Un Método Práctico para Obtener Distribuciones a Priori y su Aplicación en la Evaluación de Modelos de Confiabilidad

Dr. Humberto Gutiérrez Pulido, CUCEI, U. de Guadalajara
Dr. Víctor Aguirre-Torres, ITAM.
Dr. Andrés Christen, CIMAT.

Seminario Aleatorio, ITAM
Enero de 2004

Plan de la platica

- Conceptos básicos
- Génesis del trabajo
- Método general propuesto para definir las distribuciones a priori
- Aplicación a los modelos más usuales
- Validación del método
- Selección de modelos
- Conclusiones

Conceptos básicos

- La confiabilidad mide la duración de los productos
- Probabilidad de que el producto no falle
- Es costo generar datos (fallas)
- Experimentos censurados:
  - En el tiempo t_c cuando acabó el estudio no habían fallado (censura por la derecha), o
  - fallaron antes del tiempo t_c (censura por la izquierda), o bien,
  - se sabe que fallaron en cierto intervalo de tiempo (censura por intervalo)

Distribuciones posterior y predictiva

- La inferencia Bayesiana está basado en especificar un modelo $f(x|\boldsymbol{\theta})$ para los datos observados, X; dado un vector de parámetros $\boldsymbol{\theta}$ de valor desconocido.
- Generándose, $L(\mathrm{X}|\boldsymbol{\theta})$, que es la distribución conjunta condicional de X para un valor dado de $\boldsymbol{\theta}$.
- Se supone que $\boldsymbol{\theta}$ es una v.a. con distribución a priorio inicial $\pi(\boldsymbol{\theta})$.
- A través de $\pi(\boldsymbol{\theta})$ como se incorpora formalmente el conocimiento inicial

- La distribución posterior $\pi(\boldsymbol{\theta}\,|\,\mathbf{X})$ es la distribución condicional de $\boldsymbol{\theta}$ dado X

$$\pi(\boldsymbol{\theta}\,|\,\mathbf{X}) = \frac{L(\mathbf{X}\,|\,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{\int_{\Theta} L(\mathbf{X}\,|\,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

$$\pi(\boldsymbol{\theta}\,|\,\mathbf{X}) \propto L(\mathbf{X}\,|\,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$$

y

$$f(\mathbf{X}) = \int_{\Theta} L(\mathbf{X}\,|\,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

Cte. normalizadora de $\pi(\boldsymbol{\theta}\,|\,\mathbf{X})$

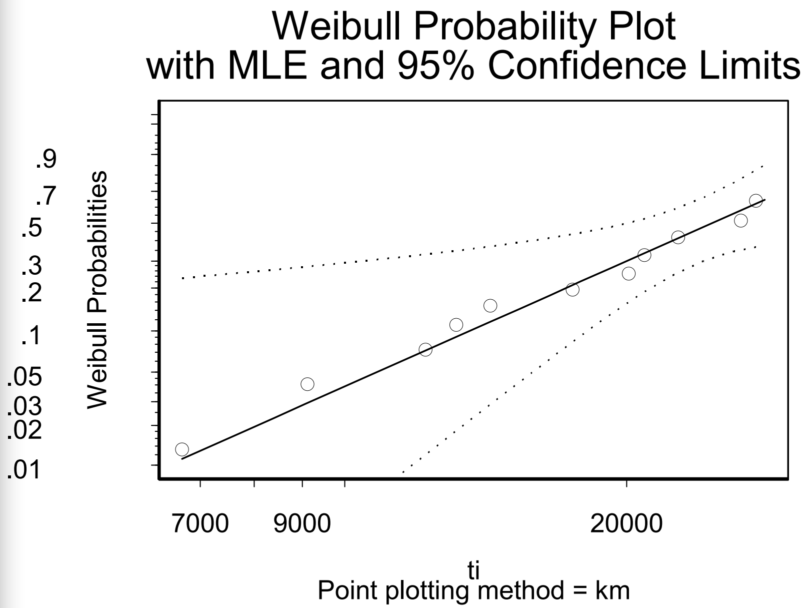
Génesis del Problema.

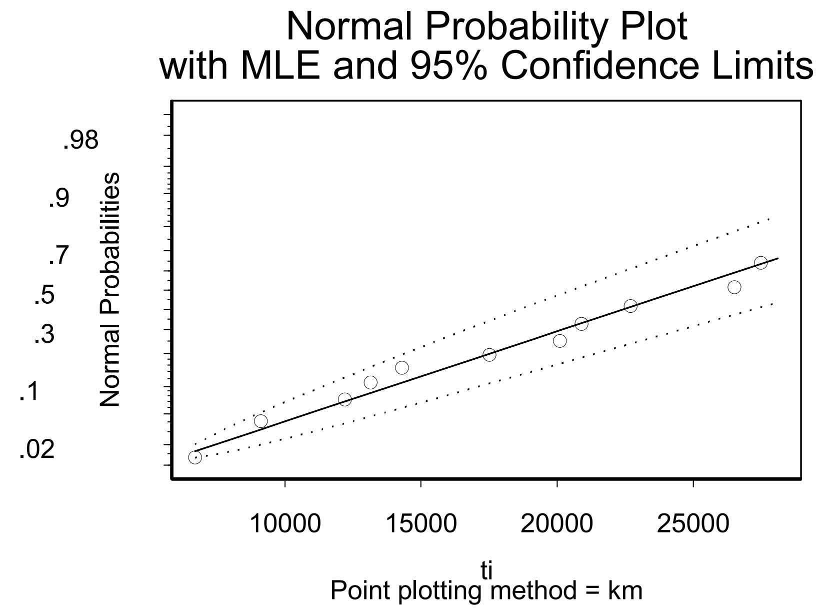
- Evaluación de la especificación de modelos comúnmente aplicados en confiabilidad, $f(x|\theta)$
- Gráficos de probabilidad, Meeker y Escobar (1998).
- Muestras censuradas y pequeñas.
- Bondad de ajuste, difícil.

