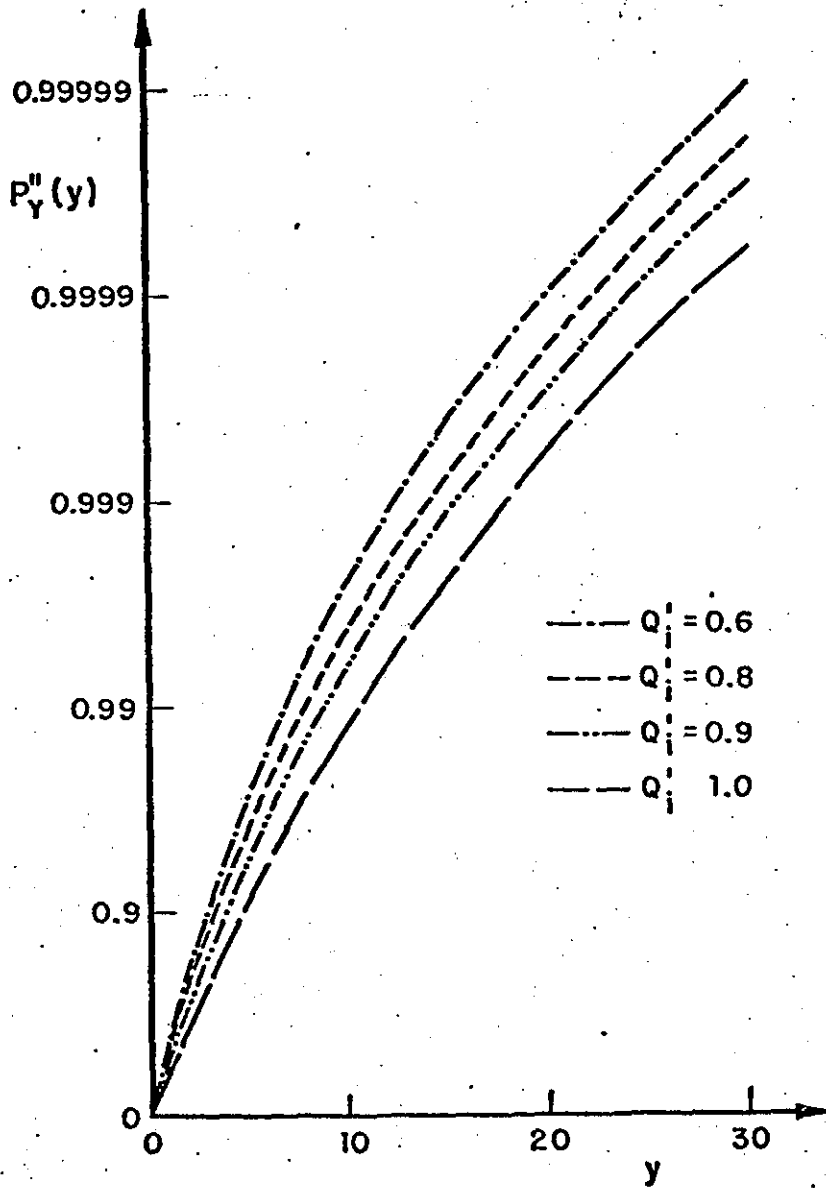


Fig 2.2 Densidades acumuladas $y_3=8$

TESIS CON
FALLA DE ORIGEN

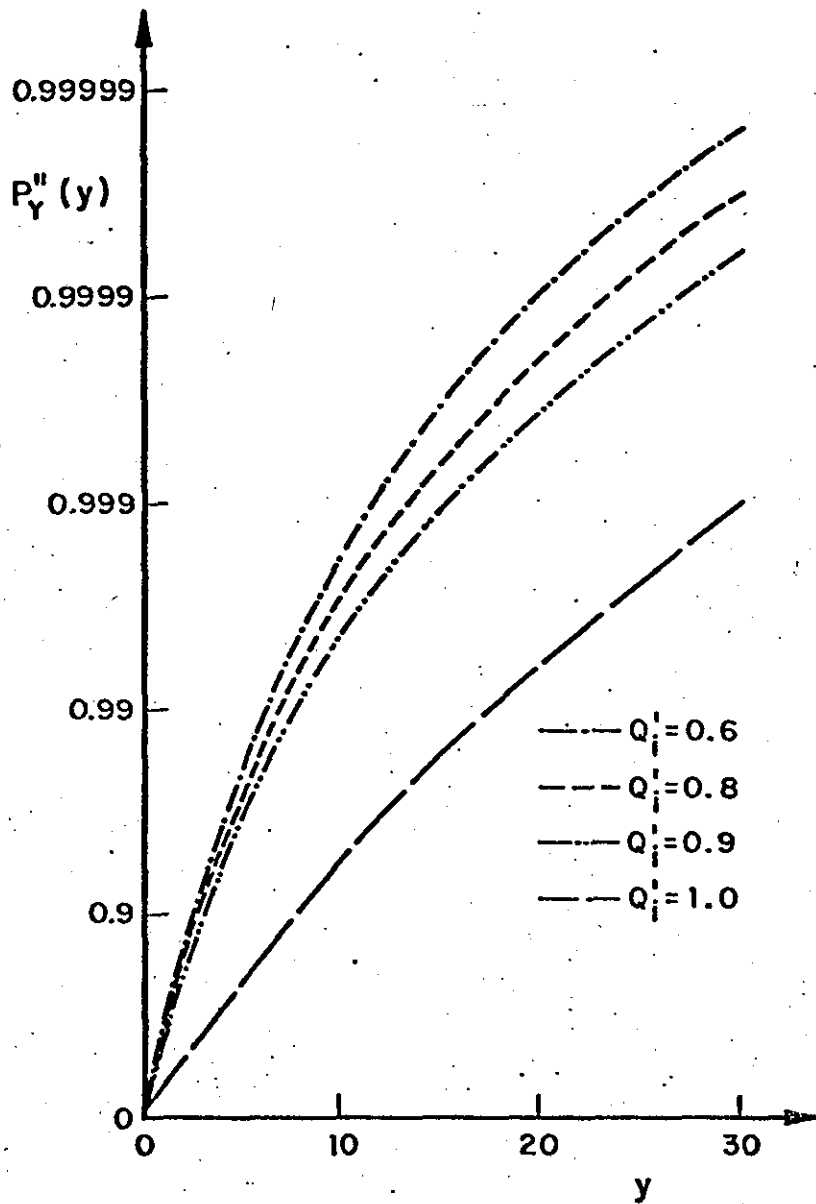


Fig 2.3 Densidades acumuladas $y_3=20$

TESIS CON
FALLA DE ORIGEN

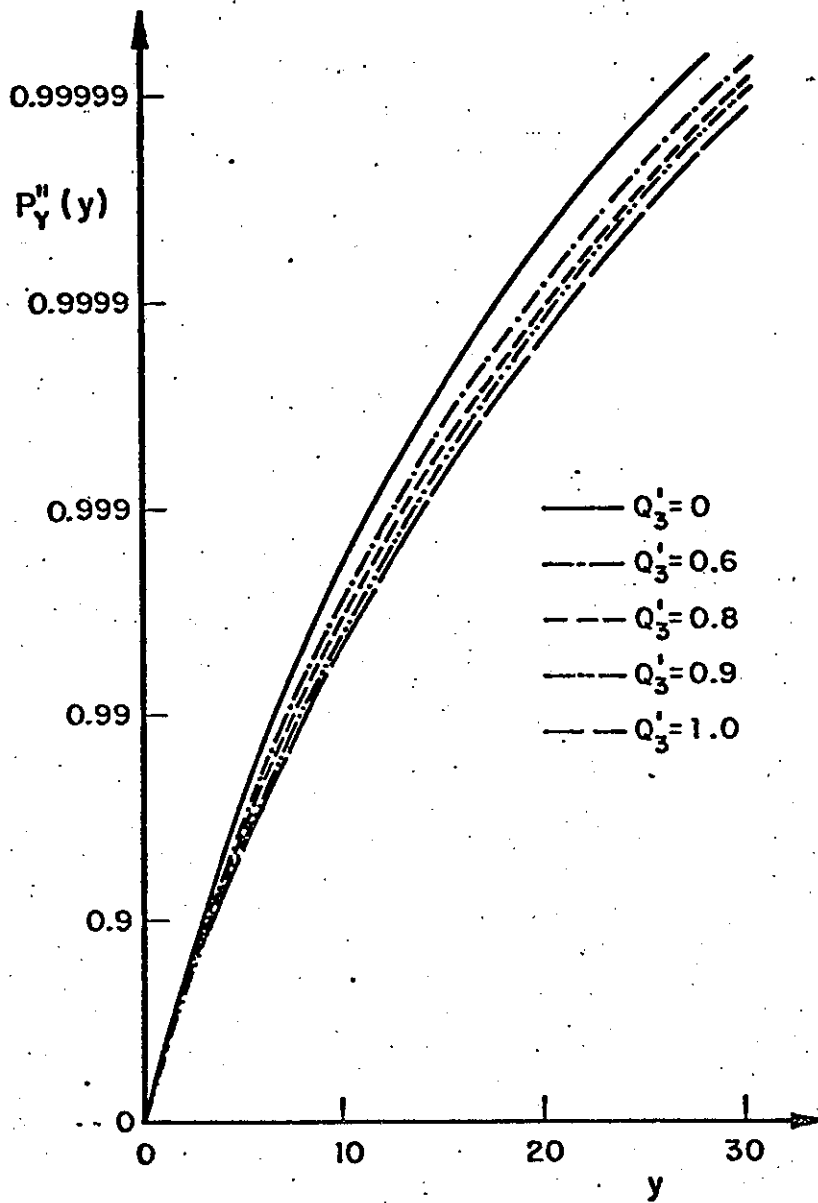


Fig 2.4 Densidades acumuladas $y_3=4$, $Q_1^1 = Q_2^1 = 1$

TESIS CON
FALLA DE ORIGEN

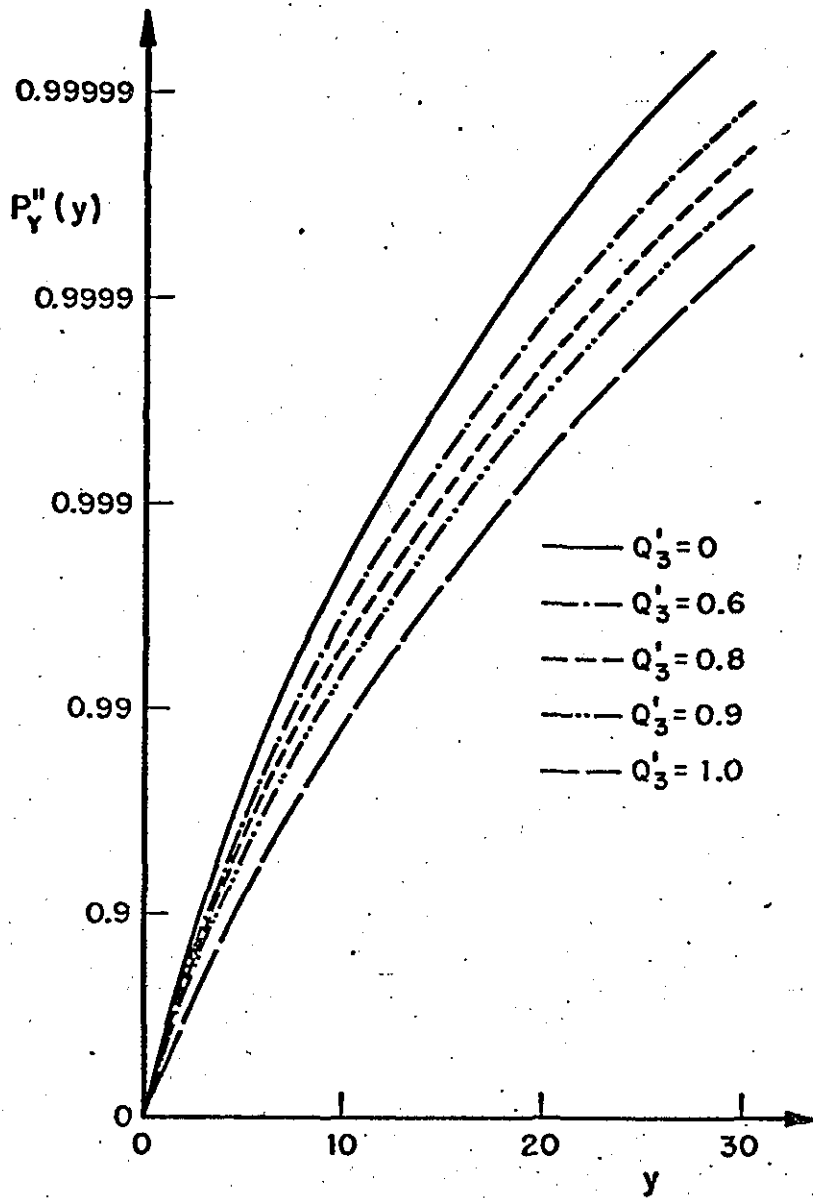


Fig 2.5 Densidades acumuladas $y_3=8$, $Q_1^1 = Q_2^1 = 1$

TESIS CON
FALLA DE ORIGEN

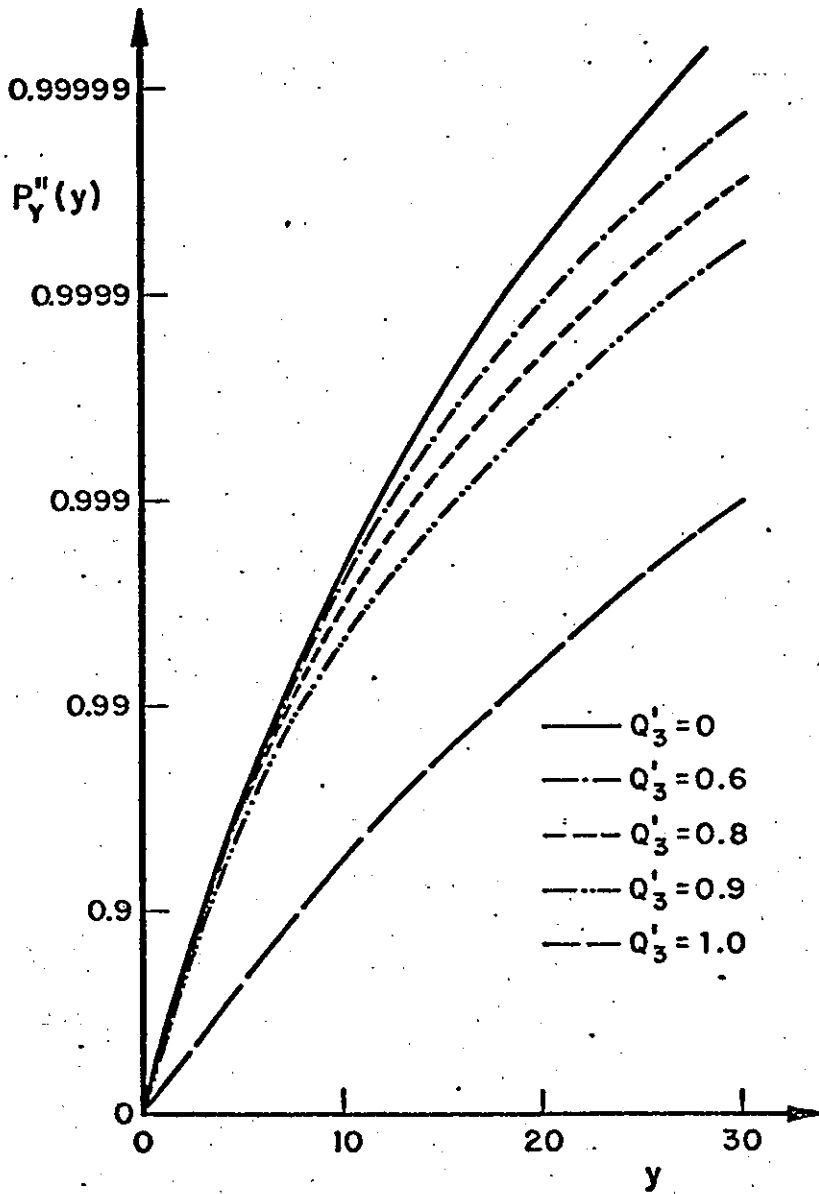


Fig 2.6 Densidades acumuladas $y_3=20$, $Q_1^i = Q_2^i = 1$

