PERT--- Program Evaluation and Review Technique

Scheduling model based on Activity on Arrow Networks:

- Activity durations are random (probabilisitc).
- Emphasis placed on events (like probability theory) and less on activities.

REVIEW OF OF BASIC PROBABILITY THEORY

EVENTS

S = Sample Space = $\{E_1, E_2, \cdots\}$ = Set of all possible outcomes (simple events E_i)

A = Compound event = $\{E_i, E_j, \cdots E_k\}$ = Set of simple events E_i, $\quad A \subseteq S$

A^C = Complement of event A = $\{E_i$ not in $A\}$

A and A^C are mutually exclusive : $A \cap A^C = 0$, and collectively exhaustive : $A \cup A^C = S$

THE THREE AXIOMS OF PROBABILITY THEORY

I. $\quad 0 \le P[A] \le 1$

II. $\quad P[S] = 1$

IIIa. $\quad \sum_{E_i \in S} P[Ei] = 1$ where E_i are mutually exclusive and collectively exhaustive

IIIb. $\quad P[A \cup B] = P[A] + P[B]$ where A and B are mutually exclusive

(either IIIa or IIIb can serve as the third axiom)

PROBABILITY OF A COMPOUND EVENT

$$P[A] = \sum_{E_i \in A} P[Ei]$$

$P[A] = 1 - P[A^C]$ Often the easiest way to calculate the probability of an event

PROBABILITY OF CONDITIONAL EVENTS

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

$$P[A \cap B] = P[A|B]\ P[B] = P[B|A]\ P[A]$$

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

PROBABILISTIC INDEPENDENCE

$P[A \cap B] = P[A]\ P[B]$ *or*

$P[A|B] = P[A]$ *or*

$P[B|A] = P[B]$

RANDOM VARIABLES

Random variables are used to describe probabilisitc processes that result in numerical outcomes.
Random variables are usually shown with capital letters from the end of the alphabet : *X, Y, Z*
Basic event : $a \le X \le b$
$P[X = x]$ = Probability that X takes on the value x, $0 \le P[X = x] \le 1$

Random variables can be discrete or continuous :

Discrete : S is enumerable

Toss of a coin : $S = \{0,1\}$

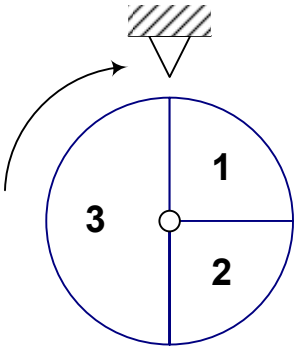
Toss of a die : $S = \{1,2,3,4,5,6\}$

Number of experiments to the first success : $S = \{1,2,3,\cdots\}$

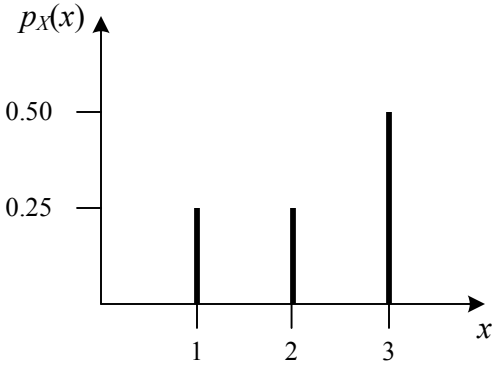
Continuous : S is a subset of the set of real numbers $\Re$

Activity Duration : $S = \{x : a \le x \le b\}$

Time to next arrival : $S = \{x : 0 \le x \le \infty\}$



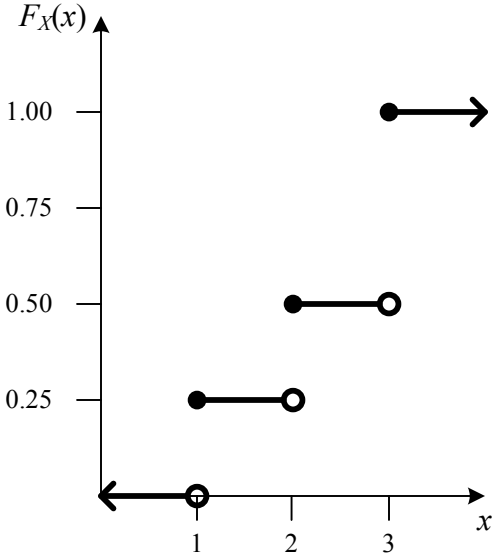
EXAMPLE OF A DISCRETE RANDOM VARIABLE



Probability Mass Function (PMF) $= p_X(x) = P[X = x]$

$$P[X = x] = p_X(x) = \begin{cases} 0.25 & x = 1 \\ 0.25 & x = 2 \\ 0.5 & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

$0 \le p_X(x) \le 1$ and $\sum_{\text{all } x} p_X(x) = 1$



Cumulative Distribution Function (CDF) $= F_X(x) = P[X \le x]$

$$F_X(x) = P[X \le x] = \begin{cases} 0 & x < 1 \\ 0.25 & 1 \le x < 2 \\ 0.5 & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

$0 \le F_X(x) \le 1$, $F_X(x)$ is a never - decreasing function

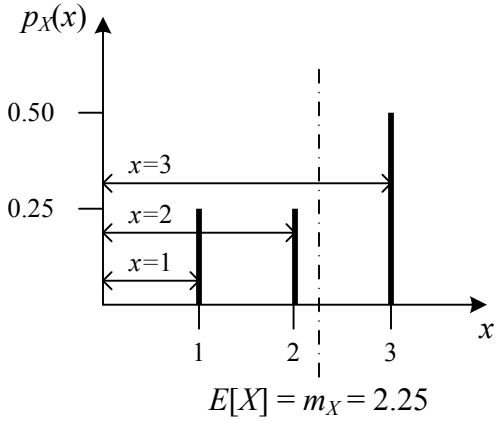
MOMENTS OF DISCRETE RANDOM VARIABLES

EXPECTED VALUE (FIRST MOMENT)

Expected Value = Mean = Centroid of probability mass = Limit of long - term average of outcomes

$$E[X] = m_X = \sum_{\text{all } x} x \cdot p_X(x)$$

$$E[X] = m_X = 1 \times 0.25 + 2 \times 0.25 + 3 \times 0.5 = 2.25$$



VARIANCE (SECOND CENTROIDAL MOMENT)

Variance = Centroidal moment of inertia of probability mass = Measure of dispersion (variability)

$$Var[X] = \sigma_X^2 = \sum_{\text{all } x} (x - E[X])^2 \cdot p_X(x)$$

$$Var[X] = \sigma_X^2 = 1.25^2 \times 0.25 + 0.25^2 \times 0.25 + 0.75^2 \times 0.5 = 0.6875$$

Theorem of parallel axes

$$Var[X] = \sigma_X^2 = E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] = E[X^2] - 2E[X]E[X] + E[(E[X])^2]$$

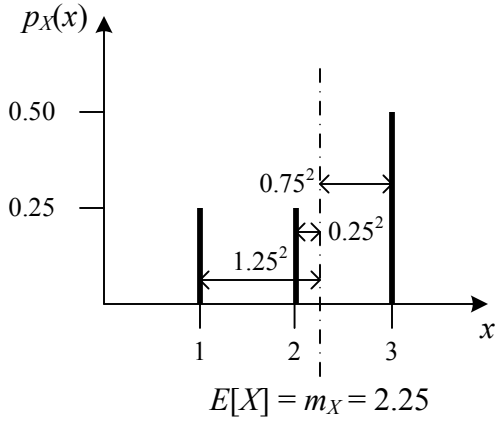
$$= E[X^2] - 2(E[X])^2 + (E[X])^2 = E[X^2] - (E[X])^2$$

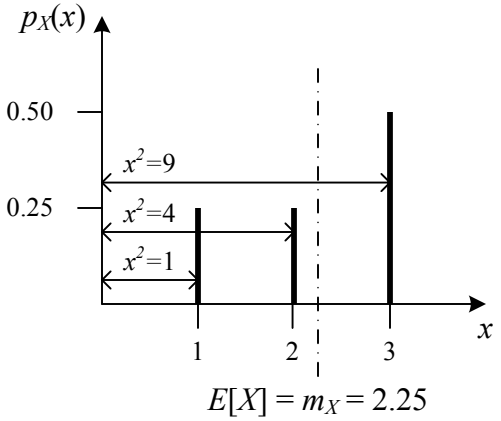
$$E[X^2] = 1^2 \times 0.25 + 2^2 \times 0.25 + 3^2 \times 0.5 = 0.25 + 1.00 + 4.5 = 5.75$$

$$Var[X] = \sigma_X^2 = E[X^2] - (E[X])^2 = 5.75 - 2.25^2 = 0.6875$$

Standard Deviation (has the same units as X)

$$SD[X] = \sigma_X = \sqrt{Var[X]} = \sqrt{0.6875} = 0.83$$

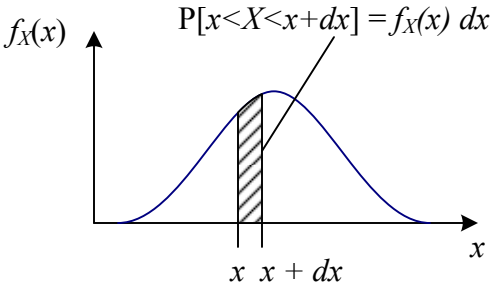




CONTINUOUS RANDOM VARIABLES

For a continuous random variable $P[X = x] = 0 \quad \Rightarrow \quad$ no probability mass exists at any point x

Instead, points x are associated with probability density : Probability Density Function (PDF) $= f_X(x)$

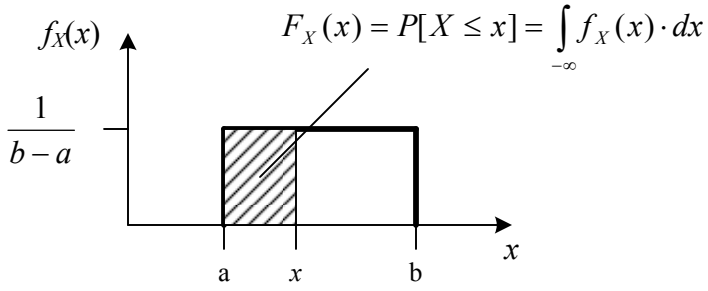


$$0 \le f_X(x), \qquad f_X(x) \text{ may be} > 1$$

$$P[S] = \int_{-\infty}^{+\infty} f_X(x) \cdot dx = 1$$

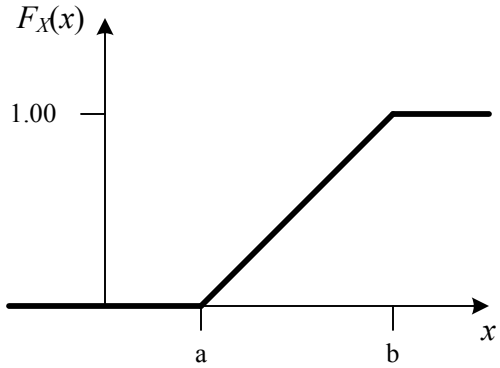
$$P[a \le X \le b] = \int_a^b f_X(x) \cdot dx$$

EXAMPLE: THE UNIFORM PDF



Uniform PDF

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

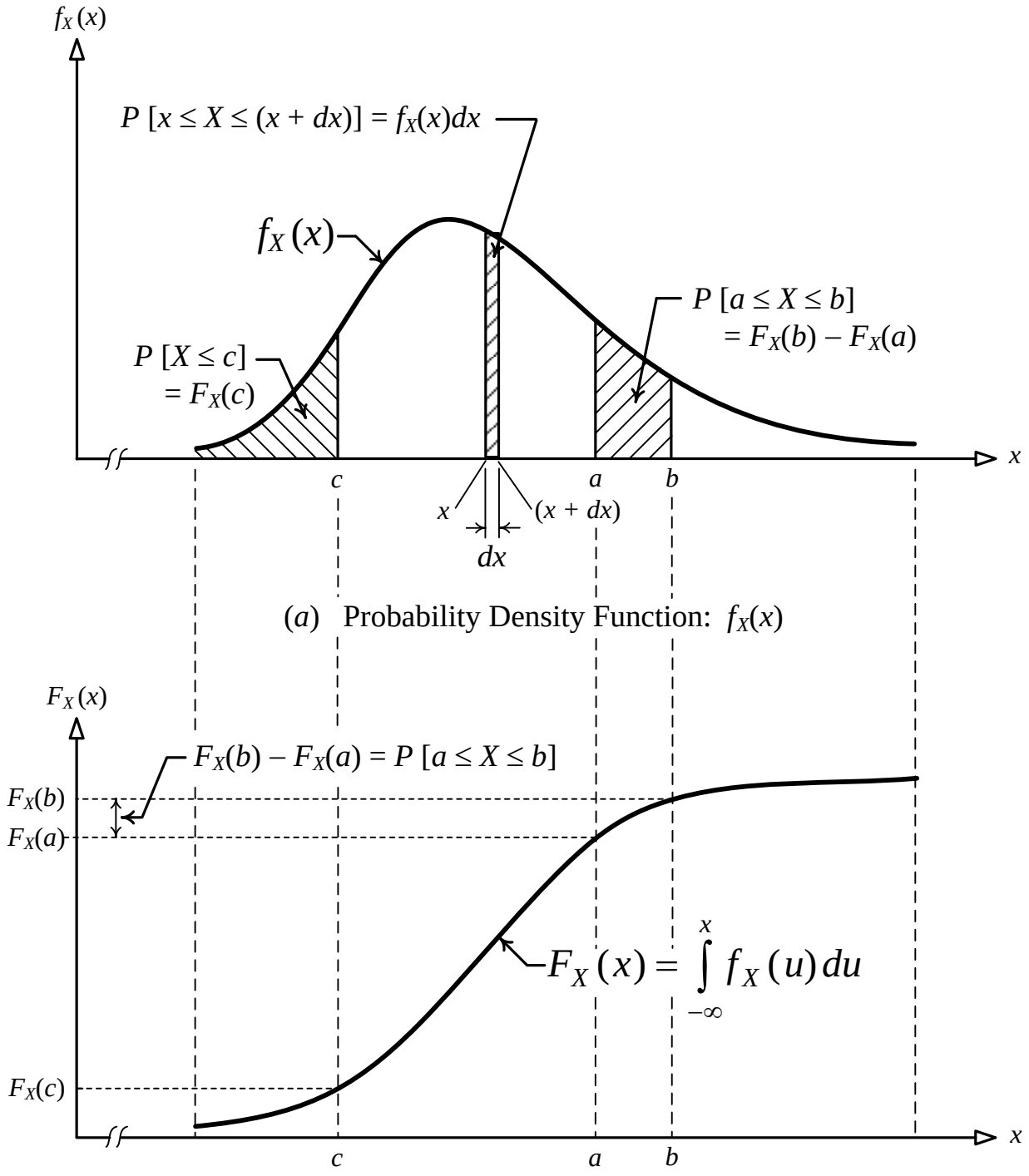


CDF of Uniform PDF

For $a \le x \le b$

$$F_X(x) = \int_a^x \frac{1}{b-a} \cdot dx = \left[\frac{x}{b-a} \right]_a^x = \frac{x-a}{b-a}$$

$$F_X(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & b \le x \end{cases}$$



(*a*) Probability Density Function: $f_X(x)$

(*b*) Cumulative Distribution Function: $F_X(x)$

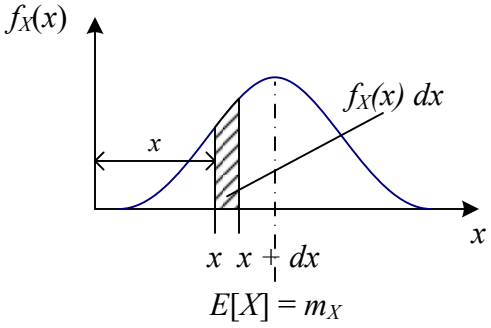
Figure 10.2

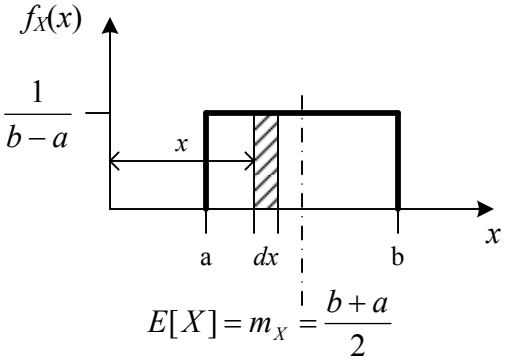
MOMENTS OF CONTINUOUS RANDOM VARIABLES

EXPECTED VALUE (FIRST MOMENT)

Expected Value = Mean = Centroid of area under PDF = Limit of long - term average of outcomes

$$E[X] = m_X = \int_{-\infty}^{+\infty} x \cdot f_X(x)dx \qquad E[X] = m_X = \int_a^b x \cdot \frac{1}{b-a}dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

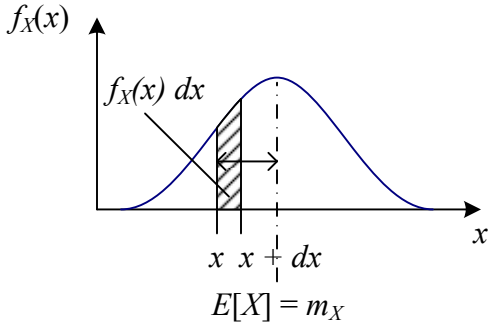


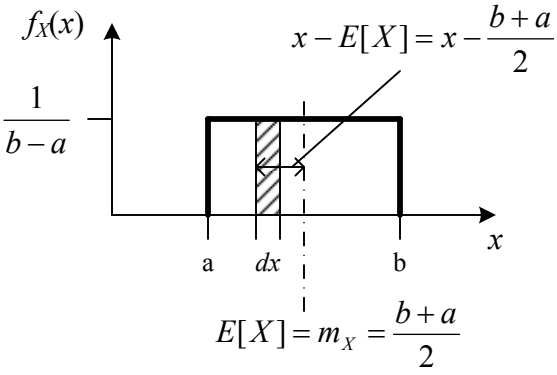


VARIANCE (SECOND CENTROIDAL MOMENT)

Variance = Centroidal momentof inertia of area under PDF = Measure of dispersion (variability)

$$Var[X] = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - E[x])^2 \cdot f_X(x)dx \qquad Var[X] = \sigma_X^2 = \int_a^b (x - \frac{b+a}{2})^2 \cdot \frac{1}{b-a}dx$$





$$Var[X] = \sigma_X^2 = E[(x - E[x])^2] = E[X^2] - (E[X])^2$$

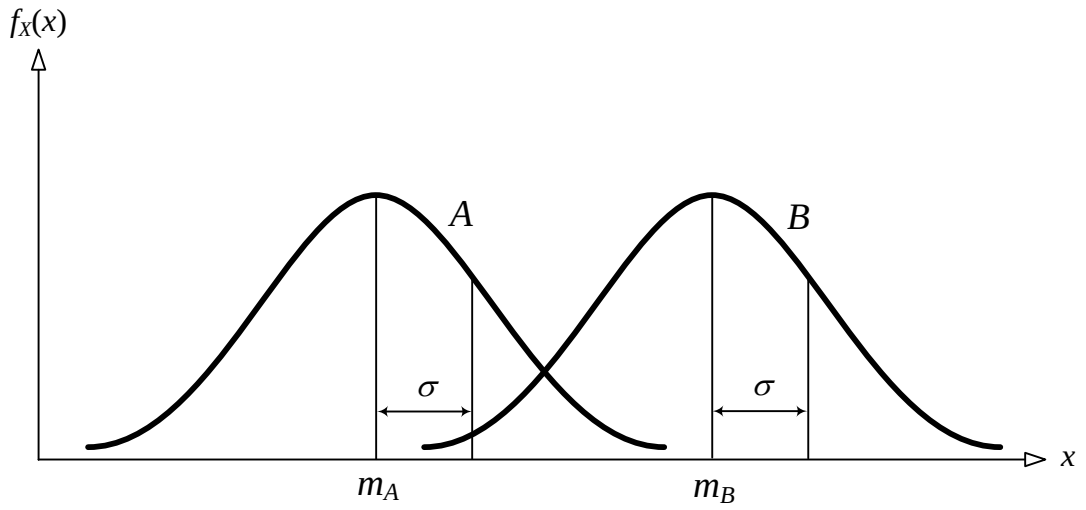
$$E[X^2] = \int_a^b \frac{x^2 dx}{b-a} = \left[\frac{x^3}{3(b-a)}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

$$Var[X] = \frac{a^2+ab+b^2}{3} - \frac{(b+a)^2}{4} = \frac{(b-a)^2}{12}$$

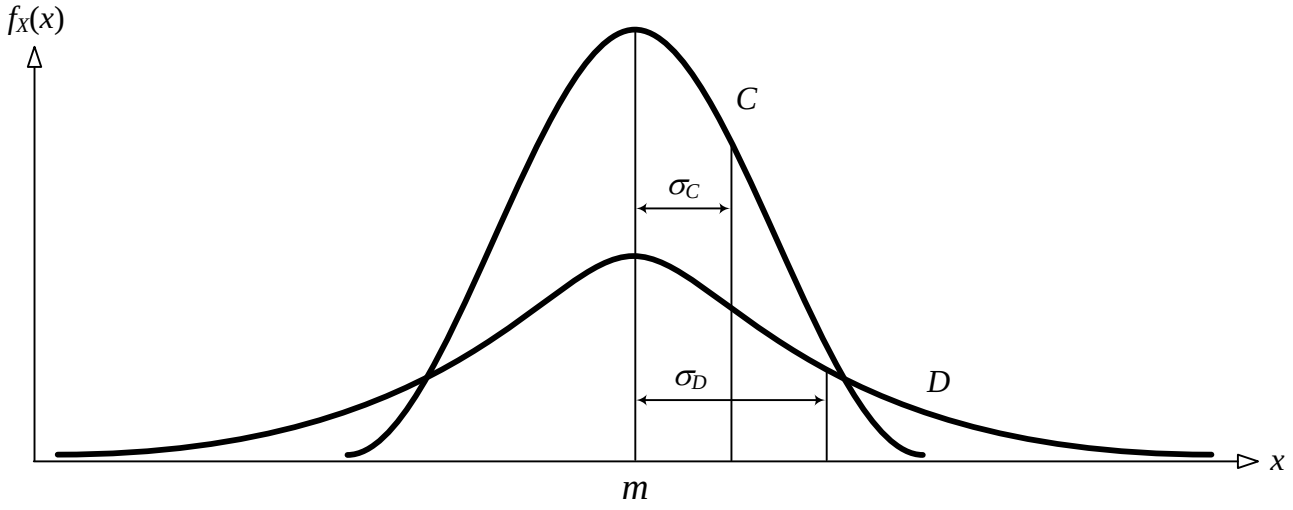
Standard Deviation (has the same units as X)

$$SD[X] = \sigma_X = \sqrt{Var[X]} = \frac{(b-a)}{2\sqrt{3}}$$

P.G. Ioannou



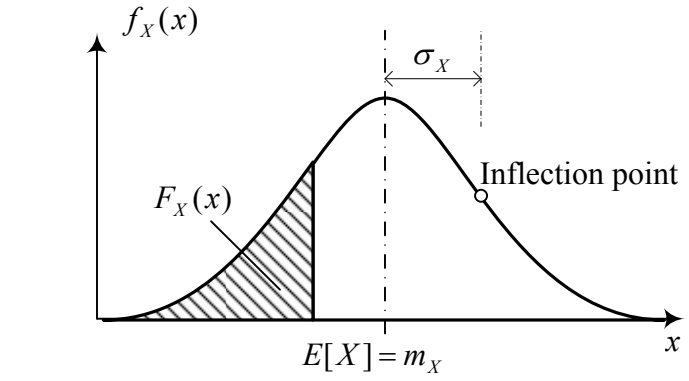
(a) Same Standard Deviation, Different Means



(b) Same Mean, Different Standard Deviations

Figure 10.4 Relationships Between Normal Distributions

THE NORMAL DISTRIBUTION

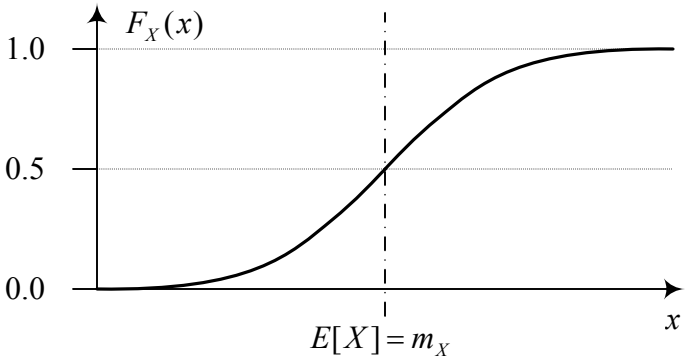


$$X \sim N(m_X, \sigma_X)$$

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-m_X}{\sigma_X})^2}$$

$F_X(x)$ does not exist. I.e., there is no function , such that $\frac{dF_X(x)}{dx} = f_X(x)$.

Thus, $F_X(x)$ can only be evaluated using numerical methods for specific values of x, m_X, and σ_X.

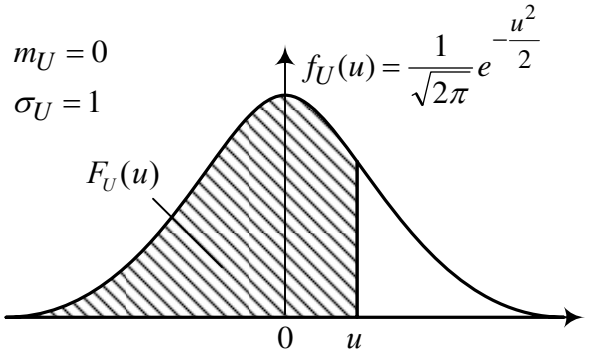


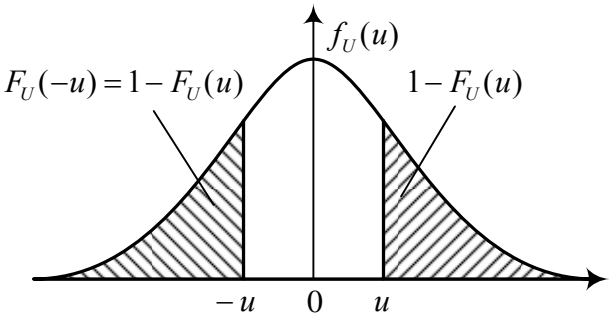
THE STANDARD NORMAL VARIATE

$$U = \frac{X - m_X}{\sigma_X}$$

The CDF $F_U(u)$ is widely available in tables. These tables usually show $F_U(u)$ only for positive values of u, because :

$$F_U(-u) = 1 - F_U(u)$$





HOW TO CALCULATE $F_X(x)$ FOR ANY m_X and σ_X

$$F_X(x) = P[X \le x] = P[\frac{X - m_X}{\sigma_X} \le \frac{x - m_X}{\sigma_X}] =$$

$$P[U \le \frac{x - m_X}{\sigma_X}] = F_U(\frac{x - m_X}{\sigma_X}) \quad \Rightarrow$$

$$\boxed{F_X(x) = F_U(\frac{x - m_X}{\sigma_X})}$$

Example : $X \sim N(6,4)$ $\quad m_X = 6,\ \sigma_X^2 = 16$

$$F_X(x) = F_U(\frac{x-6}{4})$$

$$F_X(10) = F_U(\frac{10-6}{4}) = F_U(1) = 0.8413$$

$$F_X(-1) = F_U(\frac{-1-6}{4}) = F_U(-1.75) =$$

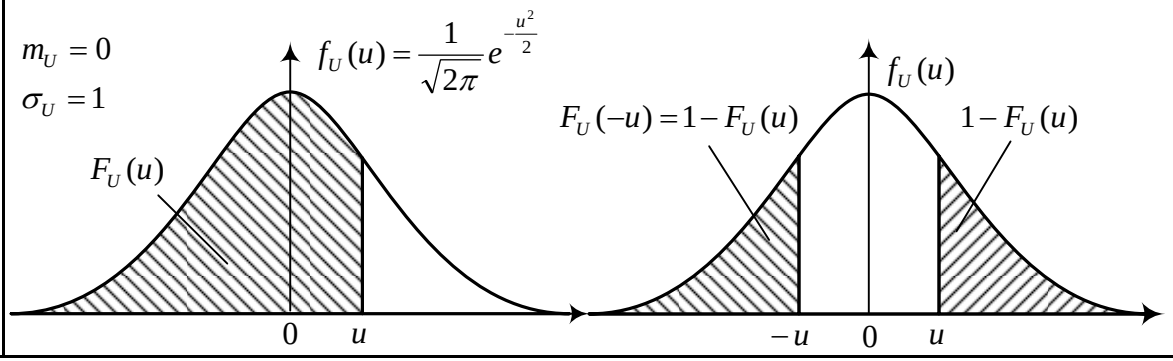
$$= 1 - F_U(1.75) = 1 - 0.96 = 0.04$$

Values of the Standardized Normal Distribution, $U \sim N(0,1)$

Table gives values of the Cumulative Distribution Function $F_U(u)$, for $u > 0$

For $-u < 0$, apply $F_U(-u) = 1 - F_U(u)$

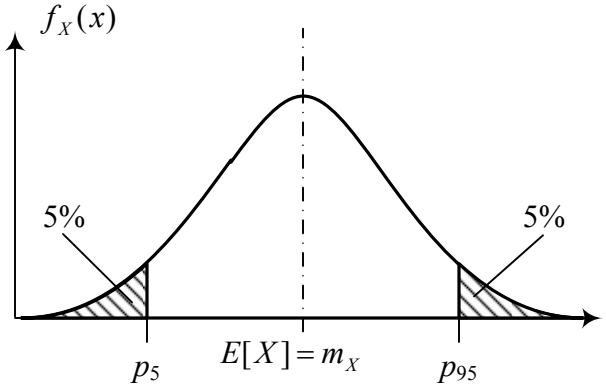
u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725									
2.1	0.98214									
2.2	0.98610									
2.3	0.98928									
2.4	0.99180									
2.5	0.99379									
3.0	0.99865									
3.5	0.999767									
4.0	0.9999683									
4.5	0.9999966									
5.0	0.99999971									
5.5	0.999999981									



