1.12. Análisis de Incertidumbre de Primer Orden Segundo Momento

En la Sección 1.11.4 se enfatizó que para cualquier variable aleatoria X su fdp $f_X(x)$ representa la manera más adecuada de expresar su carácter y comportamiento aleatorio. No obstante, en ocasiones su determinación puede ser difícil porque (1) los datos muestrales y los argumentos físicos son insuficientes para establecer la fdp, y (2) hay dificultad para medir la incertidumbre [Cornell (1972)]. Adicionalmente, puede haber dificultad en la determinación analítica de la fdp derivada $f_Z(z)$ a partir de la fdp conjunta de variables aleatorias explicativas $X_1, X_2, ..., X_k$ y de la relación funcional $Z = g(X_1, X_2, ..., X_k)$, lo cual se puede remediar de forma aproximada haciendo la derivación numéricamente mediante simulación de Montecarlo (ver Sección 1.11.4). Según Cornell (1972), en la medida en que los parámetros de diseño o variables de decisión sean insentitivos a momentos superiores a la media y la varianza, el análisis de incertidumbre de la variable aleatoria puede lograrse de manera aproximada con estos dos primeros momentos. En la Sección 1.11.4 se mostró la determinación analítica de media y varianza de variables derivadas. En esta sección se complementa lo anterior cuando esa determinación analítica no es posible, sino que es necesario recurrir a métodos aproximados. El método de Análisis de Primer Orden Segundo Momento (POSM) es el más popular de éstos para aproximar la varianza [Benjamin y Cornell (1970), Chow et al. (1988), Ang y Tang (2007)]. En este método la función relacional g se linealiza alrededor de un valor de referencia, y se obtienen el primer y el segundo momentos. El caso más simple es cuando $Z = g(X)$. Si el valor de referencia es $\overline{x}$, entonces $\overline{y} = g(\overline{x})$. Si el valor real de x es diferente de $\overline{x}$, el efecto de esta discrepancia en y se estima expandiendo $g(x)$ en una serie de Taylor alrededor de $x = \overline{x}$

$$y = g\left(\overline{x}\right) + \left(x - \overline{x}\right)\left[\frac{dg}{dx}\right]_{x=\overline{x}} + \frac{1}{2!}\left(x - \overline{x}\right)^2\left[\frac{d^2g}{dx^2}\right]_{x=\overline{x}} + ... \tag{1.293}$$

donde las derivadas se evalúan en $x = \overline{x}$. Si se desprecian los términos de orden 2 o superior, la expresión de Primer Orden resultante para el error en y es

$$y - \overline{y} \approx \left(x - \overline{x}\right)\left[\frac{dg}{dx}\right]_{x=\overline{x}} \tag{1.294}$$

La varianza de este error es

$$s_Y^2 = E\left[\left\{\left(x - \overline{x}\right)\left[\frac{dg}{dx}\right]_{x=\overline{x}}\right\}^2\right] = \left[\frac{dg}{dx}\right]_{x=\overline{x}}^2 s_X^2 \tag{1.295}$$

que queda en función de la varianza de la variable explicatoria. Consecuentemente, la desviación estándar de Y es

$$s_Y \approx \left[\frac{dg}{dx}\right]_{x=\overline{x}} s_X \tag{1.296}$$

En el caso general de k variables explicatorias, la expansión de Taylor es

$$y = g\left(\overline{x}\right) + \sum_{i=1}^{k}\left(x_i - \overline{x}_i\right)\left[\frac{\partial g\left(\mathbf{x}\right)}{\partial x_i}\right]_{\mathbf{x}=\overline{\mathbf{x}}} +$$

$$\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(x_i - \overline{x}_i\right)\left(x_j - \overline{x}_j\right)\left[\frac{\partial^2 g\left(\mathbf{x}\right)}{\partial x_i \partial x_j}\right]_{\mathbf{x}=\overline{\mathbf{x}}} + ... \tag{1.297}$$

y la aproximación de primer orden es

$$y = g\left(\overline{x}\right) + \sum_{i=1}^{k} \left(x_i - \overline{x}_i\right) \left[\frac{\partial g\left(\mathbf{x}\right)}{\partial x_i}\right]_{\mathbf{x}=\overline{\mathbf{x}}} \tag{1.298}$$

