Illustro, issn 2311-4126, 2015, vol.6, 21-35

DE LA ESTIMACIÓN DEL VALOR ESPERADO DE LA DURACIÓN DE PAQUETES DE TRABAJO CONFORMADOS POR ACTIVIDADES INDEPENDIENTES DE REALIZACIÓN SIMULTÁNEA EN PROYECTOS Y SU RIESGO

ABOUT THE ESTIMATION OF THE EXPECTED VALUE OF DURATION ON WORK PACKAGES CONFORMED BY INDEPENDENT ACTIVITIES SIMULTANIOUSLY PERFORMED ON PROJECTS AND THEIR RISKS

Ronal W. Arela Bobadilla
Universidad Católica San Pablo, Arequipa, Perú

Resumen:

El presente ensayo tiene por objetivo describir el mecanismo que opera en la estimación de la duración de un paquete de trabajo en un proyecto, compuesto por actividades independientes realizadas en forma paralela o simultánea. Se realizan operaciones estadísticas para demostrar que las estimaciones basadas en las duraciones promedio de cada actividad, sin tomar en cuenta su variación, son susceptibles a estar erradas y subestiman la duración final. Asimismo, se demuestra que el riesgo de que este tipo de paquetes de trabajo se retrase es superior al riesgo de cada tarea individual. Se realiza una simulación y se comprueban las ideas expuestas.

Palabras Clave: Gestión del cronograma, administración de proyectos, riesgo.

Summary:

The aim of this essay is to describe the mechanism, which operates in the estimation of the work packages length in a project, composed of independent activities performed simultaneously. Stadistics operations are performed to demonstrate that estimations are based on the average lenght of each activity, without taking in account their variations; they are susceptible to be wrong and underestimate their final duration. Likewise, it demonstrates the risk of delay in this type of work packages, is higher than the risk of each individual work. A simulation is done and it proves the exposed ideas.

Key words: Chronogram management, project management, risk.

Introducción

Una gran parte del que hacer en las ciencias sociales se basa en las estimaciones de cierto tipo de eventos, y nuestras decisiones dependen de la información recogida en este proceso. Incluso en nuestra vida realizamos estimaciones sobre el tiempo que demoraremos realizando ciertas tareas, estimaciones de las probabilidades de encontrar un taxi a cierta hora del día en cierto lugar, estimamos el comportamiento de cierta persona en ciertas situaciones. Nuestra vida está llena de estimaciones debido a que no somos observadores directos de los eventos, ya sea por su complejidad o porque son situaciones que aún no han sucedido.

La administración de proyectos no es ajena a este proceso de estimación. La estimación de los tiempos de duración del proyecto se realiza en base a la duración del conjunto de actividades que componen los paquetes de trabajo para entregar los "deliverables". La duración de las tareas no son exactas, sino que su duración real puede ser diferente a la duración estimada, generando demoras o tiempos muertos. Entonces, debido a la naturaleza de la duración de las tareas, estas deben ser tratadas como variables aleatorias, con una esperanza matemática, una varianza y una distribución. Entonces, el valor esperado de la duración de la tarea es susceptible de ser estimado, así como las probabilidades de que cada tarea se complete en un tiempo determinado. Sin embargo, el procedimiento aumenta su dificultad cuando se trata de tareas relacionadas. Por ejemplo, una tarea que presenta tareas precedentes o que se llevarán a cabo a continuación, o, como se verá en este ensayo, tareas que son parte de un paquete de trabajo cuyo entregable solo puede ser completado cuando todas las actividades culminen satisfactoriamente.

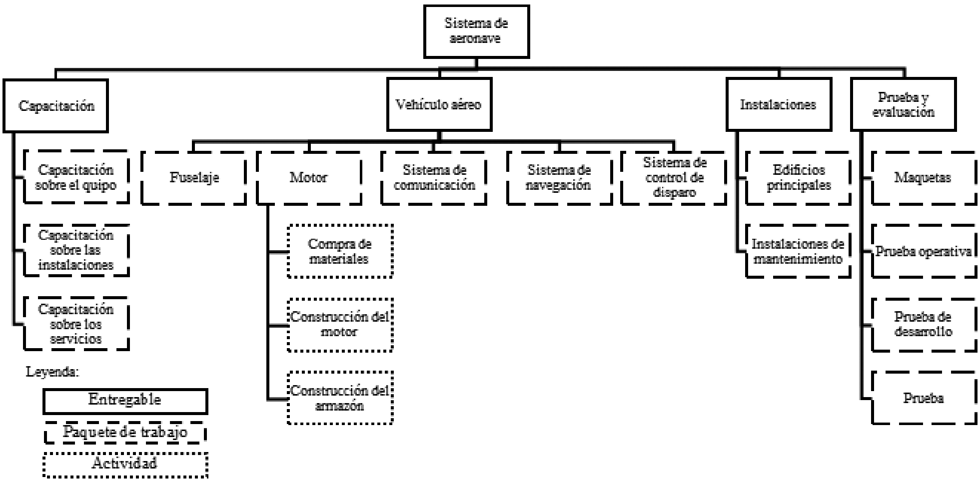
El presente trabajo busca realizar una estimación estadística de la duración necesaria para completar un paquete de trabajo compuesto por actividades que pueden llevarse a cabo paralelamente. En la primer parte, se plantean los problemas y las consecuencias de realizar estimaciones sin tomar en cuenta las propiedades estadísticas de las variables aleatorias que son las duraciones de las actividades o tareas. En la segunda, se plantea la estimación del valor esperado de la duración total del paquete de trabajo. En la tercera, se plantea una simulación Montecarlo para verificar los resultados. Luego, se muestran las conclusiones más importantes y algunos temas de discusión.

