

CAPÍTULO 11

ESPERANZA MATEMÁTICA PARCIAL

Sumario: 1. Introducción. 2. Esperanza matemática parcial en variables continuas. 2.1. Distribución Normal. 2.2. Distribución Lognormal. 2.3. Distribución Gamma. 2.4. Distribución de WEIBULL. 2.5. Distribución de GUMBEL del mínimo. 2.6. Distribución de GUMBEL del máximo. 2.7. Distribución de PARETO. 2.8. Distribución empírica dada con una serie de frecuencias. 3. Truncamiento. 4. Problema de inventario perecedero con demanda aleatoria. 5. Acumulativo de la variable. Curva ABC. Índice de GINI. 6. Apéndice. 6.1 Demostraciones de las expresiones de la esperanza matemática parcial para las distintas distribuciones. 6.1.1. Distribución Normal. 6.1.2. Distribución Lognormal. 6.1.3. Distribución Gamma. 6.1.4 Distribución de WEIBULL. 6.1.5. Distribución de PARETO. 6.1.6. Distribución de GUMBEL del mínimo. 6.1.7. Distribución de GUMBEL del Máximo. 6.2. Demostraciones de las expresiones del Índice de GINI para los modelos de PARETO y Lognormal. 6.2.1. Modelo de PARETO. 6.2.2. Modelo Lognormal. 7. Problemas.

1. INTRODUCCIÓN

Este capítulo y el siguiente tratan sobre un poderoso procedimiento de análisis, casi desconocido en nuestro medio, que permite resolver un conjunto importante de problemas, en general de naturaleza económica, que fue propuesto por el estadístico estadounidense ROBERT SCHLAIFER y desarrollado en sus obras (1967) que no fueron traducidas a nuestro idioma. Helo aquí, por primera vez en una obra en idioma español.

Existen diversas aplicaciones, muchas de ellas de naturaleza económica, en las cuales debemos tratar funciones de una variable aleatoria del siguiente tipo

$$(11.1) \quad \begin{aligned} x \leq a_1: & \quad y = b_1 + b_2x \\ a_1 < x \leq a_2: & \quad y = b_3 + b_4x \\ x > a_2: & \quad y = b_5 + b_6x \end{aligned}$$

Se dice que y es *función lineal por tramos* de x . Las abscisas $x=a_1$, $x=a_2$, etc. se denominan *puntos de corte*. Algunos ejemplos

* En Capital Federal y Gran Buenos Aires, el consumo domiciliario de electricidad se factura del siguiente modo: Hasta 300 Kwh en el bimestre, un cargo fijo de 4,46\$/bim más 0,081 \$/Kwh; más de 300 Kwh, 16,29\$/bim más 0,042 \$/Kwh (Datos del año 2005).

* En Capital Federal y Gran Buenos Aires, el consumo domiciliario de gas se factura del siguiente modo: Un cargo fijo de 7,74 \$/bimestre más 0,143651 \$/m³.

* En una obra civil, la empresa constructora tiene un plazo de n días para concluir la. Si se atrasa, tiene una penalización de b \$/día y si se adelanta recibe un premio de c \$/día.

El problema fundamental que se quiere resolver es el cálculo de la media de la función y , supuesto que conocemos la distribución de x . En ese caso, si x es una variable continua

$$(11.2) \quad E[y(x)] = \int_x y(x) f(x) dx$$

Si y está definida como en (11.1), tenemos

$$(11.3) \quad E[y(x)] = \int_{-\infty}^{a_1} (b_1 + b_2 x) f(x) dx + \int_{a_1}^{a_2} (b_3 + b_4 x) f(x) dx + \int_{a_2}^{\infty} (b_5 + b_6 x) f(x) dx =$$

$$= b_1 \int_{-\infty}^{a_1} f(x) dx + b_2 \int_{-\infty}^{a_1} x f(x) dx + b_3 \int_{a_1}^{a_2} f(x) dx + b_4 \int_{a_1}^{a_2} x f(x) dx + b_5 \int_{a_2}^{\infty} f(x) dx + b_6 \int_{a_2}^{\infty} x f(x) dx$$

Las integrales del tipo $\int_a^b f(x) dx$ se expresan en términos de la función de distribución (izquierda o derecha), ya que

$$(11.4) \quad \int_a^b f(x) dx = F(b) - F(a) = G(a) - G(b)$$

Las integrales del tipo $\int_a^b x f(x) dx$ tienen fórmulas especiales, según la distribución de la variable x . A tal efecto, definiremos:

$$(11.5) \quad H(x) = \int_{-\infty}^x x f(x) dx ; \quad J(x) = \int_x^{\infty} x f(x) dx$$

denominadas **Esperanza Matemática Parcial (Izquierda y Derecha)** respectivamente). Se cumple entre ellas la siguiente relación:

$$(11.6) \quad H(x) + J(x) = E(x) = \mu$$

Tendremos entonces, en general

$$(11.7) \quad \int_a^b x f(x) dx = H(b) - H(a) = J(a) - J(b)$$

2. ESPERANZA MATEMÁTICA PARCIAL EN VARIABLES CONTINUAS

Veremos el cálculo de las funciones $H(x)$ y $J(x)$ para los modelos fundamentales y algunas aplicaciones de interés.

2.1. Distribución Normal

$$(11.8) \quad H(x) = \int_{-\infty}^x x f(x) dx = \mu \Phi(Z) - \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{Z^2}{2}\right) ; \quad Z = \frac{x - \mu}{\sigma}$$

Ejemplo 1. Una empresa constructora ha ganado la licitación estatal para una obra civil por un monto de u\$s 600.000 y un plazo de entrega de 210 días, con una cláusula que penaliza en u\$s 2.500 cada día de atraso, y un premio de u\$s 1.500 por día de adelanto. El costo de la obra para la empresa se ha calculado en u\$s 330.000 fijos más 500 u\$s/día. Además, mediante la teoría de redes se ha determinado que la duración de los trabajos se distribuye Normalmente con media 195 días y desvío estándar 25 días. a) Calcular el beneficio esperado. b) ¿Cuál es la probabilidad de ganar más de u\$s 165.000? Graficar la función $G(B)=P(B > B)$.

a) Podemos expresar el beneficio (B) en función de la duración (x) de la obra como sigue

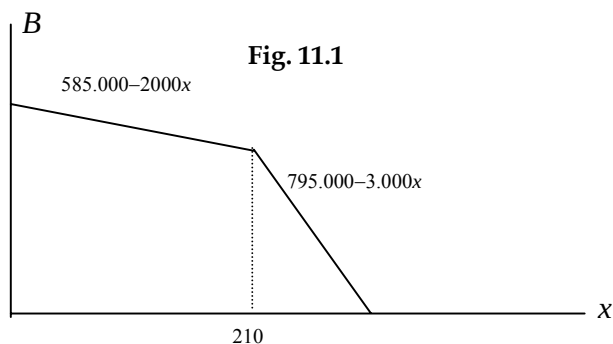
$$x \leq 210: B = 600.000 - 330.000 - 500x + 1.500(210-x)$$

$$x \geq 210: B = 600.000 - 330.000 - 500x - 2.500(x-210)$$

o sea, operando con las expresiones

$$(11.9) \quad x \leq 210: B = 585.000 - 2.000x$$

$$x \geq 210: B = 795.000 - 3.000x$$



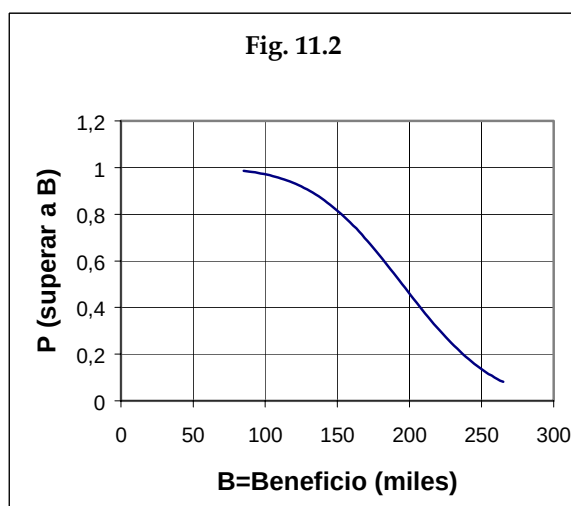
Vemos, en la Fig. 11.1, la representación de esta función lineal por tramos. Calculemos ahora su esperanza matemática. De acuerdo con (11.4) y (11.7):

$$(11.10) \quad E(B) = 585.000F(210) - 2.000H(210) + 795.000G(210) - 3.000J(210)$$

También tenemos

$$F(210) = \Phi\left(\frac{210 - \mu}{\sigma}\right) = \Phi\left(\frac{210 - 195}{25}\right) = \Phi(0,6) = 0,72575$$

$$G(210) = 1 - F(210) = 0,27425$$



$$H(210) = \mu F(210) - \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{0,6^2}{2}\right) = 133,1906$$

$$J(210) = \mu - H(210) = 61,8094$$

Reemplazando en (11.10) obtenemos $E(B) = \text{u}\$190.783$

b) De las expresiones (11.9) se deduce que a un beneficio $B=165.000$ le corresponde una duración $x=210$ días. Por lo tanto, la probabilidad pedida es

$$P(B > 165.000) = P(x < 210) = F(210) = 0,72575.$$

En la Fig. 11.2 vemos la gráfica de la probabilidad de superar un dado beneficio $G(B)$ en función de B .

Ejemplo 2. Una fábrica de conservas tiene una línea de descabezado, eviscerado y fileteado de pescado y compra diariamente la producción total de varios barcos pesqueros a 3.000 \$/tn. La línea tiene un costo fijo de 4.200 \$/día siendo su capacidad máxima de producción de 15 tn. Si se pescan menos de 10 tn tiene un costo de 1.600 \$/tn; si se pescan entre 10 y 15 tn el costo es de 1.250 \$/tn y si se pescan más de 15 tn el pescado sobrante no procesado se vende a 2.000 \$/tn, a una planta más pequeña, como un recupero de costo. Se sabe experimentalmente que la pesca diaria es una variable aleatoria con distribución Normal de media 12 tn y desvío estándar 2 tn. Calcular el costo esperado diario.

El costo diario (C) en función de la pesca (x) se expresa como sigue

$$x < 10: \quad C = 4.200 + 3.000x + 1.600x = 4.200 + 4.600x$$

$$10 \leq x \leq 15: \quad C = 4.200 + 3.000x + 1.250x = 4.200 + 4.250x$$

$$x > 15: \quad C = 4.200 + 3.000x + 1250 \times 15 - 2000(x-15) = 52.950 + 1.000x$$

El costo esperado será:

$$(11.11) \quad E(C) = 4.200F(10) + 4.600H(10) + 4.200[F(15)-F(10)] + \\ + 4.250[H(15)-H(10)] + 52.950G(15) + 1.000J(15)$$

Tenemos

$$F(10) = \Phi\left(\frac{10-12}{2}\right) = \Phi(-1) = 0,15866$$

$$F(15) = \Phi\left(\frac{15-12}{2}\right) = \Phi(1,5) = 0,93319$$

$$G(15) = 1 - F(15) = 0,06681$$

$$H(10) = \mu F(10) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left[-\frac{(-1)^2}{2}\right] = 1,41998$$

$$H(15) = \mu F(15) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left[-\frac{1,5^2}{2}\right] = 10,93924$$

$$J(15) = \mu - H(15) = 1,06076$$

Reemplazando en (11.11), obtenemos $E(C) = \$55.506,51$

Ejemplo 3. En una obra que demandará 52 días de trabajo se contrata una grúa cuya tarifa básica es de 54 \$/hora. En días normales, el tiempo de trabajo es una variable aleatoria Normal de media 10,5 hs y desvío 1 hora. Las condiciones del contrato de la grúa son las siguientes: a) En días de lluvia no se trabaja pero hay que pagar 8 hs. de trabajo; b) en los días normales, si se trabaja menos de 10 hs, se pagan 10 hs completas; si se trabaja entre 10 y 12 hs se paga la tarifa básica y si se trabaja más de 12, el excedente se paga con un recargo del 20%. Si se sabe que, en promedio, 1 de cada 7 días es lluvioso, ¿cuál es el costo esperado del servicio de grúa?

