ESTIMACION ESTADISTICA DE LA DURACION DE ACTIVIDADES EN UNA PROGRAMACION PERT

Por GUILLERMO CARRILLO VARGAS
y PEDRO GONZALEZ-HABA GONZALEZ
Ingenieros de Caminos, Canales y Puertos

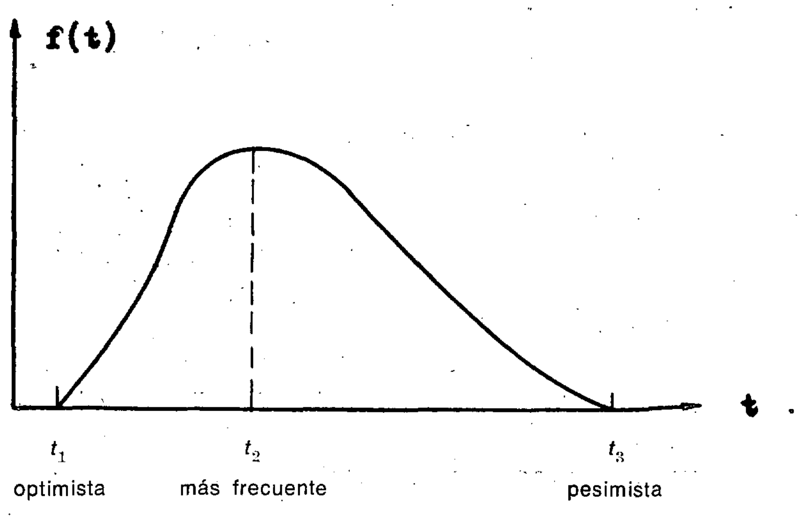
Este trabajo tiene por objeto estudiar estadísticamente los diversos procedimientos que se adoptan para la estimación de la duración de actividades en un programa PERT. Dado que dicha estimación es premisa previa de cualquier programación, sus conclusiones son aplicables en general.

La hipótesis básica aceptada en la estimación del tiempo que ocupa una actividad determinada en un programa PERT, es que dichos tiempos siguen una "distribución β''", con un recorrido que coincide con el intervalo (tiempo optimista; tiempo pesimista).

El procedimiento usual supone y exige hacer tres estimaciones del tiempo de duración de cada actividad:

— Estimación optimista — correspondiente al tiempo más corto con condiciones óptimas, prácticamente inalcanzables *(t_1)*.
— Estimación pesimista — o tiempo más largo con condiciones más desfavorables *(t_3)*.
— Estimación más probable — que corresponde al tiempo que se hubiera presentado con mayor frecuencia si la actividad se hubiera repetido numerosas veces *(t_2)*.

Partiendo de la hipótesis básica de aceptar una "distribución β", la ley de densidad de probabilidad será la siguiente:



$$f(t) = \frac{1}{B(p,q)\,(t_3 - t_1)^{p+q-1}} \cdot (t - t_1)^{p-1} \cdot (t - t_3)^{q-1} ;$$

donde;

$t =$ tiempo de duración de una actividad.

p y $q =$ parámetros a los que, de momento se impone la única condición de $p > 0$, $q > 0$.

$B(p,q) =$ el valor de la integral euleriana.

$$\int_0^1 y^{p-1}\,(1 - y)^{q-1} \cdot dy.$$

El cambio de variable: $x = \dfrac{t - t_1}{t_3 - t_1}$ transforma la función $f(t)$ en la $\varphi(x)$ que se indica a continuación:

$$f(t) \cdot dt = \varphi(x) \cdot dx ;$$

$$\varphi(x) = B(p,q) \cdot x^{p-1} \cdot (1 - x)^{q-1} ;$$

con $0 \leq x \leq 1$ y $p > 0$, $q > 0$.

Standarizada de esta forma la variable, los valores correspondientes a la media, varianza y moda de la distribución son los que siguen:

$$\text{Media} \;\ldots\ldots\ldots\; (x) : \quad \bar{x} = \frac{p}{p+q} ;$$

$$\text{Varianza} \;\ldots\ldots\; (x) : \quad \sigma^2{}_x = \frac{p \cdot q}{(p+q)^2\,(p+q+1)} ;$$

$$\text{Moda} \;\ldots\ldots\ldots\; (x) : \quad \bar{\bar{x}} = \frac{p-1}{p+q-2} .$$

El método usualmente empleado admite dos hipótesis:

— La aproximación de la desviación típica de la distribución por la sexta parte del recorrido, $t_3 - t_1$.