Percentiles	
$F(u)$	u
50.0%	0.00000
55.0%	0.12566
60.0%	0.25335
65.0%	0.38532
70.0%	0.52440
75.0%	0.67449
80.0%	0.84162
82.5%	0.93459
85.0%	1.03643
87.5%	1.15035
90.0%	1.28155
90.5%	1.31058
91.0%	1.34075
91.5%	1.37220
92.0%	1.40507
92.5%	1.43953
93.0%	1.47579
93.5%	1.51410
94.0%	1.55477
94.5%	1.59819
95.0%	1.64485
95.5%	1.69540
96.0%	1.75069
96.5%	1.81191
97.0%	1.88079
97.5%	1.95996
98.0%	2.05375
98.5%	2.17009
99.0%	2.32634
99.25%	2.43239
99.50%	2.57583
99.75%	2.80706

P.G. Ioannou, 10/30/2007

PERCENTILES, MEDIAN, MODE



PERCENTILES

The kth percentile, p_k, of a random variable X is that value x such that

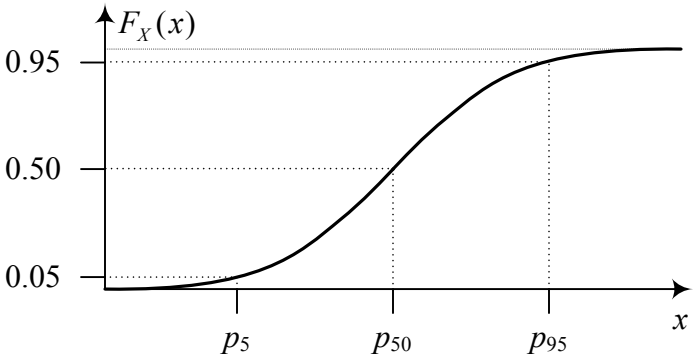
$$F_X(p_k) = k\%$$

In PERT we will find the following three percentiles useful for the definition of the Perry and Greig formulas:

$p_5 = 5^{th}$ percentile: $F_X(p_5) = 5\%$

$p_{50} = 50^{th}$ percentile: $F_X(p_{50}) = 50\%$

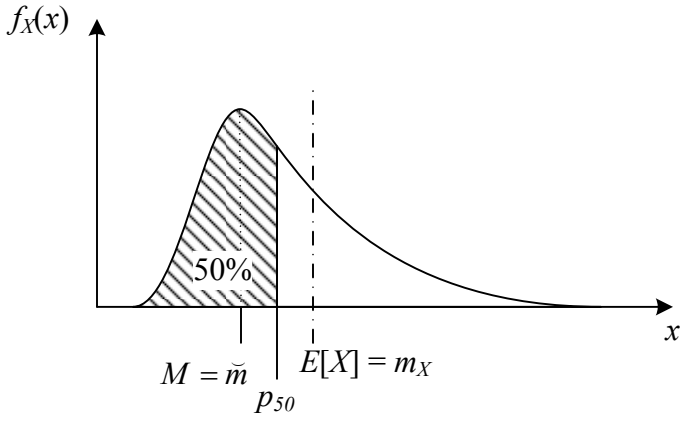
$p_{95} = 95^{th}$ percentile: $F_X(p_{95}) = 95\%$



MEDIAN

The median p_{50} of a random variable X divides the area under the PDF into two parts of equal area. By definition, the median is equal to the 50th percentile:

$$F_X(p_{50}) = 50\%$$



MODE

The mode $M = \breve{m}$ of a random variable X is that value of x where the PDF $f_X(x)$ has a maximum:

$$\left.\frac{df_X(x)}{dx}\right|_M = 0$$

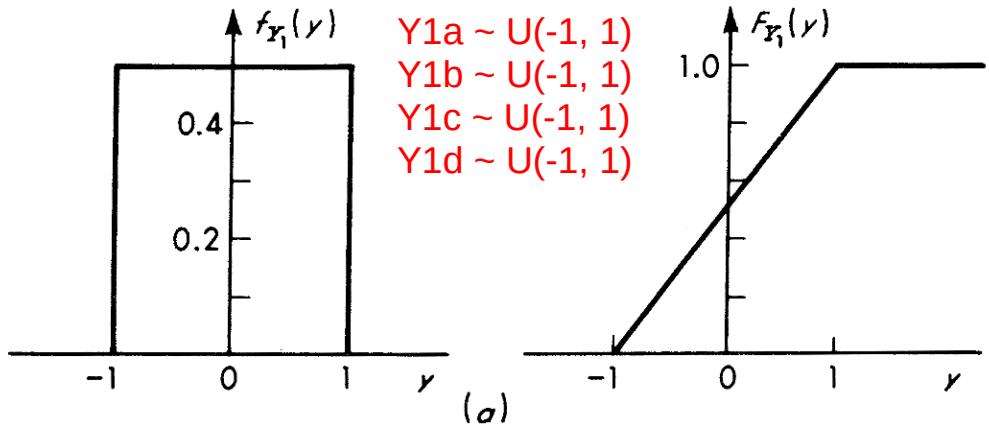
The mode is also called

"the most likely value of X".

Most distributions have a single mode (i.e., they are unimodal).

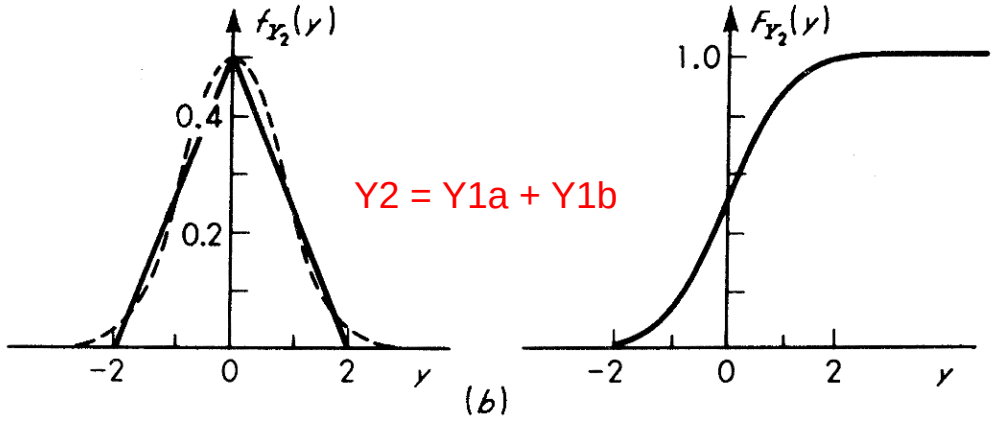
CENTRAL LIMIT THEOREM

Y1a ~ U(-1, 1)
Y1b ~ U(-1, 1)
Y1c ~ U(-1, 1)
Y1d ~ U(-1, 1)



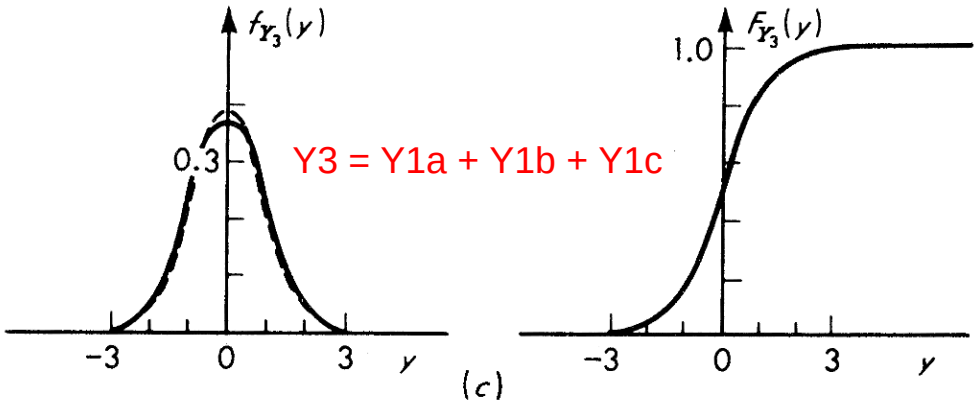
(a)

Y2 = Y1a + Y1b



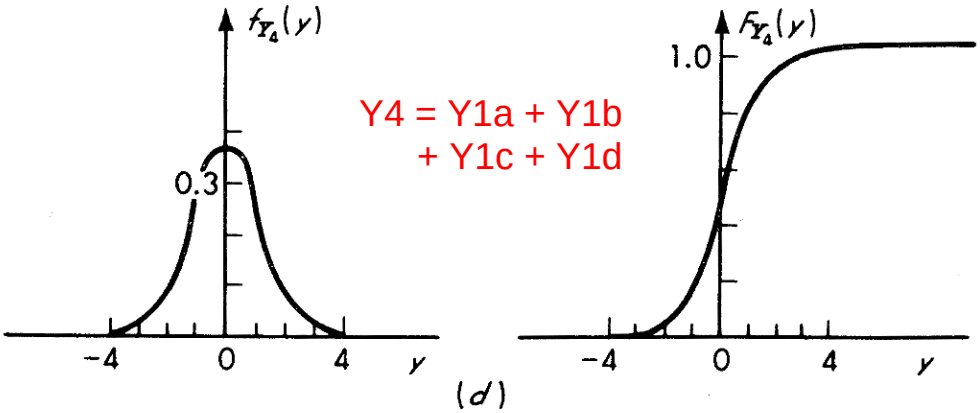
(b)

Y3 = Y1a + Y1b + Y1c



(c)

Y4 = Y1a + Y1b
+ Y1c + Y1d



(d)

Expected Value and Variance of a Sum of Random Variables

P.G. Ioannou November 7, 2007

Expected Value is a Linear Operator

(1) $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$

(2) $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$

The above is true even if $X_1, X_2, \ldots, X_n$ are not independent

Variance is Not a Linear Operator

If the random variables $X_1, X_2, \ldots, X_n$ are independent (or uncorrelated) then

(3) $Var[X_1 + X_2 + \ldots + X_n] = Var[X_1] + Var[X_2] + \ldots + Var[X_n]$

Otherwise,

(4)
$$\begin{aligned} Var[\sum_{i=1}^{n} X_i] &= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov[X_i, X_j] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{X_i X_j} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{X_i X_j}\sigma_{X_i}\sigma_{X_j} \\ &= \sum_{i=1}^{n} Var[X_i] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} Cov[X_i, X_j] \\ &= \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \rho_{X_i X_j}\sigma_{X_i}\sigma_{X_j} \end{aligned}$$

Sum all elements of a nxn square matrix

Cov[X1,X1] Cov[X1,X2].....Cov[X1,Xn]
Cov[X2,X1] Cov[X2,X2].....Cov[X2,Xn]
.....
.....
Cov[Xn,X1] Cov[Xn,X2].....Cov[Xn,Xn]

where ρ_{XY} is called the correlation coefficient

(5) $\rho_{XY} = \frac{Cov[X,Y]}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$ $\qquad -1 \le \rho_{XY} \le 1$

If X and Y are mutually independent then $\rho_{XY} = 0$

If $(0 < \rho_{XY} \le +1)$ then X and Y are positively correlated

If $(-1 \le \rho_{XY} < 0)$ then X and Y are negatively correlated

TABLE 6.3 **NORMAL DISTRIBUTION**

Normal	$N(\mu,\sigma^2)$
Possible applications	Errors of various types, e.g., in the impact point of a bomb; quantities that are the sum of a large number of other quantities (by virtue of central limit theorems)
Density (see Fig. 6.5)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ for all real numbers x
Distribution	No closed form
Parameters	Location parameter $\mu \in (-\infty,\infty)$, scale parameter $\sigma > 0$
Range	$(-\infty,\infty)$
Mean	μ
Variance	σ^2
Mode	μ
MLE	$\hat{\mu} = \bar{X}(n)$, $\hat{\sigma} = \left[\frac{n-1}{n} S^2(n)\right]^{1/2}$
Comments	1. If two jointly distributed normal random variables are uncorrelated, they are also independent. For distributions other than normal, this implication is not true in general

TABLE 6.3 (*continued*)

Normal	$N(\mu,\sigma^2)$

2. Suppose that the joint distribution of $X_1, X_2, \ldots, X_m$ is multivariate normal and let $\mu_i = E(X_i)$ and $C_{ij} = \mathrm{Cov}(X_i, X_j)$. Then for any real numbers $a, b_1, b_2, \ldots, b_m$, the random variable $a + b_1X_1 + b_2X_2 + \cdots + b_mX_m$ has a normal distribution with mean $\mu = a + \sum_{i=1}^{m} b_i\mu_i$ and variance

$$\sigma^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} b_i b_j C_{ij}$$

Note that we need *not* assume independence of the X_i's. If the X_i's *are* independent, then

$$\sigma^2 = \sum_{i=1}^{m} b_i^2 \mathrm{Var}(X_i)$$

3. The N(0,1) distribution is often called the *standard* or *unit normal distribution*
4. If $X_1, X_2, \ldots, X_k$ are independent standard normal random variables, then $X_1^2 + X_2^2 + \cdots + X_k^2$ has a chi-square distribution with k df, which is also the gamma$(k/2,2)$ distribution
5. If $X \sim N(\mu,\sigma^2)$, then e^X has the *lognormal distribution* with parameters μ and σ, denoted $LN(\mu,\sigma^2)$
6. If $X \sim N(0,1)$, if Y has a chi-square distribution with k df, and if X and Y are independent, then $X/\sqrt{Y/k}$ has a t distribution with k df (sometimes called *Student's t distribution*)
7. If the normal distribution is used to represent a nonnegative quantity (e.g., time), then its density should be truncated at $x = 0$ (see Sec. 6.8)
8. As $\sigma \to 0$, the normal distribution becomes degenerate at μ

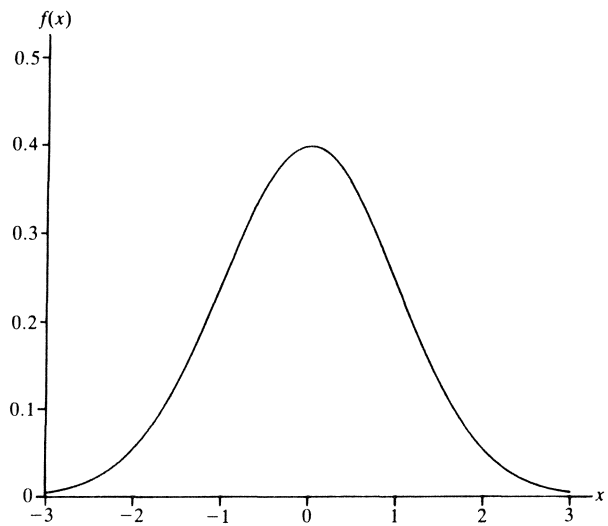


FIGURE 6.5
N(0,1) density function.

TABLE 6.3 (*continued*)

BETA DISTRIBUTION

Beta	**beta(α_1,α_2)**
Possible applications	Used as a rough model in the absence of data (see Sec. 6.9); distribution of a random proportion, such as the proportion of defective items in a shipment; time to complete a task, e.g., in a PERT network
Density (see Fig. 6.7)	$f(x)=\begin{cases}\dfrac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)} & \text{if } 0<x<1\\ 0 & \text{otherwise}\end{cases}$ where $B(\alpha_1,\alpha_2)$ is the *beta function*, defined by $$B(z_1,z_2)=\int_0^1 t^{z_1-1}(1-t)^{z_2-1}\,dt$$ for any real numbers $z_1>0$ and $z_2>0$. Some properties of the beta function: $$B(z_1,z_2)=B(z_2,z_1),\qquad B(z_1,z_2)=\frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1+z_2)}$$
Distribution	No closed form, in general. If either α_1 or α_2 is a positive integer, a binomial expansion can be used to obtain $F(x)$, which will be a polynomial in x, and the powers of x will be, in general, positive real numbers ranging from 0 through $\alpha_1+\alpha_2-1$
Parameters	Shape parameters $\alpha_1>0$ and $\alpha_2>0$
Range	[0,1]
Mean	$\dfrac{\alpha_1}{\alpha_1+\alpha_2}$
Variance	$\dfrac{\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}$
Mode	$\begin{cases}\dfrac{\alpha_1-1}{\alpha_1+\alpha_2-2} & \text{if } \alpha_1>1,\ \alpha_2>1\\ 0 \text{ and } 1 & \text{if } \alpha_1<1,\ \alpha_2<1\\ 0 & \text{if } (\alpha_1<1,\ \alpha_2\ge 1) \text{ or if } (\alpha_1=1,\ \alpha_2>1)\\ 1 & \text{if } (\alpha_1\ge 1,\ \alpha_2<1) \text{ or if } (\alpha_1>1,\ \alpha_2=1)\\ \text{does not uniquely exist} & \text{if } \alpha_1=\alpha_2=1\end{cases}$
MLE	The following two equations must be satisfied: $$\Psi(\hat\alpha_1)-\Psi(\hat\alpha_1+\hat\alpha_2)=\ln G_1,\qquad \Psi(\hat\alpha_2)-\Psi(\hat\alpha_1+\hat\alpha_2)=\ln G_2$$ where Ψ is the digamma function, $G_1=(\Pi_{i=1}^n X_i)^{1/n}$, and $G_2=[\Pi_{i=1}^n(1-X_i)]^{1/n}$ [see Gnanadesikan, Pinkham, and Hughes (1967)]; note that $G_1+G_2\le 1$. These equations could be solved numerically [see Beckman and Tietjen (1978)], or approximations to $\hat\alpha_1$ and $\hat\alpha_2$ can be obtained from Table 6.20 (see App. 6A), which was computed for particular (G_1,G_2) pairs by modifications of the methods in Beckman and Tietjen (1978)
Comments	1. The U(0,1) and beta(1,1) distributions are the same 2. If X_1 and X_2 are independent random variables with $X_i\sim$ gamma(α_i,β), then $X_1/(X_1+X_2)\sim$ beta(α_1,α_2) 3. A beta random variable X on [0,1] can be rescaled and relocated to obtain a beta random variable on $[a,b]$ of the same shape by the transformation $a+(b-a)X$

TABLE 6.3 (*continued*)

Beta	**beta(α_1,α_2)**
	4. $X\sim$ beta(α_1,α_2) if and only if $1-X\sim$ beta(α_2,α_1) 5. $X\sim$ beta(α_1,α_2) if and only if $Y=X/(1-X)$ has a Pearson type VI distribution with shape parameters α_1, α_2 and scale parameter 1, denoted PT6(α_1,α_2,1) 6. The beta(1,2) density is a left triangle, and the beta(2,1) density is a right triangle 7. $$\lim_{x\to 0} f(x)=\begin{cases}\infty & \text{if } \alpha_1<1\\ \alpha_2 & \text{if } \alpha_1=1\\ 0 & \text{if } \alpha_1>1\end{cases},\qquad \lim_{x\to 1} f(x)=\begin{cases}\infty & \text{if } \alpha_2<1\\ \alpha_1 & \text{if } \alpha_2=1\\ 0 & \text{if } \alpha_2>1\end{cases}$$ 8. The density is symmetric about $x=\frac{1}{2}$ if and only if $\alpha_1=\alpha_2$. Also, the mean and the mode are equal if and only if $\alpha_1=\alpha_2$

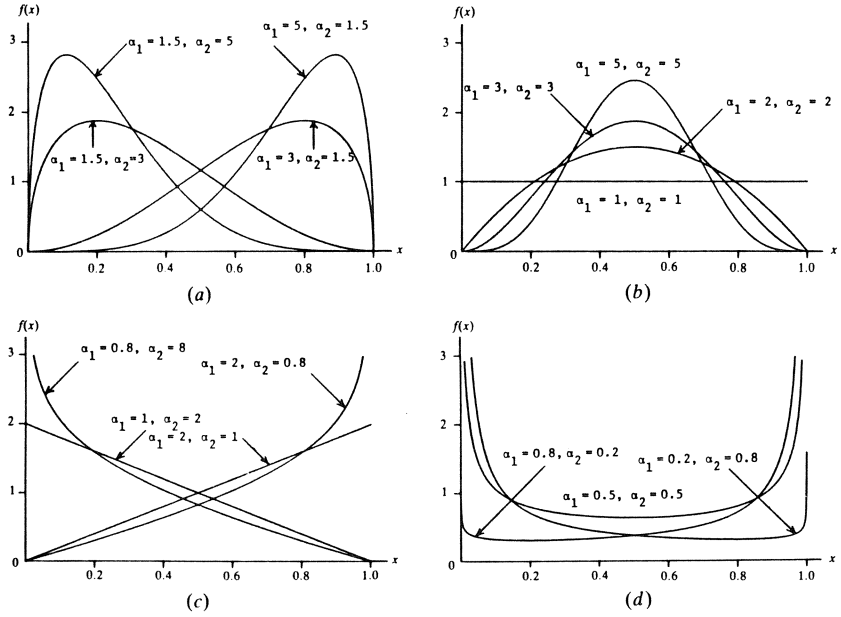
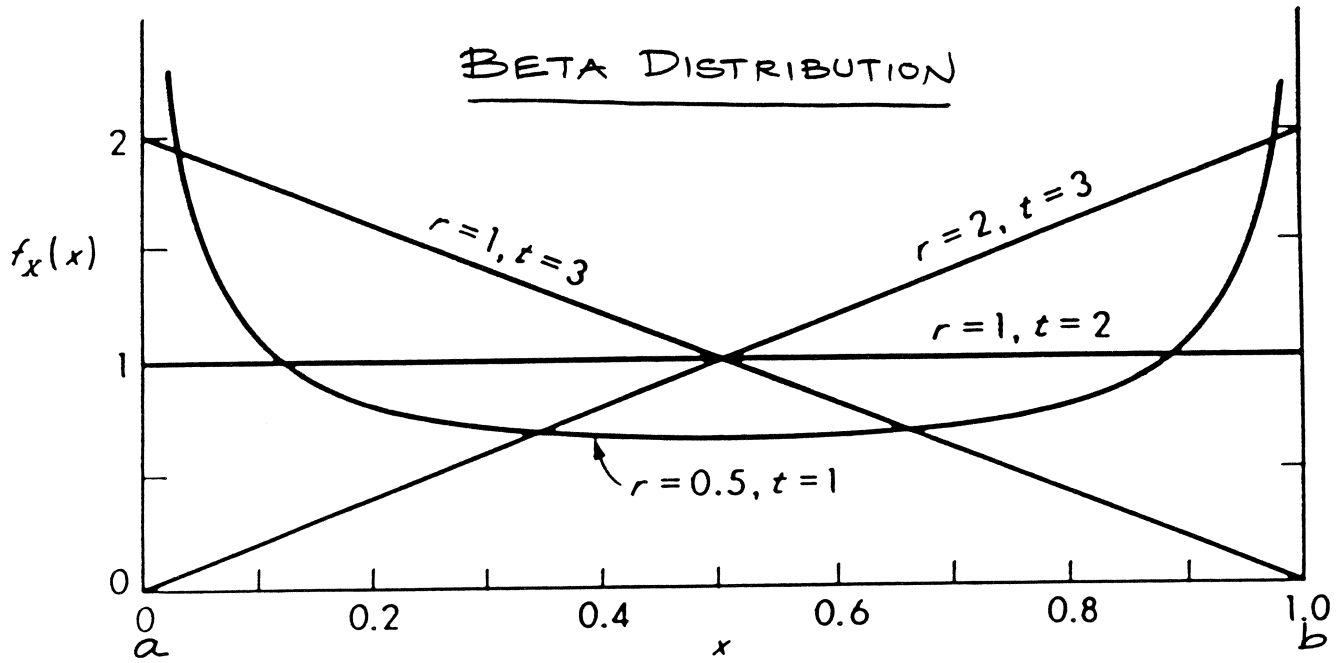


FIGURE 6.7
beta(α_1,α_2) density functions.



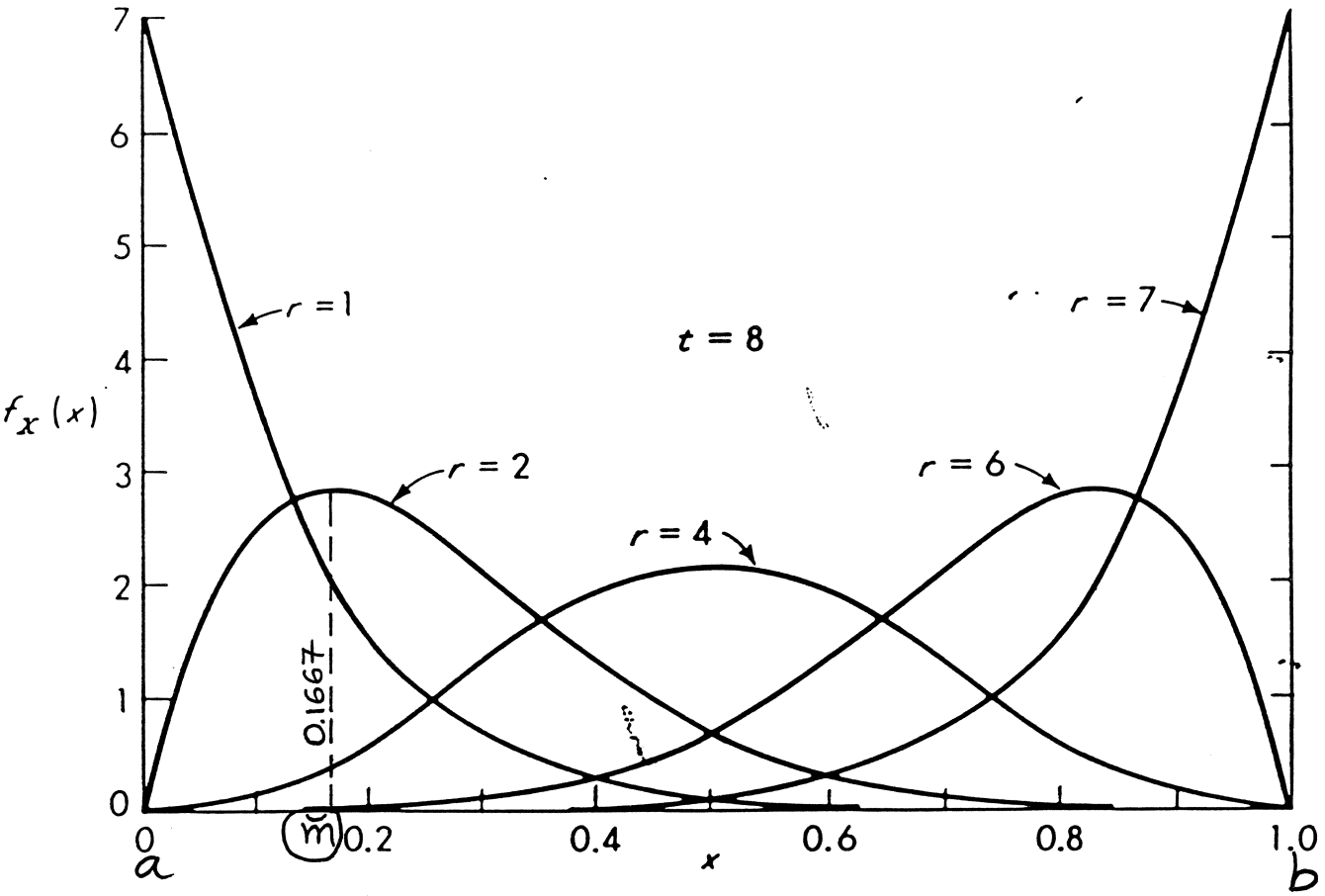


Fig. 3.4.1 Beta-distribution shapes.

$$f_X(x) = \frac{1}{B(b-a)^{t-1}} (x-a)^{r-1} (b-x)^{t-r-1} \qquad a \le x \le b$$

$$E[X] = m_X = a + \frac{r}{t}(b-a)$$

$$\mathrm{Var}[X] = \sigma_X^2 = (b-a)^2 \frac{r(t-r)}{t^2(t+1)}$$

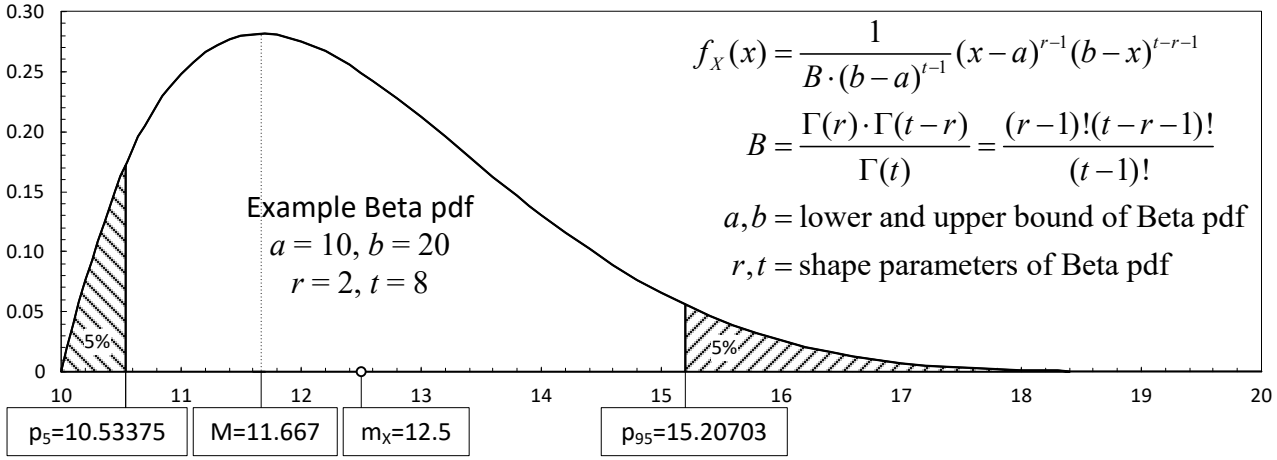
$$\breve{m} = M = \text{mode} = a + (b-a)\frac{r-1}{t-2} \quad \text{if } r>1 \text{ and } t>2$$

$$B = \frac{(r-1)!\,(t-r-1)!}{(t-1)!} \qquad r, t = \text{integers}$$

$$= \frac{\Gamma(r)\,\Gamma(t-r)}{\Gamma(t)} \qquad \text{in general}$$

$r = a_1$	$a_1 = r$
$t = a_1 + a_2$	$a_2 = t - r$

APPROXIMATION FORMULAS FOR THE MOMENTS OF THE BETA DISTRIBUTION



$$f_X(x) = \frac{1}{B \cdot (b-a)^{t-1}} (x-a)^{r-1} (b-x)^{t-r-1}$$

$$B = \frac{\Gamma(r) \cdot \Gamma(t-r)}{\Gamma(t)} = \frac{(r-1)!(t-r-1)!}{(t-1)!}$$

a, b = lower and upper bound of Beta pdf

r, t = shape parameters of Beta pdf

Example Beta parameters **Exact Moments of the Beta Distribution**

$$\left.\begin{array}{l} a = 10 \\ b = 20 \\ r = 2 \\ t = 8 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} m_X = a + \frac{r}{t}(b-a) = 10 + \frac{2}{8}(20-10) = 12.5 \\ \sigma_X^2 = (b-a)^2 \frac{r(t-r)}{t^2(t+1)} = (20-10)^2 \frac{2(8-2)}{8^2(8+1)} = 2.083 \\ \sigma_X = \sqrt{\sigma_X^2} = \sqrt{2.083} = 1.4434 \\ M = a + (b-a)\frac{r-1}{t-2} = 10 + (20-10)\frac{2-1}{8-2} = 11.667 \end{array}\right.$$

A. Original PERT Formulas **Application to Example Beta pdf**

$$\hat{m}_X = \frac{a + 4M + b}{6} \qquad \hat{m}_X = \frac{10 + 4 \times 11.667 + 20}{6} = 12.778 \quad (\approx m_X = 12.5)$$

$$\hat{\sigma}_X = \frac{b-a}{6} \qquad \hat{\sigma}_X = \frac{20-10}{6} = 1.67 \quad (\approx \sigma_X = 1.44)$$

B. Perry & Greig Formulas **Application to Example Beta pdf**

$$\hat{m}_X = \frac{p_5 + 0.95M + p_{95}}{2.95} \qquad \hat{m}_X = \frac{10.53375 + 0.95 \times 11.667 + 15.20703}{2.95} = 12.48 \quad (\approx m_X = 12.5)$$

$$\hat{\sigma}_X = \frac{p_{95} - p_5}{3.25} \qquad \hat{\sigma}_X = \frac{15.20703 - 10.53375}{3.25} = 1.438 \quad (\approx \sigma_X = 1.44)$$

(*) A symbol with a hat "^" on top is an *estimator* of the quantity under the hat.