Suponiendo que la sucesión de días normales y lluviosos puede considerarse, aproximadamente, como un proceso de BERNOULLI, el número total de días necesarios para completar la obra es una variable de PASCAL, cuyo valor esperado es $E(n)=52/P(\text{día normal})=52/(6/7)=60,667$. Habrá entonces 52 días normales y, en términos esperados, 8,667 días lluviosos. El costo esperado del servicio de la grúa será:

$$(11.12) \quad E(C) = 52.E(C_{\text{día normal}}) + 8,667.C_{\text{día lluvioso}}$$

El costo para un día lluvioso no es aleatorio, pues vale $C_{\text{día lluvioso}} = 8\text{hs} \times 54\$/\text{hora} = \$432$. El costo para un día normal depende del tiempo de trabajo (x) como sigue:

$$x < 10: \quad C_{\text{día normal}} = 54 \times 10 = 540$$

$$10 \leq x \leq 12: \quad C_{\text{día normal}} = 54x$$

$$x > 12: \quad C_{\text{día normal}} = 54 \times 12 + 54 \times 1,2(x-12) = -129,6 + 64,8x$$

El costo esperado para un día normal será

$$(11.13) \quad E(C_{\text{día normal}}) = 540F(10) + 54[H(12) - H(10)] - 129,6G(12) + 64,8J(12)$$

Tenemos

$$F(10) = \Phi\left(\frac{10-10,5}{1}\right) = \Phi(-0,5) = 0,308538$$

$$F(12) = \Phi\left(\frac{12-10,5}{1}\right) = \Phi(1,5) = 0,933193$$

$$G(12) = 1 - F(12) = 0,066807$$

$$H(10) = \mu F(10) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left[-\frac{(-0,5)^2}{2}\right] = 2,887584$$

$$H(12) = \mu F(12) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left[-\frac{1,5^2}{2}\right] = 9,669009$$

$$J(12) = \mu - H(12) = 0,830991$$

Reemplazamos en (11.13) y obtenemos $E(C_{\text{día normal}}) = \$577,9975$; finalmente, de (11.12) resulta el costo esperado total para la obra $E(C) = \$33.800$.

Ejemplo 4 (Desviación media absoluta). La desviación media absoluta de una variable se define como la esperanza de $|x-\mu|$ y, en general, esta relacionada con el desvío estándar. Calcular su expresión para una variable x con distribución Normal de media μ y desvío estándar σ .

Sea $y = |x-\mu|$. Podemos plantear la siguiente relación entre x e y :

$$x \leq \mu : y = \mu - x$$

$$x > \mu : y = x - \mu$$

Por lo tanto:

$$(11.14) \quad E(y) = E|x-\mu| = \mu F(\mu) - H(\mu) + J(\mu) - \mu G(\mu)$$

Dado que en la Normal $F(\mu) = G(\mu) = 0,5$, los términos extremos se cancelan. Además

$$H(\mu) = \mu F(\mu) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left(-\frac{Z^2}{2}\right) = \frac{\mu}{2} - \frac{\sigma}{\sqrt{2\pi}} \quad \text{pues } Z = 0 \text{ para } x = \mu$$

$$J(\mu) = \mu G(\mu) + \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left(-\frac{Z^2}{2}\right) = \frac{\mu}{2} + \frac{\sigma}{\sqrt{2\pi}}$$

Reemplazando en (11.14) obtenemos

$$(11.15) \quad E|x-\mu| = \sigma \sqrt{\frac{2}{\pi}}$$

2.2. Distribución Lognormal

$$(11.16) \quad H(x) = \int_0^x xf(x)dx = \left[\text{Exp}\left(m + \frac{D^2}{2}\right) \right] \Phi\left(\frac{\ln x - m}{D} - D\right) = \mu \Phi\left(\frac{\ln x - m}{D} - D\right)$$

Ejemplo 5. En una ciudad se planea imponer una tasa municipal a los inmuebles, del 0,1% del valor fiscal de la propiedad, con un monto mínimo de \$20. El valor fiscal de los inmuebles es una variable Lognormal con media \$25.350 y desvío estándar \$23.200. Calcular la contribución promedio por inmueble.

La contribución (C) de un inmueble, expresada en función de su valor fiscal (x) es

$$x \leq 20.000: C = 20$$

$$x > 20.000: C = 0,001x$$

Su valor esperado será

$$(11.17) \quad E(C) = 20F(20.000) + 0,001J(20.000)$$

Dado que esta variable tiene distribución Lognormal, sus parámetros serán:

$$m = \text{Ln} \left[\frac{\mu}{\sqrt{1 + (\sigma/\mu)^2}} \right] = \text{Ln} \left[\frac{25.350}{\sqrt{1 + (23.200/25.350)^2}} \right] = 9,836313$$

$$D = \sqrt{\text{Ln}[1 + (\sigma/\mu)^2]} = \sqrt{\text{Ln}[1 + (23.200/25.350)^2]} = 0,780028$$

Entonces

$$F(20.000) = \Phi \left(\frac{\text{Ln } 20.000 - m}{D} \right) = \Phi(0,08612) = 0,53431$$

$$\begin{aligned} H(20.000) &= \mu \Phi \left(\frac{\text{Ln } 20.000 - m}{D} - D \right) = 25.350 \Phi(-0,69391) = \\ &= 25.350 \times 0,24387 = 6.182,10 \end{aligned}$$

$$J(20.000) = \mu - H(20.000) = 19.167,90$$

Reemplazando en (11.17) resulta la contribución promedio $E(C) = \$29,85$.

2.3. Distribución Gamma

$$(11.18) \quad H(x) = \int_0^x x f(x) dx = r\beta F_g(x \setminus r + 1; \beta) = \mu F_g(x \setminus r + 1; \beta)$$

Ejemplo 6. La demanda diaria de un periódico tiene distribución Gamma con media 800 y desvío estándar 320 (miles de ejemplares). Cada ejemplar tiene un costo total de \$0,80 y el precio de venta neto para el editor es de \$0,70, con lo cual tendría pérdida, pero tiene un ingreso adicional de \$350.000 en concepto de avisos. Diariamente se edita 1 millón de ejemplares. Calcular el beneficio esperado diario, sabiendo que cada ejemplar sobrante tiene un valor residual de \$0,10.

El beneficio diario, expresado en función de la demanda es

$$\begin{aligned} x \leq 1.000.000: \quad B &= 0,7x - 0,8 \times 1.000.000 + 350.000 + 0,1(1.000.000 - x) \\ &= -350.000 + 0,6x \end{aligned}$$

$$x > 1.000.000: \quad B = (0,7 - 0,8) \times 1.000.000 + 350.000 = 250.000$$

El beneficio esperado será

$$(11.19) \quad E(B) = -350.000F(1.000.000) + 0,6H(1.000.000) + 250.000G(1.000.000)$$

Como la demanda tiene distribución Gamma, debemos calcular sus parámetros:

$$\beta = \frac{\sigma^2}{\mu} = \frac{320.000^2}{800.000} = 128.000; \quad r = \left(\frac{\mu}{\sigma}\right)^2 = 6,25$$

Entonces tenemos:

$$F(1.000.000) = F_g(1.000.000 \mid r=6,25; \beta=128.000) = 0,761355$$

$$G(1.000.000) = 1 - F(1.000.000) = 0,238645$$

$$H(1.000.000) = \mu F_g(1.000.000 \mid r=7,25; \beta=128.000) = 800.000 \times 0,628222 = \\ = 502.577,6$$

$$J(1.000.000) = \mu - H(1.000.000) = 297.422,4$$

Las probabilidades acumuladas de la Gamma fueron calculadas con la función =DISTR.GAMMA(x ; r ; β ; verdadero) del EXCEL. También pueden calcularse con la aproximación de WILSON-HILFERTY:

$$(11.20) \quad F_g(x \mid r; \beta) \cong \Phi \left\{ 3\sqrt{r} \left[\left(\frac{x}{\beta r} \right)^{1/3} + \frac{1}{9r} - 1 \right] \right\}$$

que arrojaría en este caso

$$F_g(1.000.000 \mid 6,25; 128.000) = 0,7619; \quad F_g(1.000.000 \mid 7,25; 128.000) = 0,6284$$

Reemplazando en (11.19), obtenemos el beneficio esperado diario $E(B) = \$94.734$.

2.4. Distribución de WEIBULL

$$(11.21) \quad H(x) = \int_0^x x f(x) dx = \beta \Gamma \left(1 + \frac{1}{\omega} \right) F_g \left[\left(\frac{x}{\beta} \right)^{\omega} \mid r = 1 + \frac{1}{\omega}; \beta' = 1 \right]$$

Ejemplo 7 (Mantenimiento preventivo). Un rodamiento de un sistema de transmisión tiene una duración variable con distribución de WEIBULL de parámetros $\beta=1,52$ (miles de horas) y $\omega=3,8$. El rodamiento funciona 8 horas por día durante los 365 días del año. Se desea establecer un tiempo T de intervalos de reemplazo a efectos de minimizar el costo total anual. Cada reemplazo tiene un costo total (con mano de obra incluida) de $C_1=\$235$ pero si el rodamiento a reemplazar falla en funcionamiento, se incurre en un costo adicional de $C_2=\$430$. Calcular el intervalo T óptimo y el costo esperado anual.

Calculemos primero el costo esperado (Y) por cada reemplazo. Sea t el tiempo hasta la falla del rodamiento:

Si $t < T$ es $Y = C_1 + C_2$ (El rodamiento en funcionamiento ha fallado)

Si $t \geq T$ es $Y = C_1$ (El rodamiento en funcionamiento no ha fallado)

Será entonces

$$E(\mathbf{Y}) = (C_1 + C_2)F(T) + C_1G(T) = C_1 + C_2F(T)$$

Si en un año hay N reemplazos, el costo esperado anual es $NE(\mathbf{Y})$ pero N es la cantidad total de horas de un año ($R=365 \times 8=2.920$) dividida por el tiempo medio entre reemplazos $[E(\mathbf{x})]$. Si el reemplazo siempre se produjera antes de que falle el rodamiento funcionando, dicho tiempo medio sería directamente T , pero debemos considerar los sucesos indeseados en que el reemplazo se produce cuando el rodamiento en funcionamiento ha fallado. Nuevamente, sea t el tiempo hasta la falla del rodamiento y x el tiempo en que el rodamiento se reemplaza. Entonces

Si $t < T$ es $x = t$ (El rodamiento en funcionamiento ha fallado)

Si $t \geq T$ es $x = T$ (El rodamiento en funcionamiento no ha fallado)

Tenemos así

$$E(\mathbf{x}) = H(T) + TG(T)$$

por lo tanto

$$(11.22) \quad \text{Costo esperado anual} = NE(\mathbf{Y}) = \frac{R}{E(\mathbf{x})} E(\mathbf{Y}) = R \left[\frac{C_1 + C_2 F(T)}{H(T) + TG(T)} \right]$$

Observemos que la expresión entre corchetes es el costo promedio por hora de funcionamiento del rodamiento.

En este caso es

$$C_1 = \$235 ; C_2 = \$430$$

$$F(T) = 1 - \text{Exp} \left[- \left(\frac{T}{\beta} \right)^\omega \right] ; G(T) = \text{Exp} \left[- \left(\frac{T}{\beta} \right)^\omega \right]$$

$$H(T) = \beta \Gamma \left(1 + \frac{1}{\omega} \right) F_g \left[\left(\frac{T}{\beta} \right)^\omega \mid r = 1 + \frac{1}{\omega}; \beta' = 1 \right]$$

Con estos datos, el valor de T que minimiza (11.22) es $T=0,994$ miles de horas = 994 horas = 124,25 días y el costo esperado anual es \$955,93.