— La existencia válida de una relación lineal entre la media y el valor más probable:

$$\bar{x} = \frac{4\bar{\bar{x}} + 1}{6} .$$

Admitiendo estos dos supuestos, y a través de las relaciones:

$$\bar{t} = t_1 + (t_3 - t_1)\,\bar{x} ;$$

$$\sigma^2{}_t = (t_3 - t_1)^2 \cdot \sigma^2{}_x ;$$

$$\bar{\bar{t}} = t_1 + (t_3 - t_1) \cdot \bar{\bar{x}} ;$$

se obtiene:

$$\bar{t} = \frac{t_1 + 4\,t_2 + t_3}{6} ; \quad \sigma_t = \frac{t_3 - t_1}{6} ;$$

fórmulas utilizadas comúnmente.

Malcolm, Roseboom, Clark y Fazar sostuvieron que la adopción de los dos supuestos mencionados anteriormente no implicaban ninguna consecuencia sobre la situación del valor modal, t_2, en relación con los extremos del intervalo $[t_1 ; t_3]$.

Sin embargo, es evidente que las hipótesis:

$$\sigma_x = \frac{1}{6} ; \quad \bar{x} = \frac{4\bar{\bar{x}} + 1}{6} ;$$

conducen a un sistema de ecuaciones inmediato:

$$\frac{p \cdot q}{(p+q)^2 \cdot (p+q+1)} = \frac{1}{36} ; \qquad \frac{p}{p+q} = \frac{4 \cdot \dfrac{p-1}{p+q-2} + 1}{6} ;$$

cuyas soluciones son:

$$p = 3 \pm \sqrt{2} ; \quad q = 3 \mp \sqrt{2} ; \quad p = q = 4 ;$$

lo cual muestra claramente que las referidas hipótesis son estadísticamente restrictivas.

Si se tiene, además, en cuenta que el coeficiente de simetría toma los valores:

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \mp \cdot \frac{1}{\sqrt{2}} ; \quad \text{o bien} \quad \beta_1 = 0 ;$$

se deduce que dicho coeficiente aparece fijado por la simple aceptación de lo establecido anteriormente, lo que no está de acuerdo con la afirmación de que no se hace hipótesis alguna sobre la posición relativa de la moda, t_2, respecto a t y t_3.

Donaldson intenta superar estas dificultades de la forma que se expone a continuación:

Supone tres estimaciones previas:

— estimación optimista
— estimación pesimista
— estimación media.

Es decir, sustituye la estimación previa del valor más probable por la del valor medio.

Supongamos que dichas estimaciones sean t_1, t_2, t_3 correspondientes, respectivamente a la optimista, media y pesimista y los valores respectivos, a, m, b.

A partir de las fórmulas anteriores, se deduce:

$$m = \frac{p \cdot b + q \cdot a}{p+q} ; \text{ y sustituyendo } a,\ m,\ b \text{ por } t_1,\ t_2,\ t_3,\ \frac{q}{p} = \frac{t_3 - t_2}{t_2 - t_1} ; \qquad [1]$$

La varianza viene dada por:

$$\sigma^2 = \frac{(b-a)^2 \cdot p \cdot q}{(p+q)^2 \cdot (p+q+1)} .$$

Haciendo las sustituciones por t_1 y t_3, y teniendo en cuenta la relación [1] obtenemos una estimación de la varianza:

$$[\sigma^2{}_t] = \frac{(t_3 - t_2)(t_2 - t_1)}{p+q+1} .$$

Donaldson hace, asimismo, la hipótesis razonable de que la curva de distribución ha de ser tangente al eje horizontal en los puntos extremos del intervalo, lo cual implica que $p > 2$, $q > 2$.

El mínimo valor de $(p + q)$ que satisface las condiciones $p > 2$, $q > 2$; y al mismo tiempo la condición [1] es:

$$2 + 2Z + \varepsilon ;$$

siendo ε cualquier cantidad infinitésima positiva y

$$Z = \text{máx.} \left[\frac{t_3 - t_2}{t_2 - t_1} \cdot \quad \frac{t_2 - t_1}{t_3 - t_2} \right] .$$

En consecuencia el estimador:

$$[\sigma^2{}_t] = \frac{(t_3 - t_2)(t_2 - t_1)}{3 + 2Z} = \frac{(t_3 - t_2)(t_2 - t_1)}{1 + \dfrac{2(t_2 - t_1)}{\min(t_3 - t_2, t_2 - t_1)}} ,$$

resulta ser un estimador por exceso de la varianza, con la condición de mínima tangencia.