Desarrollo

La dirección de proyectos aplica los conocimientos, habilidades, herramientas y técnicas a las actividades del proyecto para cumplir con los requisitos que de este se demandan (Project Management Institute, 2013).

Un campo del conocimiento importante en la gestión de los proyectos es la gestión del alcance del mismo, y dentro de este el proceso de la creación del *Work Breakdown Structure* (WBS), que es la subdivisión de los entregables y el trabajo del proyecto en componentes más pequeños y fáciles de manejar (Project Management Institute, 2013), es una pieza crítica. Las partes finales del WBS son los denominados Paquetes de Trabajo (*Work Packages*). Los paquetes de trabajo pueden dividirse a su vez en las actividades necesarias para completarlo. La Figura 1 resume el WBS, los *Work Packages* y las actividades de la construcción de un sistema de aeronave.

Figura 1.
WBS para la construcción de un sistema de aeronave



Fuente: Project Management Body of Knowledge (PMBOK).

Entonces, la duración que tome llevar a cabo los entregables del proyecto dependerá de la duración de las actividades de cada paquete de trabajo. Así, su retraso o la culminación de las actividades antes del tiempo previsto tendrán un impacto mayúsculo en la duración del proyecto entero; especialmente si las actividades pertenecen a la ruta crítica del proyecto.

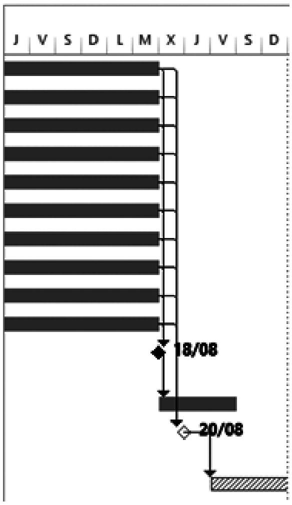
Las duraciones de las actividades de los paquetes de trabajo no son fenómenos determinísticos –fenómenos cuyo valor se repite en la realización de varios experimentos– sino fenómenos aleatorios o estocásticos que tienen una media –esperanza matemática o valor esperado– y una desviación estándar. Estas variaciones tienen un impacto en la duración del paquete de trabajo, cuya magnitud no es necesariamente intuitiva o fácil de estimar.

Savage (2009) coloca el siguiente ejemplo para ilustrar el efecto de las variaciones de la duración de las actividades en la duración total estimada del paquete de trabajo. Dado un paquete de trabajo de un proyecto de software que requiere de 10 actividades para ser culminado y que pueden realizarse en paralelo, el jefe de proyecto le preguntará a cada administrador de las actividades cuál es el tiempo estimado para su culminación. De la discusión entre el administrador del proyecto y los administradores de las actividades, se concluye que cada actividad dura entre 3 y 9 días –en el ejemplo real fueron meses; sin embargo, esto no altera en absoluto nuestras conclusiones–, siendo su valor promedio de 6 días. Entonces, el administrador del proyecto estima que la duración total del paquete de trabajo será de 6 días en promedio. Sin embargo, esta estimación es incorrecta.

Savage (2009) demuestra mediante un experimento Montecarlo que el valor esperado de la duración de este paquete de trabajo no es 6 días, sino, dada alguna desviación estándar para las actividades, de 8.6 días, mayor al valor máximo de los valores esperados de las duraciones de las actividades. Esta situación se representa en el diagrama Gantt de la Figura 2; las últimas dos filas corresponden a la estimación tomando como final estimado del paquete de trabajo al valor esperado de la duración máxima de las actividades que realizó Savage.

Figura 2.
Diagrama Gantt para 10 actividades/tareas con estimaciones de la duración total del paquete de trabajo

Nombre de tarea	Duración
Tarea 1	6 días
Tarea 2	6 días
Tarea 3	6 días
Tarea 4	6 días
Tarea 5	6 días
Tarea 6	6 días
Tarea 7	6 días
Tarea 8	6 días
Tarea 9	6 días
Tarea 10	6 días
Fin del paquete de trabajo (valor máximo de los valores esperados de las duraciones de las tareas	0 días
Tarea 11	3 días
Fin del paquete de trabajo (valor esperado del máximo de las duraciones de las tareas)	0 días
Tarea 11 inicio esperado	3 días



Fuente: Elaboración propia.

Matemáticamente, la estimación que indica que la duración total del paquete de trabajo es de 6 días, está definida como el valor máximo de los valores esperados o promedios de duración de cada actividad:

$$Duración_{p1} = \max\{E(Duración_1); E(Duración_2); \ldots ; E(Duración_k)\} \quad (1)$$

Donde k es el número de actividades que deberán ser completadas, en este caso k=10; $Duración_{p1}$ es la duración total del paquete de trabajo y $Duración_i$ es la duración de la actividad . La función implica que la duración total estará dada por aquella actividad que demore más, debido a que el paquete de trabajo culmina cuando todas las actividades sin excepción culminaron. Llamaremos a esta medida la duración máxima de los valores esperados de las actividades (*MaxVE*).