2.5. Distribución de PARETO

$$(11.23) \quad H(x) = \int_{\theta}^x x f(x) dx = \frac{\psi\theta}{\psi-1} \left[1 - \left(\frac{\theta}{x} \right)^{\psi-1} \right]$$

2.6. Distribución de GUMBEL del mínimo

$$(11.24) \quad H(x) = \int_{-\infty}^x x f(x) dx = \theta(1 - e^{-t}) + \beta \int_0^t e^{-t} \ln t \, dt$$

$$(11.25) \quad \int_0^t e^{-t} \ln t \, dt = (1 - e^{-t}) \ln t + \sum_{n=1}^{\infty} \frac{(-1)^n t^n}{n!n}$$

$$(11.26) \quad t = \text{Exp} \left(\frac{x - \theta}{\beta} \right)$$

Fórmula aproximada para la integral

$$(11.27) \quad \int_0^t e^{-t} \ln t \, dt \cong -C(1 - e^{-t}) + \frac{1}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]$$

Reemplazando en (11.24) se obtiene la fórmula aproximada

$$(11.28) \quad H(x) \cong \mu(1 - e^{-t}) + \frac{\beta}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]; \quad \mu = \theta - C\beta$$

2.7. Distribución de GUMBEL del máximo

$$(11.29) \quad H(x) = \int_{-\infty}^x x f(x) dx = \theta e^{-t} + C\beta + \beta \int_0^t e^{-t} \ln t \, dt$$

$$(11.30) \quad t = \text{Exp} \left(-\frac{x - \theta}{\beta} \right)$$

Reemplazando (11.27) en (11.29) se obtiene la fórmula aproximada

$$(11.31) \quad H(x) \cong \mu e^{-t} + \frac{\beta}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]; \quad \mu = \theta + C\beta$$

2.8. Distribución empírica dada con una serie de frecuencias

Sea la tabla de frecuencias

X	$a_1 - b_1$	$a_2 - b_2$	...	$a_k - b_k$
F_i	F_1	F_2	...	F_k
$\tau_i \cong F_i X_i$	τ_1	τ_2	...	τ_k

En esta tabla llamamos $X_i = (a_i + b_i)/2$ al punto medio de cada intervalo y τ_i al total efectivo, que se aproxima, si no es dato, como el producto de la frecuencia por el punto medio ($F_i X_i$). Llamamos j al intervalo en que se encuentra x y $h_j = b_j - a_j$ a su amplitud, entonces

$$(11.32) \quad H(x) = \frac{1}{n} \sum_{i=1}^{j-1} \tau_i + \frac{1}{n} \frac{(x + a_j)(x - a_j)}{2 h_j} F_j$$

3. TRUNCAMIENTO

El truncamiento de una variable consiste en **eliminar** un intervalo de su dominio. Algunos ejemplos

* En el Ejemplo 1, suponga que han transcurrido X_0 días, o sea que se ha eliminado el intervalo $x < X_0$ y la obra no ha concluido. Quizás desearíamos conocer la distribución del tiempo que falta para concluirla.

* Una persona tiene X_0 años de edad, es decir que se ha eliminado el intervalo $x < X_0$. ¿cuál es su esperanza de vida actual?

* En un país se desea conocer qué parte de la renta salarial pertenece al 10% superior, al 10% inferior, y qué relación hay entre ambas.

Para una variable x , con una dada distribución, supongamos que hemos eliminado el intervalo $x < X_0$ y deseamos obtener la distribución de la variable resultante después de la eliminación. El problema es bastante simple, ya que se trata de una probabilidad condicional, es decir, al haber eliminado el intervalo $x < X_0$, el nuevo dominio es su complemento, o sea $x > X_0$. Si llamamos y a la variable resultante, tendremos

$$G_y(y) = P[(x > y) \setminus (x > X_0)] = \frac{P(x > y)}{P(x > X_0)} = \frac{G_x(y)}{G_x(X_0)} \quad (y > X_0)$$

$$(11.33) \quad F_y(y) = 1 - \frac{G_x(y)}{G_x(X_0)} = \frac{F_x(y) - F_x(X_0)}{G_x(X_0)}$$

La función de densidad de la variable truncada (a la derecha) será

$$(11.34) \quad f_y(y) = \frac{f_x(y)}{G_x(X_0)} \quad (y > X_0)$$

Con el mismo razonamiento, si se elimina el intervalo $x > X_0$ (truncamiento a la izquierda), obtenemos:

$$(11.35) \quad F_y(y) = \frac{F_x(y)}{F_x(X_0)}$$

$$(11.36) \quad f_y(y) = \frac{f_x(y)}{F_x(X_0)} \quad (y < X_0)$$

Para este truncamiento a la izquierda, la media de la variable y es

$$(11.37) \quad \mu_y = E(y) = \frac{1}{F(X_0)} \int_{-\infty}^{X_0} y f_x(y) dy = \frac{H_x(X_0)}{F_x(X_0)} \quad (y < X_0)$$

Análogamente, para el truncamiento a la derecha, la media de la variable truncada es

$$(11.38) \quad \mu_y = E(y) = \frac{1}{G(X_0)} \int_{X_0}^{\infty} y f_x(y) dy = \frac{J_x(X_0)}{G_x(X_0)} \quad (y > X_0)$$

Finalmente, podemos hacer un *truncamiento a dos colas*, es decir, eliminar los intervalos $x < X_1$ y $x > X_2$, o sea que el dominio resultante sería $X_1 \leq y \leq X_2$. En ese caso, tendremos

$$(11.39) \quad F_y(y) = \frac{F_x(y) - F_x(X_1)}{F_x(X_2) - F_x(X_1)} \quad (X_1 \leq y \leq X_2)$$

$$(11.40) \quad f_y(y) = \frac{f_x(y)}{F_x(X_2) - F_x(X_1)} \quad (X_1 \leq y \leq X_2)$$

$$(11.41) \quad \mu_y = E(y) = \frac{H_x(X_2) - H_x(X_1)}{F_x(X_2) - F_x(X_1)} \quad (X_1 \leq y \leq X_2)$$

Ejemplo 8. En una planta se reciben tablas de longitud variable que se clasifican en: Categoría A, las menores de 1,5m; categoría B, las que tienen longitudes comprendidas entre 1,5 y 2m, y categoría C, las mayores de 2m. Se sabe que las proporciones son 12, 58 y 30%, respectivamente. Determinar la longitud media en cada categoría.

Calculemos los parámetros de la distribución, suponiendo que la misma es Normal:

$$\frac{1,5 - \mu}{\sigma} = Z_{0,12} = -Z_{0,88} = -1,174988; \quad \frac{2 - \mu}{\sigma} = Z_{0,7} = 0,524401$$

Se obtiene así $\mu = 1,846$ y $\sigma = 0,294$ metros.

Ahora debemos calcular:

$$H(1,5) = \mu F(1,5) - \frac{\sigma}{\sqrt{2\pi}} \exp\left[-\frac{(-1,174988)^2}{2}\right] = 0,16191$$

$$H(2) = \mu F(2) - \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{0,524401^2}{2}\right) = 1,18998$$

Las longitudes medias de las categorías serán

$$\mu_1 = \frac{H(1,5)}{F(1,5)} = \frac{0,16191}{0,12} = 1,349$$

$$\mu_2 = \frac{H(2) - H(1,5)}{F(2) - F(1,5)} = \frac{1,18998 - 0,16191}{0,58} = 1,773$$

$$\mu_3 = \frac{J(2)}{G(2)} = \frac{\mu - H(2)}{1 - F(2)} = \frac{1,846 - 1,18998}{0,3} = 2,187$$

Ejemplo 9 (Continuación del Ejemplo 1). Supongamos que ya han transcurrido 210 días y la obra no ha concluido, es decir que ha vencido el plazo de entrega estipulado en el contrato. ¿Cuál es, a partir de ahora, el tiempo esperado para concluir la obra?

Si ya han transcurrido 210 días, significa que se ha eliminado el intervalo $x < X_0 = 210$, por lo tanto el tiempo total de la obra es, con esta información:

$$E[x \mid (x > 210)] = \frac{J(210)}{G(210)} = \frac{61,8094}{0,27425} \cong 225 \text{ días}$$

El tiempo esperado, a partir de ahora, para concluir la obra será de 15 días.

Ejemplo 10 (Continuación del Ejemplo 6). Calcular: a) La *demanda* media para los días en que no se supera la tirada de 1.000.000 de ejemplares. b) Ídem para los días en que se supera la tirada. c) El número medio de ejemplares *vendidos*.

a) Para los días en que la demanda no supera la tirada de 1.000.000 de ejemplares, la media es

$$E[x \mid (x < 1.000.000)] = \frac{H(1.000.000)}{F(1.000.000)} = \frac{502.577,6}{0,761355} = 660.109$$

b) Para los días en que la demanda sí supera la tirada de 1.000.000 de ejemplares, la media es

$$E[x|(x > 1.000.000)] = \frac{J(1.000.000)}{G(1.000.000)} = \frac{297.422,4}{0,238645} = 1.246.296$$

c) Podría parecer que la respuesta es 800.000, pero ese es el número medio de ejemplares demandados, no vendidos. Los ejemplares vendidos (Q) se expresan en función de la demanda como sigue

$$\begin{aligned} x \leq 1.000.000: & \quad Q = x \\ x > 1.000.000: & \quad Q = 1.000.000 \end{aligned}$$

La venta esperada será

$$\begin{aligned} E(Q) &= H(1.000.000) + 1.000.000G(1.000.000) = 502.577,6 + 1.000.000 \times 0,238645 = \\ &= 741.223 \end{aligned}$$

Ejemplo 11. En la Argentina, en 1992, los hombres tienen una esperanza de vida al nacer de 68,44 años, con un desvío estándar de 10,12 años. Suponiendo que la distribución es de GUMBEL del mínimo, calcular la esperanza de vida de un hombre de 65 años.

La esperanza de vida (total) de un hombre de 65 años es

$$E[x|(x > 65)] = \frac{J(65)}{G(65)}$$

Calculemos entonces los parámetros de la distribución

$$\sigma = \frac{\pi}{\sqrt{6}} \beta; \quad \beta = \frac{\sqrt{6}}{\pi} \sigma = 7,89 \text{ años}$$

$$\mu = \theta - C\beta; \quad \theta = \mu + C\beta = 72,99 \text{ años}$$

$$G(65) = \text{Exp} \left[-\text{Exp} \left(\frac{65 - \theta}{\beta} \right) \right] = 0,6955$$

La esperanza matemática parcial derecha, calculada con un programa, es

$$J(65) = 51,2747$$

de donde resulta $E[x|(x > 65)] = 73,7$ años, o sea 8,7 años a partir del momento en que tiene 65 años.

En forma aproximada, podemos calcular $J(65)$ mediante las expresión (11.27)

$$(11.27) \quad H(x) \cong \mu(1 - e^{-t}) + \frac{\beta}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]$$

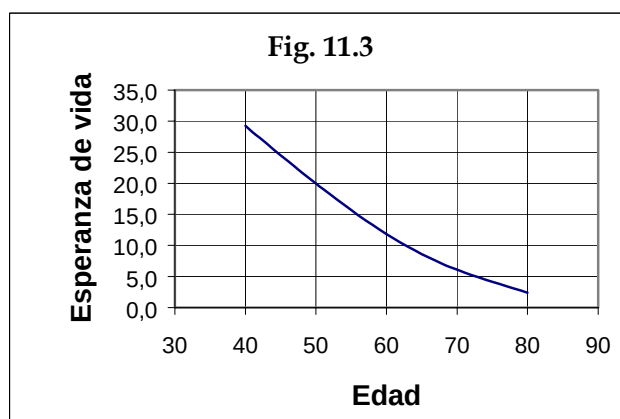
siendo $t = \text{Exp}\left(\frac{x - \theta}{\beta}\right)$

O sea, $t = 0,36327$; $H(65) \cong 17,255$; $J(65) \cong 51,18$; $E[x \setminus (x > 65)] \cong 73,6$.

En la Fig. 11.3 hemos representado gráficamente la función

$$Y(x) = E[x \setminus (x > x)] - x$$

en función de x . Se observa que es prácticamente lineal con el paso de los años.



4. PROBLEMA DE INVENTARIO PERECEDERO CON DEMANDA ALEATORIA

Existen muchos productos que tienen período de vencimiento, luego del cual pierden total o parcialmente su valor. Son ejemplos diversos alimentos, como los lácteos, el pan, las verduras, la carne, etc; también los periódicos, las flores, y otros. Supongamos que durante su período de frescura, nuestro producto es demandado aleatoriamente de acuerdo con una determinada distribución que, en la mayoría de los casos, será Normal o Gamma, cuyos parámetros se habrán determinado después de un tiempo razonable de toma de datos. Nuestro producto tiene un *costo de producción* (c) y un *precio de venta* (v); también, pasado el período de frescura, si hay sobrante, puede eventualmente tener un *valor residual* (v_r). Finalmente, puede suceder en algunos casos que, si la demanda no es totalmente satisfecha en el período de frescura, asumamos una pérdida adicional por esa causa, aparte de la venta perdida. Dicha pérdida se denomina *costo de escasez* (c_s); un caso típico es el pan: Si un cliente llega a la panadería y no hay pan, se va a otro negocio, pero si hubiera pan, compraría quizás otros productos, como facturas, que tienen un beneficio marginal muy superior al del producto principal, pero justamente, por esta razón, puede ser muy importante optimizar el suministro del mismo. Así pues, el problema fundamental que se plantea es determinar el stock óptimo para maximizar el beneficio esperado.