TESIS CON
FALLA DE ORIGEN

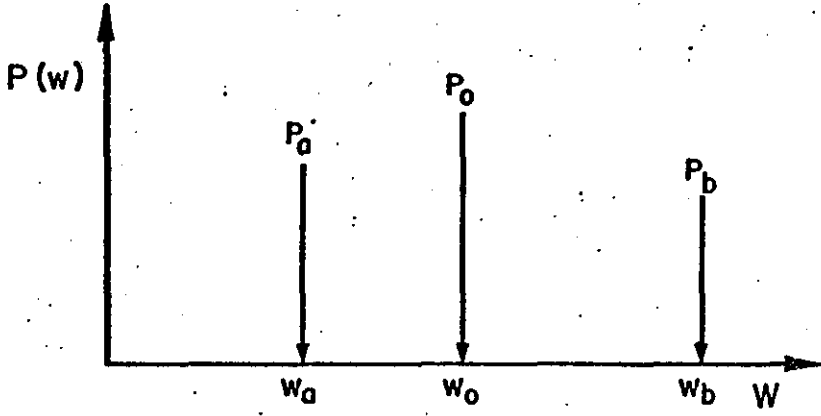


Fig 3.1 Distribución previa de W

TESIS CON
FALLA DE ORIGEN

60

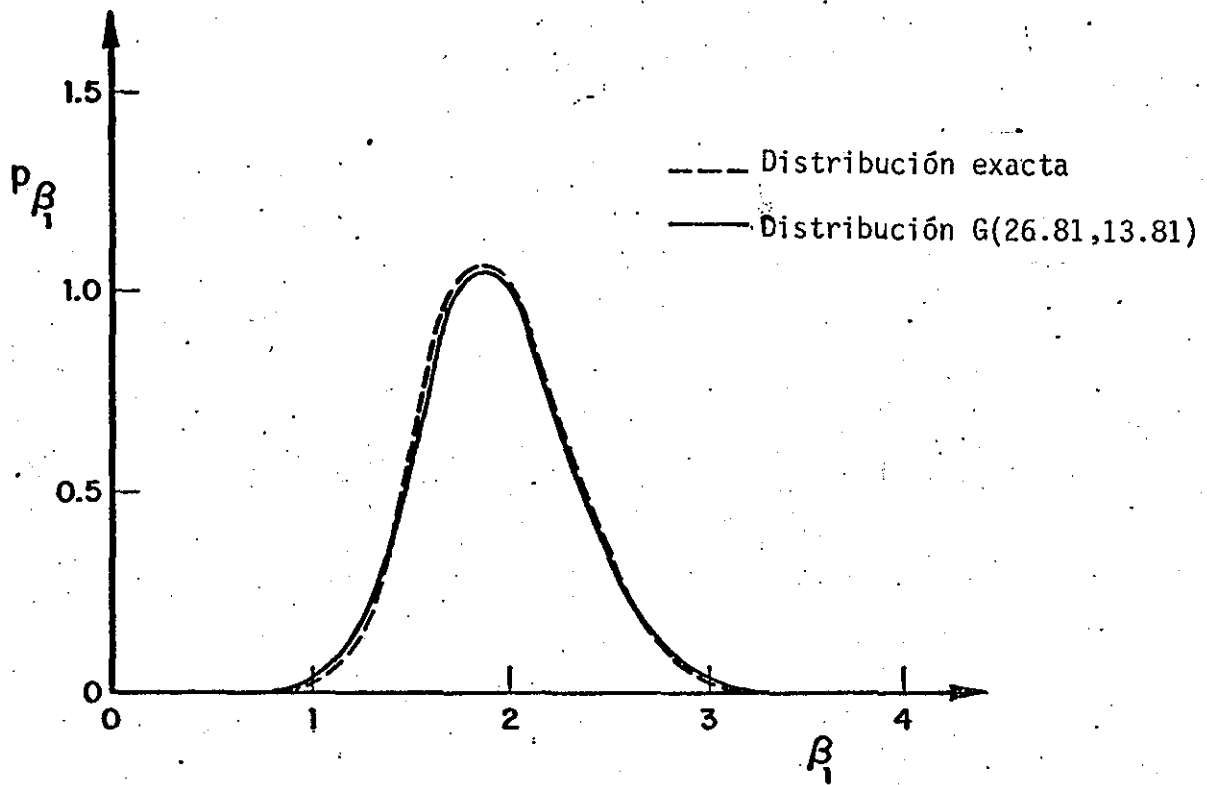


Fig 5.1 Ejemplo 5

TESIS CON
FALLA DE ORIGEN

Tabla 2.1 Credibilidades posteriores cuando $Q_1^1 = Q_2^1 = Q_3^1$

y_3	Q_i^1	Q_0''	Q_1''	Q_2''	Q_3''	Q_{12}''	Q_{13}''	Q_{23}''	Q_{123}''
4	0.9	.00108	.01117	.00526	.00489	.05281	.04645	.07969	.79866
8	0.9	.00170	.02107	.00651	.00393	.08118	.03378	.08451	.76733
20	0.9	.00414	.08011	.01628	.00278	.34656	.01703	.06909	.46401