Recordemos ahora que la definición del valor esperado (VE). El VE es el promedio de los valores que obtendríamos si pudiéramos ejecutar el paquete de trabajo y sus actividades un gran número de veces (Rolstadås, 2008). Entonces, las duraciones alrededor del VE son más probables que duraciones más alejadas.

Ahora, el valor esperado de la duración del paquete de trabajo (*VEMax*) está definido como:

$$Duración_{p2} = E(\max\{Duración_1; Duración_2; \ldots; Duración_k\}) \qquad (2)$$

Debe quedar claro que la (1) no es de ninguna forma igual a (2). Por lo tanto, los valores estimados de la duración del paquete de trabajo son diferentes entre ambas medidas.

Savage (2009) estima el *VEMax* basándose en una simulación Montecarlo, sin embargo, es necesario conocer formalmente el proceso que describe la variable aleatoria $Duración_{p2}$ condicional a las funciones de distribución de las duraciones de sus actividades.

Entonces, la definición matemática del valor esperado de una variable aleatoria-continua '*y*' viene dada por (Wooldridge, 2012):

$$E(y) = \int_{-\infty}^{\infty} y f(y)\, dy$$

Y si '*y*' fuera una variable aleatoria discreta

$$E(y) = \sum_{-\infty}^{\infty} y f(y)$$

Donde (*y)* es la función de densidad de la variable aleatoria –la letra *f* minúscula hace referencia a la función de densidad, mientras una *F* mayúscula a la función de distribución.

En este ejemplo definamos $Duración_p = y_{max}$. El valor esperado de la variable $Duración_p$ viene dado por:

$$E(y_{max}) = \int_{-\infty}^{\infty} y_{max} f(y_{max})\, dy_{max}$$

Por simplicidad, asumamos que las duraciones de las actividades y el paquete de trabajo pueden tomar valores negativos, por ello la inclusión del término –∞ en la integración. Esto no afectará nuestra comprensión de los cálculos aquí realizados. Luego veremos cómo realizar los mismos cálculos cuando la variable aleatoria solo pueda tomar valores positivos, como sucede en la vida real.

En la ecuación anterior, no conocemos la función de densidad de ; sin embargo, sabemos que:

$$y_{max} = \max\{Duración_1; Duración_2; \ldots; Duración_k\}$$

Luego, la función de distribución de y_{max} evaluada en algún punto 'y' es igual a las probabilidades de que y_{max} sea menor o igual a 'y' (Tesler, 2013):

$$F_{y_{max}}(y) = Prob[y_{max} \leq y]$$
$$F_{y_{max}}(y) = Prob[\max\{Duración_1; Duración_2; \ldots; Duración_k\} \leq y]$$
$$F_{y_{max}}(y) = Prob[Duración_1 \leq y; Duración_2 \leq y; \ldots; Duración_k \leq y]$$

Que es la función de distribución conjunta.

Debido a que las actividades son independientes entre sí; es decir, sus funciones de distribución son independientes, podemos separar la función de distribución conjunta en los productos de las funciones de distribución individuales:

$$F_{y_{max}}(y) = Prob[Duración_1 \leq y] * Prob[Duración_2 \leq y] * \ldots * Prob[Duración_k \leq y]$$
$$F_{y_{max}}(y) = F_1(y) * F_2(y) * \ldots * F_k(y) \qquad (3)$$

Si todas las funciones de distribución fueran idénticas; es decir, con la misma media y la misma deviación estándar –cabe indicar que esta condición junto a la independencia de las distribuciones se resume en la propiedad de variables aleatorias independientes e idénticamente distribuidas o i.i.d. – podríamos reducir la expresión anterior a:

$$F_{y_{max}}(y) = F(y)^k$$

Cuya función de densidad es:

$$f_{y_{max}}(y) = \frac{dF_{y_{max}}(y)}{dy} = k * F(y)^{k-1} f(y)$$

Sin embargo, en la dirección de proyectos, pocas veces tendremos actividades con igual media y desviación estándar. Así, es más sensato quedarse con la expresión (3). Entonces, la función de densidad de y_{max} está definida por

$$f_{y_{max}}(y) = \frac{dF_{y_{max}}(y)}{dy} = \frac{d[F_1(y) * F_2(y) * \ldots * F_k(y)]}{dy}$$

Siguiendo la regla de derivación de un producto de forma iterativa y colocando F_i en lugar de $F_i(y)$ y f_i en lugar de $f_i(y)$ obtenemos:

$$f_{y_{max}}(y) = \frac{dF_{y_{max}}(y)}{dy} = f_1 * F_2 * \ldots * F_k + f_2 * F_1 * F_3 * \ldots * F_k + \cdots + f_i * F_1 * \ldots * F_{i-1} * F_{i+1} * \ldots * F_k + f_k * F_1 * \ldots * F_{k-1}$$
$$f_{y_{max}}(y) = \frac{dF_{y_{max}}(y)}{dy} = \sum_{i=1}^{k}\left(f_i * \prod_{j=1;j\neq i}^{k} F_j\right)$$

Nota: ∑ significa sumatoria y ∏ significa productorio.