Ejemplo 12. Una editorial ha lanzado a la venta una nueva enciclopedia que edita en fascículos semanales cuya demanda se ha estabilizado en una distribución aproximadamente Normal de media 6.500 y desvío estándar 700 ejemplares. El costo unitario es de \$7 y se venden a \$13. No se considera valor residual ni costo de escasez. Calcular: a) La tirada semanal óptima.

b) El beneficio esperado. c) El número medio de ejemplares vendidos. d) El valor esperado de la información perfecta (VEIP).

Sea x la demanda, T la cantidad de ejemplares que se editarán (tirada) y B el beneficio. Podemos plantear la siguiente relación

$$\begin{aligned} x \leq T: \quad B &= 13x - 7T \\ x > T: \quad B &= (13-7)T = 6T \end{aligned}$$

El beneficio esperado será

$$E(\mathbf{B}) = 13H(T) - 7TF(T) + 6TG(T) = 13H(T) - 7TF(T) + 6T[1-F(T)]$$

$$(11.42) \quad E(\mathbf{B}) = 13H(T) - 13TF(T) + 6T$$

La tirada óptima será aquella que maximice el beneficio esperado. Derivemos entonces la expresión con respecto a T e igualemos a cero:

$$(11.43) \quad \frac{dE(\mathbf{B})}{dT} = 13 \frac{dH(T)}{dT} - 13F(T) - 13T \frac{dF(T)}{dT} + 6 = 0$$

Ahora

$$\frac{dF(T)}{dT} = F'(T) = f(T)$$

Además, como

$$H(T) = \int_{-\infty}^T xf(x)dx$$

es

$$\frac{dH(T)}{dT} = Tf(T)$$

Reemplazando en (11.43)

$$\frac{dE(\mathbf{B})}{dT} = 13Tf(T) - 13F(T) - 13Tf(T) + 6 = -13F(T) + 6 = 0$$

Vemos que la segunda derivada es negativa, como corresponde a un máximo. Resulta así

$$F(T) = 6/13 = 0,462$$

Como la demanda tiene distribución Normal, deberá ser

$$\Phi\left(\frac{T-\mu}{\sigma}\right) = 0,462; \quad \frac{T-\mu}{\sigma} = Z_{0,462} = -0,0954$$

por lo tanto, $T = \mu - 0,0954\sigma = 6.500 - 0,0954 \times 700 = 6.433$ ejemplares.

b) Con la expresión (11.42) calculamos el beneficio esperado

$$(11.42) \quad E(\mathbf{B}) = 13H(T) - 13TF(T) + 6T$$

$$H(T) = \mu F(T) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left(-\frac{Z^2}{2}\right) = 2.725$$

Reemplazando en (11.42), obtenemos $E(\mathbf{B}) = \$35.386$.

c) El número Q de ejemplares vendidos depende de la demanda x como sigue

$$x \leq 6.433: Q = x$$

$$x > 6.433: Q = 6.433$$

Entonces, el valor esperado de Q será

$$E(Q) = H(6.433) + 6.433G(6.433) = 6.186.$$

d) Recordemos que el *VEIP* se calcula como diferencia entre el beneficio esperado que podría obtenerse si se dispusiera (semanalmente en este caso) del dato exacto del resultado a obtener, es decir, la demanda x . Si supiéramos dicho valor de demanda, ¿qué tirada haríamos?; obviamente una tirada igual a la demanda, es decir $T=x$. En ese caso, el beneficio sería $(v-c)x$ y su valor esperado sería el *BEIP*, o sea

$$BEIP = (v-c)E(x) = (13-7) \times 6.500 = \$39.000$$

Entonces, el *VEIP* es

$$VEIP = BEIP - E(\mathbf{B}) = \$39.000 - \$35.386 = \$3.614.$$

Si recordamos el *Teorema Fundamental de la Teoría de la Decisión* (vea el Capítulo ...) podemos calcular también el *VEIP* como el arrepentimiento esperado de la estrategia óptima. Para ello, debemos expresar el arrepentimiento en función de la demanda, como sigue:

$$x \leq 6.433: A = 7(6.433 - x) = 45.031 - 7x \quad (\text{Costo de los ejemplares sobrantes}).$$

$$x > 6.433: A = (13-7)(x - 6.433) = 6x - 38.598 \quad (\text{Beneficio que hubieran arrojado los ejemplares no satisfechos}).$$

El arrepentimiento esperado será

$$E(\mathbf{A}) = 45.031F(6.433) - 7H(6.433) + 6J(6.433) - 38.598G(6.433)$$

Verifique el lector que se obtiene nuevamente el valor del *VEIP* = \$3.614.

Esta cantidad (\$3.614) es lo máximo que podría pagarse a una persona que nos informara sin error, todas las semanas, la demanda x que se producirá.

Vamos a generalizar la discusión de este Ejemplo 12. Hemos llamado:

c : Costo unitario del producto.

v : Precio de venta unitario del producto.

v_r : Valor residual unitario del producto.

c_s : Costo de escasez unitario del producto.

x : Demanda del producto, en unidades, durante su período de frescura.

T : Lote óptimo a determinar, en unidades.

B : Beneficio durante el período de frescura

El beneficio (B), expresado en función de la demanda (x) es:

$$x \leq T: B = vx - cT + v_r(T - x) = (v - v_r)x - (c - v_r)T$$

$$x > T: B = (v - c)T - c_s(x - T) = (v - c + c_s)T - c_s x$$

El beneficio esperado será

$$E(B) = (v - v_r)H(T) - (c - v_r)TF(T) + (v - c + c_s)TG(T) - c_s J(T)$$

$$= (v - v_r)H(T) - (c - v_r)TF(T) + (v - c + c_s)T [1 - F(T)] - c_s [\mu - H(T)]$$

$$= (v - v_r + c_s)H(T) - (v - v_r + c_s) TF(T) + (v - c + c_s)T - c_s \mu$$

Derivemos con respecto a T e igualemos a cero:

$$\frac{dE(B)}{dT} = (v - v_r + c_s) \frac{dH(T)}{dT} - (v - v_r + c_s) \left[F(T) + T \frac{dF(T)}{dT} \right] + (v - c + c_s) = 0$$

Ahora, dado que

$$\frac{dH(T)}{dT} = Tf(T); \quad \frac{dF(T)}{dT} = f(T)$$

resulta

$$(11.44) \quad \frac{dE(B)}{dT} = -(v - v_r + c_s)F(T) + (v - c + c_s) = 0$$

La segunda derivada

$$\frac{d^2 E(B)}{dT^2} = -(v - v_r + c_s) f(T)$$

es negativa ya que el valor residual debe ser menor que el precio de venta (y también que el costo). Por lo tanto, la solución es un máximo. Resolvemos (11.44) para $F(T)$:

$$(11.45) \quad F(T) = \frac{v - c + c_s}{v - v_r + c_s}$$

Es decir que el lote óptimo T , será el fractil de área $(v - c + c_s)/(v - v_r + c_s)$ de la distribución de la demanda.

5. ACUMULATIVO DE LA VARIABLE. CURVA ABC. ÍNDICE DE GINI¹

5.1. Acumulativo de la variable

Consideremos la siguiente situación. Un auditor interno debe controlar los registros de facturas de compra en el sistema de la empresa. No puede, ni desea, controlar todas las facturas, pues se trata de un número enorme, de varios miles. Su idea es controlar sólo las de montos significativos, mayores que un valor X_0 , de modo que su trabajo, en volumen, (en cantidad de facturas) sea razonable, pero que pueda controlar un porcentaje importante del monto total. Llamemos x_i' a los montos de las n facturas que controlará, de valores mayores que X_0 , y sea Y el monto total controlado. Tendremos

$$Y = x_1' + x_2' + \dots + x_n'$$

El valor esperado de Y es² $E(Y) = nE(x')$. Pero las variables x' son truncadas a la derecha, de valores mayores que un cierto X_0 , entonces, de acuerdo con la discusión sobre truncamiento que vimos en la Sección 3

$$E(x') = \frac{J(X_0)}{G(X_0)}$$

por lo tanto,

$$E(Y) = nE(x') = n \frac{J(X_0)}{G(X_0)}$$

Deseamos que este valor, $E(Y)$ sea una fracción importante del monto total; llamemos P a dicha fracción. Si hay una población total de N facturas, cuyo monto medio general es μ , el monto total es $N\mu$, por lo tanto, la fracción a controlar será $PN\mu$. Deberá cumplirse entonces:

$$(11.46) \quad n \frac{J(X_0)}{G(X_0)} = PN\mu$$

Suponiendo que conocemos la distribución de la variable, que en el caso de facturas de compra será en general Lognormal, seguramente conoceremos μ ; tal vez también podamos conocer N (número total de facturas) y P será un dato del problema, ya que es la fracción del monto total que queremos controlar. Nuestra incógnita fundamental es X_0 , pero en la expresión (11.46) subsiste el desconocimiento de n , o sea el número de facturas bajo control. Sin embargo, si tenemos en cuenta que, evidentemente, $G(X_0) = n/N$, pues es la fracción de facturas que se controlan con respecto al total, la ecuación se simplifica:

$$(11.47) \quad J(X_0) = P\mu$$

¹ CORRADO GINI (1884-1965), estadístico italiano.

² En el Capítulo ... sobre combinación de variables, veremos que la esperanza de una suma de variables es igual a la suma de sus esperanzas.

Ejemplo 13 (Facturas de compra). En una empresa textil se ha despedido al jefe de compras por causas fáciles de imaginar. El nuevo jefe designado deberá controlar una enorme cantidad de facturas de compra, cuyos montos son variables con distribución Lognormal sumamente asimétrica y dispersa de media \$620 y desvío estándar \$1.100. El nuevo jefe de compras controlará todas las facturas superiores a X_0 , de modo tal de controlar el 70% del monto total. Calcular: a) El valor de X_0 . b) El porcentaje de facturas que controlará.

En la distribución Lognormal, tenemos

$$J(X_0) = \mu - H(X_0) = \mu - \mu \Phi\left(\frac{\ln X_0 - m}{D} - D\right)$$

De acuerdo con (11.47), esta expresión debe ser igual a $P\mu$, por lo tanto

$$1 - \Phi\left(\frac{\ln X_0 - m}{D} - D\right) = P$$

$$\frac{\ln X_0 - m}{D} - D = Z_{1-P} = -Z_P$$

$$(11.48) \quad X_0 = \text{Exp}[m + D(D - Z_P)]$$

Calculemos los parámetros de nuestra Lognormal:

$$m = \ln\left[\frac{\mu}{\sqrt{1 + (\sigma/\mu)^2}}\right] = 5,718435; \quad D = \sqrt{\ln\left[1 + \left(\frac{\sigma}{\mu}\right)^2\right]} = 1,192715$$

Para $P=0,7$, es $Z_P = Z_{0,7} = 0,524401$, entonces, de (11.48), obtenemos $X_0 = \$675,56$.

b) El porcentaje de facturas que controlará es:

$$P(x > X_0) = 1 - \Phi\left(\frac{\ln X_0 - m}{D}\right) = 1 - \Phi(0,6683) = 1 - 0,748 = 0,252 = 25,2\%$$

La expresión de cálculo de esta probabilidad se simplifica, pues reemplazando X_0 por su igual en (11.48) se obtiene $P(x > X_0) = \Phi(Z_P - D)$, que claramente es menor que P .