4	0.8	.00929	.04279	.02013	.01873	.08992	.07908	.13568	.60436
8	0.8	.01387	.07646	.02361	.01426	.13093	.05449	.13631	.55007
20	0.8	.02373	.20414	.04148	.00708	.39248	.01929	.07824	.23355
4	0.6	.08106	.14000	.06587	.06129	.11031	.09701	.16644	.27802
8	0.6	.10585	.21890	.06758	.04083	.14056	.05850	.14633	.22144
20	0.6	.11883	.38331	.07789	.01329	.27635	.01358	.05509	.06167

Q_0 = probabilidad de que ningún valor sea válido

Q_i = probabilidad de que y_i sea válido

Q_{ij} = probabilidad de que y_i y y_j sean válidos

Q_{ijk} = probabilidad de que y_i, y_j y y_k sean válidos

Tabla 2.2 Credibilidades posteriores cuando $Q_1^1 = Q_2^1 = 1$

y_3	Q_3^1	Q_{12}''	Q_{123}''
4	1	0	1
8	1	0	1
20	1	0	1
4	0.9	.06203	.93797
8	0.9	.09567	.90433
20	0.9	.42755	.57245
4	0.8	.12952	.87048
8	0.8	.19227	.80773
20	0.8	.62693	.37307
4	0.6	.28407	.71593
8	0.6	.38829	.61171
20	0.6	.81756	.18244
4, 8, 20	0	1.000	0

TESIS CON
FALLA DE ORIGEN

TABLA 3.1 COMPARACION DE RESULTADOS AL EMPLEAR LA
 APROXIMACION DE LA EC 3.33 PARA p''_{u_0}

u / $p''_{u_0}(u)$	VALOR EXACTO	VALOR APROX.	ERROR, %
0	0.7504	0.7703	2.7