This example illustrates that,

- Given accurate estimates, a, M, and b, the PERT approximations are quite close to the exact values.
- Given accurate estimates, p_5, M, and p_{95}, the Perry and Greig approximations are almost exact.

P.G. Ioannou, 10/14/2020

MANAGEMENT SCIENCE
Vol. 21, No. 12, August 1975
Printed in U.S.A.

ESTIMATING THE MEAN AND VARIANCE OF SUBJECTIVE DISTRIBUTIONS IN PERT AND DECISION ANALYSIS†

C. PERRY AND I. D. GREIG‡

The University of New England, New South Wales

1. Introduction

The original PERT formulae [5] employ the mode and endpoints to estimate the mean and standard deviation of subjective probability distributions.[1] Though widely-used, they have met with criticism for their inaccuracy, and for their being limited to the beta distribution when there is no a priori reason why the distribution should be beta ([4], [6]). This note presents alternative, accurate and relatively distribution-free formulae for estimating the means and standard deviations of subjective probability distributions. The formulae use the 5 and 95 percentiles together with the mode. Other proposed alternatives to the original PERT formulae are critically examined, and alternatives which employ the median in place of the mode are noted.

2. Estimating the Standard Deviation

Estimating the variance of subjective probability distributions by direct means is accepted as being difficult ([1], [2]). While Clarke [1] originally used endpoints or "extremes" for this purpose, the 5 and 95 percentiles have subsequently been found to provide better estimates of the variance ([6], [7]). This may be because subjects will never generally have experienced the extremes, or known it if they have. It may also be difficult for subjects to remove "acts of God" from the other more "usual" stochastic influences in their experience, as required by the assumption of a beta distribution [4]. So Moder and Rogers [6] suggested using the 5 and 95 percentiles (p_5 and p_{95}) to find the standard deviation by the relatively distribution-free approximation of (1), but with $d = 3.20$.

$$\sigma = (p_{95} - p_5)/d, \qquad d = 3.25. \tag{1}$$

However, in uncertain situations subjective probability distributions by their very nature are unlikely to show nearby values with greatly differing probabilities of occurrence; that is, they will tend to be rounded in shape rather than peaked, or platykurtic. In this situation a more appropriate value of d is the 3.25 shown.[2] Pearson

* All Notes are refereed.

† Processed by Professor Burton V. Dean, Departmental Editor for Research and Planning; received April 1974, revised October 1974. This paper was with the authors 2 months for revision.

‡ The authors wish to thank Jock Anderson, John Dillon, Bill Griffiths and John Phillips for aid and comments.

[1] The original PERT formulae are:

$$\mu = (p_0 + 4m + p_{100})/6, \quad \sigma = (p_{100} - p_0)/6$$

where p_n is the nth percentile, and m is the mode.

[2] Consider Figure 2 in [7, p. 537] where appropriate denominators are shown for the general system of Pearson curves, defined by β_1 and β_2. The third shape factor (skewness) and the fourth shape factor (kurtosis) of Pearson curves are $\sqrt{\beta_1}$ and β_2 respectively. Thus for the relatively unskewed and platykurtic distributions in the north-west corner of the figure (for bell-shaped Pearson curves), the denominator for equation (1) is 3.25.

and Tukey [7, p. 539] present a method to refine this estimate of σ if the median is known.

3. Estimating the Mean

We have, by experimentation, found a formula for estimating the mean using a weighted average of the mode (m) and the 5 and 95 percentiles. This formula is very accurate and is sufficiently distribution-free to justify its general application, as will be shown.

$$\mu = (p_5 + 0.95\, m + p_{95})/2.95. \tag{2}$$

To derive this formula we worked with a wide variety of shapes of the standardised forms of three distributions commonly used in management science: the beta, the gamma (including the chisquare) and the lognormal. We found the compromise modal weight of 0.95 gave quite small errors for each of the distributions.

To calculate the errors for each distribution and shape, the 5 and 95 percentiles, the mode, and the true mean and standard deviation were taken variously from published formulas and tables and the first three used in (2) to estimate a mean. This was then compared to the true mean and the difference expressed as a percentage of the standard deviation. For distributions that are not extremely skewed, the estimation error is less than 5 percent, and is usually around 1 or 2 percent.

More precisely, errors less than 5 percent can be expected for all non-U beta distributions, all gamma distributions, and lognormal distributions with $\sigma \leqq 0.13$ (in the notation of [3]). Thus, any skew is allowed in the beta and gamma, and reasonable

TABLE 1

Absolute Errors in the Estimation of the Variance (Expressed as a Percentage of the Range) of the Beta Distribution

a \ b =	1	2	3	4	5	9	11	14	19	29	59
1	5.0	4.0	3.2	2.6	2.1	1.1	0.8	0.6	0.4	0.2	0.0
2		3.6	3.1	2.6	2.2	1.3	1.0	0.7	0.5	0.2	0.1
3			2.8	2.5	2.2	1.4	1.1	0.8	0.6	0.3	0.1
4				2.3	2.1	1.4	1.2	0.9	0.6	0.3	0.1
5					1.9	1.4	1.2	0.9	0.7	0.4	0.1
9						1.2	1.1	0.9	0.7	0.5	0.2
11							1.0	0.9	0.7	0.5	0.2
14	Original PERT							0.8	0.7	0.5	0.2
19	formula								0.6	0.5	0.2
29										0.4	0.2
59											0.2
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2		0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3			0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
4				0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
5					0.1	0.0	0.0	0.0	0.0	0.0	0.0
9						0.0	0.0	0.0	0.0	0.0	0.0
11							0.0	0.0	0.0	0.0	0.0
14	Alternative formula,							0.0	0.0	0.0	0.0
19	equation (1)								0.0	0.0	0.0
29										0.0	0.0
59											0.0

skew in the lognormal. In cases of marked skew the lognormal may be an inappropriate representation of subjective probability distributions for it becomes leptokurtic with skew at an increasing rate [3, p. 118]. The estimating formula is precise for all symmetric distributions such as the normal and the uniform. Thus most subjective probability distribution shapes will have been represented.

Another rather similar formula [6] advanced as an alternative to the original PERT formula is:

$$\mu = (p_5 + 4m + p_{95})/6. \tag{3}$$

For beta curves not showing extreme skew or kurtosis we found this formula gave errors in estimation up to four times as large as did the original PERT formula.

For those who do not share the preference for the mode indicated by the PERT formula's wide adoption, the median (p_{50}) as a measure of central location provides a more accurate estimate of the mean. This is achieved [7] through the extraordinarily accurate, distribution-free,

$$\mu = p_{50} + 0.185\ (p_{95} + p_5 - 2p_{50}), \tag{4}$$

which gives errors less than 0.5 percent for all bell-shaped distributions (except for cases of extreme leptokurtosis) and most often errors less than 0.1 percent.

4. Comparison of PERT and Alternative Formulae

As evidence of the total superiority of the alternative distribution-free formulae (1) and (2) over the original PERT formulae, Tables 1 and 2 show their respective ab-

TABLE 2

Absolute Errors in the Estimation of the Mean (Expressed as Percentage of the Range) of the Beta Distribution

a \ b =	1	2	3	4	5	9	11	14	19	29	59
1	0	1.1	0	1.4	2.8	6.7	7.9	9.3	10.9	12.6	14.6
2		0	0.5	1.4	2.4	5.7	6.9	8.3	10.0	11.9	14.1
3			0	0.8	1.7	4.8	6.0	7.4	9.1	11.2	13.6
4				0	0.8	3.8	5.0	6.5	8.3	10.5	13.2
5					0	3.0	4.2	5.6	7.5	9.8	12.8
9						0	1.2	2.8	4.8	7.5	11.2
11							0	1.6	3.6	6.4	10.5
14	Original PERT							0	2.1	5.0	9.5
19	formula								0	3.1	7.9
29										0	5.3
59											0
1	0	0.5	0.4	0.3	0.2	0.0	0.1	0.6	0.1	0.1	0.1
2		0	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1
3			0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1
4				0	0.0	0.1	0.1	0.1	0.1	0.1	0.1
5					0	0.0	0.1	0.1	0.1	0.1	0.1
9						0	0.0	0.0	0.1	0.1	0.1
11							0	0.0	0.0	0.1	0.1
14	Alternative formula,							0	0.0	0.0	0.0
19	equation (2)								0	0.0	0.0
29										0	0.0
59											0

solute errors for a wide range of shapes of the beta distribution (the only distribution considered to which the original formulae apply). Note the errors in estimation are expressed as a percentage of the *range* (e.g. as in [4]). The a and b of the tables refer to the beta distribution parameters. Skewness increases with the difference between a and b, and peakedness as a and b increase together. Errors for negative skew are not shown but are the mirror image of those values given. It can be seen that the alternative formulae are more accurate for virtually every shape.

5. Application

For practical purposes the nature of the underlying distribution can be ignored when applying the alternative formulae (1) and (2). They should find application wherever the procedures they replace have been used. A further use might be in decision analysis when finding the expected utility of a consequence. The basis of the procedure is that percentiles of a transformed distribution can be found by simply transforming the percentiles of the original distribution [7, p. 533], while the moments and the mode are not equivalently transformed. From estimates of p_5, m and p_{95}, the mean of the consequence is found by (2), and then the median found (if not elicited directly) by solving (4). Now that p_5, p_{50} and p_{95} of the consequence are known, their equivalents in the transformed distribution are simply found by transforming them by the utility function. Finally, expected utility is obtained by substituting the transformed percentiles in (4). Such a procedure would avoid the common practice of treating a continuous variable as discrete.

References

1. Clark, C. E., "The PERT Model for the Distribution of an Activity Time," *Operations Research*, Vol. 10, No. 3, (1962), pp. 405–406.
2. Hogarth, R. M., *Cognitive Processes and the Assessment of Subjective Probability Distributions*, ISEAD Research Paper No. 97, Fontainbleu, France, 1973.
3. Johnson, N. L. and Kotz, S., *Continuous Univariate Distributions—1*, Houghton-Mifflin Company, Boston, 1970.
4. MacCrimmon, K. R. and Ryavec, C. A., "An Analytical Study of the PERT Assumptions," *Operations Research*, Vol. 12, No. 1, (1964), pp. 16–37.
5. Malcolm, D. G., Roseboom, J. H., Clark, C. E. and Fazar, W., "Application of a Technique for Research and Development Program Evaluation," *Operations Research*, Vol. 7, No. 5, (1959), pp. 646–669.
6. Moder, J. J. and Rodgers, E. G., "Judgement Estimates of the Moments of PERT Type Distributions," *Management Science*, Vol. 15, No. 2, (October 1968), pp. B76–83.
7. Pearson, E. S. and Tukey, J., "Approximating Means and Standard Deviations Based on Distances Between Percentage Points of Frequency Curves," *Biometrika*, Vol. 52, (1965), pp. 533–546.

Expected Value and Variance of a Sum of Random Variables

P.G. Ioannou November 7, 2007

Expected Value is a Linear Operator

(1) $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$

(2) $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$

The above is true even if $X_1, X_2, \ldots, X_n$ are not independent

Variance is Not a Linear Operator

If the random variables $X_1, X_2, \ldots, X_n$ are independent (or uncorrelated) then

(3) $Var[X_1 + X_2 + \ldots + X_n] = Var[X_1] + Var[X_2] + \ldots + Var[X_n]$

Otherwise,

$$
\begin{aligned}
(4) \quad Var[\sum_{i=1}^{n} X_i] &= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov[X_i, X_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{X_i X_j} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{X_i X_j}\sigma_{X_i}\sigma_{X_j} \\
&= \sum_{i=1}^{n} Var[X_i] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} Cov[X_i, X_j] \\
&= \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \rho_{X_i X_j}\sigma_{X_i}\sigma_{X_j}
\end{aligned}
$$

where ρ_{XY} is called the correlation coefficient

$$
(5) \quad \rho_{XY} = \frac{Cov[X, Y]}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad -1 \leq \rho_{XY} \leq 1
$$

If X and Y are mutually independent then $\rho_{XY} = 0$

If $(0 < \rho_{XY} \leq +1)$ then X and Y are positively correlated

If $(-1 \leq \rho_{XY} < 0)$ then X and Y are negatively correlated

Properties of the Moments of Random Variables

P.G. Ioannou January 31, 2003

Note: Formulas involving summations apply to *discrete* random variables while formulas with integrals apply to *continuous* random variables. The symbols a and b indicate constants.

Expected Value

(1) $E[X] = m_X = \sum_{\text{all } x} x p_X(x)$

(2) $E[X] = m_X = \int_{-\infty}^{+\infty} x f_X(x) dx$

(3) $E[g(X)] = \sum_{\text{all } x} g(x)\, p_X(x)$

(4) $E[g(X)] = \int_{-\infty}^{+\infty} g(x)\, f_X(x) dx$

(5) $E[g(X)] \neq g(E[X])$ (unless $g(X)$ is a linear function of X)

(6) $E[aX] = aE[X]$

(7) $E[b] = b$

(8) $E[aX + b] = aE[X] + b$

(9) $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

(10) $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$

(11) $E[a_1X_1 + a_2X_2 + \ldots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \ldots + a_nE[X_n]$

(12)
$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx \\ &= \int_{-\infty}^{+\infty} x \{ \int_{-\infty}^{+\infty} f_{XY}(x,y) dy \} dx \\ &= \int_{-\infty}^{+\infty} \{ \int_{-\infty}^{+\infty} x f_{X|Y}(x,y) dx \} f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} E[X|Y] f_Y(y) dy \\ &= E[E[X|Y]] \end{aligned}$$

Variance

(13) $Var[X] = \sigma_X^2 = \sum_{\text{all } x} (x - E[X])^2 p_X(x)$

(14) $Var[X] = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x) dx$

(15) $Var[X] = E[(x - E[X])^2]$

(16) $Var[X] = E[X^2] - (E[X])^2$ (easiest way to calculate $Var[X]$)

(17) $Var[aX] = a^2 Var[X]$

(18) $Var[b] = 0$

(19) $Var[aX + b] = a^2 Var[X]$

Standard Deviation

(20) $\sigma_X = SD[X] = \sqrt{Var[X]} = \sqrt{\sigma_X^2}$

Coefficient of Variation

(21) $V_X = \frac{\sigma_X}{m_X}$

Variance — Covariance — Correlation Coefficient

(22) $Cov[X, Y] = \sigma_{XY} = \sum_{\text{all } x} \sum_{\text{all } y} (x - E[X])(y - E[Y]) p_{XY}(x, y)$

(23) $Cov[X, Y] = \sigma_{XY} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E[X])(y - E[Y]) f_{XY}(x, y) dx dy$

(24) $Cov[X, Y] = E[(x - E[X])(y - E[Y])]$

(25) $Cov[X, Y] = E[XY] - E[X]E[Y]$ (easiest way to calculate $Cov[X, Y]$)

(26) $\rho_{XY} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad -1 \leq \rho_{XY} \leq 1$

If X and Y are mutually independent then $\rho_{XY} = 0$

If $(0 < \rho_{XY} \leq +1)$ then X and Y are positively correlated

If $(-1 \leq \rho_{XY} < 0)$ then X and Y are negatively correlated

(27) $Cov[X, a] = 0$

(28) $Cov[X, X] = \sigma_{XX} = Var[X] = \sigma_X^2 \quad (\rightarrow \rho_{XX} = 1)$

(29) $Cov[aX + b, cY + d] = ac\,Cov[X, Y]$

(30) $\rho_{aX+b,cY+d} = \rho_{XY}$

(31) $$\begin{aligned} Var[aX + bY] &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY} \\ &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho_{XY}\sigma_X\sigma_Y \end{aligned}$$

(32) $Var[X \pm Y] = \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY}$

(33) $$\begin{aligned} Var[\sum_{i=1}^{n} X_i] &= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov[X_i, X_j] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{X_i X_j} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{X_i X_j}\sigma_{X_i}\sigma_{X_j} \\ &= \sum_{i=1}^{n} Var[X_i] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i \neq j}}^{n} Cov[X_i, X_j] \\ &= \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \rho_{X_i X_j}\sigma_{X_i}\sigma_{X_j} \end{aligned}$$

$$
\begin{aligned}
(34)\quad Var[\sum_{i=1}^{n} a_i X_i] &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov[X_i, X_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{X_i X_j} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \rho_{X_i X_j} \sigma_{X_i} \sigma_{X_j} \\
&= \sum_{i=1}^{n} a_i^2 Var[X_i] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i \neq j}}^{n} a_i a_j Cov[X_i, X_j] \\
&= \sum_{i=1}^{n} a_i^2 Var[X_i] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_i a_j \rho_{X_i X_j} \sigma_{X_i} \sigma_{X_j}
\end{aligned}
$$

The following two formulas apply if the random variables $X_1, X_2, \ldots, X_n$ are independent:

$$(35)\quad Var[X_1 + X_2 + \ldots + X_n] = Var[X_1] + Var[X_2] + \ldots + Var[X_n]$$

$$(36)\quad Var[a_1 X_1 + a_2 X_2 + \ldots + a_n X_n] = a_1^2 Var[X_1] + a_2^2 Var[X_2] + \ldots + a_n^2 Var[X_n]$$

Moments of X, Y Functions

$$(37)\quad E[g(X,Y)] = \sum_{\text{all } x}\sum_{\text{all } y} g(X,Y)\, p_{XY}(x,y)$$

$$(38)\quad E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(X,Y)\, f_{XY}(x,y)\, dx dy$$

$$(39)\quad E[g(X,Y)] \neq g(E[X], E[Y]) \qquad \text{(unless } g(X,Y) \text{ is a linear function of } X, Y)$$

Moments of a Product

$$(40)\quad E[XY] = E[X]E[Y] + Cov[X,Y]$$

$$(41)\quad E[XY] = E[X]E[Y] \qquad \text{(where } X \text{ \& } Y \text{ are independent)}$$

$$(42)\quad E[\frac{X}{Y}] \neq \frac{E[X]}{E[Y]} \qquad \text{(even if } X \text{ \& } Y \text{ are independent)}$$

$$(43)\quad Var[XY] = m_X^2 \sigma_Y^2 + m_Y^2 \sigma_X^2 + \sigma_X^2 \sigma_Y^2 \qquad \text{(where } X \text{ \& } Y \text{ are independent)}$$

$$(44)\quad V_Z^2 = V_X^2 + V_Y^2 + V_X^2 V_Y^2 \qquad \text{(where } Z = XY, \text{ and } X \text{ \& } Y \text{ are independent)}$$

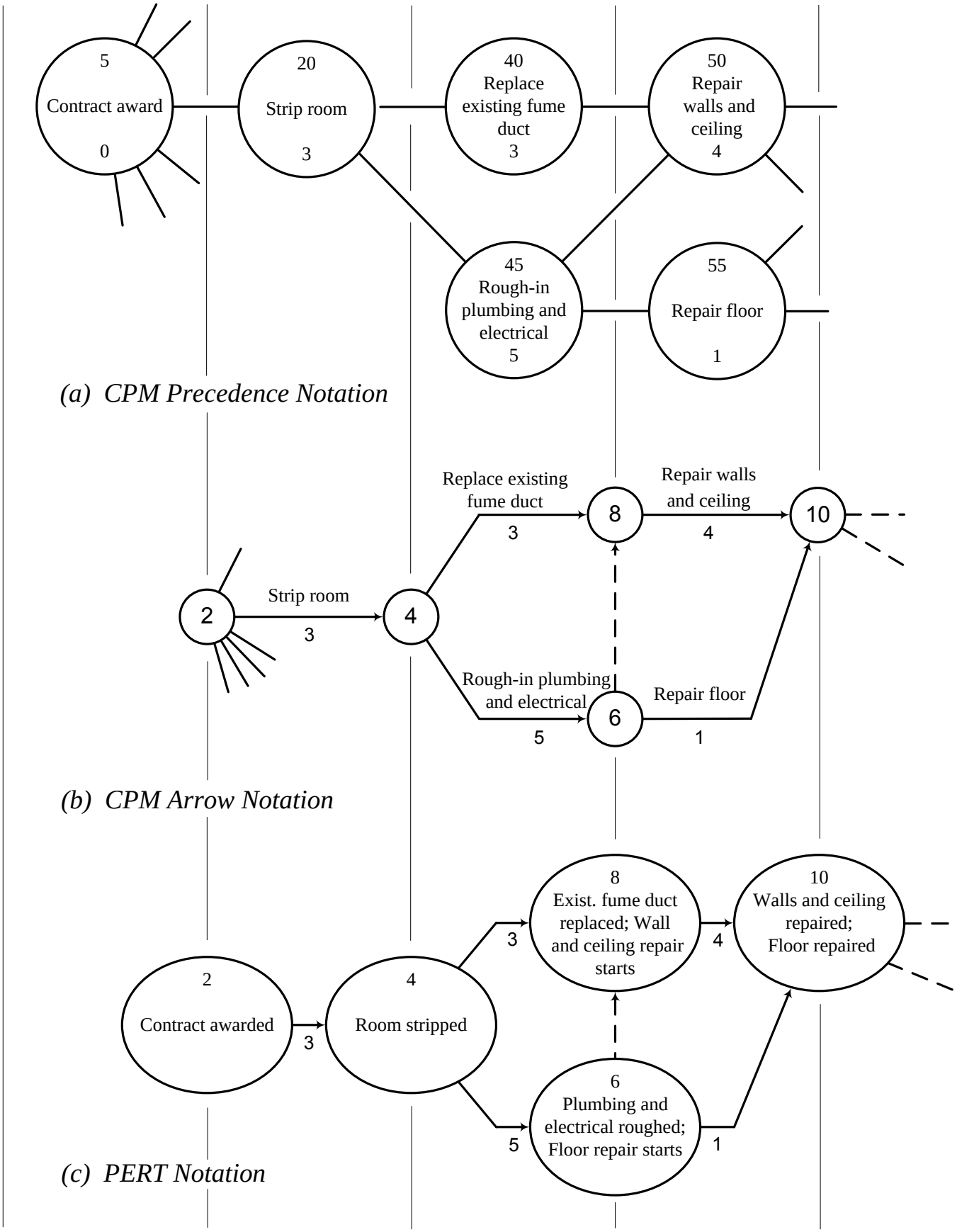


Figure 10.6 Network Comparisons

PERT Approximations

$$E[T_{ij}] \approx t_e = \frac{a + 4M + b}{6}$$

$$SD[T_{ij}] \approx \sigma = \frac{b - a}{6}$$

Path: $1-3-4-5-7$

$$\mu_{Path} = \sum t_e = 26$$

$$\sigma^2_{Path} = \sum \sigma^2 = 9$$

$$\sigma_{Path} = \sqrt{\sigma^2_{Path}} = 3$$

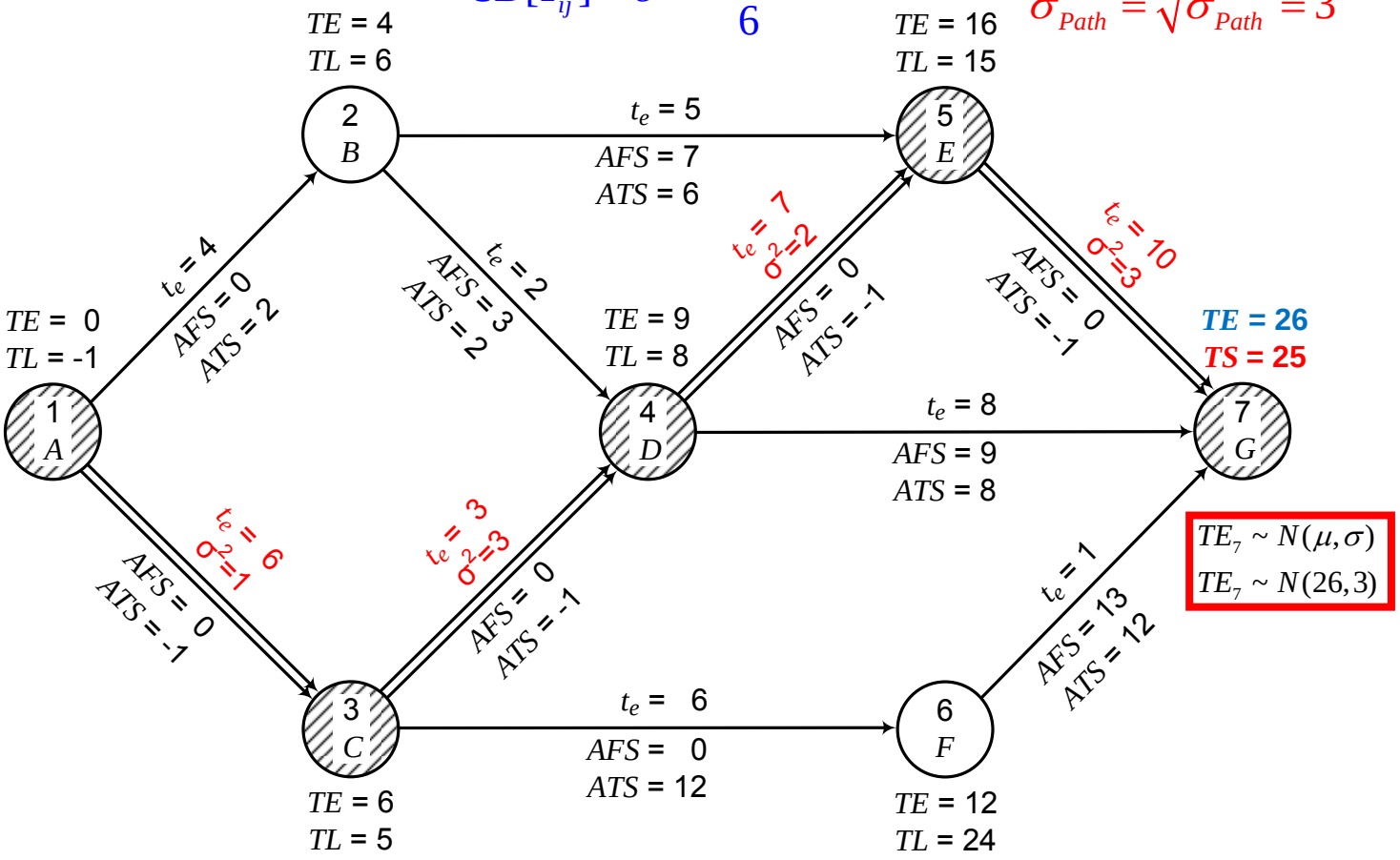


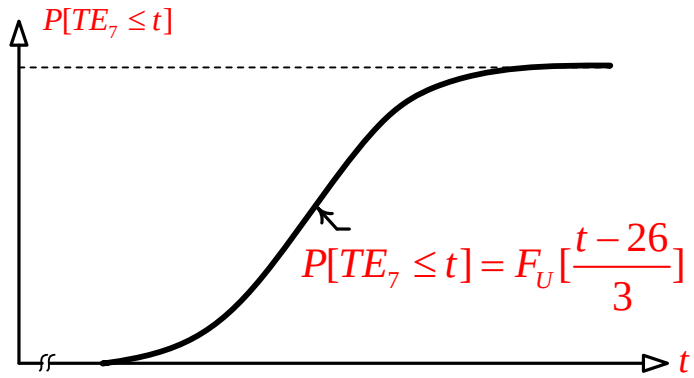
Figure 10.7 Simplified PERT Network

Example :