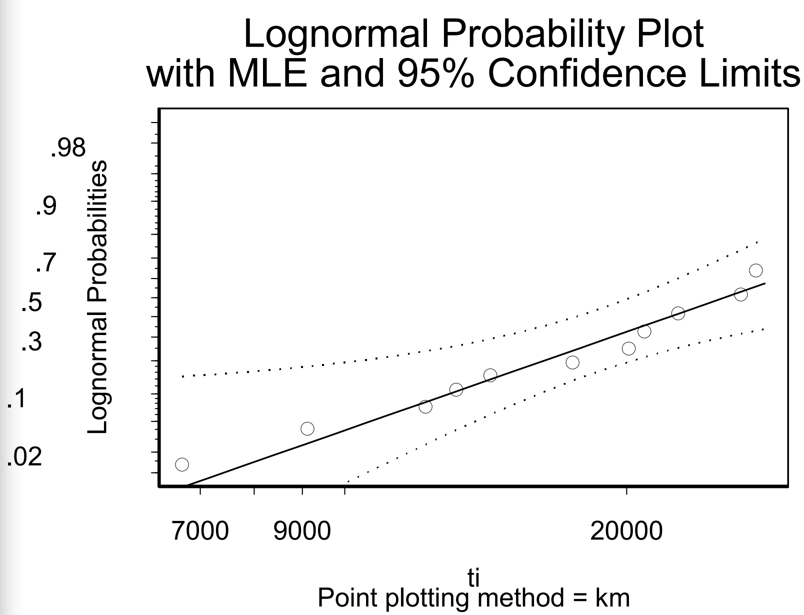
6700, 6950*, 7820*, 8790*, 9120, 9660*, 9820*, 11310*, 11690*, 11850*, 11880*, 12140*, 12200, 12870*, 13150, 13330*, 13470*, 14040*, 14300, 17520,

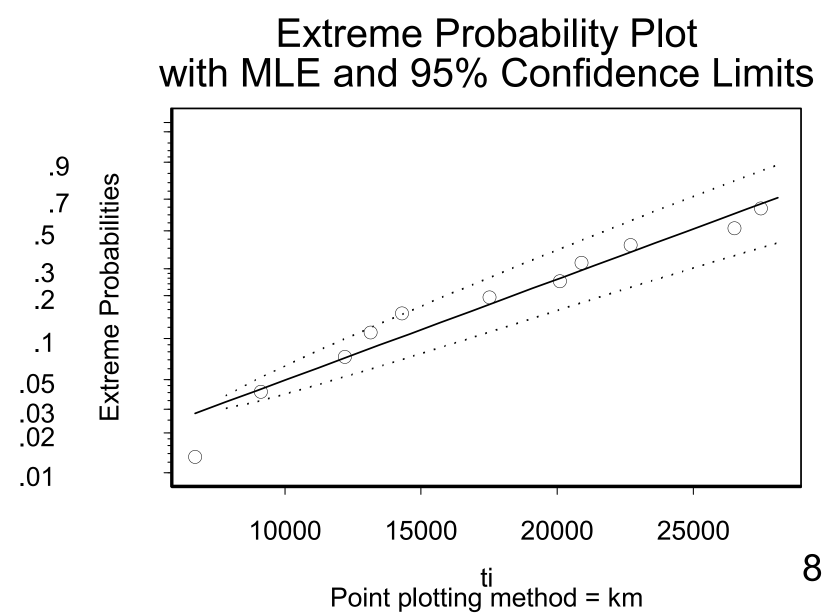
17540*, 17890*, 18420*, 18960*, 18980*, 19410*, 20100, 20100*, 20150*, 20320*, 20900, 22700, 23490*, 26510, 27410*, 27490, 27890*, 28100*

- Distancia de falla en 38 amortiguadores de vehículos. Se observaron 11 fallas.