Entonces, el valor esperado de la duración del paquete de trabajo está definido como:

$$E(y_{max}) = \int_{-\infty}^{\infty} \left(y_{max} * \sum_{i=1}^{k} \left(f_i * \prod_{j=1; j \neq i}^{k} F_j \right) \right) dy$$

Comprobación

Para simplificar el cálculo y sin perder generalización del procedimiento, realizaremos la estimación de la duración total de un paquete de trabajo compuesto por 3 actividades y no 10 como en el ejemplo de Savage. Cada actividad tiene un valor esperado, una desviación estándar y una función de distribución conocidas.

Para 3 actividades ($k = 3$), nuestra función de densidad está definida como:

$$f_{y_{max}}(y) = \frac{dF_{y_{max}}(y)}{dy} = \sum_{i=1}^{k} \left(f_i * \prod_{j=1; j \neq i}^{k} F_j \right)$$

$$f_{y_{max}}(y) = \frac{dF_{y_{max}}(y)}{dy} = f_1 F_2 F_3 + f_2 F_1 F_3 + f_3 F_1 F_2$$

Y el valor esperado de la duración del paquete de trabajo:

$$E(y_{max}) = \int_{-\infty}^{\infty} (y_{max} * (f_1 F_2 F_3 + f_2 F_1 F_3 + f_3 F_1 F_2)) dy_{max}$$

$$E(y_{max}) = \int_{-\infty}^{\infty} (y_{max} * (f_1 F_2 F_3)) dy_{max} + \int_{-\infty}^{\infty} (y_{max} * (f_2 F_1 F_3)) dy_{max} + \int_{-\infty}^{\infty} (y_{max} * (f_3 F_1 F_2)) dy_{max}$$

Dado que:

$$F_i = \int f_i$$

$$E(y_{max}) = \int_{-\infty}^{\infty} \left(y_{max} * \left(f_1 \int_{-\infty}^{y_{max}} f_2 \int_{-\infty}^{y_{max}} f_3 \right) \right) dy_{max} + \int_{-\infty}^{\infty} \left(y_{max} * \left(f_2 \int_{-\infty}^{y_{max}} f_1 \int_{-\infty}^{y_{max}} f_3 \right) \right) dy_{max} + \int_{-\infty}^{\infty} \left(y_{max} * \left(f_3 \int_{-\infty}^{y_{max}} f_1 \int_{-\infty}^{y_{max}} f_2 \right) \right) dy_{max}$$

Además, supongamos que las actividades están normalmente distribuidas de la siguiente forma:

$$Duración_{Actividad\ 1} \sim N(\mu_1; \sigma_1)$$
$$Duración_{Actividad\ 2} \sim N(\mu_2; \sigma_2)$$
$$Duración_{Actividad\ 3} \sim N(\mu_3; \sigma_3)$$

Se lee: la duración de la actividad *i* está normalmente distribuida con media μ_i y desviación estándar de σ_i.

Para resolver (4), debemos reemplazar con las funciones de densidad de cada una de las actividades. Para ello, la función de densidad normal es:

$$f_i(y) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{\left(\frac{y-\mu_i}{\sigma_i}\right)^2}{2}}$$

Y la función de distribución es la integral de la función de densidad:

$$F_i(y) = \int \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{\left(\frac{y-\mu_i}{\sigma_i}\right)^2}{2}} dy$$

Para resolver el ejemplo, supongamos los siguientes valores en días para las medias y las desviaciones estándares de cada actividad.

$$Subrutina\ 1 = x_1 \sim N(50; 10)$$
$$Subrutina\ 2 = x_2 \sim N(55; 5)$$
$$Subrutina\ 3 = x_3 \sim N(52; 8)$$

Debido a la complejidad del proceso de integración, se resolvió de forma iterativa tomando valores enteros para y *X1,X2,X3*, y se diseñó una rutina en computador para tal proceso. Así, el *VEMax* de la duración del paquete de trabajo es aproximadamente:

$$59.15\ días = E\left(y_{max}\right)$$
$$= \int_{-\infty}^{\infty}\left(y_{max} * \left(f_1 \int_{-\infty}^{y_{max}} f_2 \int_{-\infty}^{y_{max}} f_3\right)\right) dy_{max}$$
$$+ \int_{-\infty}^{\infty}\left(y_{max} * \left(f_2 \int_{-\infty}^{y_{max}} f_1 \int_{-\infty}^{y_{max}} f_3\right)\right) dy_{max}$$
$$+ \int_{-\infty}^{\infty}\left(y_{max} * \left(f_3 \int_{-\infty}^{y_{max}} f_1 \int_{-\infty}^{y_{max}} f_2\right)\right) dy_{max}$$