Al objeto de compararlo con el estimador usual

$$\left[\frac{1}{6}(t_3 - t_1) \right]^2 = 0{,}028\,(t_3 - t_1)^2 ;$$

tomamos:

$$L = \frac{t_2 - t_1}{t_3 - t_1} .$$

Si

$$0 < t_2 - t_1 \leq t_3 - t_2 ;$$

$$\frac{[\sigma^2{}_t]}{(t_3 - t_1)^2} = \frac{L^2(1 - L)}{L + 2} \; (0 < L \leq {}^1/_2) ,$$

con lo cual se forma el siguiente cuadro:

L	0,1	0,2	0,3	0,4	0,5
$\frac{[\sigma^2{}_t]}{(t_3 - t_1)^2}$	0,004	0,015	0,027	0,040	0,050

A continuación se estima el coeficiente de simetría:

$$[\beta_1] = \frac{(1 - 2L)(L + 2)^{1/2}}{(1 + L)(1 - L)^{1/2}} ; \quad (0 < t_2 - t_1 \leq t_3 - t_2) ,$$

el cual toma los valores:

L	0,1	0,2	0,3	0,4	0,5
$[\beta_1]$	1,111	0,829	0,558	0,286	0,000

Por consiguiente, el procedimiento de Donaldson, expuesto, no incurre en contradicciones estadísticas como lo hace el método usual, permite variaciones apreciables del coeficiente de simetría, e incluye el supuesto de la doble tangencia horizontal en los extremos, que parece razonable como ya se ha dicho anteriormente.

La variación fundamental con el usualmente empleado se basa en la obtención de una estimacióh directa de la media en lugar de la moda. Ello no parece ofrecer dificultades, ya que a menudo la moda se estima, por criterios puramente intuitivos, e incluso, frecuentemente la moda más que valor más frecuentes es *media* de valores más frecuentes.

Helen Coon considera, sin embargo, la posibilidad de seguir utilizando la estimación del valor modal como dato previo, dentro de la mayor rigurosidad estadística en que se mueve Donaldson.

En efecto, sean t_1 y t_3 los tiempos estimados como optimista y pesimista, respectivamente, y t'_2 la estimación del valor más frecuente:

$$\text{Moda } (t); \quad \overline{\overline{t}} = \frac{t_1 (q-1) + t_3 (p-1)}{p+q-2};$$

de donde se deduce:

$$\frac{q-1}{p-1} = \frac{t_3 - t'_2}{t'_2 - t_1}: \qquad [2]$$

El mínimo valor de *(p + q)* que satisface simultáneamente la relación (2) y la condición de tangencia, $p > 2$, $q > 2$, es:

$$3 + Z + \varepsilon;$$

siendo ε un infinitésimo positivo y

$$Z = \text{Máx}\left[\frac{t_3 - t'_2}{t'_2 - t_1}; \quad \frac{t'_2 - t_1}{t_3 - t'_2}\right]. \qquad [3]$$

Si consideramos $\varepsilon = 0$ en el estimador $[\sigma_t^2]$, de la varianza, los valores de p y q utilizados a partir de la estimación de la media son:

$$\text{Si} \quad t_3 - t_2 > t_2 - t_1; \quad p = 2 \quad q = 2\,\frac{t_3 - t_2}{t_2 - t_1}.$$

$$\text{Si} \quad t_2 - t_1 > t_3 - t_2; \quad q = 2 \quad p = 2\,\frac{t_2 - t_1}{t_3 - t_2}.$$

Partiendo de $\varepsilon = 0$ y de la estimación de la moda deducimos que:

$$\text{Si} \qquad t_3 - t'_2 > t'_2 - t_1 \, ; \quad p = 2 \qquad q = \frac{t - t'_2}{t'_2 - t_1} + 1 \, .$$

$$\text{Si} \qquad t'_2 - t_1 > t_3 - t'_2 \, ; \quad q = 2 \qquad p = \frac{t'_2 - t_1}{t_3 - t'_2} + 1 \, .$$

Con estos valores de p y q puede escribirse un estimador de la varianza $[\sigma_t{}^2]$ tal como el siguiente:

$$[\sigma^2{}_t] = \frac{2\,(t_3 - t_1)^2\,(Z+1)}{(Z+3)^2\,(Z+4)} \, ;$$

con la condición (3) y la estimación correspondiente a la media sería:

$$\overline{[t]} = \frac{2\,t \; + t \;\, (Z+1)}{Z+3} \qquad \text{si} \quad t_3 - t'_2 > t'_2 - t_1 \, ;$$

$$\overline{[t]} = \frac{(Z+1)\,t_3 + 2\,t_1}{Z+3} \qquad \text{si} \quad t'_2 - t_1 > t_3 - t'_2 \, .$$

De todo lo anterior se deduce, en consecuencia, que partiendo de la hipótesis $\varepsilon = 0$, puede usarse con la misma rigurosidad estadística, el valor modal o la media en los cálculos, si bien las fórmulas basadas en la moda no son tan simples como aquéllas con base en la media, aunque tampoco presenten una gran complicación.