Weibull Probability Plot
with MLE and 95% Confidence Limits
Weibull Probabilities
.9
.7
.5
.3
.2
.1
.05
.03
.02
.01
7000
9000
20000
ti
Point plotting method = km


Normal Probability Plot
with MLE and 95% Confidence Limits
Normal Probabilities
.98
.9
.7
.5
.3
.1
.02
10000
15000
20000
25000
ti
Point plotting method = km


Lognormal Probability Plot
with MLE and 95% Confidence Limits
Lognormal Probabilities
.98
.9
.7
.5
.3
.1
.02
7000
9000
20000
ti
Point plotting method = km


Extreme Probability Plot
with MLE and 95% Confidence Limits
Extreme Probabilities
.9
.7
.5
.3
.2
.1
.05
.03
.02
.01
10000
15000
20000
25000
ti
Point plotting method = km

Enfoque Bayesiano

- Sea $M_1, M_2, \ldots, M_m$, modelos a evaluar por Teorema de Bayes

$$P(M_j \mid \boldsymbol{X}) = \frac{VI(\boldsymbol{X} \mid M_j) Pr(M_j)}{\sum_{i=1}^{m} VI(\boldsymbol{X} \mid M_i) Pr(M_i)} \qquad j = 1,2,\ldots,m$$

donde

$$VI(\boldsymbol{X} \mid M_j) = \int_{\Theta_j} L(\boldsymbol{X} \mid \theta_j, M_j) \pi(\theta_j \mid M_j) d\theta_j$$

con $\pi(\boldsymbol{\theta}_j \mid M_j)$ la densidad a priori para el vector de parámetros $\boldsymbol{\theta}_j$

Problemas para especificar $\pi(\boldsymbol{\theta}_j|M_j)$

- $M_1(\mu,\tau)$, $M_2(\mu,\tau)$, $M_3(\mu,\tau)$, $M_4(\eta,\theta)$ y $M_5(\theta)$

- Normal (M_1).
- Log-Normal (M_2).
- Valor Extremo (M_3).
- Weibull (M_4).
- Exponencial (M_5).
- Más modelos.

- Normal-gama $(m_j, k_j, \alpha_j, \beta_j)$

$$\pi(\mu, \tau \mid \alpha_1, \beta_1, m_1, k_1) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \left(\frac{k_1}{2\pi}\right)^{1/2}$$
$$\tau^{\alpha_1 - 1/2} \exp\left[-\frac{k_1 \tau}{2}(\mu - m_1)^2 - \beta_1 \tau\right]$$

- Distribución $u(a_j, b_j, d_j, e_j)$

$$\pi(\mu, \tau \mid a_1, b_1, d_1, e_1) = \pi(\mu \mid a_1, b_1)\pi(\tau \mid d_1, e_1)$$
$$= \frac{1 \qquad\qquad 1}{(b_1 - a_1)(e_1 - d_1)}$$

Requerimientos para el Método de Obtención de a Prioris.

- Que funcione simultáneamente para los cinco o más modelos.
- Que dependa de características fácilmente comunicables de la variable de interés, X=tiempo de vida.
- Que permita el cálculo numérico de la verosimilitud integrada.

- Diferentes enfoques Bayesianos:
  - Muestras de ensayo
  - Datos históricos previos
  - Bayesiana no paramétrica
  - Muestras imaginarias de ensayo para estimar constantes normal-izadoras

Método Propuesto
Información Requerida.

- Intervalos para $M = E(X)$ y $S = \sqrt{Var(X)}$

$$[L_m, U_m] \qquad [L_s, U_s]$$

- Intervalos para dos percentiles y la mediana

$$[Lt_{p_1}, Ut_{p_1}] \quad [Lt_{p_2}, Ut_{p_2}] \quad [D_L, D_U]$$

Método Cuando se Dan Intervalos para *M* y *S*.

- Se consideran modelos con 2 $\theta = (\theta_1, \theta_2)$ parámetros como máximo donde θ_1 tiene más influencia sobre localización y θ_2 sobre escala.
- Vemos los intervalos como información sobre:

$$E(X|\theta) = h_1(\theta_1, \theta_2) \text{ y } V(X|\theta) = h_2(\theta_1, \theta_2)$$

- Sean $E(X|\theta) = E(M)$ $V(X|\theta) = E^2(S)$,

Método cuando se dan Intervalos para *M* y *S*.

- Por lo tanto $h_1(\theta_1, \theta_2) = E(M)$ $h_2(\theta_1, \theta_2) = E^2(S)$
- Resolviendo

$$\theta_j = r_j(M, S^2),\ j = 1, 2.$$

- Valores medios de *M* y *S* generan valores medios de θ_1 y θ_2 y valores extremos de *M* y *S* generan valores extremos de θ_1 y θ_2 .

Modelo Lognormal

$$E(X \mid \mu, \tau) = \exp(\mu + \frac{0.5}{\tau}) = h_1(\mu, \tau)$$
$$V(X \mid \mu, \tau) = \exp(2\mu + \tau^{-1})[\exp(\tau^{-1}) - 1] = h_2(\mu, \tau)$$

$$\mu = \log(M) - 0.5\log\left(\frac{S^2}{M^2} + 1\right) = r_1(M, S^2),$$
$$\tau = 1/\log\left(\frac{S^2}{M^2} + 1\right) = r_2(M, S^2)$$

Normal Gama$(\alpha_i, \beta_i, m_i, k_i)$

θ_2 sigue una $G(\alpha_i, \beta_i)$:

$E(\theta_2) = \alpha_i/\beta_i$ y $V(\theta_2) = \alpha_i/\beta_i^2$

$$\theta_2^* = r_2(E(M), E^2(S)) = \alpha_i/\beta_i$$

$$V(\theta_2) = \left(\frac{|\theta_2^* - r_2[E(M), U_s^2]|}{3}\right)^2 = \alpha_i/\beta_i^2$$

Resolviendo para α_i y β_i :

$$\beta_i = \frac{E(\theta_2)}{V(\theta_2)} \equiv \frac{\theta_2^*}{\left(\frac{|\theta_2^* - r_2[E(M), U_s^2]|}{3}\right)^2} \text{ y } \alpha_i = \beta_i \theta_2^*$$

$\theta_1 \sim t(m_i, 2\alpha_i, \alpha_i k_i/\beta_i)$, entonces

$E(\theta_1) = m_i$ y $V(\theta_1) = \frac{\beta_i}{k_i(\alpha_i - 1)}$.