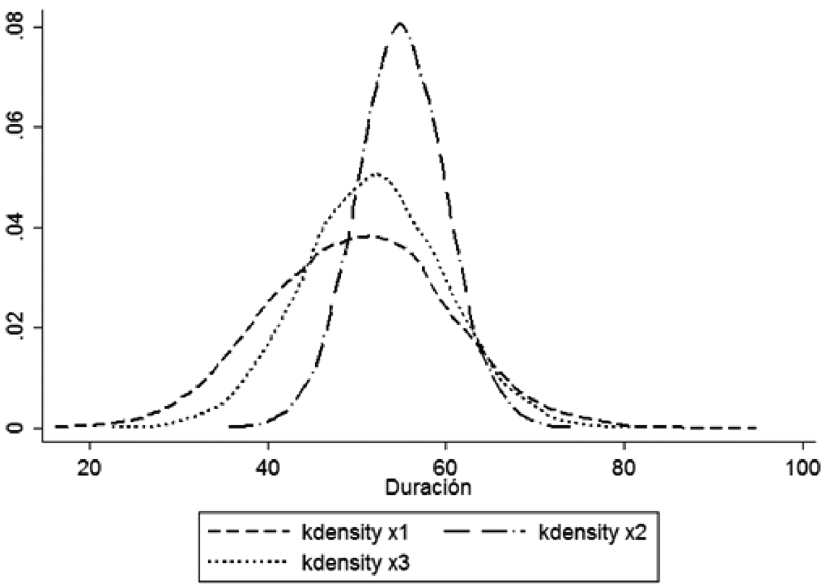
Por otro lado, el *MaxVE* es:

$$MaxVE = \max\{E(x_1); E(x_2); E(x_3)\}$$
$$MaxVE = \max\{50; 55; 52\}$$
$$MaxVE = 55$$

A partir de esto, es evidente que el *MaxVE* de las actividades no coincide con el *VEMax*. Entonces, el primer procedimiento puede generar serios problemas en la estimación de tiempos; en este caso, el *MaxVE* subestima la duración estimada del paquete de trabajo en 4.15 días.

El presente ejemplo fue simulado mediante un experimento Montecarlo. Las distribuciones de densidad de cada una de las actividades en este ejemplo pueden verse en el gráfico:

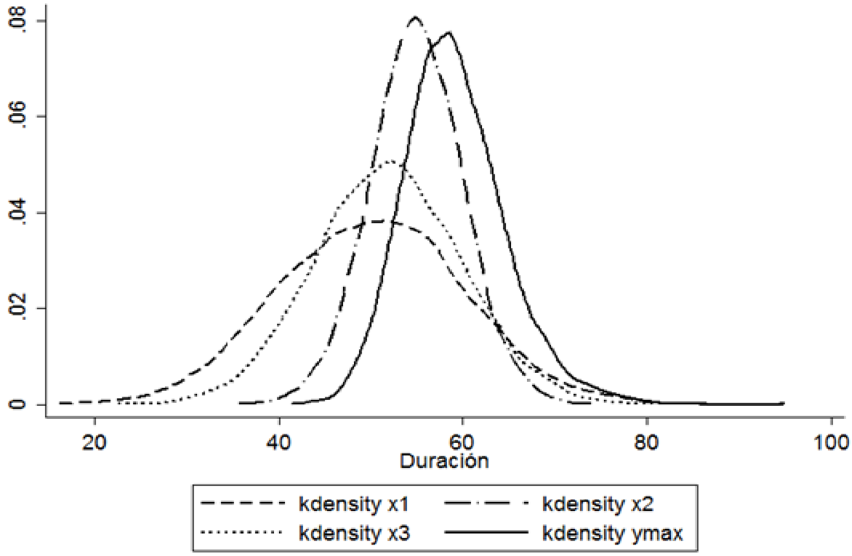
Gráfico 1.
Funciones de densidad de las actividades 1,2 y 3



Fuente: Elaboración propia.

Agregando la función de densidad de y_{max} al gráfico:

Gráfico 2.
Funciones de densidad de las actividades 1,2 y 3 y el paquete de trabajo



Fuente: Elaboración propia.

En el Gráfico 2 puede observarse que la distribución de es distinta a y e incluso puede no ser normal. El test de normalidad aplicado a esta distribución indica que existe suficiente evidencia para rechazar la hipótesis nula de normalidad:

Tabla 1.
Test de asimetría y curtosis de normalidad de y_{max}

Test	Variable	Observaciones	Pr(Asimetría)	Pr(Curtosis)	Chi2(2)	Prob>chi2
Ajustado (Royston, 1991)		10,000	0.0000	0.0000	.	0.0000
No ajustado (D'Agostino, 1990)		10,000	0.0000	0.0000	583.6	0.0000

Notas: Hipótesis nula de normalidad. El estadístico Chi2 del test ajustado no es reportado debido a que su valor es demasiado grande; sin embargo, sí se reporta su probabilidad.
Fuente: Elaboración propia.