Ejemplo 14. De acuerdo con datos del diario *Clarín* del día 24 de agosto de 2003, los salarios de la Argentina presentan la siguiente distribución de frecuencias (página siguiente). Calcular: a) A partir de la tabla, la relación de ingreso total entre el 10% superior y el 10% inferior. b) Los parámetros de la distribución Lognormal que rige esta variable y su grado de asimetría. c) A partir del modelo Lognormal, la relación de ingreso total entre el 10% superior y el 10% inferior.

a) Sea N la población total activa de nuestro país. El ingreso total de cualquier decil es $0,1N\mu_i$, siendo μ_i el ingreso medio del decil i . Entonces, la relación de ingresos entre el 10% superior y el 10% inferior será:

Grupo social	Media May 2003
Decil 1	82
Decil 2	150
Decil 3	192
Decil 4	249
Decil 5	316
Decil 6	396
Decil 7	487
Decil 8	621
Decil 9	872
Decil 10	1.975

$$\frac{0,1N\mu_{10}}{0,1N\mu_1} = \frac{1.975}{82} = 24,1$$

b) Calculemos la media y el desvío estándar de los datos de la tabla. Si tenemos una población total de N individuos, es

$$\mu = \frac{1}{N} 0,1N \sum_{i=1}^{10} \mu_i = 534$$

$$\sigma^2 = \frac{1}{N} 0,1N \sum_{i=1}^{10} \mu_i^2 - \mu^2 = 281.702$$

$$\sigma = 530,76$$

Los parámetros de la Lognormal serán

$$m = \ln \left[\frac{\mu}{\sqrt{1 + (\sigma/\mu)^2}} \right] = 5,936856; \quad D = \sqrt{\ln [1 + (\sigma/\mu)^2]} = 0,828903$$

El grado de asimetría (Capítulo 8) está dado por

$$\alpha_3 = \frac{E[(x - \mu)^3]}{\sigma^3} = [Exp(D^2) + 2] \sqrt{Exp(D^2) - 1} = 3,964$$

c) Si llamamos X_{1-h} al percentil $(1-h)$, el ingreso medio de los individuos con ingresos superiores a X_{1-h} es $J(X_{1-h})/h$. Análogamente, el ingreso medio de los individuos con ingresos inferiores a X_h es $H(X_h)/h$. La relación de ingresos será entonces

$$\mathfrak{R} = \frac{J(X_{1-h})}{H(X_h)} = \frac{\mu - H(X_{1-h})}{H(X_h)}$$

Para el caso que nos ocupa, la variable x tiene distribución Lognormal, entonces

$$(11.49) \quad \mathfrak{R} = \frac{\mu - \mu \Phi\left(\frac{\ln X_{1-h} - m}{D} - D\right)}{\mu \Phi\left(\frac{\ln X_h - m}{D} - D\right)} = \frac{1 - \Phi\left(\frac{\ln X_{1-h} - m}{D} - D\right)}{\Phi\left(\frac{\ln X_h - m}{D} - D\right)}$$

Además

$$\ln X_{1-h} = m + DZ_{1-h}; \quad \ln X_h = m - DZ_{1-h}$$

Reemplazando en (11.49):

$$(11.49') \quad \mathfrak{R} = \frac{1 - \Phi\left(\frac{m + DZ_{1-h} - m}{D} - D\right)}{\Phi\left(\frac{m - DZ_{1-h} - m}{D} - D\right)} = \frac{1 - \Phi(Z_{1-h} - D)}{\Phi(-Z_{1-h} - D)} = \frac{1 - \Phi(Z_{1-h} - D)}{1 - \Phi(Z_{1-h} + D)}$$

Reemplazando $Z_{1-h} = Z_{0,9} = 1,2816$ y $D = 0,8289$, se obtiene $\mathfrak{R} = 18,7$, en un cierto desacuerdo con respecto al valor obtenido de los datos (24,1).

5.2. Curva ABC. Índice de GINI

Volviendo a nuestro problema del auditor (Ejemplo 13), vimos la relación

$$(11.47) \quad J(X_0) = P\mu$$

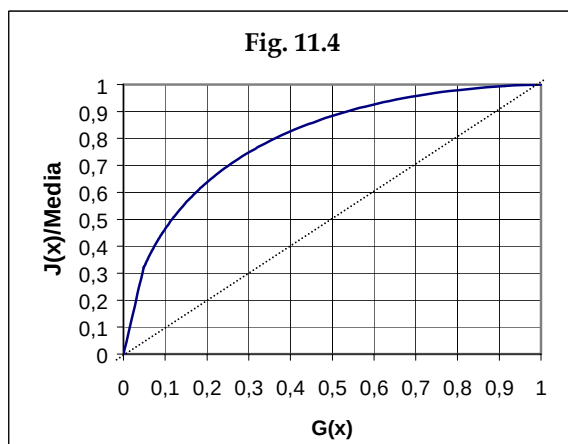
o sea

$$(11.47') \quad P = \frac{J(X_0)}{\mu}$$

Así, pues, el monto efectivo acumulado de los valores superiores a X_0 es igual a $J(X_0)/\mu$. Por otra parte, el porcentaje de individuos con valores superiores a X_0 es $G(X_0)$. En el Ejemplo 13 teníamos $G(X_0) = 0,252$ y $J(X_0)/\mu = 0,7$, es decir, que controlando un 25,2 de las facturas, las de montos superiores a X_0 , controlamos un 70% del monto total. La representación gráfica de $J(X_0)/\mu$, llevando como variable de abscisas a $G(X_0)$, es la denominada *curva ABC*, llamada también *curva de LORENZ*.

Ejemplo 15 (Continuación del Ejemplo 13 –Facturas de compra–). Trazar la curva ABC para los datos del Ejemplo 13.

Tenemos $m = 5,7184$ y $D = 1,1927$. Conviene que calculemos los valores paramétricos, tomando como parámetro a X_0 , y así obtenemos la siguiente tabla cuya representación gráfica vemos en la Fig. 11.4 (Curva ABC)



x	G(x)	J(x) / μ	x	G(x)	J(x) / μ
	0	0	1100	0,141	0,546
2200	0,049	0,321	1000	0,159	0,578
2100	0,053	0,335	900	0,182	0,612
2000	0,057	0,350	800	0,209	0,649
1900	0,062	0,366	700	0,243	0,690
1800	0,068	0,383	600	0,285	0,734
1700	0,075	0,402	500	0,339	0,781
1600	0,082	0,421	400	0,409	0,832
1500	0,091	0,443	300	0,505	0,886
1400	0,100	0,465	200	0,638	0,939
1300	0,112	0,490	100	0,825	0,983
1200	0,125	0,517	0	1,000	1,000

Así, por ejemplo, observamos que si se controla el 10% de las facturas, se controla el 46,5 del monto total.

Cuanto más empinada es la curva ABC, más concentración tiene la variable en sus valores altos, es decir, con unos pocos individuos se controla una importante masa del monto total. En el problema del auditor, lo más conveniente para él es que la curva sea lo más empinada posible. Sin embargo, si consideramos por ejemplo, los clientes de una empresa, en que la x es el monto de sus compras, una curva muy empinada es peligrosa, pues unos pocos clientes concentran un volumen muy alto de ventas; en ese caso, la pérdida de un cliente importante, puede traer un quebranto. El empinamiento de la curva ABC está en estrecha dependencia con el coeficiente de variación de la variable x , es decir, cuanto más alta es la relación σ/μ , más empinada es la curva y por ende, mayor es la concentración en los valores altos.

Para el caso de los salarios de un país, una curva muy empinada significa que un pequeño porcentaje de la población total concentra un alto porcentaje de la renta, es decir, cuanto más empinada es la curva, más desigual es la distribución de la renta.

Hay varias medidas del grado de concentración de la renta; una de ellas es la que vimos en el Ejemplo 14, esto es, la relación del ingreso total del 10% más rico al 10% más pobre.

Otra medida, muy utilizada, es denominado Índice de GINI (η), que es el doble del área comprendida entre la curva ABC y la recta a 45 grados. Siempre es un número comprendido entre 0 y 1; cuanto más cercano a 1, más desigual es la distribución de la renta. En los países escandinavos, dicho índice está en el orden de 0,3, pero en casi toda América Latina supera a 0,4. Valores superiores a 0,6 significan una distribución sumamente inequitativa.

Para todas las distribuciones, el Índice de GINI se puede calcular en función de los parámetros de forma, si los hay, o, si no, en función del coeficiente de variación. En general, a mayor coeficiente de variación, mayor es el Índice GINI.

Por el tipo de variable al que se aplica este índice, sólo tiene sentido que indiquemos, fundamentalmente, las expresiones para los modelos Lognormal y de PARETO, pero también lo daremos para otras, como una cuestión de interés teórico.

Distribución Lognormal

$$(11.50) \quad \eta = 2\Phi\left(\frac{D}{\sqrt{2}}\right) - 1$$

Distribución de PARETO

$$(11.51) \quad \eta = \frac{1}{2\psi - 1}$$

Distribución Normal

$$(11.52) \quad \eta = \frac{\sigma}{\mu\sqrt{\pi}}$$

Distribución Gamma

$$(11.53) \quad \eta = \frac{\Gamma(2r)}{2^{2r-1} r \Gamma^2(r)}$$

Distribución de WEIBULL

$$(11.54) \quad \eta = 1 - 2^{-1/\omega}$$

Distribución de GUMBEL del mínimo

$$(11.55) \quad \eta = \frac{\beta \text{Ln } 2}{\theta - C\beta} = \frac{\sigma}{\mu} \frac{\sqrt{6}}{\pi} \text{Ln } 2$$

Distribución de GUMBEL del máximo

$$(11.56) \quad \eta = \frac{\beta \text{Ln } 2}{\theta + C\beta} = \frac{\sigma}{\mu} \frac{\sqrt{6}}{\pi} \text{Ln } 2$$

Distribución empírica dada con una tabla de frecuencias

Sea la tabla de frecuencias

X	$a_1 - b_1$	$a_2 - b_2$	...	$a_k - b_k$
F_i	F_1	F_2	...	F_k
$\tau_i \cong F_i X_i$	τ_1	τ_2	...	τ_k

Los totales efectivos τ_i son los montos totales de cada intervalo que, a veces, pueden ser dato, de lo contrario se estiman multiplicando la frecuencia (F_i) del intervalo por el punto medio del mismo $[X_i = (a_i + b_i)/2]$. Llamemos

$\tau = \sum_{i=1}^k \tau_i$ al total efectivo de la población

$\varphi_i = \sum_{h=1}^i \frac{\tau_h}{\tau}$ la proporción acumulada del efectivo hasta el intervalo i

La fórmula de BROWN (1994) para el Índice de GINI es

$$(11.57) \quad \eta = \left| 1 - \frac{1}{n} \left[F_k + \sum_{i=1}^{k-1} (F_i + F_{i+1}) \varphi_i \right] \right|$$

o también, si llamamos f_i a la frecuencia relativa F_i/n , tenemos

$$(11.57') \quad \eta = \left| 1 - \left[f_k + \sum_{i=1}^{k-1} (f_i + f_{i+1}) \varphi_i \right] \right|$$

Ejemplo 16 (Continuación del Ejemplo 14). Calcular el Índice de GINI para los salarios de nuestro país en agosto 2003.

A partir de la expresión

$$(11.50) \quad \eta = 2\Phi\left(\frac{D}{\sqrt{2}}\right) - 1$$

obtenemos, para $D = 0,8289$, $\eta = 0,442$.

Calculémoslo ahora con la fórmula de BROWN (11.57'). La tabla de frecuencias es

Grupo social	$f_i = F_i/n$	$f_i + f_{i+1}$	Media	φ_i	$(f_i + f_{i+1})\varphi_i$
Decil 1	0,1	0,2	82	0,01536	0,003071
Decil 2	0,1	0,2	150	0,04345	0,008689
Decil 3	0,1	0,2	192	0,07940	0,015880
Decil 4	0,1	0,2	249	0,12603	0,025206
Decil 5	0,1	0,2	316	0,18521	0,037041
Decil 6	0,1	0,2	396	0,25936	0,051873
Decil 7	0,1	0,2	487	0,35056	0,070112
Decil 8	0,1	0,2	621	0,46685	0,093371
Decil 9	0,1	0,2	872	0,63015	0,126030
Decil 10	0,1		1.975	1,00000	

Al ser en ese caso $f_i = F_i/n = 0,1$ para todos los intervalos, tenemos

$$\eta = \left| 1 - \left[f_k + \sum_{i=1}^{k-1} (f_i + f_{i+1}) \varphi_i \right] \right| = \left| 1 - [0,1 + (0,003071 + 0,008689 + \dots + 0,126030)] \right| = 0,4687$$

Ejemplo 17 (Segundo ejemplo con salarios del país). En la página de Internet

www.cambiocultural.com.ar/actualidad/distribuciondelingreso.htm

encontramos los siguientes datos de ingresos de Argentina correspondientes al primer semestre del año 2004 (Fuente: INDEC, Encuesta permanente de hogares), para una población de $N=13.009.707$ personas con ingresos.

f_i =Fracción de la población	0,153	0,132	0,123	0,109	0,092	0,093	0,084	0,078	0,073	0,063
τ_i = Ingreso total (\$10 ⁶)	174,60	305,00	401,10	471,20	504,40	643,50	738,10	885,90	1.174,60	2.229,80