$$P[TE_7 \le 25] = F_U[\frac{25-26}{3}]$$

$$= F_U[-\frac{1}{3}] = 1 - F_U[\frac{1}{3}] =$$

$$1 - 0.63 = 37\%$$



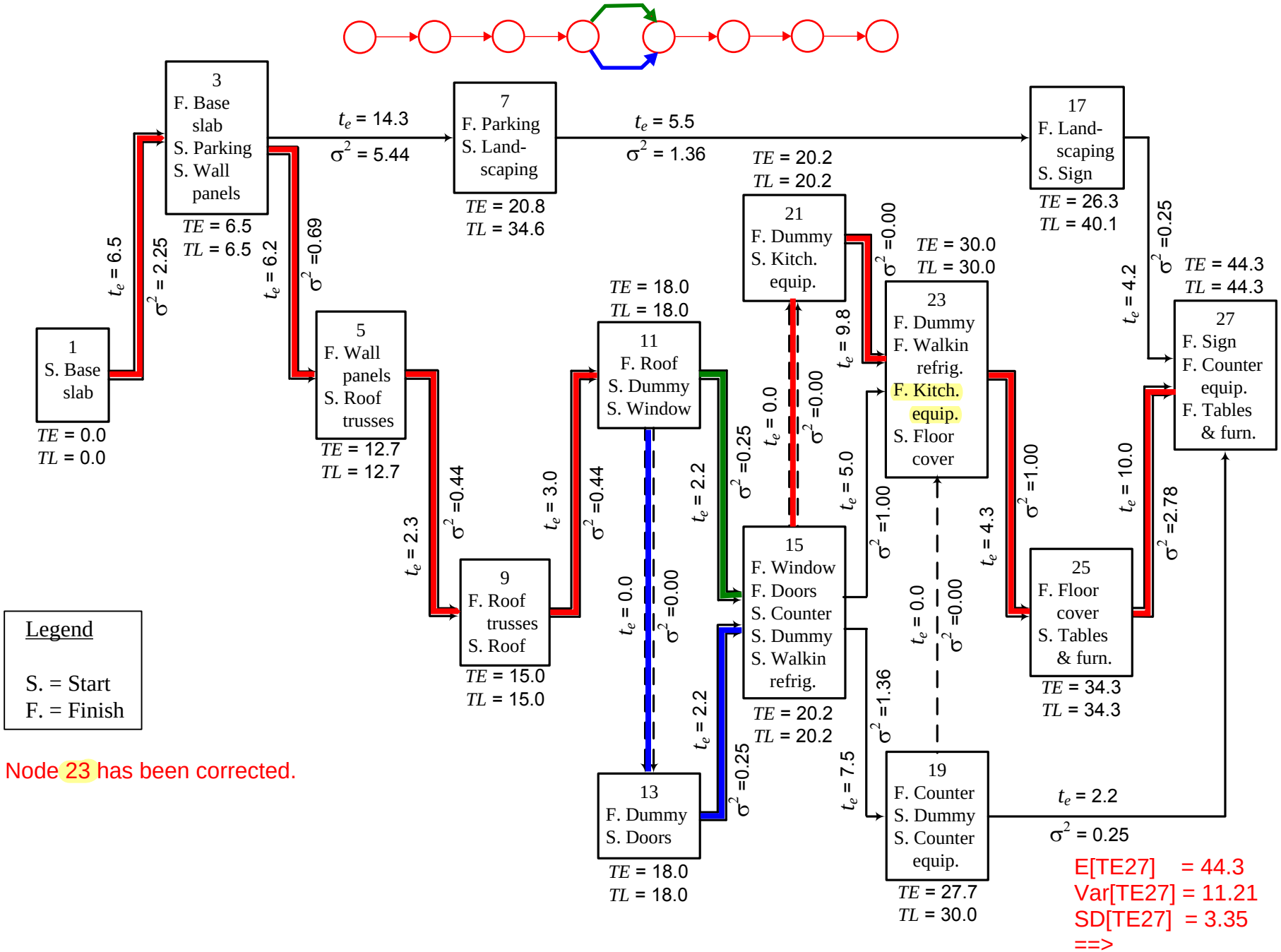


Figure 10.8 Fast Food Outlet Analysis — Typical PERT Network

PERT Approximations

$$E[T_{ij}] \approx t_e = \frac{a + 4M + b}{6}$$

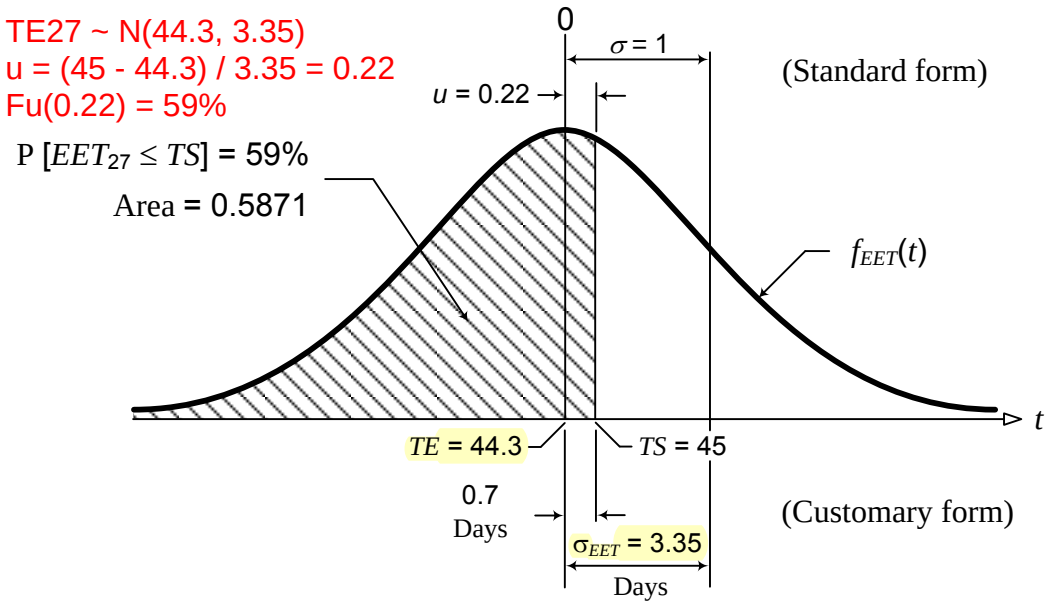
$$SD[T_{ij}] \approx \sigma = \frac{b - a}{6}$$

TABLE 10.3

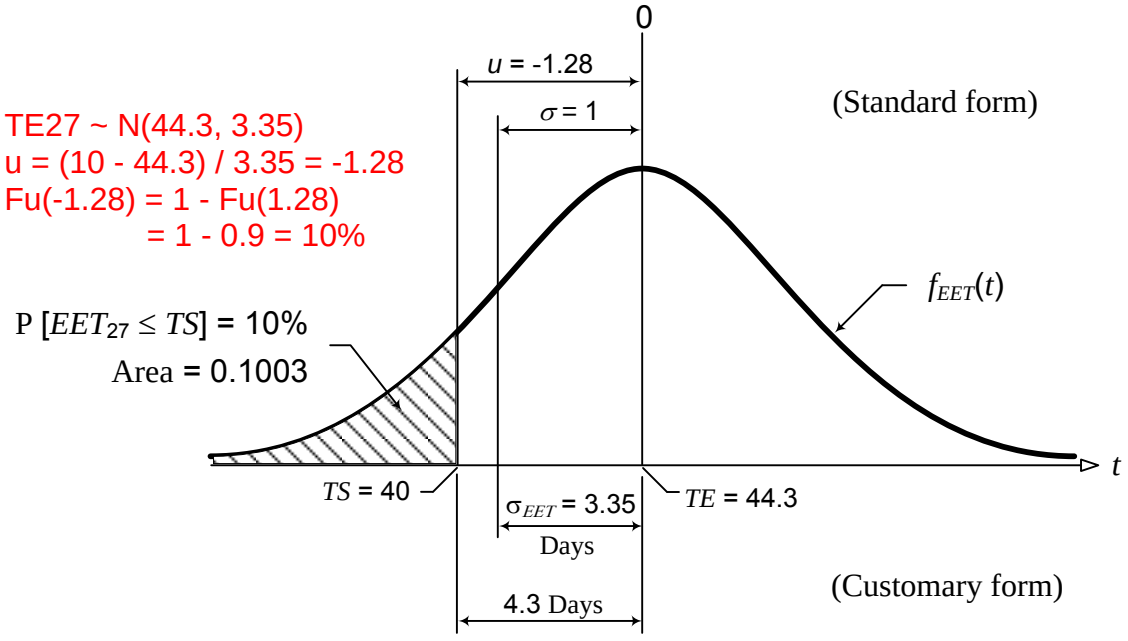
Typical PERT Computations for Fast Food Outlet Analysis

CP	Event		Activity	Durations (Days)			t_e	TE_i	$TE_I + t_e$	$TL_j - t_e$	TL_j	σ^2	*AFS*	*ATS*
	i	*j*		*a*	*m*	*b*								
*	1	3	Base slab	3	6	12	6.5	0.0	6.5	0.0	6.5	2.25	0	0
*	3	5	Wall panels	4	6	9	6.2	6.5	12.7	6.5	12.7	0.69	0	0
	3	7	Parking area	6	15	20	14.3	6.5	20.8	20.3	34.6	5.44	0	13.8
*	5	9	Roof trusses	1	2	5	2.3	12.7	15.0	12.7	15.0	0.44	0	0
	7	17	Landscaping	3	5	10	5.5	20.8	26.3	34.6	40.1	1.36	0	13.8
*	9	11	Roofing	1	3	5	3.0	15.0	18.0	15.0	18.0	0.44	0	0
*	11	13	Dummy	0	0	0	0.0	18.0	18.0	18.0	18.0	0.00	0	0
*	11	15	Windows	1	2	4	2.2	18.0	20.2	18.0	20.2	0.25	0	0
*	13	15	Doors	1	2	4	2.2	18.0	20.2	18.0	20.2	0.25	0	0
	15	19	Counter	3	8	10	7.5	20.2	27.7	22.5	30.0	1.36	0	2.3
*	15	21	Dummy	0	0	0	0.0	20.2	20.2	20.2	20.2	0.00	0	0
	15	23	Walk-in refrigerator	2	5	8	5.0	20.2	25.2	25.0	30.0	1.00	4.8	4.8
	17	27	Sign	3	4	6	4.2	26.3	30.5	40.1	44.3	0.25	13.8	13.8
	19	23	Dummy	0	0	0	0.0	27.7	27.7	30.0	30.0	0.00	2.3	2.3
	19	27	Counter equipment	1	2	4	2.2	27.7	29.9	42.1	44.3	0.25	14.4	14.4
*	21	23	Kitchen equipment	4	10	15	9.8	20.2	30.0	20.2	30.0	3.36	0	0
*	23	25	Floor coverings	2	4	8	4.3	30.0	34.3	30.0	34.3	1.00	0	0
*	25	27	Tables and furnishings	5	10	15	10.0	34.3	44.3	34.3	44.3	2.78	0	0

Var[TE27] = Sum(Var) = 11.21
SD[TE27] = 3.35

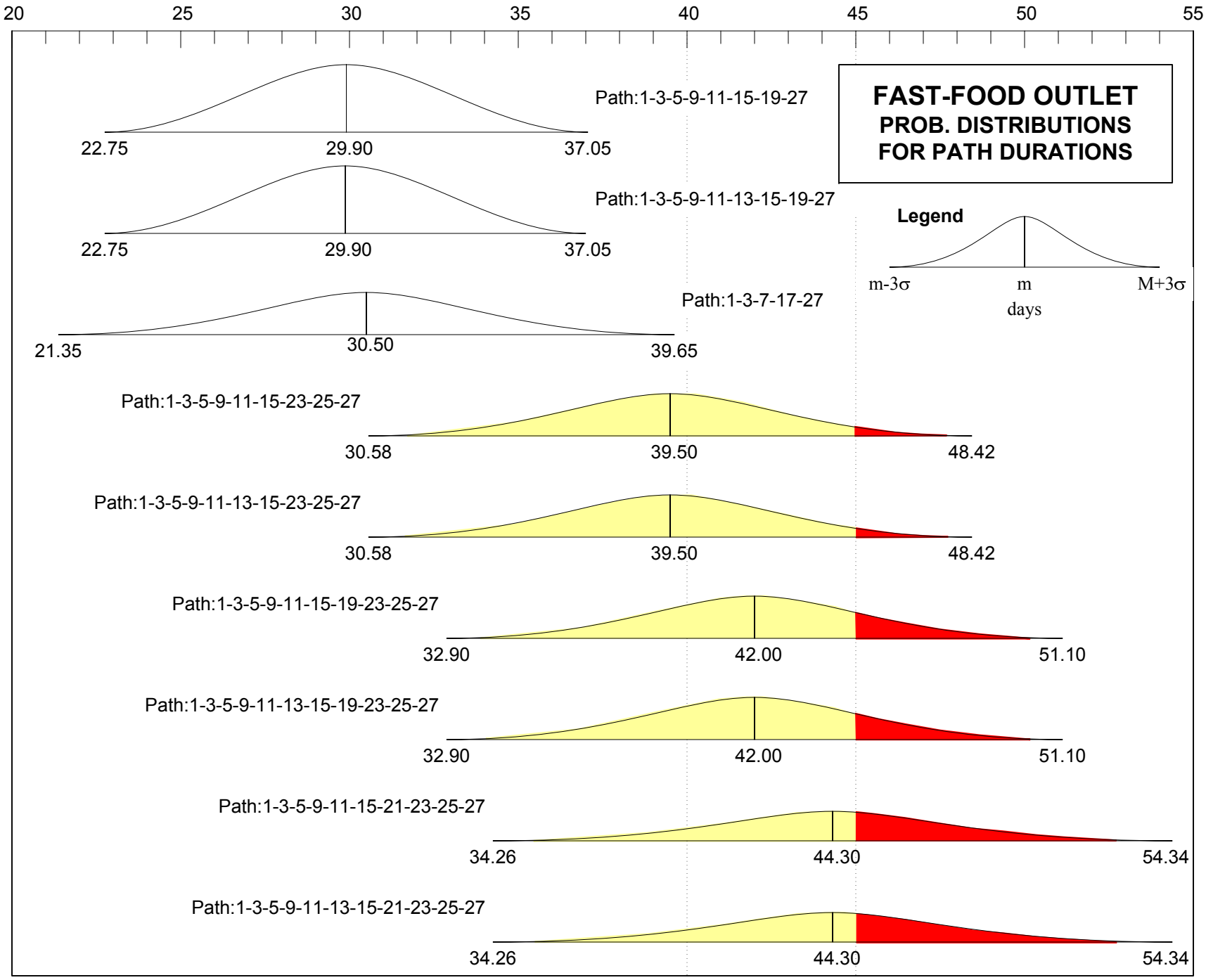


(*a*) Forty-five-day Schedule



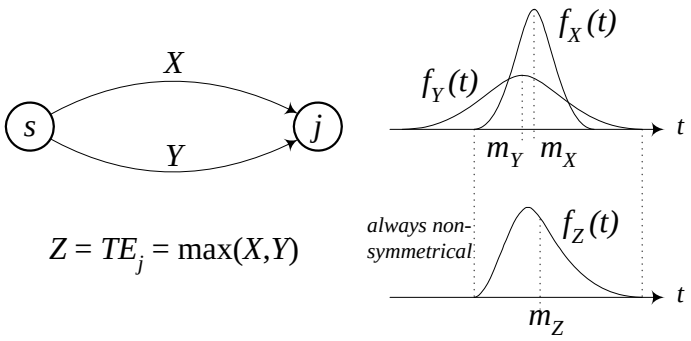
(*b*) Forty-day Schedule

Figure 10.9 Fast Food Outlet Analysis
Probability of meeting 40- and 45-day scheduled durations



P.G. Ioannou, 5/19/97

PERT: MERGE–EVENT BIAS



$Z = TE_j = \max(X,Y)$

Consider two paths, from the start event *S* to event *j*. Let the random variables *X* and *Y* represent the durations of these paths. Let the random variable *Z* represent TE_j.

The CDF of *Z* can be determined as follows ("|" means "given that"):

$$F_Z(t) = P[Z \le t] = P[(X \le t) \cap (Y \le t)] = P[Y \le t \,|\, X \le t]\, P[X \le t]$$

$$F_Z(t) = F_{Y|X \le t}(t)\, F_X(t)$$

P[A ∩ B] = P[B|A] P[A]

If the two paths *X* and *Y* are *independent*, then $F_{Y|X \le t}(t) = F_Y(t)$, and thus

$$F_Z(t) = F_X(t) F_Y(t)$$

In general, however, *X* and *Y* may share resources or depend on the same causal variables (crews, equipment, supervision, management, weather, etc.). The two paths may even share the same activities. Any of these will make X and Y *positively correlated*. I.e.,

$$F_{Y|X \le t}(t) \ge F_Y(t)$$

Thus, positive correlation increases $F_Z(t)$ and makes $F_X(t)F_Y(t)$ a lower limit:

$$F_X(t)F_Y(t) \quad \le \quad F_Z(t) \quad \le \quad \min\{F_X(t), F_Y(t)\}$$

The exact value of $F_Z(z)$ depends on the correlation coefficient ρ_{XY} as follows:

$\rho_{XY} \to 0$ $\quad F_Z(t) \to F_X(t)F_Y(t)$

$\rho_{XY} \to +1$ $\quad F_Z(t) \to \min\{F_X(t), F_Y(t)\}$

Example: P[X<45] = 60%, P[Y<45] = 80%

48% < P[Z<45] < 60%

To determine $P(Z \ge t)$, we typically compute (or find limits for) $P(Z \le t) = F_Z(t)$ first. The limits for $P(Z \ge t) = 1 - F_Z(t)$ are then given by

Example: 48% < P[Z<45] < 60%

40% < P[Z>45] < 52%

$$1 - \min\{F_X(t), F_Y(t)\} \quad \le \quad 1 - F_Z(t) \quad \le \quad 1 - F_X(t)F_Y(t)$$

Computing $P(Z \ge t)$ directly, (i.e., not as $1 - F_Z(t)$), is possible but cumbersome:

$$P[Z \ge t] = P[(X \ge t) \cup (Y \ge t)] = P[X \ge t] + P[Y \ge t] - P[(X \ge t) \cap (Y \ge t)]$$

Clearly, even if X and Y are independent, $P[Z \ge t] \ne P[X \ge t]\, P[Y \ge t]$. Computing $P(Z \ge t)$ as the product $P[X \ge t]P[Y \ge t]$ is a common mistake that should be avoided.

PNET ANALYSIS OF FAST FOOD OUTLET PROJECT

P.G. Ioannou, May 25, 2004

i	j	a	m	b	E[tij]	SD[tij]	Var[tij]
1	3	3	6	12	6.5	1.5	2.25
3	5	4	6	9	6.2	0.8	0.69
3	7	6	15	20	14.3	2.3	5.44
5	9	1	2	5	2.3	0.7	0.44
7	17	3	5	10	5.5	1.2	1.36
9	11	1	3	5	3.0	0.7	0.44
11	13	0	0	0	0.0	0.0	0.00
11	15	1	2	4	2.2	0.5	0.25
13	15	1	2	4	2.2	0.5	0.25
15	19	3	8	10	7.5	1.2	1.36
15	21	0	0	0	0.0	0.0	0.00
15	23	2	5	8	5.0	1.0	1.00
17	27	3	4	6	4.2	0.5	0.25
19	23	0	0	0	0.0	0.0	0.00
19	27	1	2	4	2.2	0.5	0.25
21	23	4	10	15	9.8	1.8	3.36
23	25	2	4	8	4.3	1.0	1.00
25	27	5	10	15	10.0	1.7	2.78

i	Path i (Πi)	E[Ti]	Var[Ti]	SD[Ti]	E[Ti] -3SD[Ti]	E[Ti] +3SD[Ti]
1	1-3-5-9-11-15-21-23-25-27	44.33	11.22	3.35	34.28	54.38
2	1-3-5-9-11-13-15-21-23-25-27	44.33	11.22	3.35	34.28	54.38
3	1-3-5-9-11-15-19-23-25-27	42.00	9.22	3.04	32.89	51.11
4	1-3-5-9-11-13-15-19-23-25-27	42.00	9.22	3.04	32.89	51.11
5	1-3-5-9-11-15-23-25-27	39.50	8.86	2.98	30.57	48.43
6	1-3-5-9-11-13-15-23-25-27	39.50	8.86	2.98	30.57	48.43
7	1-3-7-17-27	30.50	9.31	3.05	21.35	39.65
8	1-3-5-9-11-15-19-27	29.83	5.69	2.39	22.67	36.99
9	1-3-5-9-11-13-15-19-27	29.83	5.69	2.39	22.67	36.99
	Max E[Ti]:	44.33		Range:	34.28	54.38

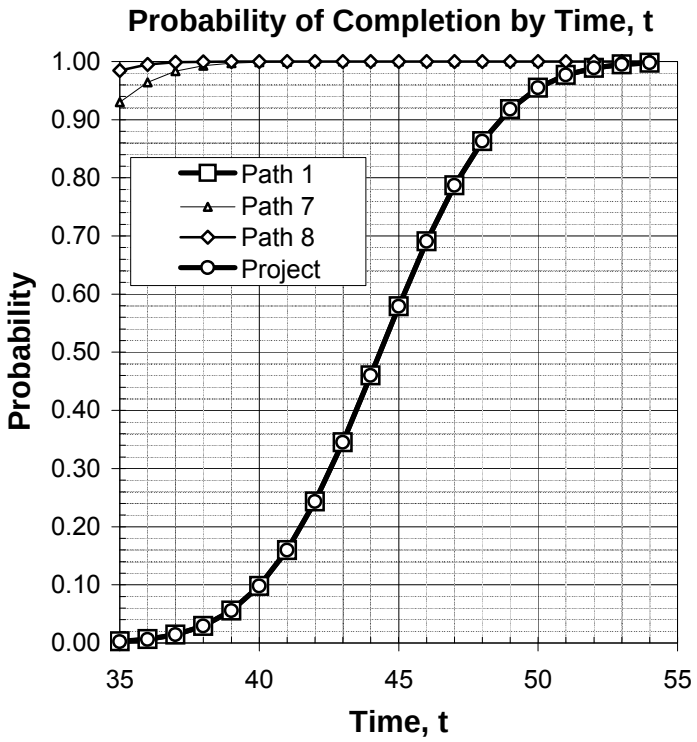
$$\rho_{ij} = \frac{\sum_{k \in (\Pi_i \cap \Pi_j)} \sigma_k^2}{\sigma_{T_i} \sigma_{T_j}}$$

I.e., the numerator is the sum of the Var[tij] of all activities that belong to both Paths i and j.

Paths

1,2,3,4,5,6,7,8,9	ρ12	0.98	Eliminate Paths 2,3,4,5,6.
	ρ13	0.77	They are represented by Path 1
	ρ14	0.75	because ρ1j > 0.5, j=2,3,4,5,6.
	ρ15	0.79	Paths 8 and 9 could also be
	ρ16	0.76	eliminated ρ18, ρ19 ~= 0.5).
	ρ17	0.22	They have been retained
	ρ18	0.51	to illustrate more PNET steps
	ρ19	0.48	as shown below with ρ78, etc.
7,8,9	ρ78	0.31	Cannot eliminate Paths 8 or 9.
	ρ79	0.31	They are not represented by 7.
8,9	ρ89	0.96	Eliminate Path 9, repres. by 8.

*Note that Paths 7,8,9 are "too short" and could have been eliminated before computing ρij to save time and energy. The calculation of ρ1j would then leave just Path 1 and ignore all other paths (i.e., as in the textbook).



t	P[Ti <= t] NORMDIST(t, E[Ti], SD[Ti], 1) i=1	i=7	i=8	PNET P[Project<= t]
(1)	(2)	(3)	(4)	(5) = (2)*(3)*(4)
35	0.00	0.93	0.98	0.00
36	0.01	0.96	1.00	0.01
37	0.01	0.98	1.00	0.01
38	0.03	0.99	1.00	0.03
39	0.06	1.00	1.00	0.06
40	0.10	1.00	1.00	0.10
41	0.16	1.00	1.00	0.16
42	0.24	1.00	1.00	0.24
43	0.35	1.00	1.00	0.35
44	0.46	1.00	1.00	0.46
45	0.58	1.00	1.00	0.58
46	0.69	1.00	1.00	0.69
47	0.79	1.00	1.00	0.79
48	0.86	1.00	1.00	0.86
49	0.92	1.00	1.00	0.92
50	0.95	1.00	1.00	0.95
51	0.98	1.00	1.00	0.98
52	0.99	1.00	1.00	0.99
53	1.00	1.00	1.00	1.00
54	1.00	1.00	1.00	1.00

JOURNAL OF
THE ENGINEERING
MECHANICS DIVISION

ANALYSIS OF ACTIVITY NETWORKS UNDER UNCERTAINTY[a]

By Alfredo H-S. Ang,[1] M. ASCE, Jamal Abdelnour,[2] and Amal A. Chaker[3]

INTRODUCTION

Organizational activity networks are useful in the planning and scheduling of projects, including construction, maintenance and repair, and research and development. The most commonly used method is the well-known *critical-path method* (CPM). The CPM is based on the assumption that the durations of the activities in a project network are deterministic; accordingly, the project duration estimated with CPM would naturally also be deterministic.

In most projects, the durations of the various activities are not always predictable; e.g., in construction, there are many uncontrollable factors such as weather and resource availability that invariably introduce uncertainties into the estimation of the activity durations. As a consequence, the required project durations are difficult to predict with certainty. Properly then the activity durations should be modeled as random variables, and evaluation of the project time considered as a problem of probabilistic network analysis.

Methods such as PERT (5) consider the activity durations as random variables; however, the required project duration is usually determined solely on the basis of the mean critical path (i.e., the network path with the longest expected time). Unfortunately, this invariably underestimates the required completion time for a given network (4). Monte Carlo simulation (8) of course, may be used, e.g., GERTS (6); however, for large networks this would require consider-

Note.—Discussion open until January 1, 1976. To extend the closing date one month, a written request must be filed with the Editor of Technical Publications, ASCE. This paper is part of the copyrighted Journal of the Engineering Mechanics Division, Proceedings of the American Society of Civil Engineers, Vol. 101, No. EM4, August, 1975. Manuscript was submitted for review for possible publication on September 9, 1974.

[a]Presented at the October 21-25, 1974, ASCE Environmental Engineering Meeting, held at Kansas City, Mo. (Preprint 2372).

[1]Prof. of Civ. Engrg., Univ. of Illinois at Urbana-Champaign, Urbana, Ill.

[2]Asst. Vice Pres., Trans Mediterranean Airways, Bierut, Lebanon; formerly Research Asst. in Civ. Engrg., Univ. of Illinois at Urbana-Champaign, Urbana, Ill.

[3]Presidential Asst., Algerian State Oil Co., Alger, Algeria; formerly, Research Asst. in Civ. Engrg., Univ. of Illinois at Urbana-Champaign, Urbana, Ill.

able expense that may not be necessary. For certain special types (reducible) of networks, a "network reduction" method has been proposed (7); however, for more general networks, such as those found in construction, this method would neither be effective nor practical. This limited review of related works serves to indicate that there is considerable interest in and need for probabilistic analysis of activity networks.

A practical method for evaluating activity networks under uncertainty is needed to provide the probabilistic information required for making proper decisions. Developed herein is a simple and practical probabilistic method for evaluating or predicting the completion times of project networks, referred to herein as the *probabilistic network evaluation technique* (PNET) (1).

Probability of Project Completion

Network Paths and Associated Probabilities.—Consider a project network with a specified number of activities (arrow notation is implied throughout). Let $\pi_1, \pi_2, \ldots, \pi_n$ be the set of all possible paths from the start node to the end node of the network and let $T_1, T_2, \ldots, T_n$ be the durations of the respective paths.

Completion of the project obviously requires completion of all the activities, and thus of all the possible paths in the network; then if T is the actual completion time of the project, we have

$$(T \leq t) = (T_1 \leq t) \cap (T_2 \leq t) \cap \ldots \cap (T_n \leq t) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

and by virtue of de Morgan's law

$$(T > t) = (T_1 > t) \cup (T_2 > t) \cup \ldots \cup (T_n > t) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

in which $\cap$ stands for intersection; and $\cup$ for union of events.

From Eq. 1, we see that completion of a project in time t, i.e., the event $(T \leq t)$, will involve all the paths in the network. Theoretically, then, the probability of completing a project in time t, denoted $p(t)$, is

$$p(t) \equiv P(T \leq t) = P(T_1 \leq t, T_2 \leq t, \ldots, T_n \leq t) \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

whereas the probability that the project duration will exceed t, denoted $q(t)$, is

$$q(t) \equiv P(T > t) = P(T_1 > t \cup T_2 > t \cup \ldots \cup T_n > t) \quad \ldots\ldots\ldots\ldots (4)$$

Clearly, $p(t) + q(t) = 1.0$.

In practical terms, the completion-time probability, $p(t)$, has the following implication. A small $p(t)$ means that the required completion-time is likely to exceed the target time t, whereas a large $p(t)$ means that the project is likely to be completed within t.