Como θ_1 es el parametro con más influencia sobre $E(X|\boldsymbol{\theta})$, entonces

$$E(\theta_1) = r_1(E(M), E^2(S)) \equiv \theta_1^*$$

$$r_1(E(M), E^2(S)) = m_i$$

$$V(\theta_1) = \left(\frac{|\theta_1^* - r_1[L_m, E^2(S)]|}{3}\right)^2 = \frac{\beta_i}{k_i(\alpha_i - 1)}$$

Resolviendo

$$m_i = \theta_1^* \quad \text{y} \quad k_i = \frac{9\beta_i}{(\alpha_i - 1)\left(|\theta_1^* - r_1[L_m, E^2(S)]|\right)^2}$$

Priori Uniforme

$$
\begin{aligned}
a_t &= \min\left\{r_1\left(L_m, E^2(S)\right),\ r_1\left(U_m, E^2(S)\right)\right\}\\
b_t &= \max\left\{r_1\left(L_m, E^2(S)\right),\ r_1\left(U_m, E^2(S)\right)\right\}\\
d_t &= \min\left\{r_2\left(E(M), U_s^2\right),\ r_2\left(E(M), L_s^2\right)\right\}\\
e_t &= \max\left\{r_2\left(E(M), U_s^2\right),\ r_2\left(E(M), L_s^2\right)\right\}
\end{aligned}
$$

Implementación.

Densidad de Probabilidades

M_1	$\left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left[-\frac{\tau}{2}(x-\mu)^2\right]$
M_2	$\left(\frac{\tau}{2\pi}\right)^{1/2} x^{-1} \exp\left[-\frac{\tau}{2}(\log(x)-\mu)^2\right]$
M_3	$\tau \exp\left[\tau(x-\mu) - \exp\left(\tau(x-\mu)\right)\right]$
M_4	$\frac{\lambda}{\eta}\left(\frac{x}{\eta}\right)^{\lambda-1} \exp\left(-\left(\frac{x}{\eta}\right)^{\lambda}\right)$
M_5	$\lambda \exp\left[-\lambda(x-\gamma)\right] \quad 0 < \lambda < x$

Implementación a Priori Normal-Gamma.

	β_i	α_i	m_i	k_i
M_1	$\frac{1}{E^2(S)\left[\frac{\frac{1}{E^2(S)}-\frac{1}{[U_s]^2}}{3}\right]^2}$	$\frac{\beta_1}{E^2(S)}$	$E(M)$	$\frac{36\beta_1}{(U_m-L_m)^2(\alpha_1-1)}$
M_2	$\frac{\frac{9}{\log\left(\frac{E^2(S)}{E^2(M)}+1\right)}}{\log^2\left[\left(\frac{U_s^2}{E^2(M)}+1\right)\left(\frac{E^2(S)}{E^2(M)}+1\right)^{-1}\right]}$	$\frac{\beta_2}{\log\left(\frac{E^2(S)}{E^2(M)}+1\right)}$	$\log\left(\frac{E^2(M)}{(E^2(S)+E^2(M))^{1/2}}\right)$	$\frac{9\beta_2}{(\alpha_2-1)\log^2\left(\frac{E(M)}{L_m}\right)}$
M_3	$\frac{9\pi}{\sqrt{6}E(S)\left[E(\tau)-\frac{\pi}{\sqrt{6}U_s}\right]^2}$	$\frac{\pi\beta_3}{\sqrt{6}E(S)}$	$E(M)+0.45E(S)$	$\frac{9\beta_3}{(\alpha_3-1)[E(M)-L_m]^2}$
M_4	$\frac{9r_2[\lambda,E(M),E(S)]}{[r_2[\lambda,E(M),E(S)]-r_2[\lambda,E(M),L_s]]^2}$	$\beta_4 r_2[\lambda, E(M), E(S)]$	$\frac{E(M)}{\Gamma\left[1+\frac{1}{r_2[\lambda,E(M),E(S)]}\right]}$	$\frac{9\beta_4(\alpha_4-1)^{-1}}{(m_4-r_1[\lambda,L_m,E(S)])^2}$
M_5	$\frac{\frac{3\gamma}{E(M)-\gamma}}{\left(\frac{1}{L_m-\gamma}-\frac{1}{U_m-\gamma}\right)^2}$	$\frac{\beta_5}{E(M)-\gamma}$		