Calcular: a) La media y el desvío estándar del ingreso de Argentina. b) La relación de ingresos totales entre el 10% más rico y el 10% más pobre y el Índice de GINI, suponiendo que la distribución es Lognormal. c) El Índice de GINI con la fórmula de BROWN.

a) El cálculo de la media es inmediato pues sólo hay que sumar los ingresos totales y dividir por el total de la población, o sea

$$\mu = \frac{1}{N} \sum_{i=1}^k \tau_i = \frac{1}{13.009.707} (174,6 + 305,0 + \dots + 2.229,8) \times 10^6 = \$578,66$$

Para calcular la varianza, conviene obtener la media de cada intervalo, dividiendo el ingreso total por la cantidad de individuos, o sea:

$$X_i = \frac{\tau_i}{F_i} = \frac{\tau_i}{N f_i}$$

Para cada intervalo, debemos calcular el producto $F_i X_i^2$, o sea

$$F_i X_i^2 = \frac{\tau_i^2}{F_i}$$

La varianza será entonces

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k F_i X_i^2 - \mu^2$$

Resulta así la varianza y el desvío estándar:

$$\sigma^2 = 408.626,64 ; \sigma = \$639,24$$

También podremos calcular, para cada intervalo, la fracción de ingreso acumulado (φ_i) para luego aplicar la fórmula de BROWN. Tendremos la tabla siguiente

i	f_i	$f_i + f_{i+1}$	$F_i = N f_i$	Frec acum.	τ_i (\$10 ⁶)	Ingreso acum	φ_i	$(f_i + f_{i+1}) \varphi_i$	$F_i X_i^2$
1	0,153	0,285	1.990.485	1.990.485	174,60	174,60	0,0232	0,0066	15.315.443.221
2	0,132	0,255	1.717.281	3.707.766	305,00	479,60	0,0637	0,0162	54.169.934.915
3	0,123	0,232	1.600.194	5.307.960	401,10	880,70	0,1170	0,0271	100.538.565.949
4	0,109	0,201	1.418.058	6.726.018	471,20	1.351,90	0,1796	0,0361	156.572.890.531
5	0,092	0,185	1.196.893	7.922.911	504,40	1.856,30	0,2466	0,0456	212.566.503.438
6	0,093	0,177	1.209.903	9.132.814	643,50	2.499,80	0,3321	0,0588	342.252.436.766
7	0,084	0,162	1.092.815	10.225.629	738,10	3.237,90	0,4301	0,0697	498.521.350.823
8	0,078	0,151	1.014.757	11.240.386	885,90	4.123,80	0,5478	0,0827	773.405.662.636
9	0,073	0,136	949.709	12.190.095	1.174,60	5.298,40	0,7038	0,0957	1.452.745.167.204
10	0,063		819.612	13.009.707	2.229,80	7.528,20	1,0000		6.066.294.832.189

b) Calculemos los parámetros de la Lognormal

$$m = \ln \left[\frac{\mu}{\sqrt{1 + (\sigma/\mu)^2}} \right] = 5,96189; \quad D = \sqrt{\ln[1 + (\sigma/\mu)^2]} = 0,89312$$

La relación de ingresos entre el 10% más pobre y el 10% más rico está dada por la expresión (11.49’):

$$(11.49') \quad R = \frac{1 - \Phi(Z_{0,9} - D)}{1 - \Phi(Z_{0,9} + D)} = \frac{1 - \Phi(1,28155 - 0,89312)}{1 - \Phi(1,28155 + 0,89312)} = \frac{1 - \Phi(0,38843)}{1 - \Phi(2,17467)} =$$

$$= \frac{1 - 0,65115}{1 - 0,98517} = 23,53$$

El Índice de GINI está dado por la expresión (11.50):

$$(11.50) \quad \eta = 2\Phi\left(\frac{D}{\sqrt{2}}\right) - 1 = 2\Phi\left(\frac{0,89312}{\sqrt{2}}\right) - 1 = 2\Phi(0,6315) - 1 = 0,472$$

c) La fórmula de BROWN para el Índice de GINI es la (11.57’):

$$(11.57') \quad \eta = \left| 1 - \left[f_k + \sum_{i=1}^{k-1} (f_i + f_{i+1})\varphi_i \right] \right|$$

De la tabla tenemos:

$$\sum_{i=1}^{k-1} (f_i + f_{i+1})\varphi_i = 0,0066 + 0,0162 + \dots + 0,0957 = 0,4386$$

siendo además $f_k = 0,063$. de donde resulta $\eta = 0,4984$.

Ejemplo 18 (Tercer ejemplo con salarios del país). El jueves 9 de febrero de 2006, diario *Clarín*, página 14, publicó un completo artículo titulado *La brecha de ingresos, más amplia*. Ofrece los siguientes datos que el INDEC difundió el día anterior: 1) El 10% más rico de la población total tuvo ingresos 31 veces superiores a los del 10% más pobre. 2) La misma relación, para los percentiles 80 y 20 es de 15,3 veces. 3) El ingreso medio del 10% más pobre es de \$106. 4) El ingreso medio del 10% más rico es de \$3.268. 5) El ingreso medio global es de \$878. Estos datos comprenden a **toda la población con ingresos**, sea que trabajen en el sector formal (en blanco), en el informal (en negro), son asalariados, hacen changas, son profesionales, patrones, cuentapropistas, jubilados, rentistas o tienen un plan social. A partir de la información sobre los percentiles 90 y 10, se desea calcular: a) El parámetro D de la distribución Lognormal que rige esta variable. b) Verificar, con dicho modelo, el valor de la media $\mu = \$878$. c) El desvío estándar σ , el parámetro m y los percentiles 10 y 90 de los salarios. d) El Índice de GINI. e) Verificar, con el modelo Lognormal, la relación de ingresos de 15,3 entre el 20% más rico y el 20% más pobre.

a) Según vimos en el Ejemplo 14, la relación de ingresos entre el fractil $(1-h)$ y el fractil h está dada por la expresión

$$(11.49') \quad \mathfrak{R} = \frac{1 - \Phi(Z_{1-h} - D)}{1 - \Phi(Z_{1-h} + D)}$$

Esta relación vale, según el enunciado $3.268/106=30,83$ para $h=0,1$, es decir que tenemos

$$30,83 = \frac{1 - \Phi(Z_{0,9} - D)}{1 - \Phi(Z_{0,9} + D)}$$

Resolvemos esta ecuación con la función SOLVER del EXCEL y obtenemos $D = 0,96829$.

b) Podemos calcular la media μ global de la población a partir del dato de que el 10% inferior tiene un salario medio de $\mu_1 = \$106$. Esta es la media de la variable truncada a la izquierda, o sea:

$$\mu_1 = \frac{H(X_1)}{F(X_1)} = \frac{\mu \Phi\left(\frac{\ln X_1 - m}{D} - D\right)}{\Phi\left(\frac{\ln X_1 - m}{D}\right)}$$

No conocemos, todavía, el valor de X_1 , pero como el denominador de esta expresión es 0,1, tenemos

$$\frac{\ln X_1 - m}{D} = Z_{0,1} = -Z_{0,9}$$

luego

$$\mu_1 = \frac{\mu \Phi(Z_h - D)}{h} = \frac{\mu \Phi(-Z_{1-h} - D)}{h} = \frac{\mu [1 - \Phi(Z_{1-h} + D)]}{h}$$

Poniendo $\mu_1 = \$106$ y $h=0,1$, resulta la media global $\mu = \$866,75$, en razonable acuerdo con el valor indicado en el artículo de \$878. Para el percentil superior, se tiene

$$\mu_2 = \frac{J(X_2)}{G(X_2)} = \frac{\mu \left[1 - \Phi\left(\frac{\ln X_2 - m}{D} - D\right) \right]}{1 - \Phi\left(\frac{\ln X_2 - m}{D}\right)}$$

No conocemos, todavía, el valor de X_2 , pero como el denominador de esta expresión es 0,1, tenemos

$$\frac{\ln X_2 - m}{D} = Z_{0,9}$$

luego

$$\mu_2 = \frac{\mu [1 - \Phi(Z_{1-h} - D)]}{h}$$

Poniendo $\mu_2 = \$3.268$ y $h=0,1$, resulta nuevamente el valor $\mu = \$866,75$

c) Para calcular el desvío estándar σ , partimos de

$$D = \sqrt{\ln \left[1 + \left(\frac{\sigma}{\mu} \right)^2 \right]} = 0,96829$$

Por lo tanto

$$\frac{\sigma}{\mu} = \sqrt{\exp(D^2) - 1} = 1,24651 ; \sigma = \$1.080,42$$

El parámetro m será entonces

$$m = \ln \left[\frac{\mu}{\sqrt{1 + \left(\frac{\sigma}{\mu} \right)^2}} \right] = 6,29596$$

Ahora podemos calcular los percentiles 10 y 90 de los salarios:

$$X_1 = X_{0,1} = \exp(m + Z_{0,1}.D) = \exp(m - Z_{0,9}.D) = \$157$$

$$X_2 = X_{0,9} = \exp(m + Z_{0,9}.D) = \$1.876$$

d) El Índice de GINI está dado por

$$\eta = 2\Phi\left(\frac{D}{\sqrt{2}}\right) - 1 = 0,506$$

e) Aplicamos nuevamente la expresión (11.49')

$$(11.49') \quad \mathfrak{R} = \frac{1 - \Phi(Z_{1-h} - D)}{1 - \Phi(Z_{1-h} + D)}$$

Poniendo $h=0,2$, obtenemos el valor de $\mathfrak{R} = 15,7$, en buen acuerdo con el 15,3 dado en el artículo.

6. APÉNDICE

6.1 DEMOSTRACIONES DE LA EXPRESIONES DE LA ESPERANZA MATEMÁTICA PARCIAL PARA LAS DISTINTAS DISTRIBUCIONES

6.1.1 Distribución Normal [Expresión (11.8)]

Dado que

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{Exp}\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

es

$$H(x) = \int_{-\infty}^x xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x x \text{Exp}\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

Ponemos $Z = \frac{x-\mu}{\sigma}$ entonces $x = \mu + \sigma Z$; $dx = \sigma dZ$

$$\begin{aligned} H(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z (\mu + \sigma Z) \text{Exp}\left(-\frac{Z^2}{2}\right) dZ = \\ &= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^Z \text{Exp}\left(-\frac{Z^2}{2}\right) dZ + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^Z Z \text{Exp}\left(-\frac{Z^2}{2}\right) dZ \end{aligned}$$

En el último miembro, la primera integral, dividida por $\sqrt{2\pi}$, es $\Phi(Z)$ y la última es elemental y vale $-\text{Exp}(-Z^2/2)$. Resulta entonces la expresión (11.8)

$$(11.8) \quad H(x) = \int_{-\infty}^x xf(x)dx = \mu \Phi(Z) - \frac{\sigma}{\sqrt{2\pi}} \text{Exp}\left(-\frac{Z^2}{2}\right) \quad ; \quad Z = \frac{x-\mu}{\sigma}$$

6.1.2. Distribución Lognormal [Expresión (11.16)]

Dado que

$$f(x) = \frac{1}{xD\sqrt{2\pi}} \text{Exp}\left[-\frac{1}{2}\left(\frac{\text{Ln } x - m}{D}\right)^2\right]$$

es

$$H(x) = \int_0^x xf(x)dx = \frac{1}{D\sqrt{2\pi}} \int_0^x \text{Exp}\left[-\frac{1}{2}\left(\frac{\text{Ln } x - m}{D}\right)^2\right] dx$$

Ponemos $Z = \frac{\text{Ln } x - m}{D}$ entonces $x = \text{Exp}(m+DZ)$; $dx = D\text{Exp}(m+DZ)dZ$

$$H(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z \text{Exp}\left(-\frac{Z^2}{2}\right) \text{Exp}(m + DZ) dZ = \frac{\text{Exp}(m)}{\sqrt{2\pi}} \int_{-\infty}^Z \text{Exp}\left(DZ - \frac{Z^2}{2}\right) dZ$$