Es interesante subrayar que estos últimos métodos continúan siendo estadísticamente restrictivos, por cuanto que en uno se establece la hipótesis de $p = 2$ ó $q = 2$, según la posición de la media a uno u otro lado del punto medio del intervalo, deduciendo el otro parámetro de la relación:

$$\frac{q}{p} = \frac{t_3 - t_2}{t_2 - t_1} \, ;$$

y análogamente, algo parecido podría decirse del otro procedimiento.

En el mismo sentido puede interpretarse que el tomar valores estrictos $p = 2$ ó $q = 2$, trae como consecuencia el que no sea horizontal la tangente en el extremo considerado, ya que la condición es $p > 2$, $q > 2$. No obstante, puede suponerse un aumento infinitesimal que automáticamente daría la tangencia horizontal, o bien prescindir de esa condición en un extremo. Pero lo que resulta innegable es que ambos establecen una libertad mucho más amplia respecto a las rígidas imposiciones del método usual, permitiendo diversos valores del coeficiente de simetría, lo que, sin duda, hace que sean más representativos de la realidad.

A continuación se incluye un ejemplo tratado por los distintos métodos que se han expuesto, y que puede servir de ilustración complementaria.

Ejemplo:

Supongamos una actividad a la que se le asignan unas estimaciones de duración optimista, más frecuente, media y pesimista de 60, 70, 75 y 100, respectivamente.

Según todo lo anterior, se tendría:

I. *Método usual.*

Media, $\bar{t}$; 73,3.

Desviación, $\sigma(t)$; 6,7.

$p = 1{,}8$; $q = 4{,}2$ aproximativos de $p = 3 - \sqrt{2}$, $q = 3 + \sqrt{2}$.

Coeficiente de simetría, β_1; 0,707.

II. *Método de Donaldson.*

Media, $[\bar{t}]$; 75,0.

Desviación, $\sigma(t)$; 6,4.

$p = 2$; $q = 3{,}4$.

Coeficiente de simetría, β_1; 0,558.

III. *Método de Coon.*

Media, $\bar{t}$; 73,3.

Desviación, $\sigma(t)$; 7,1.

$p = 2$; $q = 4$.

Coeficiente de simetría, 0,590

Conclusión.

Dado que normalmente, en cualquier programación, las estimaciones de duración de actividades se refieren a tiempos medios, podría parecer, de todo lo expuesto anteriormente, y especialmente con referencia al ejemplo incluido, que es indiferente adoptar uno u otro procedimiento debido a los estrechos límites en que los mismos se mueven, y abundando en el tema incluso podría suponerse un contrasentido el hecho de que, en el método de Donaldson, el valor a adoptar sería una estimación previa sin que intervengan en absoluto los tiempos optimistas y pesimistas.

Sin embargo, en una contratación rigurosa — como es lo usual en los tiempos actuales en todo el mundo —, donde las penalidades por incumplimiento de plazos son fuertes, adquiere verdadera importancia, desde el punto de vista empresarial, la rigurosidad estadística en la estimación de los tiempos de duración de actividades, de donde se deducirá con mayor garantía la probabilidad del cumplimiento de las mismas.

Igualmente podría decirse, desde el punto de vista administrativo, cuando es urgente dar solución a un problema.

Entonces será cuando deba elegirse cuidadosamente qué tiempo medio y desviación presenta superior valor estadístico, de tal forma que al determinar el tiempo ofertado

$$t_{of} = \bar{t} + \lambda \cdot \sigma(t) ;$$

(donde λ es una función del riesgo de incumplimiento) coincida con el menor número de restricciones posibles y, por consiguiente, sea más representativo de la realidad.

Con referencia al ejemplo anterior, según las distribuciones correspondientes, podría establecerse el siguiente cuadro:

METODO	$\frac{t_{of}}{\bar{t} + \sigma}$	Probabilidad de cumplimiento.	$\frac{t_{of}}{\bar{t} + 1{,}5\,\sigma}$	Probabilidad de cumplimiento.	$\frac{t_{of}}{\bar{t} + 2\,\sigma}$	Probabilidad de cumplimiento.
Usual	80	70 %	83,4	80 %	86,7	85 %
Donaldson	81,4	78 %	84,6	90 %	87,8	98 %
Coon	80,4	73 %	83,9	87 %	87,5	96,5 %

Del cual fácilmente puede deducirse que el estudio estadístico de la duración de actividades, conduce a reducir el tiempo ofertado con mayor probabilidad de cumplimiento y superior rigor estadístico.

BIBLIOGRAFIA

D. G. Malcolm, J. H. Roseboom, C. E. Clark and W. Fazar: *Application of a Technique for Research and Development Program Evaluation.*

C. E. Clark: *The Pert Model for the distribution of an Activity Time.*

F. E. Grubbs *Attempts to Validate Certain Pert Statistics.*

H. Cramér: *Mathematical Methods of Statistics.*