Clearly, the events $(T_1 > t), (T_2 > t), \ldots, (T_n > t)$ are not mutually exclusive. Moreover, the different paths in a network may have common activities and thus would be correlated (as shown subsequently); for this reason, the events $(T_1 \leq t), (T_2 \leq t), \ldots, (T_n \leq t)$ are not statistically independent. Thus the determination of $p(t)$ or $q(t)$ requires the calculation of multiple joint probabilities.

Eq. 2 can be expressed as the union of the following mutually exclusive events:

$$(T > t) = (T_1 > t) \cup (T_1 \leq t, T_2 > t) \cup \ldots \cup (T_1 \leq t, T_2 \leq t, \ldots, T_{n-1} \leq t, T_n > t) \quad \ldots\ldots (5)$$

$$\text{then} \quad p(t) = 1 - q(t) = 1 - [P(T_1 > t) + P(T_1 \leq t, T_2 > t) + \ldots + P(T_1 \leq t, T_2 \leq t, \ldots, T_{n-1} \leq t, T_n > t)] \quad \ldots\ldots (6)$$

Individual Paths.—Consider the durations of the individual activities in a network as random variables τ_k, $k = 1, 2, \ldots$ with means and standard deviations μ_k and σ_k, respectively; and assume that the activities are mutually statistically independent.

The duration T_i of path π_i is clearly the sum of the durations of the activities comprising path π_i; thus

$$T_i = \sum_{k \epsilon \pi_i} \tau_k \quad \ldots\ldots (7)$$

Accordingly, the mean duration of path π_i is

$$\mu_{T_i} = \sum_{k \epsilon \pi_i} \mu_k \quad \ldots\ldots (8)$$

and by virtue of statistical independence, the corresponding variance is

$$\sigma^2_{T_i} = \sum_{k \epsilon \pi_i} \sigma^2_k \quad \ldots\ldots (9)$$

Although the individual activities are statistically independent, two different paths may be correlated as a result of common activities. The correlation between any two paths π_i and π_j can be easily shown to be

$$\rho_{ij} = \frac{\sum_{k \epsilon (\pi_i \cap \pi_j)} \sigma^2_k}{\sigma_{T_i} \sigma_{T_j}} \quad \ldots\ldots (10)$$

in which $k \epsilon (\pi_i \cap \pi_j)$ connotes those activities common to paths π_i and π_j. According to Eq. 10, the correlation between paths is always positive.

Bounds and Approximations

The previous analyses indicate that there is no simple way of determining the completion-time probability, $p(t)$; a known simple way is to consider just the mean critical path, leading to the PERT method. However, PERT invariably underestimates the project duration.

Bounds on Completion-Time Probability.—The completion-time probability, $p(t)$, of a project, or the complementary probability, $q(t)$, is bounded. These bounds are important because, aside from defining the limits of $p(t)$, these also provide the bases for the development of the probabilistic network evaluation technique, described subsequently.

From Eq. 1, we have

$$(T_1 \leq t) \cap (T_2 \leq t) \cap \ldots \cap (T_n \leq t) \subset (T_i \leq t); \quad \text{for any } i \quad \ldots\ldots (11)$$

Therefore $P(T_1 \leq t, \ldots, T_n \leq t) \leq P(T_i \leq t)$; for any i . . . . . (12)

and in particular $p(t) \leq \min_i [P(T_i \leq t)]$ . . . . . (13)

The equality in Eq. 13 holds when all the paths in the network are perfectly correlated.

On the other hand, since the correlation between any two paths π_i and π_j is always positive, $\rho_{ij} \geq 0$, we have

$$P(T_i \leq t \mid T_j \leq t) \geq P(T_i \leq t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (14)$$

thus $P(T_i \leq t, T_j \leq t) \geq P(T_i \leq t)\, P(T_j \leq t)$ . . . . . (15)

And by induction

$$P(T_1 \leq t, T_2 \leq t, \ldots, T_n \leq t) \geq P(T_1 \leq t)\, P(T_2 \leq t) \ldots P(T_n \leq t) \quad \ldots\ldots \quad (16)$$

which means $p(t) \geq \prod_{i=1}^{n} P(T_i \leq t)$ . . . . . (17)

In this case, equality holds when all the paths are statistically independent. The right-hand sides of Eqs. 13 and 17 therefore are, respectively, the upper and lower bounds of $p(t)$. Thus

$$\prod_{i=1}^{n} P(T_i \leq t) \leq p(t) \leq \min_i P(T_i \leq t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (18)$$

According to the previous bounds, therefore, the assumption of statistical independence among all the paths gives conservative estimate, whereas assuming perfect correlations among the paths would yield unconservative results; extending the jargon of PERT, these may be called, respectively, the *most pessimistic* and *most optimistic* predictions of the completion time. In this sense, PERT always gives the "most" optimistic estimate of a project time, i.e., the resulting $p(t)$ is too high; this is also obvious from Eq. 6 if π_1 is assumed to be the mean critical path.

Probabilistic Network Evaluation Technique—PNET

From Eq. 6, we see that successively better estimates of the project completion-time probability can be obtained by considering more paths in the evaluation of $p(t)$. In particular, if the paths, $\pi_1, \pi_2, \ldots, \pi_n$, are arranged in decreasing order of mean durations, i.e., such that

$$\mu_{T_1} \geq \mu_{T_2} \geq \ldots \geq \mu_{T_n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (19)$$

then the additional paths in Eq. 6 may be limited to the major paths (defined as paths with long mean durations). However, unless the major paths are all statistically independent (or perfectly correlated), the improvements suggested in Eq. 6 would still require evaluation of the joint probabilities which are generally not simple to perform.

An approximate method which will be called the *probabilistic network evaluation technique (PNET)* is developed herein for estimating $p(t)$ in which all the major paths are considered, but evaluation of the joint probabilities is avoided. The method is an obvious improvement over the PERT method, and under certain

conditions it reduces automatically to the PERT method.

Basis of PNET Method.—The proposed PNET method is based on the following observations: (1) The paths with long mean durations and high coefficients of variation will have the greatest significance on $p(t)$; (2) according to Eq. 13, if several paths are each highly correlated with a major path, then these paths are "represented" by the same major path, i.e., the completion-time probability associated with these several paths is represented approximately by that of this major path alone; and (3) on the other hand, according to Eq. 17, if several paths have low mutual correlations, the completion-time probability associated with these paths can be approximated with the product of the respective path probabilities.

Therefore, on the basis of these observations, if all the major paths in a network can be divided into subgroups in accordance with their correlations, such that within each subgroup the paths are highly correlated and thus are "represented" by one path, whereas between the "representative" paths the mutual correlations are low, then the completion-time probability of the project would be given approximately as the product of the completion-time probabilities of the individual "representative" paths. This can indeed be done conveniently for any network, resulting in the following PNET algorithm for determining the project completion-time probability, $p(t)$.

PNET Algorithm

1. Generate the major paths sequentially in order of decreasing mean path durations, starting with the mean critical path, π_I. This process may be done systematically and formally (9). The major paths may be limited to those whose mean path durations are at least a certain percentage of the mean critical path duration.

2. Evaluate the standard deviations for each of the major paths, using Eq. 9.

3. Identify the "representative" paths as follows. Using Eq. 10, calculate the correlations ρ_{Ii} between the critical path, π_I, and each of the other major paths, π_i. Those paths with $\rho_{Ii} > 0.5$ (assuming that $\rho = 0.5$ represents the transition between high and low correlation) are "represented" by π_I.

For those paths with $\rho_{Ii} \leq 0.5$, consider these to be a subset of paths $\{\pi_j; j = 1, 2, \ldots, k\}$ and designate the path within this subset with the longest mean duration to be π_{II}. Calculate the correlations ρ_{IIj} between π_{II} and π_j within this subset. Again, those paths among π_j with $\rho_{IIj} > 0.5$ are "represented" by π_{II}, whereas for those paths with $\rho_{IIj} \leq 0.5$ the procedure is repeated to determine the subsequent "representative" paths.

The process then leads to the set of "representative" paths, π_I, π_{II}, π_{III}, $\ldots$, π_N.

4. Calculate the probability of the "representative" paths, i.e., $P(T_I \leq t)$, $P(T_{II} \leq t)$, $\ldots$, $P(T_N \leq t)$, assuming a convenient probability distribution for T_I, T_{II}, $\ldots$, T_N and using the mean and variance evaluated in steps 1 and 2.

The project completion-time probability, $p(t)$, then is given approximately by the product

$$p(t) \simeq P(T_I \leq t)\, P(T_{II} \leq t) \ldots P(T_N \leq t) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (20)$$

Observe that if all the major paths are highly correlated with π_1, then Eq. 20 becomes $p(t) \simeq P(T_1 \leq t)$ and thus the PNET method reduces to the PERT method.

The probability of the individual paths $P(T_i \leq t)$ may be calculated on the assumption of normal distributions for T_i. If there are a large number of activities in a path, the path duration will tend toward a normal variate (by virtue of the central-limit theorem); moreover, in the range of probabilities of interest, say $0.05 \leq p(t) \leq 0.95$, the distribution would not make much difference on the calculated probabilities. The proposed PNET method, however, is independent of the distribution function, and thus is equally valid if other distributions for T_i are used.

Illustrative Applications

Three examples are presented subsequently to demonstrate the validity of the previous PNET method and to illustrate its application in construction project evaluations.

Road Pavement Project.—The project involves the paving of 2.2 miles of roadway pavement and the construction of appurtenant drainage structures, excavation to grade, placement of macadam shoulders, erection of guardrails, and landscaping. The problem is shown in Fig. 1, and is taken from Ref. 3.

The various activities of the project are described in Table 1, where the respective mean durations and corresponding standard deviations are also listed. The project network is shown in Fig. 2. All the nine paths of the network are listed in Table 2, arranged in order of decreasing mean durations; the standard deviations of these paths are also given in Table 2, calculated in accordance with Eq. 9. From Table 2, it should be clear, on the basis of the relative mean durations, that the first five paths would be sufficient for estimating $p(t)$; these five paths are, therefore, the major paths in this example.

Using Eq. 10, we identify the "representative" paths among the first five major paths as follows. Compute ρ_{1i}, obtaining

$$\rho_{12} = 0.05; \quad \rho_{13} = 0.06; \quad \rho_{14} = 0.07; \quad \rho_{15} = 0.74 \ldots\ldots\ldots\ldots\ldots \quad (21)$$

Therefore, π_5 is "represented" by π_1. Among π_2, π_3, and π_4, which have $\rho_{1i} < 0.5$, compute ρ_{2j}, obtaining $\rho_{23} = 0.79$ and $\rho_{24} = 0.69$. Therefore, π_3 and π_4 are "represented" by π_2.

Thus, π_1 and π_2 are the "representative" paths, and the completion-time probability of the project (assuming normal distributions for T_1 and T_2) is

$$p(t) \simeq P(T_1 \leq t)\, P(T_2 \leq t) = \Phi\left(\frac{t-61}{5}\right) \Phi\left(\frac{t-57}{9}\right) \quad \ldots\ldots\ldots \quad (22)$$

in which $\Phi(\cdot)$ is the cumulative probability of the standard normal distribution. This result, for all t, is shown in Fig. 3; for comparison, the results of a 1,000-sample Monte Carlo simulation is also shown.

In Fig. 3, the completion-time probabilities determined with the PERT method, i.e., $P(T_1 \leq t)$, are also shown. This illustrates that the PERT method consistently overestimates the completion-time probability (tantamount to underestimating

TABLE 1.—Activities and Estimated Durations (Pavement Project)

Description of activities (1)	Mean duration, in days (2)	Standard deviation, in days (3)
1. Dummy	0	0.
2. Set-up batch plant	2	0.5
3. Order and deliver paving mesh	5	1.0
4. Deliver rebars for double barrel culvert	6	1.5
5. Move in equipment	3	0.5
6. Deliver rebars for small box culvert	7	4.0
7. Build double barrel culvert-	10	2.0
8. Clear and grub from station 42-station 100	3	1.0
9. Clear and grub from station 100-station 158	7	1.5
10. Build box culvert at station 127	5	2.0
11. Build box culvert at station 138	3	1.5
12. Cure double barrel culvert	9	2.0
13. Move dirt between station 42 and station 100	5	1.5
14. Start moving dirt between station 100 and station 158	3	0.5
15. Cure box culvert at station 127	9	4.5
16. Cure box culvert at station 138	6	2.0
17. Order and stockpile paving material	2	0.5
18. Place subbase from station 42-station 100	7	1.73
19. Finish moving dirt between station 100 and station 158	5	2.0
20. Pave from station 42-station 100	10	2.0
21. Place subbase from station 100-station 158	7	3.31
22. Cure pavement from station 42-station 100	6	1.5
23. Pave from station 100-station 158	10	4.5
24. Cure pavement from station 100-station 158	6	1.5
25. Place shoulders from station 42-station 100	3	1.0
26. Place shoulders from station 100-station 158	3	1.0
27. Place guardrail and landscape	5	1.5
28. Dummy	0	0.

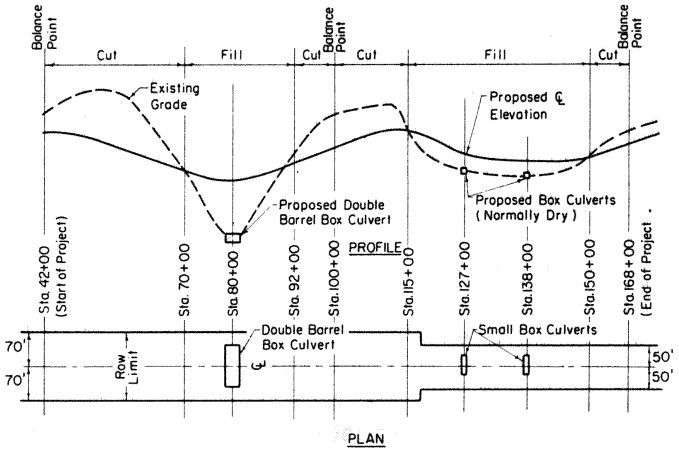


FIG. 1.—Plan and Profile of Pavement Project

TABLE 2.—Ordered Paths and Duration Statistics (Pavement Project)

Path (1)	Activities in path (2)	Mean duration, in days (3)	Standard deviation, in days (4)
π_1	4, 7, 12, 13, 18, 20, 22, 25, 27 ...	61	5.0
π_2	6, 10, 15, 19, 21, 23, 24, 26, 27 ...	57	9.0
π_3	6, 11, 16, 19, 21, 23, 24, 26, 27 ...	52	7.94
π_4	5, 9, 14, 19, 21, 23, 24, 26, 27 ...	49	6.54
π_5	5, 8, 13, 18, 20, 22, 25, 27 ...	42	4.0
π_6	3, 28, 20, 22, 25, 27 ...	29	3.24
π_7	3, 1, 23, 24, 26, 27 ...	29	5.19
π_8	2, 17, 28, 20, 22, 25, 27 ...	28	3.16
π_9	2, 17, 1, 23, 24, 26, 27 ...	28	5.12

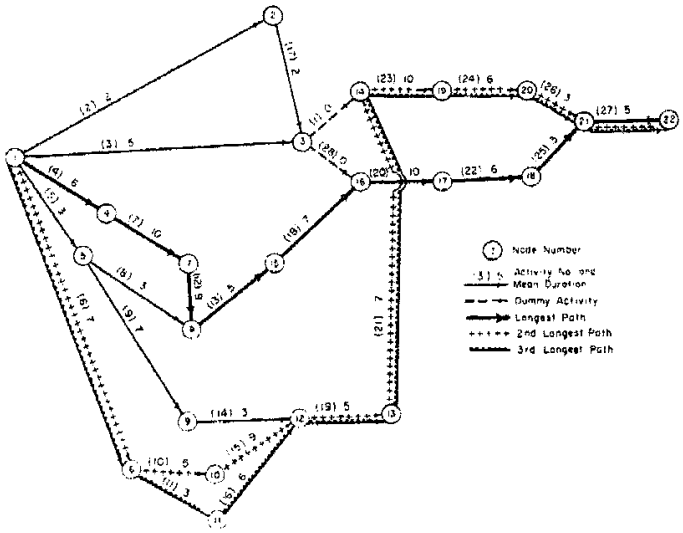


FIG. 2.—Network of Pavement Project

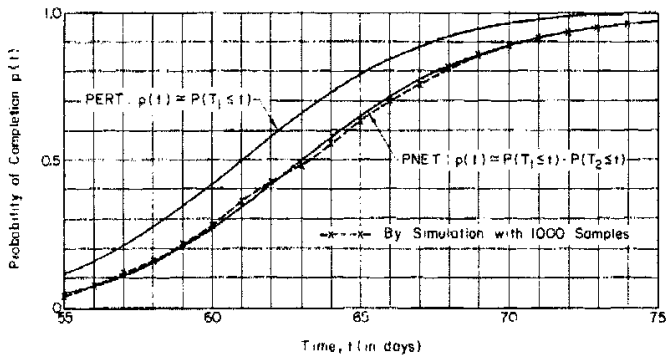


FIG. 3.—Completion-Time Probability Function (Pavement Project)

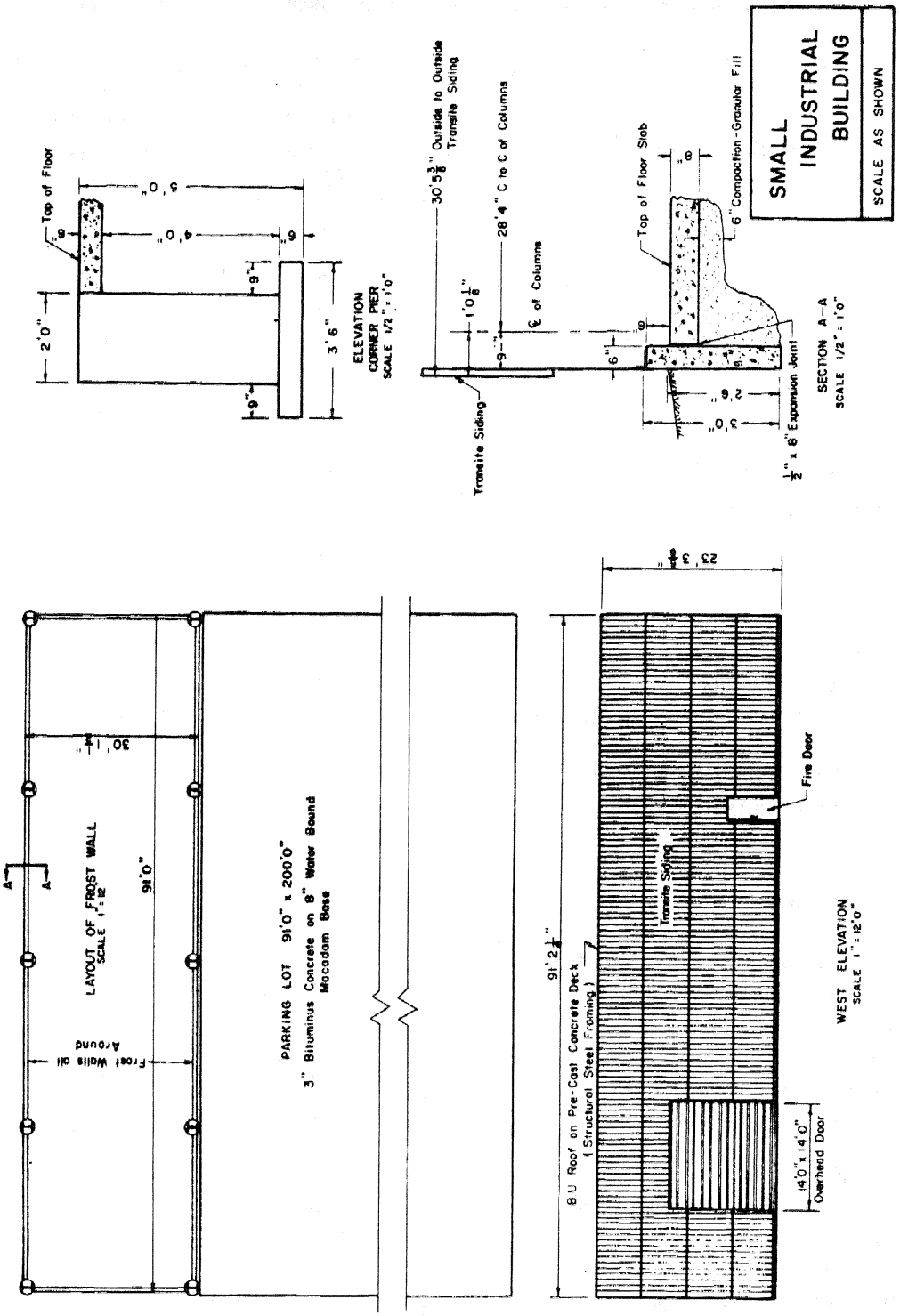


FIG. 4.—General Plan of Industrial Building Project

the required project duration). For example, in this case, the PNET method (substantiated by simulation) estimated that the project will be completed in 63 days with a probability of 50%; PERT, however, would give a corresponding probability of 66%.

TABLE 3.—Activities and Estimated Durations (Industrial Building Project)

Description of activities (1)	Mean duration, in days (2)	Standard deviation in days (3)
1. Mobilization	32	3.2
2. Move in	2	0.5
3. Initial layout	2	0.5
4. Dummy	0	0
5. Site rough grading	2	0.5
6. Layout of piers	1	0.5
7. Excavate piers	2	1.0
8. Dummy	0	0
9. Order and deliver rebars	40	12
10. Form and rebars piers	2	0.5
11. Pour piers	2	0.5
12. Cure piers	4	0.8
13. Strip piers	1	0.1
14. Dummy	0	0
15. Dummy	0	0
16. Excavate frost walls	1	0.5
17. Order and deliver structural steel columns	60	12
18. Erect structural steel columns	5	1
19. Order and deliver precast roof deck	30	6
20. Form and mesh frost walls	3	0.9
21. Pour frost walls	1	0.3
22. Cure frost walls	4	0.4
23. Strip frost walls	1	0.1
24. Backfill	2	0.5
25. Grade and compact gravel for floor	2	0.2
26. Rebar floor and set screeds	2	0.5
27. Pour and finish floor	2	0.5
28. Dummy	0	0
29. Excavate and grade parking	2	0.2
30. Stone base for parking	1	0.2
31. Dummy	0	0
32. Set roof deck	5	1.5
33. Hang siding and waterproof roof	6	1.2
34. Hang doors	4	1.2
35. Clean up	2	0.5
36. Bituminous surface in parking	3	0.3

Industrial Building Project.—This project involves the construction of a single-story industrial building with an adjoining parking lot (2), as shown in Fig. 4. The building is comprised of reinforced concrete piers, frost walls,

structural steel columns, and a precast roof deck. The various activities are described in Table 3, with the means and standard deviations of the respective durations also listed. The corresponding project network is shown in Fig. 5. In Table 4 are listed the first 10 paths arranged in decreasing mean path durations; also shown are the corresponding standard deviations, which were calculated in accordance with Eq. 9.

In this case, we identify the "representative" paths as follows. Compute ρ_{Ii} obtaining

$$\rho_{12} = 0.99; \quad \rho_{13} = \rho_{14} = \rho_{15} = \rho_{16} = \rho_{17} = \rho_{18} = 0;$$

$$\rho_{19} = 0.006; \quad \rho_{1(10)} = 0.006 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \quad (23)$$

Therefore, only path π_2 is "represented" by π_1. For paths π_3-π_{10}, which have $\rho_{Ii} < 0.5$, we compute

TABLE 4.—Ordered Paths and Duration Statistics (Industrial Building Project)

Path (1)	Activities in path (2)	Mean duration, in days (3)	Standard deviation in days (4)
π_1	17,18,32,33,35	78	12.20
π_2	17,18,32,34,35	76	12.20
π_3	9, ...,[a] 13,15,20, ..., 28,36	69	12.12
π_4	9, ... , 13,15,20, ... , 27,31,35	68	12.14
π_5	1,2,5,7,8,10, ... , 13,15,20, ... , 28,36	67	3.85
π_6	1,3, ..., 5,7,8,10, ..., 13,15,20, ..., 28,36	67	3.85
π_7	1,3,6, ... , 8,10, ... , 13,15,20, ... , 28,36	66	3.85
π_8	9, ... , 13,15,20, ... , 24,29,30,36	66	12.09
π_9	1,2,5,7,8,10, ... , 13,15,20, ... , 27,31,35	66	3.71
π_{10}	1,3, ... , 5,7,8,10, ... , 13,15,20, ... , 27,31,35	66	3.87

[a] Includes all intervening activities.

$$\rho_{34} \simeq 1.00; \quad \rho_{35} = \rho_{36} = \rho_{37} = 0.07; \quad \rho_{38} \simeq 1.00; \quad \rho_{39} = \rho_{3(10)} = 0.07 \quad \ldots \quad (24)$$

Thus, π_4 and π_8 are "represented" by π_3. Continuing, we have

$$\rho_{56} = \rho_{57} = \rho_{59} = \rho_{5(10)} \simeq 0.99 \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \quad (25)$$

Therefore, π_6, π_7, π_8, and π_{10} are "represented" by π_5. Thus, the "representative" paths in this case are π_1, π_3, and π_5, and the completion-time probability is

$$p(t) \simeq P(T_1 \leq t)\, P(T_3 \leq t)\, P(T_5 \leq t)$$

$$= \Phi\left(\frac{t-78}{12.20}\right)\Phi\left(\frac{t-69}{12.12}\right)\Phi\left(\frac{t-67}{3.85}\right) \ldots \ldots \ldots \ldots \ldots \quad (26)$$

the result is portrayed in Fig. 6. The results of a 10,000-sample Monte Carlo simulation and of the PERT method are also shown in Fig. 6 for comparison.

Apartment Complex (Large-Scale Project).—The accuracy of the PNET method has also been verified for large-project networks. An example of this is the construction of two multistory apartment buildings. The project involves over

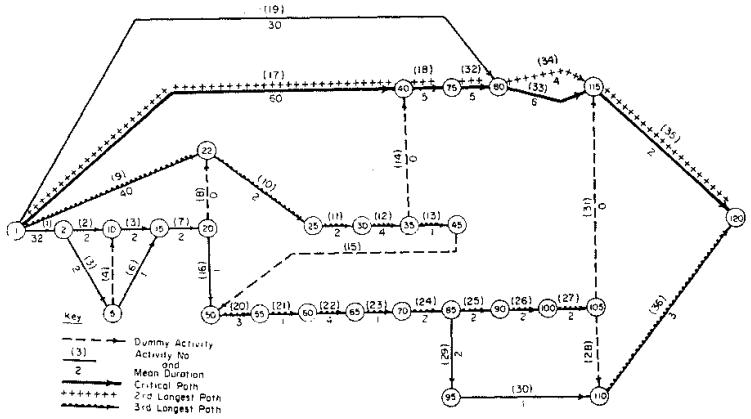


FIG. 5.—Network of Industrial Building Project

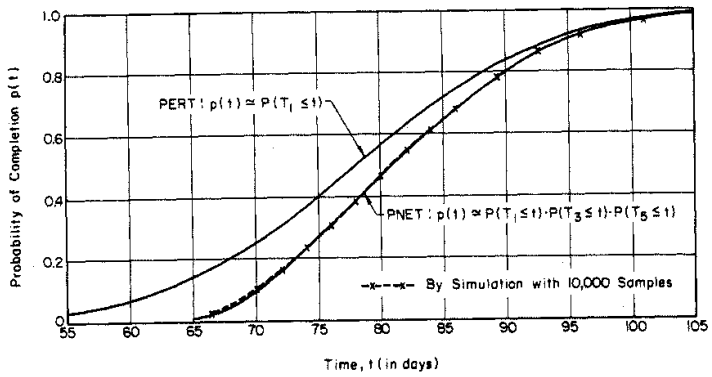


FIG. 6.—Completion-Time Probability Function (Industrial Building)

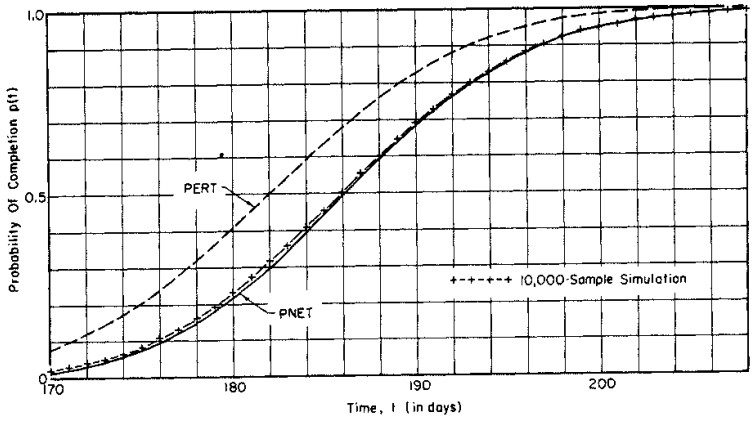


FIG. 7.—Completion-Time Probability Function (Apartment Complex)

160 activities, with over 8,000 paths among which the first 240 major paths were considered in the PNET analysis. Many of the correlations between the major paths are between 0.4 and 0.7. Despite the size of the network (over

8,000 possible paths), and the number of major paths considered, there are only two representative paths in the network. The resulting completion-time probabilities are summarized in Fig. 7.

Results of a Monte Carlo simulation, with 10,000 samples, is also shown in Fig. 7, together with the corresponding solution by PERT.

A simple computer program was developed for the PNET solution; the program generates all the possible paths in decreasing order of mean path durations, identify the representative paths, and perform the calculations required for the determination of $p(t)$. The total execution time was 15 sec on an IBM 360/75 system. The corresponding Monte Carlo simulation using the GERTS III program (6) took 38 min on the same system.

This problem should serve to illustrate the practicality of the PNET method for large-scale projects, and highlight its advantages over existing methods for the evaluation of project times, in the presence of uncertainties. This example also serves to illustrate the validity of Eq. 20 even for intermediate correlations between paths, and the reasonableness of using a transitional correlation of 0.5 for the identification of the representative paths in a network.