Completamos el cuadrado en la última integral

$$DZ - \frac{Z^2}{2} = \frac{D^2}{2} - \frac{1}{2}(Z - D)^2$$

y queda

$$\begin{aligned} H(x) &= \frac{\text{Exp}(m + D^2/2)}{\sqrt{2\pi}} \int_{-\infty}^Z \text{Exp}\left[-\frac{1}{2}(Z - D)^2\right] dZ = \frac{\text{Exp}(m + D^2/2)}{\sqrt{2\pi}} \int_{-\infty}^{Z-D} \text{Exp}\left(-\frac{u^2}{2}\right) du \\ &= \text{Exp}\left(m + \frac{D^2}{2}\right) \Phi(Z - D) \end{aligned}$$

o sea finalmente, la expresión (11.16)

$$(11.16) \quad H(x) = \int_0^x x f(x) dx = \left[\text{Exp}\left(m + \frac{D^2}{2}\right) \right] \Phi\left(\frac{\ln x - m}{D} - D\right) = \mu \Phi\left(\frac{\ln x - m}{D} - D\right)$$

6.1.3. Distribución Gamma [Expresión (11.18)]

Dado que

$$f(x) = \frac{1}{\beta \Gamma(r)} \left(\frac{x}{\beta}\right)^{r-1} e^{-x/\beta}$$

es

$$\begin{aligned} H(x) &= \frac{1}{\beta \Gamma(r)} \int_0^x x \left(\frac{x}{\beta}\right)^{r-1} e^{-x/\beta} dx = \frac{\beta}{\Gamma(r+1)} \int_0^x \left(\frac{x}{\beta}\right)^r e^{-x/\beta} dx = \\ &= r\beta \left[\frac{1}{\beta \Gamma(r+1)} \int_0^x \left(\frac{x}{\beta}\right)^r e^{-x/\beta} dx \right] \end{aligned}$$

La expresión dentro del corchete es la $F_g(x|r+1;\beta)$, con lo cual resulta la expresión (11.18):

$$(11.18) \quad H(x) = r\beta F_g(x|r+1;\beta)$$

6.1.4. Distribución de WEIBULL [Expresión (11.21)]

Dado que

$$f(x) = \frac{\omega x^{\omega-1}}{\beta^\omega} \text{Exp} \left[-\left(\frac{x}{\beta} \right)^\omega \right]$$

es

$$H(x) = \frac{\omega}{\beta^\omega} \int_0^x x x^{\omega-1} \text{Exp} \left[-\left(\frac{x}{\beta} \right)^\omega \right] dx = \omega \int_0^x \left(\frac{x}{\beta} \right)^\omega \text{Exp} \left[-\left(\frac{x}{\beta} \right)^\omega \right] dx$$

Ponemos en la integral

$$y = \left(\frac{x}{\beta} \right)^\omega ; \quad x = \beta y^{1/\omega} ; \quad dx = \frac{\beta}{\omega} y^{1/\omega-1} dy$$

y queda

$$H(x) = \omega \int_0^y y e^{-y} \frac{\beta}{\omega} y^{1/\omega-1} dy = \beta \int_0^y y^{1/\omega} e^{-y} dy = \beta \Gamma \left(1 + \frac{1}{\omega} \right) \left[\frac{1}{\Gamma \left(1 + \frac{1}{\omega} \right)} \int_0^y y^{1/\omega} e^{-y} dy \right]$$

La expresión entre corchetes es la $F_g(y \mid r=1+1/\omega; \beta'=1)$. Resulta entonces la expresión (11.21):

$$(11.21) \quad H(x) = \int_0^x x f(x) dx = \beta \Gamma \left(1 + \frac{1}{\omega} \right) F_g \left[\left(\frac{x}{\beta} \right)^\omega \mid r = 1 + \frac{1}{\omega}; \beta' = 1 \right]$$

6.1.5. Distribución de PARETO [Expresión (11.23)]

Dado que

$$f(x) = \frac{\psi \theta^\psi}{x^{\psi+1}}$$

es

$$H(x) = \int_\theta^x x \frac{\psi \theta^\psi}{x^{\psi+1}} dx = \psi \theta^\psi \int_\theta^x x^{-\psi} dx = \frac{\psi \theta^\psi}{-\psi+1} x^{-\psi+1} \Big|_\theta^x = \frac{\psi \theta^\psi}{-\psi+1} (x^{-\psi+1} - \theta^{-\psi+1})$$

Obtenemos así la expresión (11.23)

$$(11.23) \quad H(x) = \frac{\psi \theta}{\psi-1} \left[1 - \left(\frac{\theta}{x} \right)^{\psi-1} \right]$$

6.1.6. Distribución de GUMBEL del mínimo [Expresiones (11.24) y (11.27)]

Dado que

$$f(x) = \frac{1}{\beta} \text{Exp}\left(\frac{x-\theta}{\beta}\right) \text{Exp}\left[-\text{Exp}\left(\frac{x-\theta}{\beta}\right)\right]$$

es

$$H(x) = \int_{-\infty}^x x f(x) dx = \frac{1}{\beta} \int_{-\infty}^x x \text{Exp}\left(\frac{x-\theta}{\beta}\right) \text{Exp}\left[-\text{Exp}\left(\frac{x-\theta}{\beta}\right)\right] dx$$

$$\text{Llamamos } t = \text{Exp}\left(\frac{x-\theta}{\beta}\right); \quad x = \theta + \beta \text{Ln } t; \quad dx = \beta \frac{dt}{t}$$

y queda

$$H(x) = \frac{1}{\beta} \int_0^t (\theta + \beta \text{Ln } t) t e^{-t} \beta \frac{dt}{t} = \theta \int_0^t e^{-t} dt + \beta \int_0^t e^{-t} \text{Ln } t dt$$

En el último miembro, la primera integral es elemental y vale $1 - e^{-t}$, entonces resulta la expresión (11.24)

$$(11.24) \quad H(x) = \theta(1 - e^{-t}) + \beta \int_0^t e^{-t} \text{Ln } t dt$$

Como la última integral, que llamaremos $I(t)$, no es elemental, vamos a obtener su expresión mediante un desarrollo en serie. Podemos poner

$$e^{-t} \text{Ln } t = \text{Ln } t \sum_{n=0}^{\infty} \frac{(-1)^n t^n}{n!}$$

La primitiva de $t^n \text{Ln } t$ es

$$\int t^n \text{Ln } t dt = \frac{t^{n+1}}{n+1} \text{Ln } t - \frac{t^{n+1}}{(n+1)^2}$$

Esta primitiva se anula en $t = 0$, por lo tanto, entre 0 y t adopta la misma expresión. Tenemos así

$$I(t) = \int_0^t e^{-t} \text{Ln } t dt = \text{Ln } t \sum_{n=0}^{\infty} \frac{(-1)^n t^{n+1}}{n!(n+1)} - \sum_{n=0}^{\infty} \frac{(-1)^n t^{n+1}}{n!(n+1)^2} = \text{Ln } t \sum_{n=0}^{\infty} \frac{(-1)^n t^{n+1}}{(n+1)!} + \sum_{n=1}^{\infty} \frac{(-1)^n t^n}{n!n}$$

En el último miembro, la primera serie es

$$\sum_{n=0}^{\infty} \frac{(-1)^n t^{n+1}}{(n+1)!} = t - \frac{t^2}{2!} + \frac{t^3}{3!} - \dots = 1 - e^{-t}$$

Resulta entonces

$$(11.25) \quad I(t) = \int_0^t e^{-t} \ln t \, dt = (1 - e^{-t}) \ln t + \sum_{n=1}^{\infty} \frac{(-1)^n t^n}{n!n}$$

$$(11.24') \quad H(x) = \theta(1 - e^{-t}) + \beta \int_0^t e^{-t} \ln t \, dt = (\theta + \beta \ln t)(1 - e^{-t}) + \beta \sum_{n=0}^{\infty} \frac{(-1)^n t^n}{n!n}$$

Esta es la expresión exacta para $H(x)$. Vamos ahora a obtener una aproximación para la integral $I(t)$. Para ello, recordemos la aproximación de WILSON-HIFERTY para la función de distribución de la distribución Gamma:

$$F_g(t \setminus r; \beta) = \frac{1}{\beta \Gamma(r)} \int_0^t \left(\frac{t}{\beta} \right)^{r-1} e^{-t/\beta} dt \cong \Phi \left\{ 3\sqrt{r} \left[\left(\frac{t}{r\beta} \right)^{1/3} + \frac{1}{9r} - 1 \right] \right\}$$

Para $\beta = 1$ es

$$\frac{1}{\Gamma(r)} \int_0^t t^{r-1} e^{-t} dt \cong \Phi \left\{ 3\sqrt{r} \left[\left(\frac{t}{r} \right)^{1/3} + \frac{1}{9r} - 1 \right] \right\} = \Phi \left(3r^{1/6} t^{1/3} + \frac{r^{-1/2}}{3} - 3r^{1/2} \right)$$

$$(11.58) \quad \int_0^t t^{r-1} e^{-t} dt \cong \Gamma(r) \Phi \left(3r^{1/6} t^{1/3} + \frac{r^{-1/2}}{3} - 3r^{1/2} \right)$$

Dado que la derivada de $\Phi(Z)$ es

$$\frac{d\Phi(Z)}{dZ} = \frac{1}{\sqrt{2\pi}} \text{Exp} \left(-\frac{Z^2}{2} \right)$$

la derivada de (11.58) con respecto a r es

$$\begin{aligned} \int_0^t t^{r-1} e^{-t} \ln t \, dt &\cong \Gamma'(r) \Phi \left(3r^{1/6} t^{1/3} + \frac{r^{-1/2}}{3} - 3r^{1/2} \right) + \\ &+ \Gamma(r) \frac{1}{\sqrt{2\pi}} \left\{ \text{Exp} \left[-\frac{1}{2} \left(3r^{1/6} t^{1/3} + \frac{r^{-1/2}}{3} - 3r^{1/2} \right)^2 \right] \right\} \left(\frac{3r^{-5/6}}{6} t^{1/3} - \frac{r^{-3/2}}{6} - \frac{3}{2} r^{-1/2} \right) \end{aligned}$$

Ahora evaluamos esta expresión en $r = 1$:

$$\int_0^t e^{-t} \ln t \, dt \cong \Gamma'(1) \Phi \left(3t^{1/3} + \frac{1}{3} - 3 \right) + \Gamma(1) \frac{1}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{1}{6} - \frac{3}{2} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]$$

Reemplacemos

$$\Gamma(1) = 1 ; \quad \Gamma'(1) = -C = -0,577215665 ;$$

Además

$\Phi\left(3t^{1/3} + \frac{1}{3} - 3\right) = \Phi\left(3t^{1/3} - \frac{8}{3}\right)$ es la aproximación de $F_g(t; r=1; \beta=1) = 1 - e^{-t}$

Entonces tenemos la expresión (11.27)

$$(11.27) \quad I(t) = \int_0^t e^{-t} \ln t \, dt \cong -C(1 - e^{-t}) + \frac{1}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]$$

Llevando este resultado a (11.24)

$$\begin{aligned} H(x) &= \theta(1 - e^{-t}) + \beta \int_0^t e^{-t} \ln t \, dt \cong \\ &\cong \theta(1 - e^{-t}) - C\beta(1 - e^{-t}) + \frac{\beta}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right] \end{aligned}$$

obtenemos finalmente la expresión (11.28)

$$(11.28) \quad H(x) \cong (\theta - C\beta)(1 - e^{-t}) + \frac{\beta}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]$$

6.1.7. Distribución de GUMBEL del máximo [Expresiones (11.29) y (11.31)]

Dado que

$$f(x) = \frac{1}{\beta} \text{Exp} \left(-\frac{x - \theta}{\beta} \right) \text{Exp} \left[-\text{Exp} \left(\frac{x - \theta}{\beta} \right) \right]$$

es

$$H(x) = \int_{-\infty}^x x f(x) dx = \frac{1}{\beta} \int_{-\infty}^x x \text{Exp} \left(-\frac{x - \theta}{\beta} \right) \text{Exp} \left[-\text{Exp} \left(\frac{x - \theta}{\beta} \right) \right] dx$$

$$\text{Llamamos } t = \text{Exp} \left(-\frac{x - \theta}{\beta} \right); \quad x = \theta - \beta \ln t; \quad dx = -\beta \frac{dt}{t}$$

y queda

$$(11.59) \quad H(x) = \frac{1}{\beta} \int_{\infty}^t (\theta - \beta \ln t) t e^{-t} \left(-\beta \frac{dt}{t} \right) = \theta \int_t^{\infty} e^{-t} dt - \beta \int_t^{\infty} e^{-t} \ln t \, dt$$

En el último miembro, la primera integral es elemental y vale e^{-t} . La última integral puede expresarse como sigue

$$\int_t^{\infty} e^{-t} \ln t \, dt = \int_0^{\infty} e^{-t} \ln t \, dt - \int_0^t e^{-t} \ln t \, dt = -C - \int_0^t e^{-t} \ln t \, dt = -C - I(t)$$

Resulta entonces, reemplazando en (11.59), la expresión (11.29)

$$(11.29) \quad H(x) = \theta e^{-t} - \beta \left(-C - \int_0^t e^{-t} \ln t \, dt \right) = \theta e^{-t} + C\beta + \beta \int_0^t e^{-t} \ln t \, dt$$

Luego, con (11.25)

$$(11.25) \quad I(t) = \int_0^t e^{-t} \ln t \, dt = (1 - e^{-t}) \ln t + \sum_{n=1}^{\infty} \frac{(-1)^n t^n}{n!n}$$

obtenemos la expresión exacta

$$(11.29') \quad H(x) = \theta e^{-t} + C\beta + \beta(1 - e^{-t}) \ln t + \beta \sum_{n=1}^{\infty} \frac{(-1)^n t^n}{n!n}$$

Finalmente, la fórmula (11.31)

$$(11.31) \quad H(x) \cong \mu e^{-t} + \frac{\beta}{\sqrt{2\pi}} \left(\frac{t^{1/3}}{2} - \frac{5}{3} \right) \text{Exp} \left[-\frac{1}{2} \left(3t^{1/3} - \frac{8}{3} \right)^2 \right]; \quad \mu = \theta + C\beta$$

se obtiene reemplazando, en (11.29), la integral $I(t)$ por su aproximación (11.27).