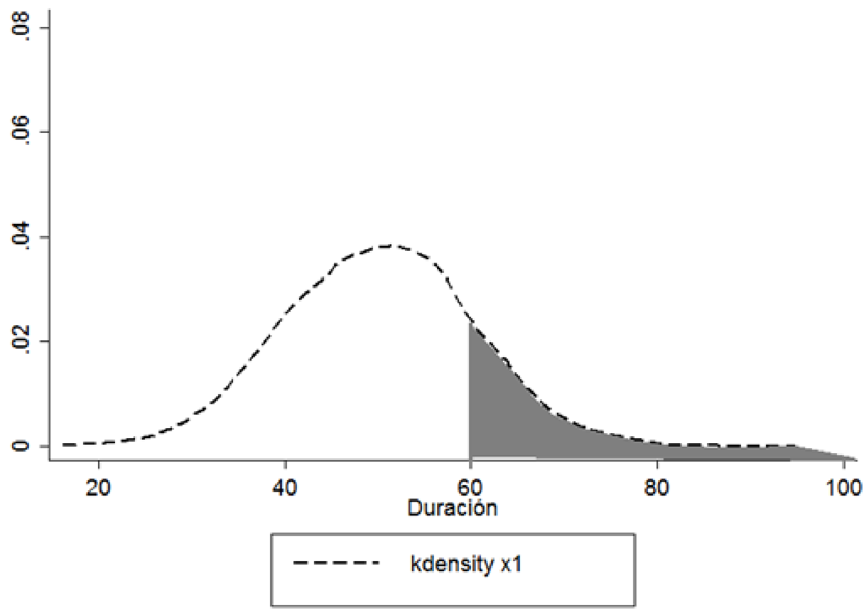
Queda entonces en evidencia que la distribución de los valores máximos de la combinación de 3 variables aleatorias normalmente distribuidas, independientes y con media y desviación estándar distinta, no está necesariamente distribuida de forma normal. Esta distribución se profundizará al analizar el riesgo asociado en la duración del paquete de trabajo. Sin embargo, hasta aquí debe quedar claro que las estimaciones realizadas con el *VEMax* no son iguales a la duración promedio o valor esperado del paquete de trabajo (*VEMax*).

Riesgo de retraso en la realización del paquete de trabajo

En una situación real, resulta útil conocer las probabilidades de que el paquete de trabajo dure más o menos que un valor determinado. En este ejemplo, interesará conocer cuál es la probabilidad de que nuestro paquete de trabajo dure más de 59 días.

Para entender el mecanismo que opera en la estimación de las probabilidades, fijémonos en la actividad 1. Si deseamos conocer las probabilidades de que la duración de la actividad 1 se extienda por más de 60 días, deberemos de estimar el área que se encuentra entre 60 días y el infinito positivo.

Gráfico 3.
Función de densidad de la duración de la actividad 1, estimación de probabilidades



Fuente: Elaboración propia.

Entonces, calcular las probabilidades de que la duración de la actividad se ubique dentro de un rango de valores, como 40 y 50 días, o por encima de un valor, como 60 días, implica calcular el área bajo la curva de la función de densidad entre estos valores. Formalmente, el área bajo la curva se calcula encontrando la integral de la función de densidad entre dichos puntos. Así:

$$Prob[x_1 \geq 60] = 1 - \int_{-\infty}^{60} f(x_1)dx_1$$

ó

$$Prob[x_1 \geq 60] = \int_{60}^{\infty} f(x_1)dx_1$$

Calcular las probabilidades que el paquete de trabajo tiene de terminar entre dos fechas o por encima de una fecha crítica (f_c), implica calcular el área bajo la curva de la función de densidad de la duración total del paquete de trabajo y_{max}:

$$Prob[y \geq fc] = \int_{fc}^{\infty} \left(f_1 \int_{-\infty}^{y} f_2 \int_{-\infty}^{y} f_3 \right) + \int_{fc}^{\infty} \left(f_2 \int_{-\infty}^{y} f_1 \int_{-\infty}^{y} f_3 \right) + \int_{fc}^{\infty} \left(f_3 \int_{-\infty}^{y} f_1 \int_{-\infty}^{y} f_2 \right)$$

O, usando la expresión (3)

$$Prob[y \geq fc] = 1 - F_1(fc) * F_2(fc) * F_3(fc)$$

Así, la probabilidad de que la duración de nuestro paquete de trabajo sea mayor a 59 días se calcula como:

$$Prob[y \geq 59] = 1 - F_1(59) * F_2(59) * F_3(59)$$
$$Prob[y \geq 59] = 1 - \int_{-\infty}^{59} f_1(y) * \int_{-\infty}^{59} f_2(y) * \int_{-\infty}^{59} f_3(y)$$
$$Prob[x_1 \geq 59] = 1 - (0.82)(0.79)(0.81)$$
$$Prob[x_1 \geq 59] = 1 - 0.52$$
$$Prob[x_1 \geq 59] = 0.48$$

Es decir, aunque la actividad 1 tenga una probabilidad de 18% (100 - 82) de acabar por encima de 59 días, la actividad 2 una probabilidad de 21% y la actividad 3 de 19%, la probabilidad de que el paquete de trabajo acabe en más de 59 días es de 48%, un valor, por mucho, mayor a las estimaciones individuales. Entonces, las probabilidades individuales no son buenos referentes por si solos para estimar las probabilidades de duración total del paquete de trabajo.

Puede llegarse a esta conclusión al observar las gráficas de densidad para las 4 variables en el Gráfico 2. Aquí, la densidad de la variable y_{max} –duración total del paquete de trabajo– se encuentra por encima de cada una de las funciones de densidad de las actividades para valores elevados de duración. Por lo tanto, entre estos valores el área bajo la curva –probabilidad– es mayor que las probabilidades individuales.

Variables aleatorias no-negativas

Hasta ahora, hemos analizado un ejemplo con la duración de las actividades como variables aleatorias con una media y desviación estándar y cuyo valor oscila entre y , con el fin de simplificar el análisis. Esto no ha sido un problema, debido a que la media de las actividades en nuestro ejemplo ha sido bastante alta y la desviación estándar relativamente baja, lo que ha implicado que sea poco probable que una actividad tenga una duración negativa (x_1<0).