Implementación a Priori Uniforme

	a_i	b_i	d_i	e_i
M_1	L_m	U_m	$\frac{1}{(U_s)^2}$	$\frac{1}{(L_s)^2}$
M_2	$\log(L_m)$ $-0.5\log\left(\frac{E^2(S)}{E^2(M)}+1\right)$	$\log(U_m)$ $-0.5\log\left(\frac{E^2(S)}{E^2(M)}+1\right)$	$\frac{1}{\log\left(\frac{U_s^2}{E^2(M)}+1\right)}$	$\frac{1}{\log\left(\frac{L_s^2}{E^2(M)}+1\right)}$
M_3	$L_m + 0.45004E(S)$	$U_m + 0.45004E(S)$	$\frac{\pi}{\sqrt{6}U_s}$	$\frac{\pi}{\sqrt{6}L_s}$
M_4	$\frac{L_m}{\Gamma\left[1+\frac{1}{r_2[\lambda, E(M), E(S)]}\right]}$	$\frac{U_m}{\Gamma\left[1+\frac{1}{r_2[\lambda, E(M), E(S)]}\right]}$	$\min\{r_1[\lambda, E(M), L_s],$ $r_1[\lambda, E(M), U_s]\}$	$\max\{r_1[\lambda, E(M), L_s],$ $r_1[\lambda, E(M), U_s]\}$
M_5	$\frac{1}{U_m-\gamma}$	$\frac{1}{L_m-\gamma}$		

Método Cuando se Dan Intervalos para Percentiles.

- Hay que diferenciar si el modelo para el tiempo de vida es simétrico o no.
- Se transforman los intervalos de percentiles a intervalos para M y S.
- Se tienen resultados para Normal-Gama y Uniforme.

Verificación de la Implementación.

- Una manera de verificar que la metodología está funcionando es obteniendo la densidad predictiva a priori:

$$f_j(x)=\int_{\Theta_j} f(x\,|\,\theta_j, M_j)\pi(\theta_j\,|\,M_j)d\theta_j$$

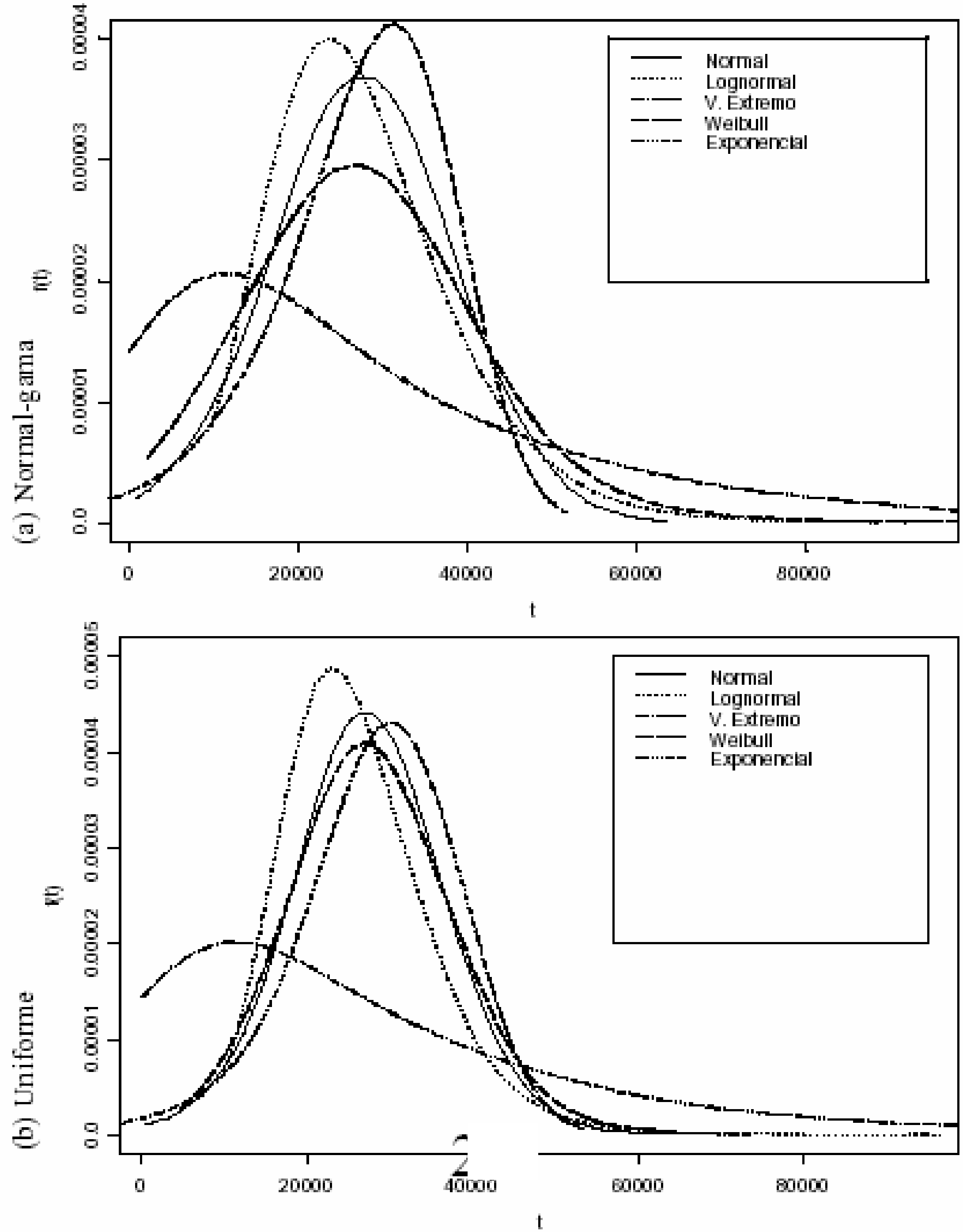
Verificación de la Implementación.