6.2. DEMOSTRACIONES DE LAS EXPRESIONES DEL ÍNDICE DE GINI PARA LOS MODELOS DE PARETO Y LOGNORMAL

La fórmula general es

$$(11.60) \quad \eta = 2I - 1$$

donde

$$(11.61) \quad I = \int_0^1 R(u) du$$

$R(u)$ es la función cuya representación gráfica es la curva ABC, que vincula a $u = G(x)$ (Función de distribución derecha) con $R(x) = J(x)/\mu$.

6.2.1. Modelo de PARETO

En este caso es

$$u = G(x) = (\theta/x)^\psi$$

$$R = J(x)/\mu = (\theta/x)^{\psi-1} = u^{1-1/\psi}$$

$$I = \int_0^1 R(u) du = \int_0^1 u^{1-1/\psi} du = \frac{\psi}{2\psi-1}$$

$$\eta = 2I - 1 = \frac{1}{2\psi-1}$$

6.2.2. Modelo Lognormal

En este caso es

$$u = 1 - \Phi\left(\frac{\ln x - m}{D}\right)$$

$$R = \frac{J(x)}{\mu} = 1 - \Phi\left(\frac{\ln x - m}{D} - D\right) = 1 - \Phi(Z_{1-u} - D) = \Phi(Z_u + D)$$

$$I = \int_0^1 R(u) du = \int_0^1 \Phi(Z_u + D) du$$

Ponemos $y = Z_u$, entonces $u = \Phi(y)$, de donde

$$du = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) \Phi(y + D) dy$$

Para calcular esta integral, la derivamos con respecto a D (derivación bajo el signo integral):

$$\frac{dI}{dD} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y+D)^2}{2}\right] dy = \frac{1}{2\pi} \exp\left(-\frac{D^2}{4}\right) \int_{-\infty}^{\infty} \exp\left[-\left(y + \frac{D}{2}\right)^2\right] dy$$

La última integral vale $\sqrt{\pi}$, entonces

$$\frac{dI}{dD} = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{D^2}{4}\right)$$

La integral buscada I , se calcula integrando esta expresión con respecto a D , con límite inferior arbitrario, límite superior igual a D y agregando una constante de integración. Tenemos así

$$\begin{aligned} I &= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^D \exp\left(-\frac{D^2}{4}\right) dD + C = \frac{\sqrt{2\pi}}{2\sqrt{\pi}} \int_{-\infty}^D \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{D}{\sqrt{2}}\right)^2\right] dD + C = \\ &= \frac{1}{\sqrt{2}} \int_{-\infty}^D \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{D}{\sqrt{2}}\right)^2\right] dD + C \end{aligned}$$

En la última integral ponemos $t = D/\sqrt{2}$ o sea $dD = \sqrt{2}dt$ y queda

$$I = \frac{1}{\sqrt{2}} \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} \text{Exp}\left(-\frac{t^2}{2}\right) \sqrt{2} dt + C = \Phi(t) + C = \Phi\left(\frac{D}{\sqrt{2}}\right) + C$$

$$\eta = 2I - 1 = 2\Phi\left(\frac{D}{\sqrt{2}}\right) + C'$$

Para $D = 0$ debe ser $\eta = 0$, por lo tanto, $C' = -1$ (o sea $C = 0$). Finalmente

$$\eta = 2\Phi\left(\frac{D}{\sqrt{2}}\right) - 1$$

7. PROBLEMAS

11.1) En una planta siderúrgica hay un horno que trabaja con un quemador a fuel-oil cuya demanda semanal sigue una ley Normal de media 40 toneladas y desvío estándar 6 toneladas. A tal efecto, hay un tanque de 45 toneladas que es completado semanalmente por camiones cisterna cuyo flete total cuesta \$1.500; el combustible se paga \$2.300 por tonelada y si alguna semana la demanda supera la capacidad del tanque, se incurre en una pérdida de \$5.000. a) ¿Cuál es el costo esperado por combustible? b) Calcular el consumo medio de combustible. c) ¿Cuál es la demanda media para las semanas en que no se supera la capacidad del tanque? d) ¿Cuál es la demanda media para las semanas en que sí se supera la capacidad del tanque?

11.2) Una línea de ferrocarril estudia la forma de optimizar el servicio de transporte de pasajeros en viajes de tipo "rápido". Cada vagón tiene 60 asientos y está prohibido que viajen pasajeros de pie. El tren parte actualmente con 10 vagones estén o no ocupados y puede, por lo tanto, llevar a lo sumo 600 pasajeros. El costo de movimiento del tren se compone de: 1) 650 \$/vagón; 2) \$1.600 por la locomotora y furgón; 3) 10 \$/pasajero. Se evalúa además en \$25 el lucro cesante por cada pasajero que no consigue pasaje por haberse completado el tren. El precio de cada pasaje es de \$ 40 y se ha logrado establecer que la demanda se puede ajustar razonablemente bien a una distribución Normal de media 550 y desvío 75 pasajes. a) Calcular el beneficio esperado. b) Indicar cómo debe procederse para maximizar el beneficio esperado. c) Calcular el número medio de pasajes vendidos. d) Calcular el número medio de pasajes vendidos para los viajes en que la demanda no supere la capacidad del tren. e) Idem para los viajes en que la demanda sí supere la capacidad del tren.

11.3) La fabricación de un producto químico se lleva a cabo en un reactor con una capacidad óptima de 5000kg de materia prima o en otro más pequeño de 1600kg. La cantidad de materia prima, que se recibe de otra planta de la misma empresa, es variable día a día con distribución Normal de media 5200kg y desvío 400kg. Cuando se reciben menos de 5300kg, se procesan en el reactor grande (aun cuando por encima de 5000kg deben tomarse algunas precauciones especiales que agregan un costo fijo adicional). Por encima de 5300kg resulta imposible procesar el excedente en el reactor grande por lo cual se procesan 5000kg en él y el resto en el más pequeño. El costo de producción para el reactor grande está compuesto por un costo fijo de \$200 y un costo variable de 3\$ por kg de M.P., pero si se procesan más de 5000kg debe agregarse un costo fijo extra de \$360. Para el reactor pequeño los costos son \$160 y 3,2 \$ por kg, respectivamente. a) Calcular el costo esperado diario de producción. b) De los días en que se supera la capacidad óptima del reactor grande, ¿en qué porcentaje no se supera su capacidad máxima? c) ¿Cuál es la probabilidad de que en un día cualquiera el costo supere los \$16.000?

11.4) La valuación fiscal de los inmuebles de una ciudad es variable con media \$65.000 y desvío \$52.000. Se ha de implantar un impuesto de emergencia por única vez del 1% del valor de la propiedad con un mínimo de \$500 y un máximo de \$2.000. Hay en total 50.000 inmuebles. Calcular el monto que se espera recaudar por el impuesto. (Advierta que la distribución de la valuación fiscal es Lognormal).

11.5) En un bar de una Facultad pública la consumición se cobra por medio de un ticket que es presentado por el consumidor al salir del bar. Si pierde el ticket, debe pagar el importe de 35\$. Dados los diferentes estratos sociales que concurren a esa Facultad, los consumos presentan una fuerte dispersión, con una media de 14\$ y un desvío de 25\$. Aunque no es muy frecuente, hay personas cuyo consumo supera los 35\$ y, en ese caso, aplicando nuestros conocidos conceptos de viveza criolla, "pierden" el ticket y pagan sólo 35\$. Calcular el importe promedio pagado por los consumidores. (Advierta que los consumos tienen distribución Lognormal).

11.6) Demostrar que, para cualquier distribución, la desviación media absoluta está dada por $E(|x-\mu|)=2\mu F(\mu)-2H(\mu)$ y aplicar este resultado a las distribuciones Exponencial y de PARETO.

11.7) La demanda semanal de combustible en una planta industrial tiene distribución Gamma de media 3.500 litros y desvío 1.500 litros. Para satisfacer dicha demanda se cuenta actualmente con un tanque de 5.000 litros que es llenado semanalmente. Así, el costo semanal de combustible se compone de: 1) \$480 por el flete del camión cisterna; 2) 0,20\$/litro por el combustible necesario para llenar el tanque; 3) 1,2\$/litro faltante por pérdida ocasionada si la demanda supera la capacidad del tanque. a) Calcular el costo esperado. b) Calcular el consumo medio de combustible. c) Calcular la demanda media para las semanas en que ésta no supera la capacidad del tanque.

11.8) En una industria química, la demanda semanal de uno de los productos tiene distribución Gamma de media 400 y desvío 250kg. El costo total de fabricación es de 125\$/kg y el precio de venta de 210\$/kg. En caso de existir remanente, se vende á 100\$/kg, ya que el producto es altamente perecedero. a) Calcular la cantidad a preparar semanalmente. b) Calcular el beneficio esperado. c) ¿Cuál es la cantidad media vendida? d) ¿Cuál es la cantidad media vendida para las semanas que hay sobrante?

11.9) En una empresa los salarios tienen distribución de PARETO de valor mínimo $\theta = \$825$ y media $\mu = \$1.283,33$. Se decide otorgar un aumento de sueldo del 25% a los que ganan menos que la media y del 15% a los que ganan más. Calcular: a) El aumento promedio otorgado. b) El nuevo salario promedio después del aumento. c) El salario promedio, antes y después del aumento, para los que ganan menos que la media. d) El salario promedio, antes y después del aumento, para los que ganan más que la media.

11.10) La duración de un rodamiento tiene distribución de WEIBULL de media $4,385 \times 10^8$ y desvío estándar $1,0425 \times 10^8$ revoluciones. El rodamiento trabaja 8 horas por día a 1000 rpm. Hace 2 años que se instaló un nuevo rodamiento y se desea calcular cuánto tiempo más se espera que dure.

11.11) En un sistema de inyección de plástico, la bomba inyectora tiene una delicada válvula de duración variable con distribución de WEIBULL de media 78 semanas y desvío estándar 18,5 semanas. Si la válvula falla en servicio, el costo por parada de producción es elevado y se ha establecido en \$5.750. El reemplazo de la válvula tiene un costo total (materiales y mano de obra) de \$2.780. Es deseable entonces, efectuar el mantenimiento preventivo, reemplazando la válvula en momento oportuno, si es posible antes de que falle, pero dada la aleatoriedad de su duración, puede suceder que se dé el suceso indeseado, esto es, que falle y haya que reemplazarla incurriendo en el elevado costo mencionado. a) Calcular el tiempo T óptimo de reemplazo de la válvula y el costo total anual (52 semanas) asociado a esta política de mantenimiento. b) Si hace un año que se instaló una nueva válvula y no ha fallado, ¿cuál es su expectativa de vida actual?

11.12) En la Argentina, la vida de los caballos de carrera tiene distribución de GUMBEL del mínimo, de parámetro $\theta = 21,2$ años y media 18,3 años. En un importante haras se desempeña un valioso semental que acaba de cumplir 17 años y se desea calcular su expectativa de vida actual.