σ_{u_0}	0.09626	0.09418	-2.2
$2\sigma_{u_0}$	0.02262	0.02241	0.9
$3\sigma_{u_0}$	0.00737	0.007519	2.0
$4\sigma_{u_0}$	0.002999	0.003115	3.9

TABLA 3.2 COMPARACION DE RESULTADOS AL EMPLEAR LA
 APROXIMACION DE LA EC 3.33 PARA p''_{y_0}

y / $p''_{y_0}(y)$	VALOR EXACTO	VALOR APROX.	ERROR, %
0	0.7504	0.7520	0.2
σ_y	0.08659	0.08578	-0.9
$2\sigma_y$	0.02034	0.02005	-1.4
$3\sigma_y$	0.006742	0.006714	-0.4
$4\sigma_y$	0.002758	0.002791	1.2

Tabla 5.1 Catálogo de temblores para el ejemplo 4

	Región 1		Región 2	
	M_i	\hat{M}_i	M_i	\hat{M}_i
1	5.27	1.27	4.90	0.90
2	4.74	0.74	4.84	0.84
3	4.69	0.69	4.82	0.82
4	4.68	0.68	4.41	0.41
5	4.49	0.49	4.40	0.40
6	4.46	0.46	4.38	0.38
7	4.40	0.40	4.31	0.31
8	4.39	0.39	4.30	0.30
9	4.24	0.24	4.19	0.19
10	4.19	0.19	4.15	0.15
11	4.16	0.16	4.13	0.13
12	4.05	0.05	4.10	0.10
Σ		5.76		4.93

Tabla 5.2 Catálogo de temblores para el ejemplo 5

	Región 1		Región 2	
	M_i	\hat{M}_i	M_i	\hat{M}_i
1	5.07	1.07	4.68	0.68
2	4.72	0.72	4.54	0.54
3	4.68	0.68	4.15	0.15
4	4.64	0.64	4.08	0.08
5	4.57	0.57	4.07	0.07
6	4.56	0.56		
7	4.55	0.55		
8	4.36	0.36		
9	4.31	0.31		
10	4.25	0.25		
11	4.20	0.20		
Σ		5.91		1.52