Sin embargo, en situaciones reales la duración de una tarea o actividad no puede ser negativa.

Una forma de lidiar con la no negatividad de la duración de las actividades es asumir que parten de una distribución log-normal. Esta distribución está definida solo para valores positivos, lo que sirve a nuestros propósitos.

Esta distribución es usada frecuentemente en variables similares, como el tiempo de falla (*time to failure*) que describe el tiempo en que una máquina deje de funcionar

Una variable aleatoria cuya distribución es log-normal puede definirse como

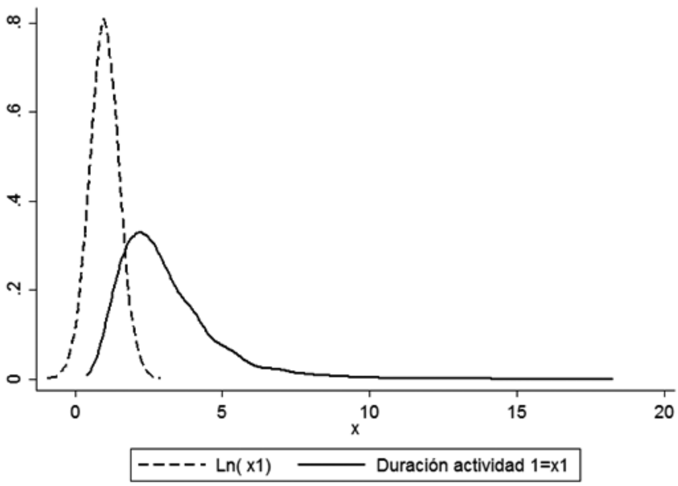
$$x_1 \sim LN(\mu_{l1}; \sigma_{l1})$$
$$f(x_1) = \left(\frac{1}{x_1 \sigma_{l1} \sqrt{2\pi}}\right) * e^{-\frac{1}{2}\left(\frac{\ln x_1 - \mu_{l1}}{\sigma_{l1}}\right)^2}$$

Su característica principal es que su logaritmo natural se distribuye normalmente:

$$\ln x_1 \sim N(\mu_1; \sigma_1)$$

Gráficamente, la función de densidad log-normal tiene una mayor área para valores pequeños y una larga cola hacia la derecha. En el Gráfico 4 se observa la función de densidad log-normal y su logaritmo que se distribuye normalmente.

Gráfico 4.
Funciones de densidad log-normal y normal



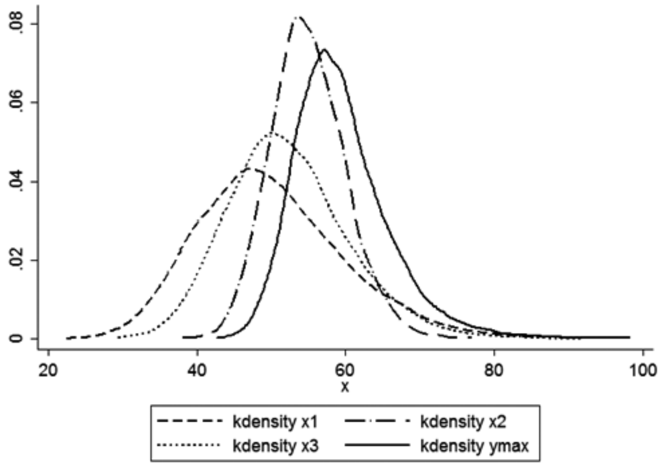
Fuente: Elaboración: propia.

Volviendo a nuestro ejemplo, las funciones de densidad log-normal para X_1, X_2, y X_3 están definidas por:

$$x_1 \sim LN(50;10)$$
$$x_2 \sim LN(55;15)$$
$$x_3 \sim LN(52;8)$$

Cuyos gráficos de densidades, además del gráfico de la duración total del paquete de trabajo son:

Gráfico 5.
Funciones de densidad log-normales de las actividades 1,2 y 3 y el paquete de trabajo



Fuente: Elaboración propia.

Entonces, el valor esperado de la duración del paquete de trabajo viene dado por la misma ecuación que definimos en secciones anteriores (4). Al reemplazar las funciones de densidad y distribución por las respectivas log-normales de X_1, X_2, y X_3 tenemos que:

$$E(y_{maxLN}) = 59.38$$

Al igual que el caso inicial de distribución normal de las duraciones de las actividades, en este nuevo caso de distribución log-normal el valor esperado de la duración del paquete de trabajo ($VEMax_{LN}$) es mayor que el máximo de los valores esperados de las actividades ($MaxVE_{LN}$).

Por otro lado, al comparar el caso de distribución normal de las actividades con la log-normal, debe indicarse que existen mayores diferencias entre el $VEMax_{LN}$ y el $VEMax_{N}$ cuando las actividades tienen una media cercana a o y una desviación estándar que permite que la variable normalmente distribuida tome valores negativos frecuentemente.