The examples presented herein are admittedly simple as real construction projects go; however, such simple projects are necessary for the purpose of illustrating the applications of the method. Moreover, these examples are sufficient to demonstrate the effectiveness of the proposed PNET method for the analysis of project networks under conditions of uncertainty.

Accuracy of PNET

The only means of evaluating the accuracy of the proposed PNET algorithm is with Monte Carlo simulation. Such simulations were performed for all the problems illustrated herein (Figs. 3, 6, and 7). It can be observed that in all cases, the comparisons are excellent, indicating that the proposed PNET method gives reliable project completion-time probability for the entire range of probabilities of interest.

It may be emphasized that if the paths in a network are either perfectly correlated or uncorrelated, the PNET method is theoretically exact by virtue of Eqs. 13 and 17. For a general network with intermediate path correlations, the solution is only approximate; the calculated probability will be too high or too low relative to the correct completion-time probability depending on the transitional correlation, ρ_o, used in the identification of the representative paths. A transitional correlation of $\rho_o = 1.0$ will give the most pessimistic result, i.e., lowest $p(t)$; whereas using $\rho_o = 0$ will yield the most optimistic result, i.e., highest $p(t)$, as in the case of PERT. With a transitional correlation of $\rho_o = 0.5$, PNET solutions have consistently proved to be in excellent agreement with large-sample Monte Carlo simulations.

Summary and Conclusions

A simple method for evaluating the completion-time probabilities of project networks is presented. The method can be used to estimate completion times of projects in the presence of uncertainty; and thus permit the consideration of network decision problems in terms of risk and probability. In contrast to

the methods of PERT and CPM where only the critical path is considered, PNET includes all possible paths in a network; despite these, however, the PNET algorithm is conceptually as simple to apply as PERT, requiring calculation of the individual path probabilities only, and can be easily programmed for machine processing.

Under certain conditions, the PNET method is theoretically exact. Moreover, for general networks the validity of the proposed PNET method has been verified by results of large-sample Monte Carlo simulations; in all cases, very close agreement was obtained for the entire range of probabilities of interest. The PNET method is a direct analytical procedure, in contrast to the Monte Carlo simulation in which repeated sample solutions are required. For this reason, the method is quite efficient; relative to Monte Carlo simulation, order-of-magnitude savings in computational time is involved.

The proposed PNET method should provide a convenient and effective tool for evaluating activity networks under uncertainty. In its present basic form, the method yields the probability of project completion within a prescribed target time, or the required project time at a prescribed probability. The method may be extended to consider the effects of uncertainty in general activity and decision networks.

Acknowledgments

The results presented here are part of a study supported by the National Science Foundation under research grant GK-43013.

Appendix I.—References

1. Ang, A. H. S., Abdelnour, J., and Chaker, A. A., "Probabilistic Evaluation of Project Networks," *International Symposium on Systems Engineering*, Purdue University, Lafayette, Ind., Oct., 1972, pp. 8–13.
2. Brand, J. D., Meyer, W. L., and Shaffer, L. R., "The Resource Scheduling Problem in Construction," *Civil Engineering Studies, Construction Research Series No. 5*, University of Illinois, Urbana, Ill., June, 1964.
3. Brooks, A. C., Leahy, J. P., and Shaffer, L. R., "A Man-Machine CPM System for Decision Making in the Construction Industry," *Civil Engineering Studies, Construction Research Series No. 9*, University of Illinois, Urbana, Ill., June, 1967.
4. MacCrimmon, K. R., and Ryavec, C. A., "Analytic Study of the PERT Assumptions," *Operations Research*, Vol. 12, 1964, pp. 16–37.
5. Malcolm, D. G., et al., "Applications of a Technique for R and D Program Evaluation," *Operations Research*, Vol. VII, No. 5, Sept.-Oct., 1959.
6. Pritsker, A. A. B., and Burgess, R. R., "The GERT Simulation Programs," Department of Industrial Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Va., 1970.
7. Ringer, L. J., "Numerical Operators for Statistical PERT Critical Path Analysis," *Management Science*, Vol. 16, No. 2, Oct., 1969.
8. Van Slyke, R. M., "Monte Carlo Methods and the PERT Problem," *Operations Research*, Vol. 11, No. 5, Sept.-Oct., 1963.
9. Yen, J. Y., "Finding the K Shortest Loopless Paths in a Network," *Management Science*, Vol. 17, No. 11, July, 1971.

Appendix II.—Notation

The following symbols are used in this paper:

$p(t)$ = probability of completing project within target time t;

T_i = duration of path i;
t = target time;
μ_{T_i} = mean duration of path i;
π_i = network path i;
$\pi_I, \pi_{II}, \ldots, \pi_N$ = "representative" paths in a network;
ρ_{ij} = correlation coefficient between paths i and j;
σ_{T_i} = standard deviation of path i duration.

Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivannan, eds.

PROJECT SCHEDULING USING STATE-BASED PROBABILISTIC DECISION NETWORKS

Photios G. Ioannou

Civil & Environmental Engineering Department
University of Michigan
Ann Arbor, Michigan 48109-2125, U.S.A.

Julio C. Martinez

Charles E. Via, Jr. Department of Civil Engineering
Virginia Tech
Blacksburg, Virginia 24061-0105, U.S.A.

ABSTRACT

This paper presents probabilistic CPM scheduling using an add-on program to the STROBOSCOPE simulation system that allows dependence and correlation between activity durations as well as control over precedence and the selection of paths. Activity durations and sequencing can be defined in terms of the dynamic information that becomes available as a project evolves and includes the actual start date and duration of activities already started. An example highway project illustrates the modeling power of this approach.

1 INTRODUCTION

It is widely recognized that construction planning and scheduling using CPM can be substantially enhanced and become more effective and more realistic by recognizing ahead of time the uncertainty in activity durations and their dependence on the dynamic state of the project as it evolves. Several tools have been developed for this purpose. PERT was a first step by allowing probabilistic activity durations. Monte-Carlo simulation can tackle the merge-event bias problem (Van Slyke 1963). GERT (Pritsker 1977) can additionally model uncertainty in the precedence of activities. VERT (Moeller & Digman 1981), MUD (Carr 1979) and DYNASTRAT (Morua Padilla 1986) can model correlation between activity durations and to a limited extent recognize the state of the project in progress.

None of the existing tools, however, have the necessary flexibility and power to model uncertainty in the duration of activities as a true function of the state of the project, nor can they model the underlying process-level operations through concurrent simulation. The CPM add-on for STROBOSCOPE presented here overcomes these limitations and serves as an example of how STROBOSCOPE can be used as a development vehicle for special-purpose simulators and more sophisticated project-level modeling tools.

2 STROBOSCOPE AND THE CPM ADD-ON

STROBOSCOPE is an acronym for STate and ResOurce Based Simulation of COnstruction ProcEsses. It is a simulation programming language designed specifically for modeling construction operations based on three-phase activity scanning and activity cycle diagrams. The STROBOSCOPE language is described in (Martinez 1996). Example applications can be found in (Ioannou & Martinez 1996a, 1996b, 1996c) and (Martinez & Ioannou 1994, 1995).

STROBOSCOPE can be extended via add-ons written in a high-level compiled language such as FORTRAN or C++. STROBOSCOPE can also be controlled from other applications via OLE automation. A STROBOSCOPE add-on is a 32-bit MS Windows dynamic-link library that extends the STROBOSCOPE language with new statements, functions, and variables. The code within the add-on can call back into the simulation engine using STROBOSCOPE's Application Programming Interface (API) described in (Martinez 1996).

The CPM add-on described in this paper is loaded into STROBOSCOPE with the `LOADADDON` statement:

```
LOADADDON C:\…\CpmAddOn.dll;
```

Once loaded, its statements, functions and variables can be used as if they were a standard part of the language. The new statements are shown in Table 1 and the new state variables are listed in Table 2. Of these, the second argument to CPMACTIVITY in Table 1 is of particular importance because it defines an activity's duration. This argument can be any allowable expression and can include functions that sample from probability distributions as well as variables that access the state of project or the state of any concurrent process simulation. Thus, the expression used to define the activity duration can be as complex as necessary, calling any number of functions and variables. The parameters of functions that sample from probability distributions can also include expressions or variables to allow easy modeling of conditional and correlated distributions.

Table 1: Statements Registered by the CPM Add-On

CPMACTIVITY *ActivityName* [*Duration*] [*Cost*]; Defines a CPM activity with the given name. Its cost and duration are sampled from the given expressions.
PRECEDENCE *PredecessorName SuccessorName*; Indicates that the successor cannot start until the predecessor is finished.
DURATION *ActivityName Duration*; (Re)defines the duration expression for CPMActivity.
CPMACTCOST *ActivityName Cost*; (Re)defines the cost expression for CPMActivity.
CPMREPLICATE *Number of Replications*; Simulates the CPM network the number of times indicated and produces a report as described in the CPMREPORT statement.
DOCPM; Performs a single forward and backward CPM pass.
CPMREPORT; Prints the 90% Confidence Interval on project duration and cost, and a report showing the average duration, cost, early and late dates, floats, and criticality for each CPM Activity.

Table 2: Variables defined by the CPM Add-On

Variable Prototype	Value Returned
CPMActivity.Duration	The sampled activity duration used in the current run.
CPMActivity.EarlyStart CPMActivity.EarlyFinish CPMActivity.LateStart CPMActivity.LateFinish	The early/late start/finish for CPMActivity in the current run.
CPMActivity.Started CPMActivity.GoingOn CPMActivity.Finished	TRUE if CPMActivity (has started)/(is going on)/(has finished) in the current run.
CPMActivity.FreeFloat CPMActivity.TotalFloat	The activity's free/total float in the current run.
CPMActivity.Critical	TRUE if CPMActivity is critical in the current run.

3 EXAMPLE HIGHWAY PROJECT

A highway with the profile and plan shown in Figure 1 will serve as an example for using the CPM add-on. This project is 11,600 ft (2.2 mile) long and has been adapted from (Brand et al 1964). Table 3 shows the names of the activities, the three-point estimates for the (optimistic, most-likely, and pessimistic) activity durations (a, M, b), the mean duration based on the PERT approximation (a+4*M+b)/6, the variance, and the standard deviation of the activity duration given by the PERT approximation (b-a)/6. The last column lists the direct predecessors of each activity and defines the connectivity of the network.

The construction work for this highway is naturally divided into two sections based on the location of the balance points for earthmoving (cut and fill). As shown in Figure 1, the western section extends from Station 42 to Station 100 and the eastern section extends from Station 100 to Station 158. Within each section the amount of cut equals the amount of fill. Assuming that the necessary resources are available, the majority of the work in each section can be performed independently and simultaneously while minimizing haul distances. The division of the work is also reflected in the definition of activities shown in Table 3. They include material and equipment procurement, cut-and-fill earthmoving, the construction of one double-barrel and two single-barrel concrete box culverts, the placement of subbase, paving, landscaping, and guardrail installation.

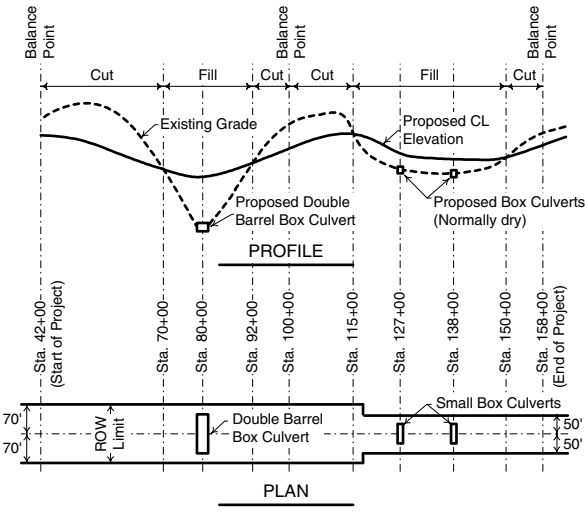


Figure 1: Plan and Profile for Highway Construction

The precedence network for this project is shown in Figure 2. Notice that activity names begin with the corresponding activity numbers from Table 3 (for easy identification) and cannot include spaces.

3.1 PERT Network Analysis

We shall begin the analysis of this project by using the PNET methodology to compute the cumulative probability of project completion (Ang et al 1975).

The basis of this approach are the following two inequalities that provide a lower and upper bound for the true cumulative probability of project completion:

$$\prod_I P[T_I \le t] \;\le\; P[T \le t] \;\le\; \min_I P[T_I \le t] \qquad (1)$$

Table 3: Activity List and Duration Data in Days for Highway Construction Project

No.	Activity Description	a	M	b	Mean	Var	SD	Preceded by
2	Setup batch plant	0.5	2	3.5	2	0.25	0.5	
3	Order & deliver paving mesh	2	5	8	5	1	1	
4	Deliver rebar for double barrel culvert	2.5	5.5	11.5	6	2.25	1.5	
5	Move in equipment	1.5	3	4.5	3	0.25	0.5	
6	Deliver rebar for small box culvert	1	4	25	7	16	4	
7	Build double barrel culvert	4	10	16	10	4	2	4
8	Clear & grub from Sta. 42 to Sta. 100	1	2.5	7	3	1	1	5
9	Clear & grub from Sta. 100 to Sta. 158	2.5	7	11.5	7	2.25	1.5	5
10	Build box culvert at Sta. 127	1	4	13	5	4	2	6
11	Build box culvert at Sta. 138	0.5	2	9.5	3	2.25	1.5	6
12	Cure double barrel culvert	3	9	15	9	4	2	7
13	Move dirt between Sta. 42 & Sta. 100	2.5	4	11.5	5	2.25	1.5	8, 12
14	Start moving dirt between Sta. 100 & Sta. 158	1.5	3	4.5	3	0.25	0.5	9
15	Cure box culvert at Sta. 127	1.5	6	28.5	9	20.3	4.5	10
16	Cure box culvert at Sta. 138	2	5	14	6	4	2	11
17	Order & stockpile paving material	0.5	2	3.5	2	0.25	0.5	2
18	Place subbase from Sta. 42 to Sta. 100	3.6	6.1	14	7	3	1.73	13
19	Finish moving dirt between Sta. 100 & Sta. 158	1	4	13	5	4	2	14, 15, 16
20	Pave from Sta. 42 to Sta. 100	4	10	16	10	4	2	3, 17, 18
21	Place subbase from Sta. 100 to Sta. 158	2	4.53	21.87	7	11	3.31	19
22	Cure pavement from Sta. 42 to Sta. 100	2.5	5.5	11.5	6	2.25	1.5	20
23	Pave from Sta.100 to Sta. 158	3	6.75	30	10	20.3	4.5	3, 17, 21
24	Cure pavement from Sta. 100 to Sta.158	2.5	5.5	11.5	6	2.25	1.5	23
25	Place shoulders from Sta. 42 to Sta. 100	1	2.5	7	3	1	1	22
26	Place shoulders from Sta. 100 to Sta.158	1	2.5	7	3	1	1	24
27	Place guardrail & landscape	2.5	4	11.5	5	2.25	1.5	25, 26

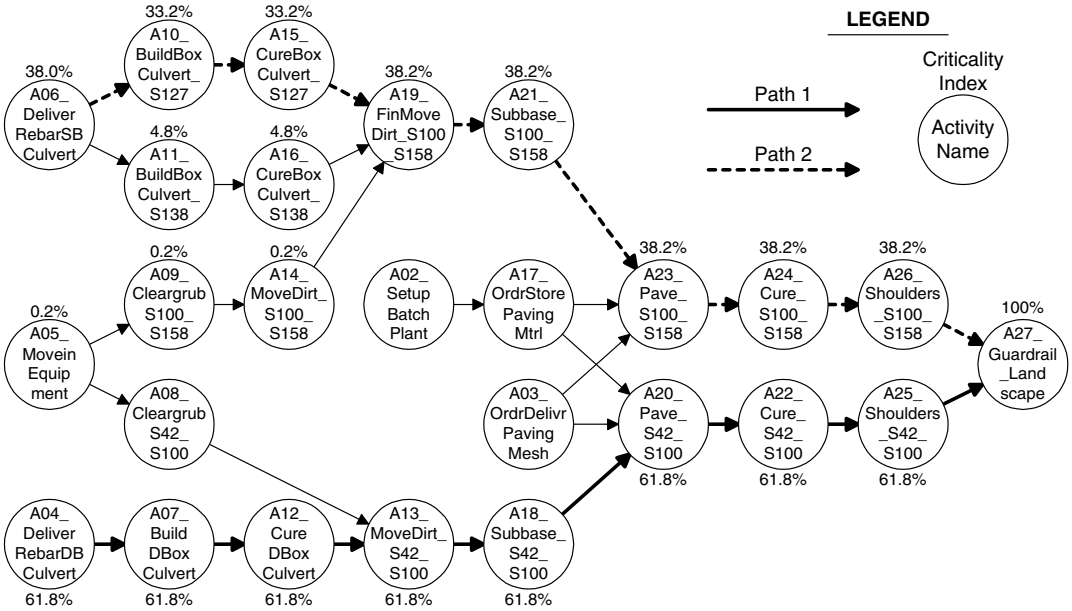


Figure 2: Precedence Network for Highway Construction Project

Here the random variable T represents the duration of the project and T_I represents the duration of path I.

For the standard PERT problem where a network has a single dominant path, the above upper and lower bounds coincide (i.e., they give the exact same values). When no single dominant path exists, however, any one of several competing paths (with overlapping probability density functions) may determine the duration of the project. In this case, it is not clear ahead of time as to which of these paths will actually be the longest, and thus prove to be critical. This phenomenon is known as *merge-event bias* and the true probability for completion of the project is somewhere between the lower and upper bounds given by Eq. (1). Whether it is closer to the upper or the lower bound depends on the degree of correlation between the durations of the competing paths. If path durations are independent then the true probability equals the lower bound. As positive correlation increases, the true probability moves closer to the upper bound.

Positive correlation between path durations is most often due to the fact that several paths share the same activities. When shared activities are long or short, all paths to which they belong tend to be long or short.

The basic idea in PNET is to eliminate from the lower bound in Eq. (1) *those paths that share enough activities to be highly correlated with (and thus represented by) other (longer) paths in the network*.

This procedure is best illustrated with an example. We begin by constructing a list of all paths in the network and arranging them in descending order of their mean duration. In this case, the network for the highway project consists of 9 paths ordered as shown in Table 4:

Table 4: Paths Ordered by their Expected Duration

Path I	Activities h in path I	Mean	SD
1	4,7,12,13,18,20,22,25,27	61	5.00
2	6,10,15,19,21,23,24,26,27	57	9.00
3	6,11,16,19,21,23,24,26,27	52	7.93
4	5,9,14,19,21,23,24,26,27	49	6.59
5	5,8,13,18,20,22,25,27	42	4.00
6	3,20,22,25,27	29	3.24
7	3,23,24,26,27	29	5.17
8	2,17,20,22,25,27	28	3.16
9	2,17,23,24,26,27	28	5.12

The standard deviation of path duration, SD, is the square root of the variance computed by assuming independence between the duration of activities h that belong to each path I:

$$\sigma_I^2 = \sum_{h \in I} \sigma_h^2 \tag{2}$$

A useful shortcut at this point is to eliminate from further consideration all paths that are clearly too short to assume the role of critical path, such as 6, 7, 8, and 9.

The fact that paths 1-5 are contenders, however, does not mean that they are independent. It is easy to show that the correlation coefficient between the durations of two paths I and J that share activities is given by:

$$\rho_{IJ} = \frac{\sum_{k \in (I \cap J)} \sigma_k^2}{\sigma_I \sigma_J} \tag{3}$$

The numerator in this expression is the sum of the variances of the activities k that belong to both paths I and J. The denominator is the product of the standard deviations of the durations of paths I and J from Table 4.

This expression is first used to compute the correlation coefficient between path 1 and every other path:

$$\rho_{12} = 0.05001 \qquad \rho_{13} = 0.05671$$
$$\rho_{14} = 0.06825 \qquad \rho_{15} = 0.73755$$

Paths with correlation coefficients greater than 0.5 are *represented* by path 1 and can be eliminated. In this case we can eliminate path 5 but not paths 2, 3, and 4.

The pairwise calculation of correlation coefficients is now repeated by substituting path 2 (the longest remaining path) in the role of path 1:

$$\rho_{23} = 0.79432 \qquad \rho_{24} = 0.68632$$

Since both correlation coefficients exceed 0.5 we can eliminate paths 3 and 4 as represented by path 2.

Thus, only paths 1 and 2 are long enough to be critical and at the same time sufficiently independent to be included in the calculation of the lower bound. The two paths are shown with heavier precedence links in the network of Figure 2 for easy identification.

In conclusion, the probability of project completion according to PNET can be approximated by:

$$P[T \le t] \approx P[T_1 \le t] \cdot P[T_2 \le t] \tag{4}$$

A graph of this approximation appears in Figure 3. The effect of merge-event bias is also shown in this figure as the difference between the PNET curve and that given by standard PERT with a dominant longest path:

$$P[T \le t] \approx P[T_1 \le t] \tag{5}$$

Clearly, Eq. (5) is the same as the upper bound in Eq. (1) and can lead to significant overstatement of the true probability of completing the work by a given deadline. In fact, the lower and upper bounds defined by Eq. (1) when considering all paths 1-5 define an envelope for the true probability of project completion given by:

$$\prod_{I=1}^{5} P[T_I \le t] \;\le\; P[T \le t] \;\le\; P[T_1 \le t] \tag{6}$$

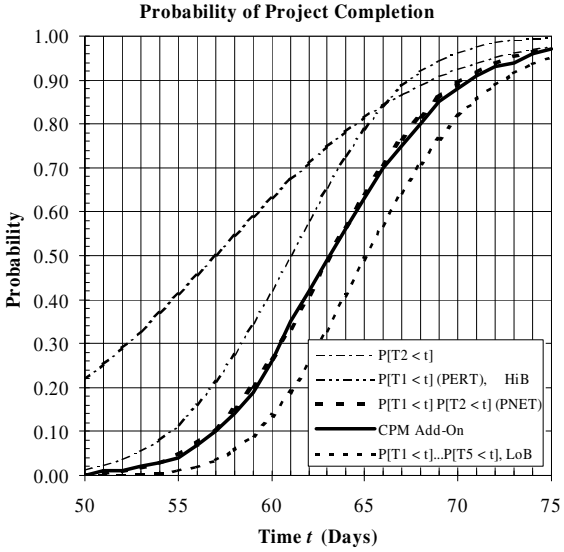


Figure 3: Probability of Project Completion

As shown in Figure 3 this envelope is quite wide and does not provide an accurate estimate for the probability of project completion. By eliminating three of the five terms in the lower bound of Eq. (5), the PNET estimate given by Eq. (4) is not only within this envelope but is also almost identical to the STROBOSCOPE simulation results, a fact that illustrates PNET's accuracy.

3.2 Basic Simulation of Highway Network

At the simplest level a probabilistic CPM network can be analyzed using *ProbSched*, a graphical pre- and post-processor for the CPM add-on that eliminates the need to write any simulation source-code manually. *ProbSched* has been implemented as an add-on to Visio (a programmable and extensible drawing program for Windows 95/NT) that facilitates the creation of intelligent schematic drawings by using preprogrammed drag-and-drop graphics. *ProbSched* includes a custom library of predefined graphical shapes that can be dragged and dropped onto the drawing page where they can be sized, positioned, and connected to form an activity network as shown in Figure 2. Double-clicking any of the activity nodes activates a custom dialog box for the specification of the appropriate distributions and data for sampling the duration and cost of the corresponding activity. Right-clicking anywhere on the drawing page activates a menu for specifying global parameters, such as the seed and the number of replications. The same menu also allows starting the simulation. This activates STROBOSCOPE, generates the necessary source code and passes it to the simulation engine through OLE automation, runs the specified number of replications, receives the simulation results, and plots them as a statistical bar chart in a separate Visio drawing.

While *ProbSched* provides all the necessary facilities for basic simulation modeling of probabilistic networks, it does not currently allow for more advanced techniques that require access to source-code level programming. In cases that need greater control, such as some of the examples presented below, it is necessary to write the source code for a simulation model using the STROBOSCOPE IDE. For example, the basic simulation input file for the highway construction project consists of statements such as the following:

```
VARIABLE nReplications 1000;
SEED 9111964; STREAMS 27; / If needed
LOADADDON CpmAddOn.dll;
CPMACTIVITY A02_SetupBatchPlant
   sPert[0.5,2,3.5,2];
CPMACTIVITY A03_DeliverPavingMesh
   sPert[2,5,8,3];            \ etc.
PRECEDENCE A02_SetupBatchPlant
   A17_ProcurePavingMtrl;
PRECEDENCE A03_DeliverPavingMesh
   A23_Pave_S100_S158;     \ etc.
CPMREPLICATE nReplications;
```

This source-code consists almost entirely of statements defined by the CPM add-on. The only standard STROBOSCOPE statements are VARIABLE, SEED, STREAMS, and LOADADDON. When this model is processed by the STROBOSCOPE simulation engine on a 400 MHz Pentium II under Windows NT 4, it produces the output shown in Figure 4.

As shown at the top of this report, STROBOSCOPE has performed 1,000 replications of this project. In each replication it sampled a duration for each activity (based on the corresponding probabilistic expression) and used these to perform the standard CPM calculations. Thus, in each replication it performed a forward pass to compute the Early Start Date (ESD) and Early Finish Date (EFD) and then a backward pass to compute the Late Finish Date (LFD) and the Late Start Date (LFD). Based on these, it determined each activity's Total Float (TF) and Free Float (FF) and identified the critical path. This process was repeated 1,000 times, and the average of the results is reported in Figure 4. For example, the column "Time" shows the average duration of each activity based on 1,000 samples. As expected, these values are very close to the expected activity durations shown in Table 3. The column "%Crit" shows each activity's *Criticality Index* (i.e., the fraction of the number of times out of a 1,000 the activity was critical).

More detailed statistics can be obtained by appending the REPORT statement at the end of the source-code file.

```
Number of replications performed : 1000
Average Project Duration         : 63.61
Std. Dev. of Project Duration    : 5.66
90% Confidence Interval for the Expected Project Duration: [63.315 - 63.905]
```

Activity	Time	ESD	LSD	EFD	LFD	FF	TF	%Crit
A02_SetupBatchPlant	2.00	0	33.56	2.00	35.56	0	33.56	0.0%
A03_OrderDeliverPavingMesh	5.02	0	32.58	5.02	37.60	26.90	32.58	0.0%
A04_DeliverRebarDBCulvert	5.97	0	2.83	5.97	8.80	0	2.83	61.8%
A05_MoveinEquipment	3.00	0	14.09	3.00	17.09	0	14.09	0.2%
A06_DeliverRebarSBCulvert	7.17	0	5.73	7.17	12.90	0	5.73	38.0%
A07_BuildDBoxCulvert	9.90	5.97	8.80	15.87	18.70	0	2.83	61.8%
A08_CleargrubS42_S100	3.01	3.00	24.76	6.02	27.77	18.93	21.76	0.0%
A09_CleargrubS100_S158	7.01	3.00	17.67	10.02	24.68	0	14.67	0.2%
A10_BuildBoxCulvert_S127	5.05	7.17	13.36	12.21	18.41	0	6.20	33.2%
A11_BuildBoxCulvert_S138	2.95	7.17	18.67	10.11	21.61	0	11.50	4.8%
A12_CureDBoxCulvert	9.07	15.87	18.70	24.94	27.77	0	2.83	61.8%
A13_MoveDirt_S42_S100	4.94	24.94	27.77	29.88	32.71	0	2.83	61.8%
A14_MoveDirt_S100_S158	2.98	10.02	24.68	13.00	27.66	9.04	14.67	0.2%
A15_CureBoxCulvert_S127	9.25	12.21	18.41	21.47	27.66	0.57	6.20	33.2%
A16_CureBoxCulvert_S138	6.05	10.11	21.61	16.16	27.66	5.88	11.50	4.8%
A17_ProcurePavingMtrl	2.04	2.00	35.56	4.04	37.60	27.89	33.56	0.0%
A18_Subbase_S42_S100	7.02	29.88	32.71	36.90	39.73	0	2.83	61.8%
A19_FinMoveDirt_S100_S158	5.00	22.04	27.66	27.04	32.67	0	5.62	38.2%
A20_Pave_S42_S100	10.00	36.90	39.73	46.90	49.73	0	2.83	61.8%
A21_Subbase_S100_S158	7.08	27.04	32.67	34.12	39.75	0	5.62	38.2%
A22_Cure_S42_S100	5.97	46.90	49.73	52.87	55.70	0	2.83	61.8%
A23_Pave_S100_S158	9.92	34.12	39.75	44.04	49.66	0	5.62	38.2%
A24_Cure_S100_S158	5.98	44.04	49.66	50.02	55.64	0	5.62	38.2%
A25_Shoulders_S42_S100	2.96	52.87	55.70	55.83	58.66	2.83	2.83	61.8%
A26_Shoulders_S100_S158	3.01	50.02	55.64	53.03	58.66	5.62	5.62	38.2%
A27_Guardrail_Landscape	4.95	58.66	58.66	63.61	63.61	0	0	100.0%

```
Execution Time = 2.894 seconds
```

Figure 4: Simulation Output for Highway Construction Project

Table 5: Path Criticality Indices

I	Activities *h* in path *I*	Mean	SD	% Crit.
1	4,7,12,13,18,20,22,25,27	61	5.00	61.8
2	6,10,15,19,21,23,24,26,27	57	9.00	33.2
3	6,11,16,19,21,23,24,26,27	52	7.93	4.8
4	5,9,14,19,21,23,24,26,27	49	6.59	0.2

This produces the standard STROBOSCOPE report that includes the minimum, maximum, average, and standard deviation for each variable shown in Figure 4.