El valor esperado es

$$\mu_Y = E\left[Y\right] \approx m_Y \approx g\left(\overline{\mathbf{x}}\right) \tag{1.299}$$

La varianza se puede aproximar como

$$\sigma_Y^2 = \mathrm{Var}\left[Y\right] \approx \mathrm{Var}\left[g\left(\overline{\mathbf{x}}\right)\right] + \mathrm{Var}\left\{\sum_{i=1}^{k} \left[\frac{\partial g\left(\mathbf{x}\right)}{\partial x_i}\right]_{\mathbf{x}=\overline{\mathbf{x}}} \left(x_i - \overline{x}_i\right)\right\} \tag{1.300}$$

Dado que el primer término de esta ecuación es cero cuando $\mathbf{x} = \overline{\mathbf{x}}$, ésta se simplifica a

$$\sigma_Y^2 = \mathrm{Var}\left[Y\right] \approx \mathrm{Var}\left\{\sum_{i=1}^{k} \left[\frac{\partial g\left(\mathbf{x}\right)}{\partial x_i}\right]_{\mathbf{x}=\overline{\mathbf{x}}} \left(x_i - \overline{x}_i\right)\right\} \tag{1.301}$$

lo cual equivale a

$$\sigma_Y^2 = \mathrm{Var}\left[Y\right] \approx \sum_{i=1}^{k} a_i^2 \sigma_{X_i}^2 + 2 \sum_{i<}^{k} \sum_{j}^{k} a_i a_j \mathrm{Cov}\left[X_i, X_j\right] \tag{1.302}$$

donde

$$a_i = \left[\frac{\partial g\left(\mathbf{x}\right)}{\partial x_i}\right]_{\mathbf{x}=\overline{\mathbf{x}}} \tag{1.303}$$

el cual se denomina el coeficiente de sensibilidad y representa la tasa de cambio del valor de la función $g(x)$ en $\mathbf{x} = \overline{\mathbf{x}}$.

Cuando X_i y X_j no son correlacionadas, la Ecuación 1.302 se reduce a

$$\sigma_Y^2 = \mathrm{Var}\left[Y\right] \approx \sum_{i=1}^{k} a_i^2 \sigma_{X_i}^2 \tag{1.304}$$

Las Ecuaciones 1.302 y 1.304 se pueden denominar dicientemente "la propagación de la incertidumbre"[Ang y Tang (2007)].

La aproximación de primer orden para la media puede ser reemplazada por la de mayor precisión de segundo orden, sobre todo cuando la relación funcional es no lineal. En este caso, el valor esperado se estima como

$$\mu_Y = E\left[Y\right] \approx m_Y \approx g\left(\overline{\mathbf{x}}\right) + \sum_{i\leq}^{k} \sum_{j}^{k} \left[\frac{\partial^2 g\left(\mathbf{x}\right)}{\partial x_i \partial x_j}\right]_{\mathbf{x}=\overline{\mathbf{x}}} \mathrm{Cov}\left[X_i, X_j\right] \tag{1.305}$$

el cual, cuando X_i y X_j no son correlacionadas, se reduce a

$$\mu_Y = E\left[Y\right] \approx m_Y \approx g\left(\overline{\mathbf{x}}\right) + \sum_{i=1}^{k} b_i \sigma_{X_i}^2 \tag{1.306}$$

donde

$$b_i = \left[\frac{\partial^2 g\left(\mathbf{x}\right)}{\partial x_i^2}\right]_{\mathbf{x}=\overline{\mathbf{x}}} \tag{1.307}$$

En la práctica se usan la aproximación de primer orden (Ecuación 1.299) o de segundo orden (Ecuaciones 1.305 y 1.306) para estimar la media de Z, y la aproximación de primer orden para su varianza (Ecuaciones 1.302 y 1.304), ya que dependen solamente de las medias y varianzas de X_i [Benjamin y Cornell (1970), Ang y Tang (2007)].