- Si esas densidades son similares y cumplen los requisitos impuestos sobre *M* y *S* será indicación de que se alcanzó el objetivo.
- Algunos casos tienen predictiva analítica, los otros se pueden obtener por simulación.

Aplicación: Experimento sobre Amortiguadores.

- 38 unidades experimentales.
- Censura aleatoria por la derecha.
- X=Tiempo de vida en millas.
- Intervalo para *M* [20,000;35,000].
- Intervalo para *S* [5,000;15,000].

Aplicación: Experimento sobre Amortiguadores

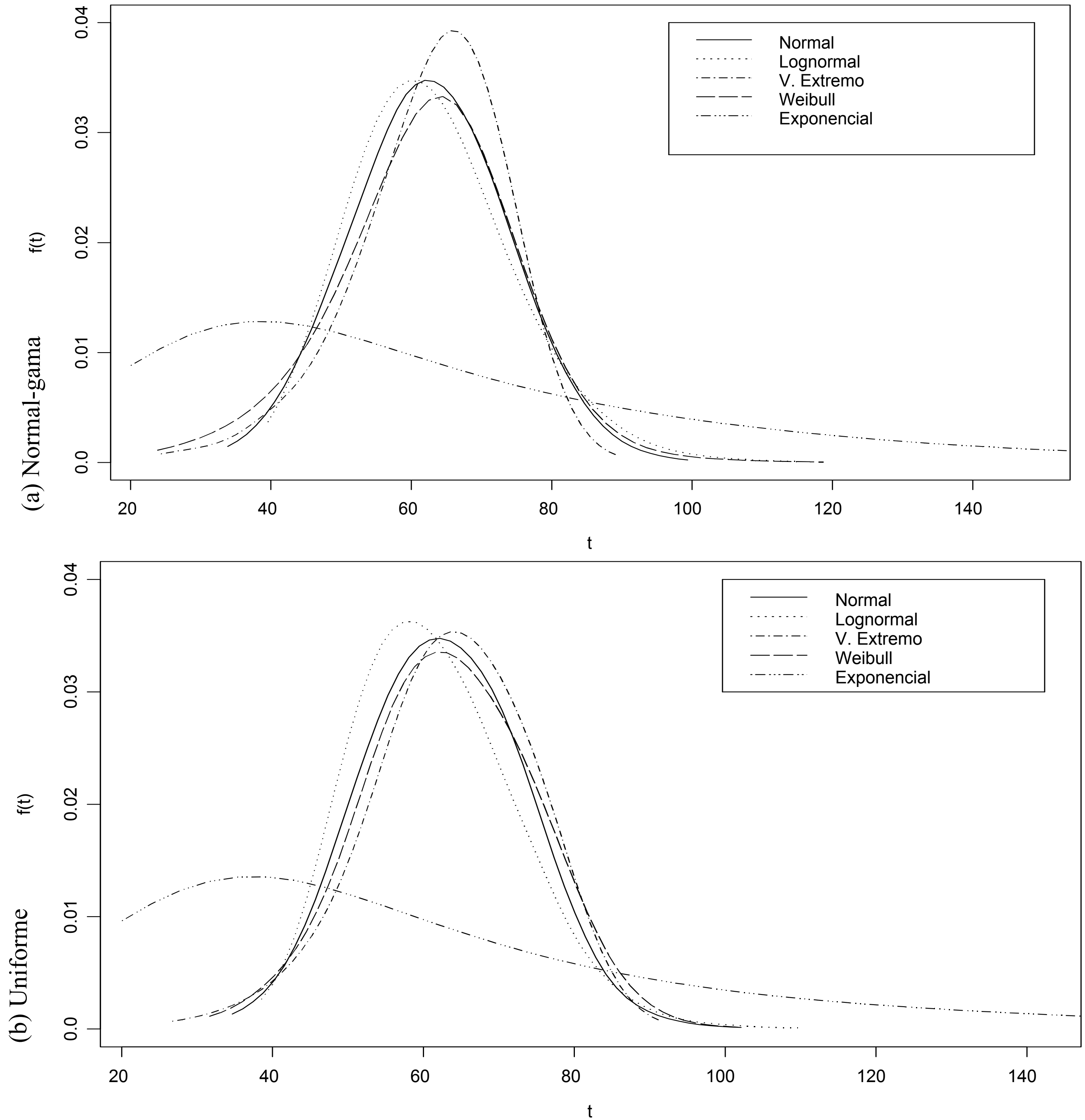


Aplicación: Experimento sobre Amortiguadores:

Intervalo para *M* [20,000;35,000].