When *ProbSched* is used to initiate the simulation process, STROBOSCOPE returns the results shown in Figure 4 back to the Visio add-on, which in turn uses them to create automatically the statistical bar chart shown in Figure 5.

The Criticality Indices listed in Figure 4 also appear next to each activity in Figure 2. An examination of this network indicates that there is no dominant path and that each of paths 1 through 4 served as the critical path in at least some replications as shown in Table 5.

The fact that there is no single dominant critical path explains why no activity (except 27) has TF = 0. Activity 25, for example, belongs to path 1 and was critical in 618

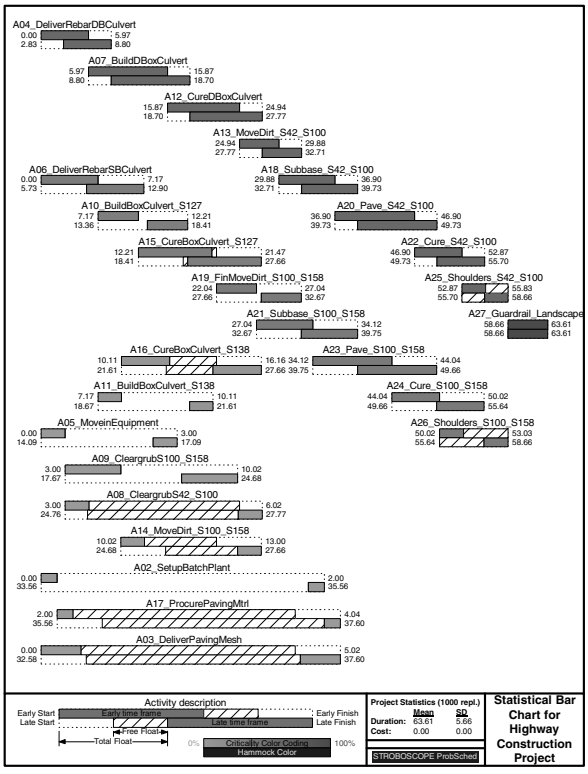


Figure 5: Statistical Bar Chart

of the 1,000 replications (TF = 0). In the remaining 382 runs it was non-critical and had TF > 0. Thus, the reported TF=2.83 is the average of 618 zeros and 382 non-zero values. This shows that in general the TF has a *mixed* probability distribution with a probability mass at TF = 0 and probability density for TF > 0.

3.3 Dynamic Scheduling Decisions

A simple illustration of a typical problem that is very difficult to solve analytically, but can be modeled quite easily with the CPM add-on, is that of precedence links that could point in either direction. It is not unusual in a highway project, for example, for the contractor to have only one paving machine. In this case, the two paving activities (20 and 23) cannot overlap and must be performed serially, one after the other. The problem is that we do not know beforehand which can be started first and which must follow. Merging the two paving activities into one does not solve the problem because it is not necessary to finish both subbase activities (in the east and west sections) in order to start paving, nor is it necessary to finish all paving before curing and shoulder work can begin.

This is an example of a dynamic decision that must be made while the project is going on. A simple way to model this using the CPM add-on is by *redefining* the durations of the two paving activities (X, Y) to add the extra time one of them must wait until the *other* activity (the one that has already started) finishes:

```
VARIABLE XWaitsForY 'Y.Started?
          Max[Y.EarlyFinish-SimTime,0]:0';
VARIABLE YWaitsForX 'X.Started?
          Max[X.EarlyFinish-SimTime,0]:0';
DURATION X XWaitsForY+Pert[aX,MX,bX];
DURATION Y YWaitsForX+Pert[aY,MY,bY];
```

Inserting the above code right before the statement CPMREPLICATE produces the results shown in Table 6 under the heading DL (dynamic links). Table 6 also shows the results for static precedence links as well as for the base case of two pavers to allow comparison.

Table 6: Effect of Correlation on Project Duration

	Two Pavers	One Paver		
		DL	23->20	20->23
Replications	1000	1000	1000	1000
Average Proj. Dur.	63.61	67.11	68.96	71.09
Std. Dev. Proj. Dur.	5.66	6.39	7.92	6.55
90% CI Lo Bound	[63.31	[66.78	[68.54	[70.75
90% CI Hi Bound	63.90]	67.44]	69.37]	71.43]

3.4 Correlated Activity Durations

Another typical problem is that of modeling correlation between activity durations caused by common factors such as the weather, site conditions, the use of the same equipment and crews, etc. Correlated activity durations can also be modeled easily with the CPM add-on. For example, in the case of two activities X and Y that are normally distributed with correlation coefficient ρ (*rho*), we can first sample X and then Y as follows:

```
DURATION X Normal[mX,sX];
DURATION Y Normal[mY+(X.Duration-mX)*
             rho*sY/sX,sY*Sqrt[1-rho^2]];
```

Using the highway project as an example, it is interesting to investigate the effect of correlation between two activities that belong to parallel paths versus that caused by two correlated activities that follow each other serially. Table 7 shows the effect of perfect positive correlation between the two subbase activities 18 (path 1) and 21 (path 2), and between subbase activity 18 (or 21) and the following paving activity 20 (or 23).

Table 7: Effect of Correlation on Project Duration

$\rho_{18,21}$ =	0	1	0	1
$\rho_{18,20} = \rho_{21,23}$ =	0	0	1	1
Replications	1000	1000	1000	1000
Average Proj. Dur.	63.43	63.27	64.1	63.07
Std. Dev. Proj. Dur.	5.75	5.91	6.71	7.2
90% CI Lo Bound	[63.13	[62.96	[63.75	[62.70
90% CI Hi Bound	63.73]	63.57]	64.45]	63.45]

The introduction of positive correlation between the two subbase activities makes paths 1 and 2 to tend to vary together and this reduces the average project duration from 63.43 to 63.27 and raises the standard deviation a little from 5.75 to 5.91. In contrast, positive correlation between a subbase activity and the paving activity that follows increases the variance for both paths 1 and 2 and this increases the average project duration from 63.43 to 64.10 and the standard deviation from 5.75 to 6.71. Having both types of correlation increases the variance for both paths 1 and 2 but also makes them move together (in lock-step) even more. This reduces the average duration from 63.43 to 63.07 and increases the standard deviation from 5.75 to 7.20.

4 CONCLUSION

The CPM add-on and its source code in C++ are included with STROBOSCOPE and are available from

http://grader.engin.umich.edu and http://strobos.ce.vt.edu. The CPM add-on illustrates STROBOSCOPE's powerful API and the ease of extending the language. It is also a solid starting point for more sophisticated project-level analysis tools. Moreover, the CPM add-on is especially very useful for teaching and research as a probabilistic scheduling tool. In fact, the CPM add-on was the primary tool for verifying a heuristic scheduling approach in a recently completed Ph.D. dissertation (Wang 1996).

ACKNOWLEDGMENTS

The authors wish to thank the National Science Foundation (Grants CMS-9415105, CMS-9733267) for supporting portions of the work presented here. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

REFERENCES

Ang, A.H., J. Abdelnoor, and A.A. Chaker. 1975. Analysis of Activity Networks under Uncertainty. *ASCE Journal of the Engineering Mechanics Division*, (101)EM4:373-386.

Brand, J.D., W.L. Meyer, and L.R. Shaffer. 1964. The Resource Scheduling Problem in Construction. *Civil Engineering Studies, Construction Research Series No. 5*. Univ. of Illinois, Urbana, Illinois.

Carr, R.I. 1979. Simulation of Construction Project Duration. *Journal of the Construction Division*, ASCE, (105)CO2:117-128.

Ioannou, P.G. and J.C. Martinez. 1996a. Comparison of Construction Alternatives Using Matched Simulation Experiments. *Journal of Construction Engineering and Management*, ASCE, (122)3:231-241.

Ioannou, P.G. and J.C. Martinez. 1996b. Simulation of Complex Construction Processes. *Proceedings, 1996 Winter Simulation Conference*. Society for Computer Simulation, San Diego, CA, 1321-1328.

Ioannou, P.G. and J.C. Martinez. 1996c. Scaleable Simulation Models for Construction Operations. *Proc. 1996 Winter Simulation Conference*. Society for Computer Simulation, San Diego, CA, 1329-1336.

Martinez, J.C. 1996. STROBOSCOPE: State and Resource Based Simulation of Construction Processes. Ph.D. Dissertation, Department of Civil and Environ. Engineering, University of Michigan, Ann Arbor, MI.

Martinez, J.C. and P.G. Ioannou. 1994. General Purpose Simulation with Stroboscope. *Proceedings, 1994 Winter Simulation Conference*, Society for Computer Simulation, San Diego, CA, 1159-1166.

Martinez, J.C. and P.G. Ioannou. 1995. Advantages of the Activity Scanning Approach in the Modeling of Complex Construction Processes. *Proceedings of the 1995 Winter Simulation Conference*, Society for Computer Simulation, San Diego, CA, 1024-1031.

Moeller, G.L. and L.A. Digman. 1981. Operations Planning With VERT, *Operations Research*, (29)4: 676-697.

Morua-Padilla, E. 1986. Resource Strategies for Dynamic Construction Project Management. Ph.D. Dissertation, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI.

Pritsker A.A.B. 1977. *Modeling and Analysis using Q-GERT Networks*. Halsted Press, John Wiley & Sons, New York, New York.

Van Slyke, R. M. 1963. Monte-Carlo Methods and the PERT Problem. *Operations Research* (11)5:839-860.

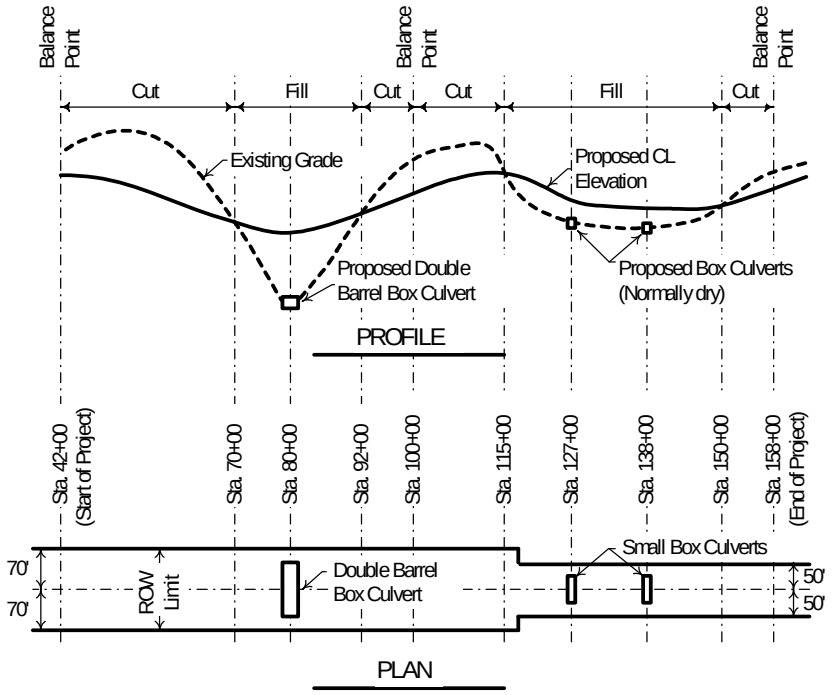
Wang, W. C. 1996. Model for Evaluating Networks Under Correlated Uncertainty – NETCOR. Ph.D. Dissertation, Department of Civil Engineering, University of California, Berkeley, CA.

AUTHOR BIOGRAPHIES

PHOTIOS G. IOANNOU is Associate Professor in the Dept. of Civil and Environmental Engineering at the Univ. of Michigan. He has received a Civil Engineer's degree from the National Technical University, Athens, Greece, in 1979; and a SMCE and Ph.D. in Civil Engineering from MIT in 1981 and 1984. From 1989-1995 he has served as Chairman of the Computing in Construction Technical Committee of the ASCE. He co-developed the UM-CYCLONE construction process simulation system with R.I. Carr, supervised the design and development of COOPS by L.Y. Liu, and chaired J.C. Martinez's Ph.D. dissertation on STROBOSCOPE. His research interests are primarily focused on the areas of decision support systems and construction process modeling.

JULIO C. MARTINEZ is an Assistant Professor in the Via Department of Civil Engineering at Virginia Tech. He received his Ph.D. in Civil Engineering at the University of Michigan in 1996; an MSE in Construction Engineering and Management from the University of Michigan in 1993; an MS in Civil Engineering from the University of Nebraska in 1987; and a Civil Engineer's degree from Universidad Catolica Madre y Maestra (Santiago, Dominican Republic) in 1986. He designed and implemented the STROBOSCOPE simulation language as part of his Ph.D. dissertation research. His research interests include discrete event simulation, scheduling of complex and risky projects, and construction management information systems.

PNET & SIMULATION ANALYSIS OF HIGHWAY PAVEMENT PROJECT



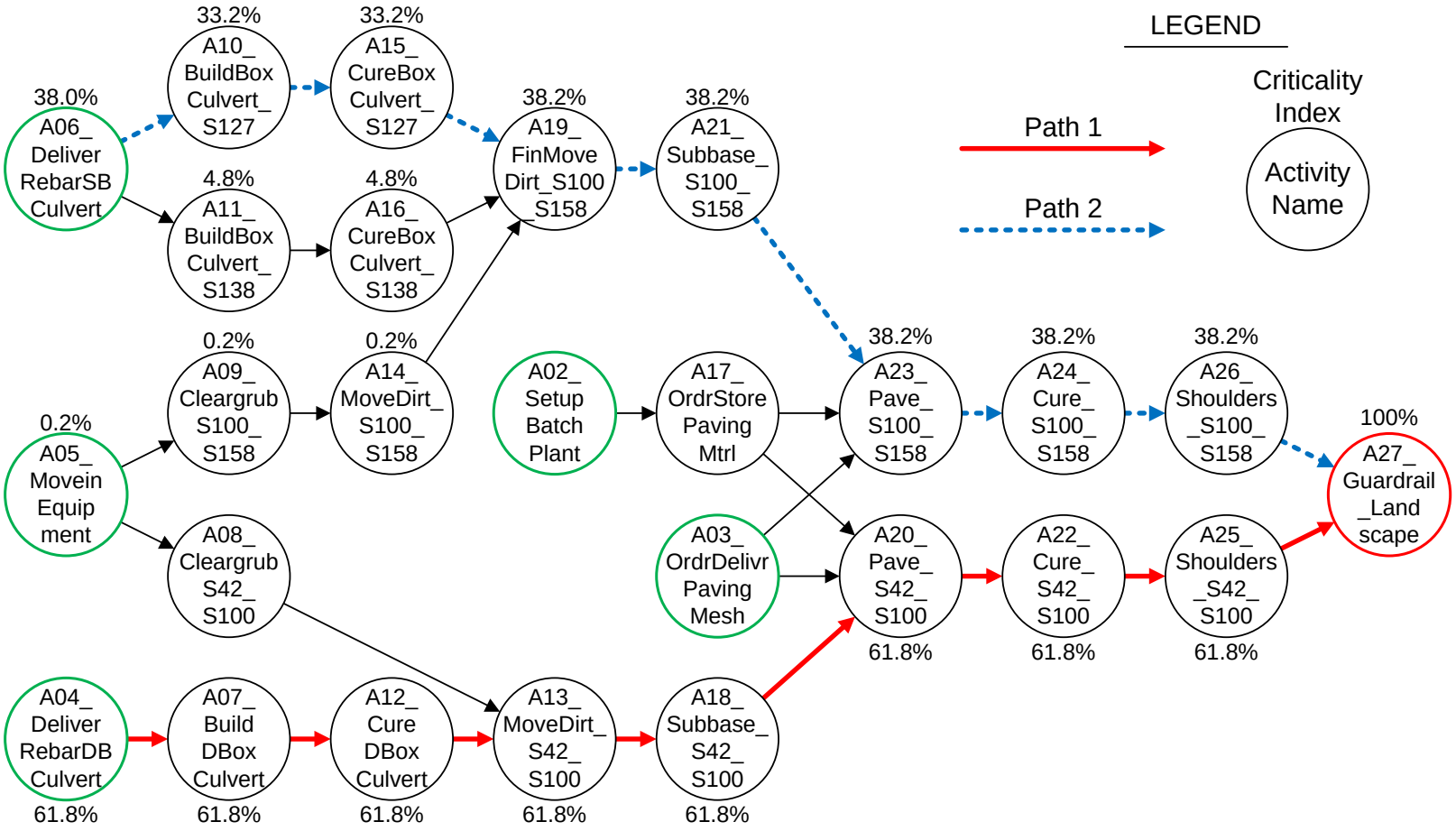
PERT Approximations

$$Mean = m = \frac{a + 4M + b}{6}$$

$$SD = \sigma = \frac{b - a}{6}$$

No.	Activity Description	a	M	b	Mean	Var	SD	Preceded by
2	Setup batch plant	0.5	2	3.5	2	0.25	0.5	
3	Order & deliver paving mesh	2	5	8	5	1	1	
4	Deliver rebar for double barrel culvert	2.5	5.5	11.5	6	2.25	1.5	
5	Move in equipment	1.5	3	4.5	3	0.25	0.5	
6	Deliver rebar for small box culvert	1	4	25	7	16	4	
7	Build double barrel culvert	4	10	16	10	4	2	4
8	Clear & grub from Sta. 42 to Sta. 100	1	2.5	7	3	1	1	5
9	Clear & grub from Sta. 100 to Sta. 158	2.5	7	11.5	7	2.25	1.5	5
10	Build box culvert at Sta. 127	1	4	13	5	4	2	6
11	Build box culvert at Sta. 138	0.5	2	9.5	3	2.25	1.5	6
12	Cure double barrel culvert	3	9	15	9	4	2	7
13	Move dirt between Sta. 42 & Sta. 100	2.5	4	11.5	5	2.25	1.5	8, 12
14	Start moving dirt between Sta. 100 & Sta. 158	1.5	3	4.5	3	0.25	0.5	9
15	Cure box culvert at Sta. 127	1.5	6	28.5	9	20.3	4.5	10
16	Cure box culvert at Sta. 138	2	5	14	6	4	2	11
17	Order & stockpile paving material	0.5	2	3.5	2	0.25	0.5	2
18	Place subbase from Sta. 42 to Sta. 100	3.6	6.1	14	7	3	1.73	13
19	Finish moving dirt between Sta. 100 & Sta. 158	1	4	13	5	4	2	14, 15, 16
20	Pave from Sta. 42 to Sta. 100	4	10	16	10	4	2	3, 17, 18
21	Place subbase from Sta. 100 to Sta. 158	2	4.53	21.87	7	11	3.31	19
22	Cure pavement from Sta. 42 to Sta. 100	2.5	5.5	11.5	6	2.25	1.5	20
23	Pave from Sta.100 to Sta. 158	3	6.75	30	10	20.3	4.5	3, 17, 21
24	Cure pavement from Sta. 100 to Sta.158	2.5	5.5	11.5	6	2.25	1.5	23
25	Place shoulders from Sta. 42 to Sta. 100	1	2.5	7	3	1	1	22
26	Place shoulders from Sta. 100 to Sta.158	1	2.5	7	3	1	1	24
27	Place guardrail & landscape	2.5	4	11.5	5	2.25	1.5	25, 26

PNET & SIMULATION ANALYSIS OF HIGHWAY PAVEMENT PROJECT



PNET & SIMULATION ANALYSIS OF HIGHWAY PAVEMENT PROJECT

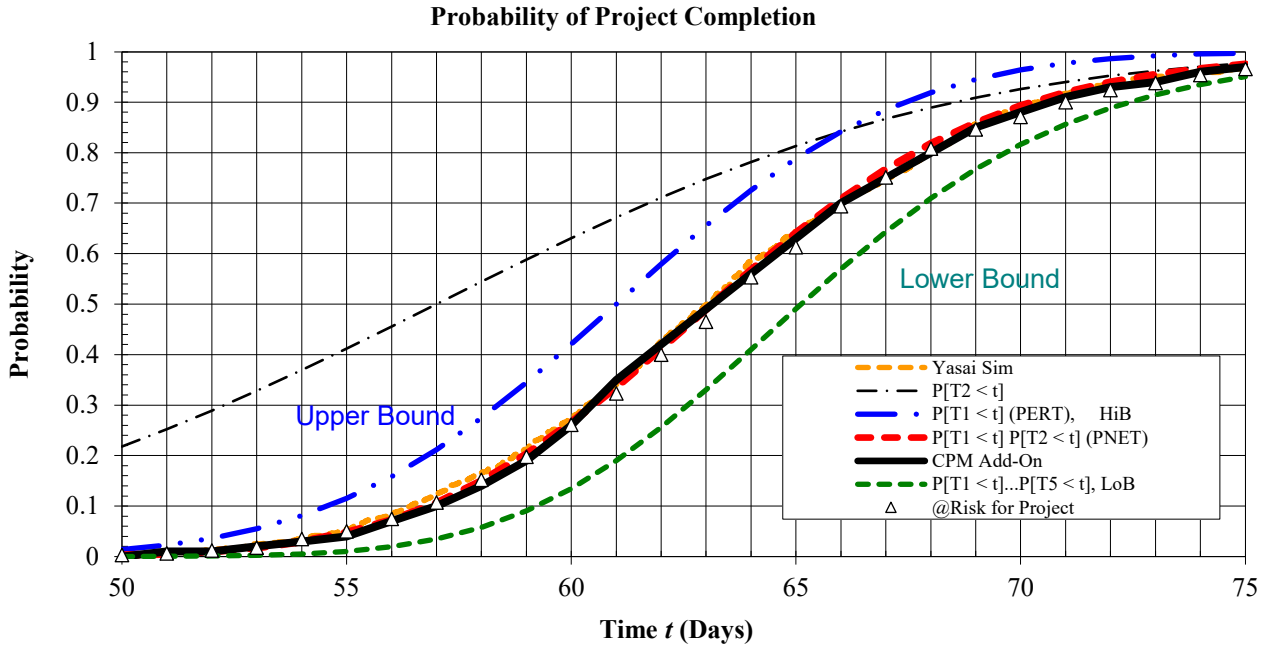
Path	Activities in path	Mean	Var	SD
1	4,7,12,13,18,20,22,25,27	61	25.00	5.00
2	6,10,15,19,21,23,24,26,27	57	80.96	9.00
3	6,11,16,19,21,23,24,26,27	52	62.96	7.93
4	5,9,14,19,21,23,24,26,27	49	43.46	6.59
5	5,8,13,18,20,22,25,27	42	16.00	4.00
6	3,20,22,25,27	29	10.50	3.24
7	3,23,24,26,27	29	26.75	5.17
8	2,17,20,22,25,27	28	10.00	3.16
9	2,17,23,24,26,27	28	26.25	5.12

Path Correlation

1 vs. (2,3,4,5)	r12	0.05	<0.5
(6,7,8,9 are too short)	r13	0.0567	<0.5
	r14	0.0683	<0.5
	r15	0.7376	>0.5
Remove Path 5. It is represented by Path 1 (keep 2,3,4)			
2 vs. (3,4)	r23	0.7943	>0.5
	r24	0.6863	>0.5
Remove Paths 3,4. They are represented by Path 2.			

Conclusion: Use only paths 1 & 2
to construct a lower bound: P[T1<t]*P[T2<t]
to approximate P[Project Duration < t] = P[T < t]
The results are shown in the table, where:
PNET: P[T<t] ~= P[T1<t]*P[T2<t]
CPM Add-On = Stroboscope simulation results

t	P[T1 < t]	P[T2 < t]	PNET	CPM Add-On
50	0.01	0.22	0.003	0.000
51	0.02	0.25	0.006	0.010
52	0.04	0.29	0.010	0.010
53	0.05	0.33	0.018	0.020
54	0.08	0.37	0.030	0.030
55	0.12	0.41	0.047	0.040
56	0.16	0.46	0.072	0.070
57	0.21	0.50	0.106	0.100
58	0.27	0.54	0.149	0.140
59	0.34	0.59	0.203	0.190
60	0.42	0.63	0.265	0.260
61	0.50	0.67	0.336	0.350
62	0.58	0.71	0.412	0.420
63	0.66	0.75	0.490	0.490
64	0.73	0.78	0.567	0.560
65	0.79	0.81	0.641	0.630
66	0.84	0.84	0.708	0.700
67	0.88	0.87	0.767	0.750
68	0.92	0.89	0.817	0.800
69	0.95	0.91	0.859	0.850
70	0.96	0.93	0.892	0.880
71	0.98	0.94	0.919	0.910
72	0.99	0.95	0.939	0.930
73	0.99	0.96	0.954	0.940
74	1.00	0.97	0.966	0.960
75	1.00	0.98	0.975	0.970

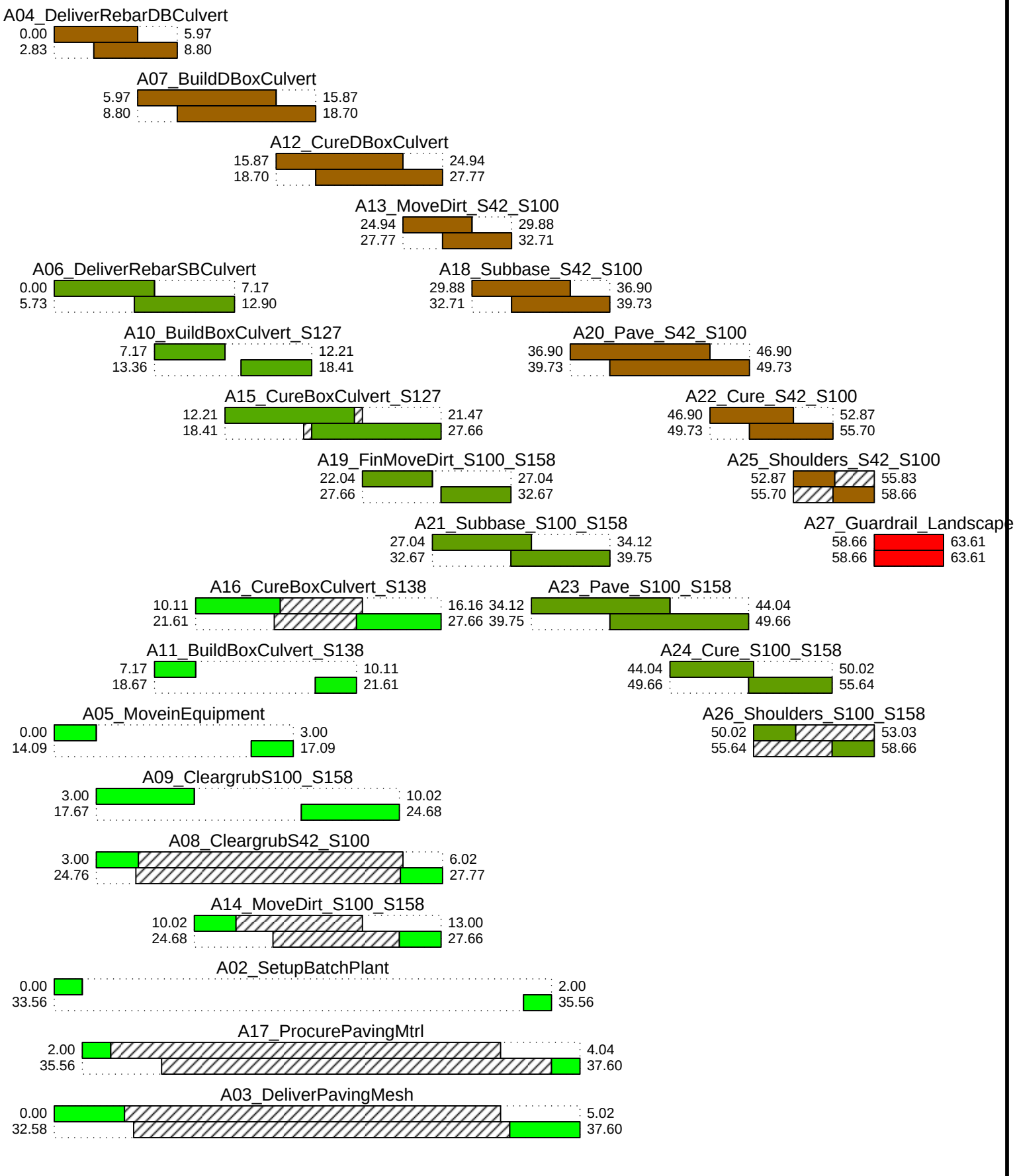


Number of replications performed : 1000
Average Project Duration : 63.61
Std. Dev. of Project Duration : 5.66
90% Confidence Interval for the Expected Project Duration: [63.315 - 63.905]