Ejemplo 1.43 *La ecuación de balance hídrico de un embalse para el mes i es* $S_{i+1} = S_i + V_{P_i} + V_{Q_i} - V_{E_i} - V_{R_i}$*, donde* S_i *es el almacenamiento en el embalse al comienzo del mes i,* S_{i+1} *es el almacenamiento en el embalse al final del mes i y* V_{X_i} *es el volumen mensual de la variable* X*, siendo* P*,* Q*,* E *y* R *las variables de lluvia sobre el embalse, caudal afluente a éste, evaporación desde el espejo de agua del embalse y caudal regulado por éste, respectivamente. Supóngase que se conoce al comienzo del mes,* $S_i = 20$*, y el volumen de salida,* $V_{R_i} = 10$*. Todas las cantidades de este ejemplo están expresadas en millones de* m^3*. Se supone además que los volúmenes de lluvia, evaporación y escorrentía de entrada son variables aleatorias independientes entre sí. De éstas se tienen estimativos de sus medias y desviaciones estándar así:* $\overline{V}_{P_i} = 1$*,* $\overline{V}_{E_i} = 3$*,* $\overline{Q}_{P_i} = 8$*,* $s_{P_i} = 0{,}5$*,* $s_{E_i} = 1$ *y* $s_{Q_i} = 2$*. Con la información suministrada es posible estimar la media y la desviación estándar del almacenamiento en el embalse al final del mes. Las aproximaciones de primer y segundo orden para el valor esperado (ver Ecuaciones 1.299 y 1.306) conducen al mismo resultado, es decir*

$$E\left[S_{i+1}\right] = S_i + E\left[V_{P_i}\right] + E\left[V_{Q_i}\right] - E\left[V_{E_i}\right] - V_{R_i}$$

$$= 20 + 1 + 8 - 3 - 10 = 16$$

Para la varianza de S_{i+1}*, la aproximación de primer orden (ver Ecuación 1.304) es*

$$Var[S_{i+1}] = Var[V_{P_i}] + Var[V_{Q_i}] + Var[V_{E_i}]$$

$$= 0{,}5^2 + 2^2 + 1^2 = 5{,}25$$

que lleva a $\sigma_{V_{S_{i+1}}} = 2{,}29$.

Es posible pensar que la suposición de independencia entre la precipitación, la escorrentía y la evaporación no sea adecuada. Si después de analizar los datos disponibles se encontrase que existe correlación entre estas tres variables, así: $\rho_{(V_{P_i},V_{Q_i})} = 0{,}8$*,* $\rho_{(V_{P_i},V_{E_i})} = -0{,}4$ *y* $\rho_{(V_{Q_i},V_{E_i})} = -0{,}3$*, el estimativo del valor esperado de* S_{i+1} *queda igual, pero el de la varianza es (ver Ecuación 1.302)*

$$Var[S_{i+1}] = Var[V_{P_i}] + Var[V_{Q_i}] + Var[V_{E_i}]$$

$$+2Cov(V_{P_i}, V_{Q_i}) - 2Cov(V_{P_i}, V_{E_i}) - 2Cov(V_{Q_i}, V_{E_i})$$

que a su vez es igual a

$$Var[S_{i+1}] = Var[V_{P_i}] + Var[V_{Q_i}] + Var[V_{E_i}]$$

$$+2\rho_{(V_{P_i},V_{Q_i})}\sigma_{V_{P_i}}\sigma_{V_{Q_i}} - 2\rho_{(V_{P_i},V_{E_i})}\sigma_{V_{P_i}}\sigma_{V_{E_i}} - 2\rho_{(V_{Q_i},V_{E_i})}\sigma_{V_{Q_i}}\sigma_{V_{E_i}}$$

$$= 0{,}5^2 + 2^2 + 1^2 + 2 \times 0{,}8 \times 0{,}5 \times 2$$

$$-2 \times (-0{,}4) \times 0{,}5 \times 0{,}1 - 2 \times (-0{,}3) \times 2 \times 1 = 8{,}45$$

lo cual significa que $\sigma_{V_{S_{i+1}}} = 2{,}91$