La probabilidad de que la duración del paquete de trabajo se prolongue por más de 59 días será de:

$$Prob_{LN}[x_1 \geq 59] = 1 - F_{LN1}(59) * F_{LN2}(59) * F_{LN3}(59)$$
$$Prob_{LN}[x_1 \geq 59] = 0.47$$

Así como en el caso de distribución normal, las probabilidades de que el paquete de trabajo culmine en más de 59 días son mayores a las probabilidades individuales de las actividades, y en este caso son de 47%.

Conclusiones

La duración de las actividades de un proyecto son variables aleatorias. Así, sean discretas o continuas, su duración oscila alrededor de una media en una magnitud proporcional a su desviación estándar.

La duración máxima de los valores esperados (*MaxVE*) de las tareas o actividades simultáneas de un paquete de trabajo no es igual al valor esperado de las duraciones máximas de las actividades (*VEMax*). Esta última forma de medir la duración del paquete de trabajo se basa en el cálculo del valor esperado de la duración total. El valor esperado es el valor promedio alrededor del cual oscilarán las duraciones totales si realizáramos el paquete de trabajo varias veces. Así, bajo la primera medición (*MaxVE*), estimaremos incorrectamente la duración más probable del paquete de trabajo lo que presentará serios problemas en la gestión del cronograma del proyecto.

La estimación adecuada de los riesgos de la duración de un proyecto pasa por utilizar las funciones de densidad y distribución de las actividades correctamente. El mal manejo de los cálculos y las propiedades estadísticas de las funciones de densidad y distribución puede traer serias consecuencias en la gestión del cronograma.

Por otro lado, la probabilidad de que el paquete de trabajo se prolongue por encima de una duración determinada es superior a las probabilidades individuales que tiene cada una de las actividades. Entonces, una estimación del riesgo en base a las funciones de distribución individuales de cada actividad es incoherente con el riesgo real que tiene el paquete de trabajo en culminar luego de la fecha de evaluación.

La distribución log-normal cumple con la característica de no-negatividad de las duraciones de las actividades. Esta distribución es usada frecuentemente en variables similares, como el tiempo de falla (*time to failure*) que describe el tiempo en que una máquina deje de funcionar.

Discusión

Se ha propuesto que las duraciones de las actividades son variables aleatorias. Su definición parte por conocer su valores esperado y su desviación estándar. Resulta importante establecer una metodología para la estimación de estos valores de forma coherente.

El proceso de estimación del valor esperado del paquete de trabajo (*VEMax*) es matemáticamente tedioso, o incluso su cálculo exacto puede llegar a ser imposible, debido a que no es posible realizar la doble integración de la función de densidad normal. Las estimaciones Montecarlo se presentan como una buena alternativa para aproximarnos al valor esperado. Sin embargo, no permite resolver problemas como encontrar el punto mínimo de duración en función a alguna característica de las actividades como la media o la varianza. Por otro lado, las propiedades de la función de densidad de la duración del paquete de trabajo no fueron discutidas a profundidad. Aunque su interpretación parece difícil y tediosa, en realidad es intuitiva e interesante. Se deja este análisis para trabajos futuros.

Se ha discutido este fenómeno cuando las duraciones de las actividades son variables aleatorias independientes. Cuando las duraciones de las actividades tengan algún grado de dependencia; por ejemplo, que compartan el uso de un recurso, los cálculos serán distintos a los presentados en este ensayo. Así, queda para la posteridad el análisis de actividades cuya duración no es independiente. Si bien la distribución log-normal cumple con el requisito de no negatividad de la duración de las actividades, otros propiedades podrían no ajustarse al fenómeno estadístico que es la duración de una actividad. Además, la interpretación de su logaritmo natural, que se distribuye normalmente, no ha sido tomada en cuenta.

Nuestro análisis parte de que se conocen las actividades para la realización del paquete de trabajo. Sin embargo, las actividades aquí descritas pudieran ser susceptibles a dividirse en sub-actividades, y tal vez todas de realización paralela o simultánea. En este sentido, la estimación de los tiempos usando el máximo de los valores esperados (*MaxVE*) para cada nivel de sub-actividades podría tener un efecto acumulativo en las estimaciones de las actividades de niveles superiores. Este efecto no ha sido analizado en este ensayo. Sin embargo, el cálculo del valor estimado de la duración del paquete de trabajo (*VEMax*) aplicado de forma iterativa para nivel de actividades, resuelve este problema.

Referencias

D'Agostino, R. B. (1990). A suggestion for using powerful and informative tests of normality. "American Statistician", 44, 316-321.

Project Management Institute. (2013). "Project Management Body of Knowledge". Pensilvania: Project Management Institute Inc.

Rolstadås, A. (2008). "Applied Project Management". Trondheim: Tapir Academic Press.

Royston, P. (1991). Comment on sg3.4 and an improved D'Agostino test. "Stata Technical Bulletin", 1, 110-112.

Savage, S. L. (2009). "The flaw of averages". New Jersey: John Wiley & Sons, Inc.

Tesler, G. (2013). "Max of n variables & ling repeats". University of California, San Diego. (Documento en formato pdf). Obtenido el 21 de abril del 2013 de http://www.math.ucsd.edu/~gptesler/283/slides/longrep_f13-handout.pdf

Wooldridge, J. (2012). "Introductory econometrics: A modern approach". Cengage Learning.

Recibido: 17-08-2015
Aceptado: 08-10-2015