Intervalo para *S* [5,000;15,000].

	NG		U	
	Media	D. Est.	Media	D. Est.
M_1	27526	10468	27360	9007
M_2	27695	10603	26125	8760
M_3	27395	10435	28240	9720
M_4	27811	13104	27892	9710
M_5	27545	27805	26777	27469



Intervalos: [50,75] [5,15].

Selección de modelos

$$P(M_j \mid \boldsymbol{X}) = \frac{VI(\boldsymbol{X} \mid M_j)Pr(M_j)}{\sum_{i=1}^{m} VI(\boldsymbol{X} \mid M_i)Pr(M_i)} \qquad j = 1,2,\ldots,m$$

$$VI(\boldsymbol{X} \mid M_j) = \int_{\Theta_j} L(\boldsymbol{X} \mid \theta_j, M_j)\pi(\theta_j \mid M_j)d\theta_j$$

Sin censura y a priori N-G, la VI está dada por:

a) Para M_1 :

$$\frac{\beta_1^{\alpha_1}\Gamma(\frac{n}{2}+\alpha_1)\,(k_1)^{1/2}}{(k_1+n)^{1/2}\Gamma(\alpha_1)(2\pi)^{n/2}h_1^{\frac{n}{2}+\alpha_1}}$$

con $h_1 = \frac{1}{2}k_1 n\frac{(m_1-\overline{x}_n)^2}{k_1+n} + \beta_1 + \frac{s_n}{2}$,

$s_n = \sum_{i=1}^{n}(x_i-\overline{x})^2$

b) Para M_2 :

$$\frac{\beta_2^{\alpha_2}\Gamma(\frac{n}{2}+\alpha_2)\,(k_2)^{1/2}}{(k_2+n)^{1/2}\Gamma(\alpha_2)(2\pi)^{n/2}h_2^{\frac{n}{2}+\alpha_2}\prod_{i=1}^{n}x_i}$$

con $h_2 = \frac{1}{2}k_1 n\frac{(m_2-\overline{w}_n)^2}{k_1+n} + \beta_2 + \frac{u_n}{2}$.

Modelo Exponencial

a) Utilizando a priori $G(\alpha_5, \beta_5)$

$$f(\mathbf{X}|M_5, \gamma) = \frac{\beta_5^{\alpha_5}}{\Gamma(\alpha_5)} \frac{\Gamma(n + \alpha_5)}{(s_5)^{n+\alpha_5}}$$

donde $s_5 = \sum_{i=1}^{n} x_i + \beta_5 - n\gamma.$

b) Suponiendo una a priori $\mathbf{u}(a_5, b_5)$:

$$\frac{u_5 \Gamma(n+1)}{v^{n+1}} [G(b_5; n+1, v) - G(a_5; n+1, v)]$$

donde $v = \sum_{i=1}^{n} x_i - n\gamma$

- Aproximación de Laplace:

$$\int_{\Theta} L(\mathbf{X}|\boldsymbol{\theta}, M_j)\pi(\boldsymbol{\theta}|M_j)d\boldsymbol{\theta}$$

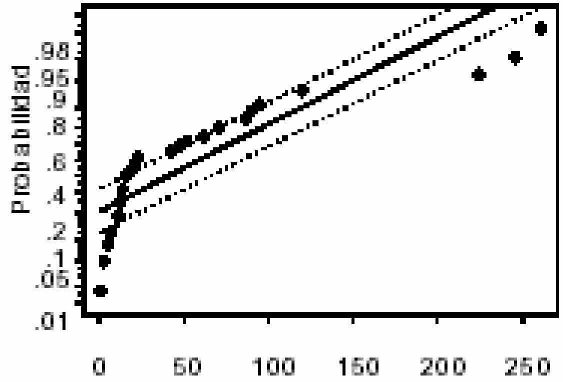
$$\cong (2\pi)^{d/2} \mid \Sigma \mid^{1/2} L(\mathbf{X}|\widetilde{\boldsymbol{\theta}}, M_j) \Pr(\widetilde{\boldsymbol{\theta}}|M_j)$$

- Monte Carlo:

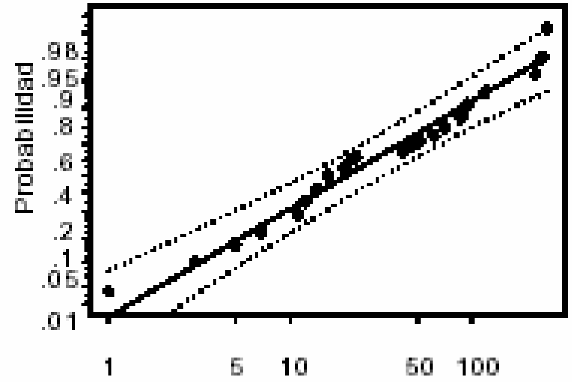
$$\widehat{f}(\mathbf{X}|M_j) = \frac{1}{m}\sum_{i=1}^{m} L(\mathbf{X}|\boldsymbol{\theta}_j^{(i)}, M_j)$$

- En Prochan(1963) se presentan tiempos de falla del equipo de aire acondicionado de un aeroplano:

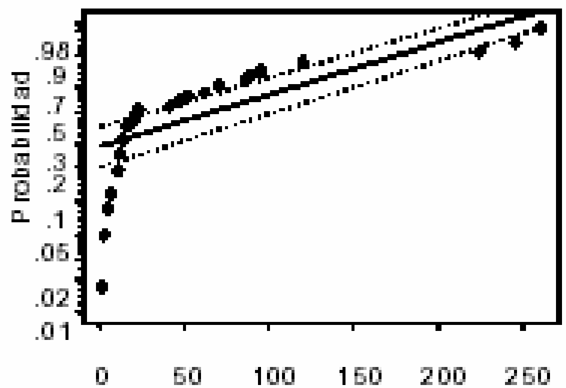
1 3 5 7 11 11 11 12 14 14 14 16 16 20 21 23 42 47 52 62 71 71 87 90 95 120 120 225 246 261



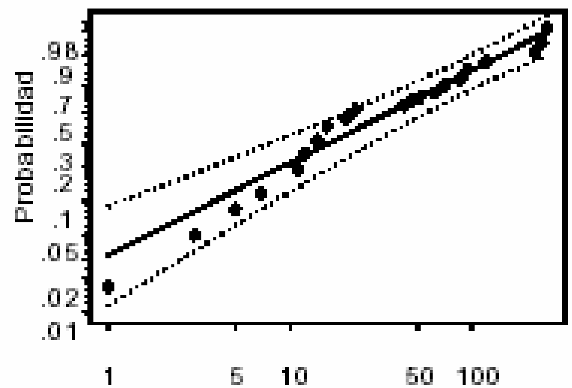
(a) Gráfica normal



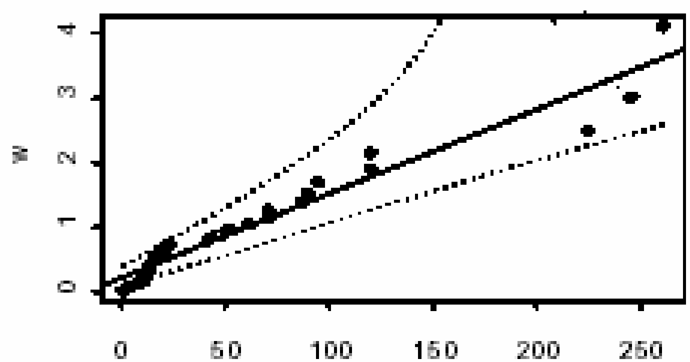
(b) Gráfica lognormal



(c) Gráfica valor extremo



(d) Gráfica Weibull



(e) Gráfica exponencial

	Paired probabilities										Total	
	M_1		M_2		M_3		M_4		M_5		$\Pr(M_j\|\mathbf{X})$	
Prior	NG	U	NG	U	NG	U	NG	U	NG	U	NG	U
M_1			0	0	1.0	1.0	0	0	0	0	0	0
M_2	1.0	1.0			1.0	1.0	0.01	0.03	0.01	0.03	0.01	0.01
M_3	0	0	0	0			0	0	0	0	0	0
M_4	1.0	1.0	0.99	0.97	1.0	1.0			0.58	0.52	0.57	0.51
M_5	1.0	1.0	0.99	0.97	1.0	1.0	0.42	0.48			0.42	0.48

Prior	Paired probabilities, amortiguadores										Total	
	M_1		M_2		M_3		M_4		M_5		$\Pr(M_j \mid \mathbf{X})$	
Model	NG	U	NG	U	NG	U	NG	U	NG	U	NG	U
M_1			0.91	0.85	0.77	0.55	0.50	0.38	1.0	1.0	0.42	0.28
M_2	0.09	0.15			0.26	0.18	0.09	0.10	1.0	1.0	0.04	0.05
M_3	0.23	0.45	0.74	0.82			0.23	0.34	1.0	1.0	0.12	0.23
M_4	0.50	0.62	0.91	0.90	0.77	0.66			1.0	1.0	0.41	0.45
M_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			0.00	0.00

Conclusiones

- Se obtuvo un método para especificar distribuciones a priori, de modelos comúnmente usados en confiabilidad, basado en intervalos para *M* y *S* del tiempo de falla.
- Se puede aplicar también en el caso que se provea información sobre percentiles.

Conclusiones.

- El procedimiento se puede extender a otros modelos de confiabilidad.
- El procedimiento permite la elección entre a priori Normal-Gama o Uniforme.
- Resulta conveniente en la aplicación de determinación de tiempos de garantía.

Conclusiones
Validación de Modelos

- Se tiene un procedimiento Bayesiano:
  - Requiere poca información inicial
  - No depende de resultados asintóticos
  - Es aplicable a cualquier tipo y nivel de censuramiento
  - Se obtuvo analíticamente la solución para algunos casos

Reportes Técnicos.

- Gutiérrez, H., Aguirre, V. y Christen, J. “A Practical Method for Obtaining Prior Distributions in Reliability”. Reporte Técnico DE-C03.12, Departamento de Estadística, ITAM.
- “Bayesian Model Evaluation in Reliability”, Reporte Técnico DE-C03.17, Departamento de Estadística, ITAM.
- humpulido@yahoo.com