Activity	Time	ESD	LSD	EFD	LFD	FF	TF	%Crit
A02_SetupBatchPlant	2.00	0	33.56	2.00	35.56	0	33.56	0.0%
A03_OrderDeliverPavingMesh	5.02	0	32.58	5.02	37.60	26.90	32.58	0.0%
A04_DeliverRebarDBCulvert	5.97	0	2.83	5.97	8.80	0	2.83	61.8%
A05_MoveinEquipment	3.00	0	14.09	3.00	17.09	0	14.09	0.2%
A06_DeliverRebarSBCulvert	7.17	0	5.73	7.17	12.90	0	5.73	38.0%
A07_BuildDBoxCulvert	9.90	5.97	8.80	15.87	18.70	0	2.83	61.8%
A08_CleargrubS42_S100	3.01	3.00	24.76	6.02	27.77	18.93	21.76	0.0%
A09_CleargrubS100_S158	7.01	3.00	17.67	10.02	24.68	0	14.67	0.2%
A10_BuildBoxCulvert_S127	5.05	7.17	13.36	12.21	18.41	0	6.20	33.2%
A11_BuildBoxCulvert_S138	2.95	7.17	18.67	10.11	21.61	0	11.50	4.8%
A12_CureDBoxCulvert	9.07	15.87	18.70	24.94	27.77	0	2.83	61.8%
A13_MoveDirt_S42_S100	4.94	24.94	27.77	29.88	32.71	0	2.83	61.8%
A14_MoveDirt_S100_S158	2.98	10.02	24.68	13.00	27.66	9.04	14.67	0.2%
A15_CureBoxCulvert_S127	9.25	12.21	18.41	21.47	27.66	0.57	6.20	33.2%
A16_CureBoxCulvert_S138	6.05	10.11	21.61	16.16	27.66	5.88	11.50	4.8%
A17_ProcurePavingMtrl	2.04	2.00	35.56	4.04	37.60	27.89	33.56	0.0%
A18_Subbase_S42_S100	7.02	29.88	32.71	36.90	39.73	0	2.83	61.8%
A19_FinMoveDirt_S100_S158	5.00	22.04	27.66	27.04	32.67	0	5.62	38.2%
A20_Pave_S42_S100	10.00	36.90	39.73	46.90	49.73	0	2.83	61.8%
A21_Subbase_S100_S158	7.08	27.04	32.67	34.12	39.75	0	5.62	38.2%
A22_Cure_S42_S100	5.97	46.90	49.73	52.87	55.70	0	2.83	61.8%
A23_Pave_S100_S158	9.92	34.12	39.75	44.04	49.66	0	5.62	38.2%
A24_Cure_S100_S158	5.98	44.04	49.66	50.02	55.64	0	5.62	38.2%
A25_Shoulders_S42_S100	2.96	52.87	55.70	55.83	58.66	2.83	2.83	61.8%
A26_Shoulders_S100_S158	3.01	50.02	55.64	53.03	58.66	5.62	5.62	38.2%
A27_Guardrail_Landscape	4.95	58.66	58.66	63.61	63.61	0	0	100.0%

Execution Time = 2.894 seconds

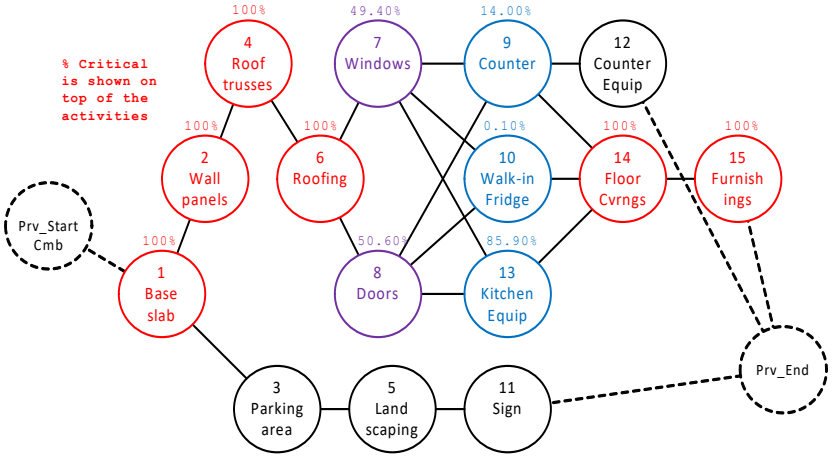
Figure 4 – Simulation Output for Highway Construction Project


A04_DeliverRebarDBCulvert
0.00 5.97
2.83 8.80
A07_BuildDBoxCulvert
5.97 15.87
8.80 18.70
A12_CureDBoxCulvert
15.87 24.94
18.70 27.77
A13_MoveDirt_S42_S100
24.94 29.88
27.77 32.71
A06_DeliverRebarSBCulvert
0.00 7.17
5.73 12.90
A18_Subbase_S42_S100
29.88 36.90
32.71 39.73
A10_BuildBoxCulvert_S127
7.17 12.21
13.36 18.41
A20_Pave_S42_S100
36.90 46.90
39.73 49.73
A15_CureBoxCulvert_S127
12.21 21.47
18.41 27.66
A22_Cure_S42_S100
46.90 52.87
49.73 55.70
A19_FinMoveDirt_S100_S158
22.04 27.04
27.66 32.67
A25_Shoulders_S42_S100
52.87 55.83
55.70 58.66
A21_Subbase_S100_S158
27.04 34.12
32.67 39.75
A27_Guardrail_Landscape
58.66 63.61
58.66 63.61
A16_CureBoxCulvert_S138
10.11 16.16
21.61 27.66
A23_Pave_S100_S158
34.12 44.04
39.75 49.66
A11_BuildBoxCulvert_S138
7.17 10.11
18.67 21.61
A24_Cure_S100_S158
44.04 50.02
49.66 55.64
A05_MoveinEquipment
0.00 3.00
14.09 17.09
A26_Shoulders_S100_S158
50.02 53.03
55.64 58.66
A09_CleargrubS100_S158
3.00 10.02
17.67 24.68
A08_CleargrubS42_S100
3.00 6.02
24.76 27.77
A14_MoveDirt_S100_S158
10.02 13.00
24.68 27.66
A02_SetupBatchPlant
0.00 2.00
33.56 35.56
A17_ProcurePavingMtrl
2.00 4.04
35.56 37.60
A03_DeliverPavingMesh
0.00 5.02
32.58 37.60
Activity description
Early Start
Late Start
Early time frame
Late time frame
Early Finish
Late Finish
Free Float
Total Float
0%
Criticality Color Coding
Hammock Color
100%
Project Statistics (1000 repl.)
Mean SD
Duration: 63.61 5.66
Cost: 0.00 0.00
STROBOSCOPE ProbSched
Statistical Bar Chart for Highway Construction Project

Simulating PERT Networks Using STROBOSCOPE
Example: FAST FOOD OUTLET ANALYSIS

Shown below is the Stroboscope input file ("CpmAddOn_FFO_Example.str")

```
VARIABLE nReplications 1000;     / Number of project replications (runs, samples)
SEED 9111964;                    / Comment out to get different results
STREAMS 15;                      / Each Activity has its own stream of random numbers
LOADADDON CpmAddOn.dll;          / This add-on allows easy modeling of PERT networks

/ The CpmAddOn.dll uses Precedence network notation and defines automatically
/ a dummy START (Prv_StartCmb) and a dummy END Activity (Prv_End).
/ Thus, the example precedence network below need not define these explicitly
/ (to have a single start and a single finish node).

/ Syntax for activity duration distribution: sPert[a,M,b,stream]

CPMACTIVITY BaseSlab sPert[3,6,12,1];
CPMACTIVITY WallPanels sPert[4,6,9,2];
CPMACTIVITY ParkingArea sPert[6,15,20,3];
CPMACTIVITY RoofTrusses sPert[1,2,5,4];
CPMACTIVITY Landscaping sPert[3,5,10,5];
CPMACTIVITY Roofing sPert[1,3,5,6];
CPMACTIVITY Windows sPert[1,2,4,7];
CPMACTIVITY Doors sPert[1,2,4,8];
CPMACTIVITY Counter sPert[3,8,10,9];
CPMACTIVITY WIFridge sPert[2,5,8,10];
CPMACTIVITY Sign sPert[3,4,6,11];
CPMACTIVITY CounterEq sPert[1,2,4,12];
CPMACTIVITY KitchenEq sPert[4,10,15,13];
CPMACTIVITY FloorCvrngs sPert[2,4,8,14];
CPMACTIVITY Furnishings sPert[5,10,15,15];

PRECEDENCE BaseSlab WallPanels;
PRECEDENCE BaseSlab ParkingArea;
PRECEDENCE WallPanels RoofTrusses;
PRECEDENCE ParkingArea Landscaping;
PRECEDENCE RoofTrusses Roofing;
PRECEDENCE Roofing Windows;
PRECEDENCE Roofing Doors;
PRECEDENCE Windows Counter;
PRECEDENCE Doors Counter;
PRECEDENCE Windows WIFridge;
PRECEDENCE Doors WIFridge;
PRECEDENCE Landscaping Sign;
PRECEDENCE Counter CounterEq;
PRECEDENCE Windows KitchenEq;
PRECEDENCE Doors KitchenEq;
PRECEDENCE WIFridge FloorCvrngs;
PRECEDENCE Counter FloorCvrngs;
PRECEDENCE KitchenEq FloorCvrngs;
PRECEDENCE FloorCvrngs Furnishings;
```




```
BINCOLLECTOR FinishByTime* 20 35 55;        / To collect data about project duration
ONEND Prv_End COLLECT FinishByTime SimTime; / Collect data about project duration
CPMREPLICATE nReplications;                 / Run simulation to collect statistics
REPORT;                                     / Produce report with BINCOLLECTOR stats
```

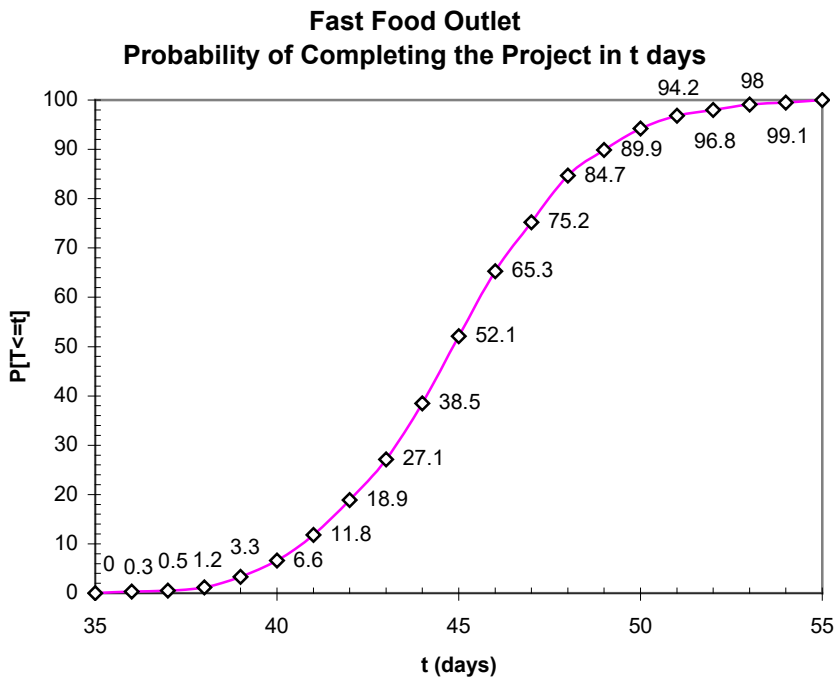
Excerpts from the corresponding Stroboscope Output File

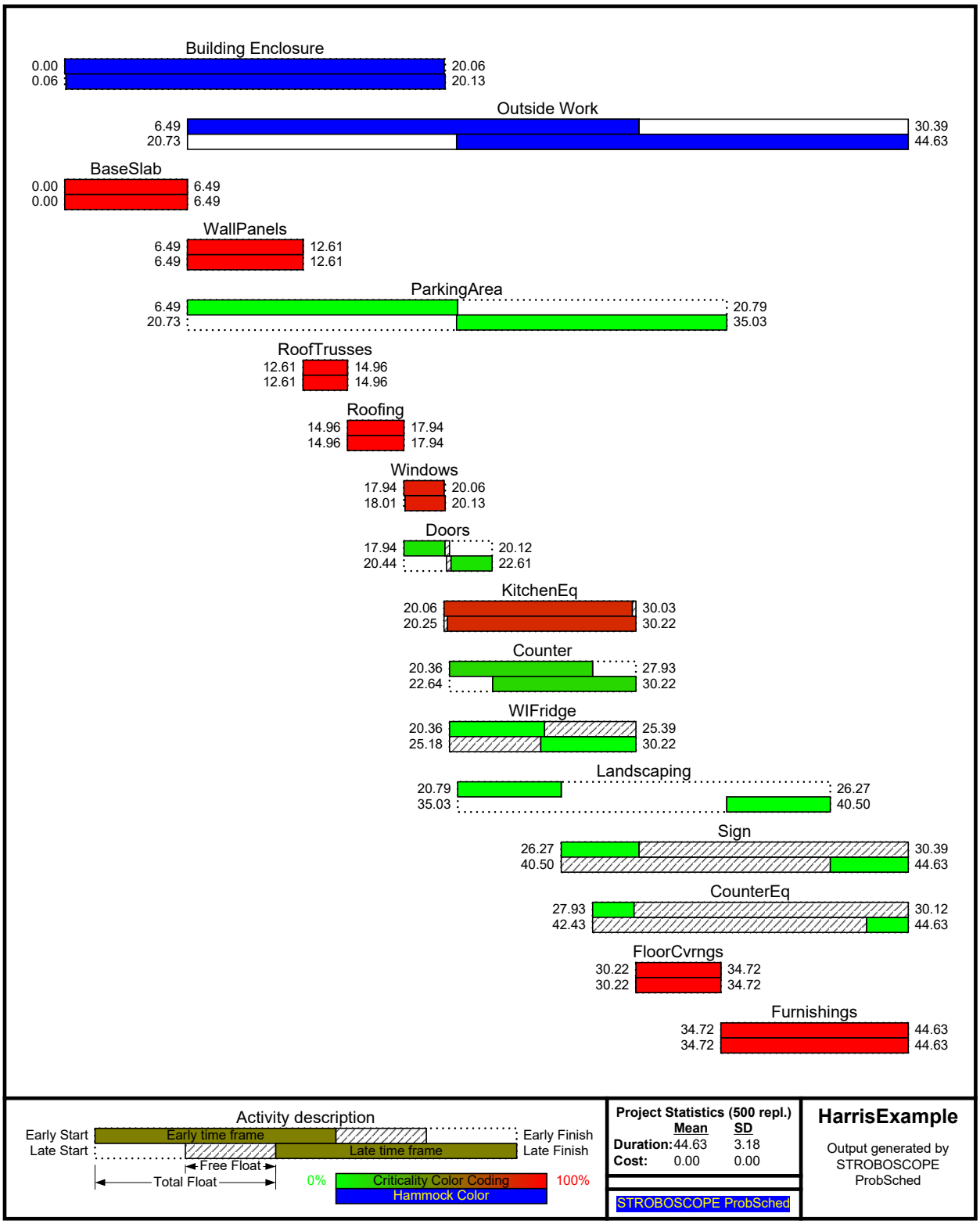
```
Stroboscope Model CpmAddOn_FFO_Example.str (65573280)

Number of replications performed : 1000
Average Project Duration         : 44.86
Std. Dev. of Project Duration    : 3.22
90% CI on project duration       : [44.70,45.03]
Average Project Cost             : 0.00
Std. Dev. of Project Cost        : 0.00
```

CPM Activity	Time	ESD	LSD	EFD	LFD	FF	TF	%Critic	Cost
BaseSlab	6.61	0.00	0.00	6.61	6.61	0.00	0.00	100.00%	0.00
WallPanels	6.16	6.61	6.61	12.77	12.77	0.00	0.00	100.00%	0.00
RoofTrusses	2.31	12.77	12.77	15.09	15.09	0.00	0.00	100.00%	0.00
Roofing	2.99	15.09	15.09	18.08	18.08	0.00	0.00	100.00%	0.00
FloorCvrngs	4.36	30.56	30.56	34.91	34.91	0.00	0.00	100.00%	0.00
Furnishings	9.95	34.91	34.91	44.86	44.86	0.00	0.00	100.00%	0.00
Doors	2.18	18.08	18.36	20.25	20.54	0.28	0.28	50.60%	0.00
Windows	2.19	18.08	18.35	20.26	20.54	0.27	0.27	49.40%	0.00
KitchenEq	9.88	20.54	20.67	30.42	30.56	0.14	0.14	85.90%	0.00
Counter	7.53	20.54	23.03	28.06	30.56	0.00	2.49	14.00%	0.00
WIFridge	4.98	20.54	25.58	25.51	30.56	5.04	5.04	0.10%	0.00
ParkingArea	14.40	6.61	20.82	21.01	35.22	0.00	14.21	0.00%	0.00
Landscaping	5.49	21.01	35.22	26.50	40.71	0.00	14.21	0.00%	0.00
Sign	4.15	26.50	40.71	30.65	44.86	14.21	14.21	0.00%	0.00
CounterEq	2.16	28.06	42.71	30.22	44.86	14.64	14.64	0.00%	0.00

Detailed statistics collected by FinishByTime

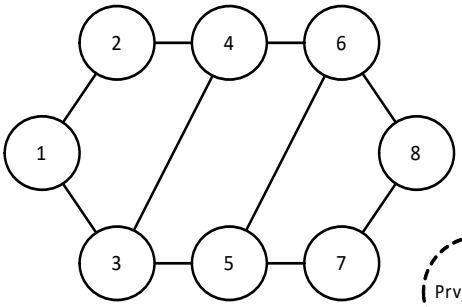
Range	Hits	%Hits
<= 35.00	0	0.00
<= 36.00	3	0.30
<= 37.00	5	0.50
<= 38.00	12	1.20
<= 39.00	33	3.30
<= 40.00	66	6.60
<= 41.00	118	11.80
<= 42.00	189	18.90
<= 43.00	271	27.10
<= 44.00	385	38.50
<= 45.00	521	52.10
<= 46.00	653	65.30
<= 47.00	752	75.20
<= 48.00	847	84.70
<= 49.00	899	89.90
<= 50.00	942	94.20
<= 51.00	968	96.80
<= 52.00	980	98.00
<= 53.00	991	99.10
<= 54.00	995	99.50
<= 55.00	1000	100.00
> 55.00	0	0.00




Building Enclosure
0.00 20.06
0.06 20.13
Outside Work
6.49 30.39
20.73 44.63
BaseSlab
0.00 6.49
0.00 6.49
WallPanels
6.49 12.61
6.49 12.61
ParkingArea
6.49 20.79
20.73 35.03
RoofTrusses
12.61 14.96
12.61 14.96
Roofing
14.96 17.94
14.96 17.94
Windows
17.94 20.06
18.01 20.13
Doors
17.94 20.12
20.44 22.61
KitchenEq
20.06 30.03
20.25 30.22
Counter
20.36 27.93
22.64 30.22
WIFridge
20.36 25.39
25.18 30.22
Landscaping
20.79 26.27
35.03 40.50
Sign
26.27 30.39
40.50 44.63
CounterEq
27.93 30.12
42.43 44.63
FloorCvrngs
30.22 34.72
30.22 34.72
Furnishings
34.72 44.63
34.72 44.63
Activity description
Early Start
Late Start
Early time frame
Late time frame
Early Finish
Late Finish
Free Float
Total Float
0%
Criticality Color Coding
Hammock Color
100%
Project Statistics (500 repl.)
Mean SD
Duration: 44.63 3.18
Cost: 0.00 0.00
STROBOSCOPE ProbSched
HarrisExample
Output generated by
STROBOSCOPE
ProbSched

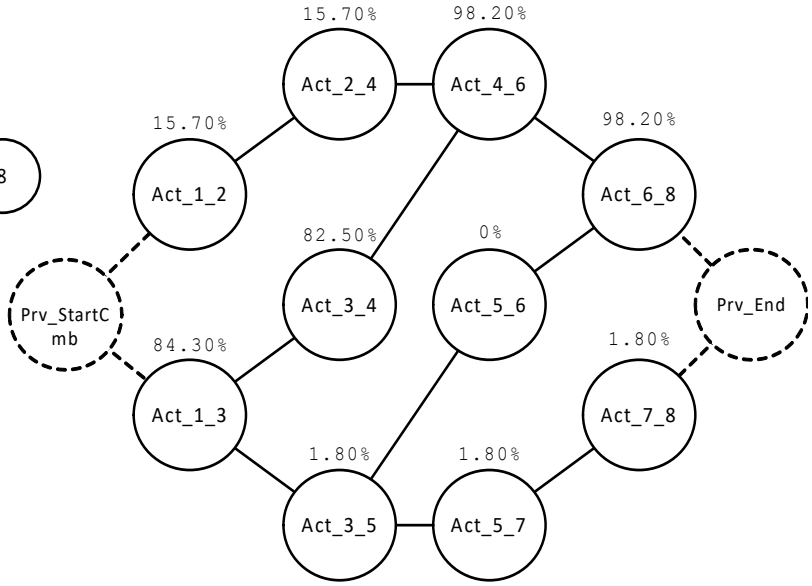
Simulating PERT Networks Using STROBOSCOPE
Example Using Perry & Greig Formulas

Original Activity on Arrow Network



Three critical paths during simulation
15.70%: 1-2-4-6-8
82.50%: 1-3-4-6-8
01.80%: 1-3-5-7-8

Equivalent Activity on Node Network



Shown below is the Stroboscope input file (PertP&G.STR)

```
VARIABLE nReplications 1000;
SEED 1964911;
STREAMS 9;
LOADADDON CpmAddOn.dll;

CPMACTIVITY Act_1_2 sPertpg[6,12,18,1];
CPMACTIVITY Act_1_3 sPertpg[13,16,19,2];
CPMACTIVITY Act_2_4 sPertpg[7,10,19,3];
CPMACTIVITY Act_3_4 sPertpg[13,14,15,4];
CPMACTIVITY Act_3_5 sPertpg[6,9,12,5];
CPMACTIVITY Act_4_6 s Pertpg[4,6,8,6];
CPMACTIVITY Act_5_6 2;
CPMACTIVITY Act_5_7 sPertpg[6,7,8,7];
CPMACTIVITY Act_6_8 sPertpg[2,4,6,8];
CPMACTIVITY Act_7_8 sPertpg[1,2,3,9];

PRECEDENCE Act_1_2 Act_2_4;
PRECEDENCE Act_1_3 Act_3_4;
PRECEDENCE Act_1_3 Act_3_5;
PRECEDENCE Act_2_4 Act_4_6;
PRECEDENCE Act_3_4 Act_4_6;
PRECEDENCE Act_3_5 Act_5_6;
PRECEDENCE Act_3_5 Act_5_7;
PRECEDENCE Act_4_6 Act_6_8;
PRECEDENCE Act_5_6 Act_6_8;
PRECEDENCE Act_5_7 Act_7_8;

BINCOLLECTOR TE_8* 10 35 45;            / To collect data about project duration, TE_8
ONEND Prv_End COLLECT TE_8 SimTime;     / Collect a sample of actual project duration
BINCOLLECTOR TE_4* 10 25 35;            / To collect data about Time to Event 4, TE_4
ONSTART Act_4_6 COLLECT TE_4 SimTime;   / Collect a sample of actual time to event 4
CPMREPLICATE nReplications;             / Run simulation to collect statistics
REPORT;                                 / Produce report including BINCOLLECTOR stats
```

Excerpts from the corresponding Stroboscope Output File

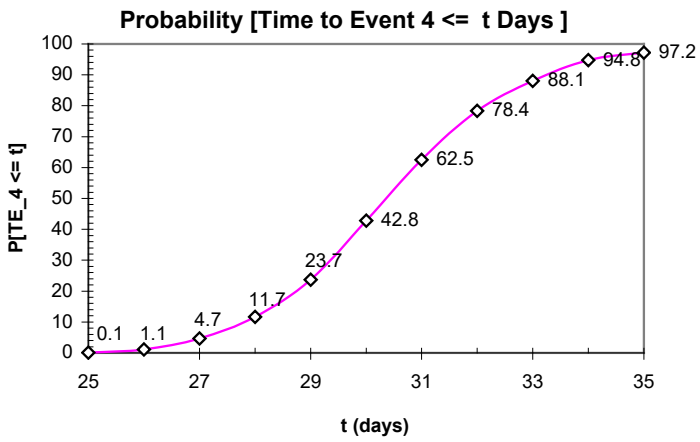
```
Stroboscope Model PertP&G.str (1936472576)

Number of replications performed : 1000
Average Project Duration         : 40.60
90% CI on project duration       : [40.46,40.75]
```

CPM Activity	Time	ESD	LSD	EFD	LFD	FF	TF	%Critic
Act_7_8	1.97	32.07	38.63	34.05	40.60	6.56	6.56	1.80%
Act_1_2	12.08	0.00	6.39	12.08	18.47	0.00	6.39	15.70%
Act_1_3	16.00	0.00	0.47	16.00	16.48	0.00	0.47	84.30%
Act_2_4	12.03	12.08	18.47	24.11	30.51	6.37	6.39	15.70%
Act_3_4	14.01	16.00	16.50	30.01	30.51	0.47	0.49	82.50%
Act_3_5	9.07	16.00	22.54	25.07	31.61	-0.00	6.54	1.80%
Act_4_6	6.01	30.49	30.51	36.50	36.51	0.00	0.02	98.20%
Act_5_6	2.00	25.07	34.51	27.07	36.51	9.43	9.44	0.00%
Act_5_7	7.00	25.07	31.63	32.07	38.63	0.00	6.56	1.80%
Act_6_8	4.09	36.50	36.51	40.59	40.60	0.02	0.02	98.20%

Collector	Samples	Mean	SD	Sum	Min	Max
ProjectDur	1000	40.60	2.83	40602.76	32.55	52.45
TE_4	1000	30.49	2.22	30489.05	24.98	39.59
TE_8	1000	40.60	2.83	40602.76	32.55	52.45

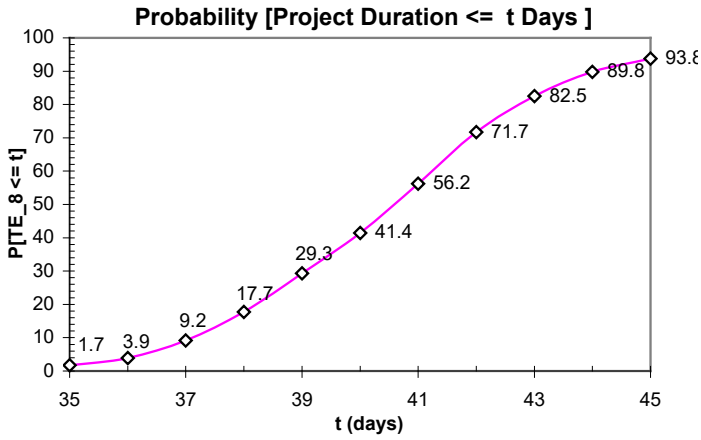
Detailed statistics collected by TE_4

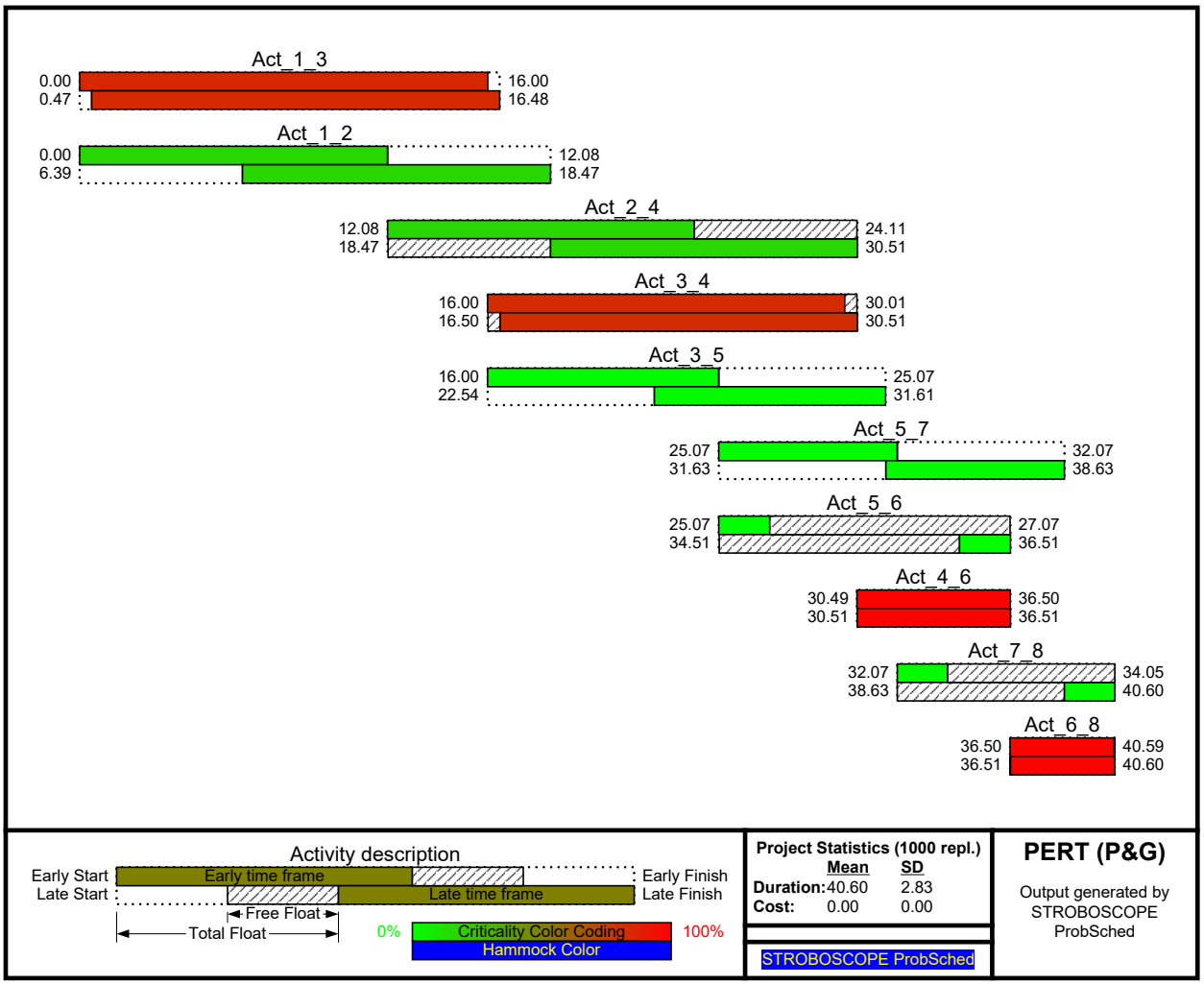
Range		Hits	%Hits
<=	25	1	0.1
<=	26	11	1.1
<=	27	47	4.7
<=	28	117	11.7
<=	29	237	23.7
<=	30	428	42.8
<=	31	625	62.5
<=	32	784	78.4
<=	33	881	88.1
<=	34	948	94.8
<=	35	972	97.2
>	35	28	2.8

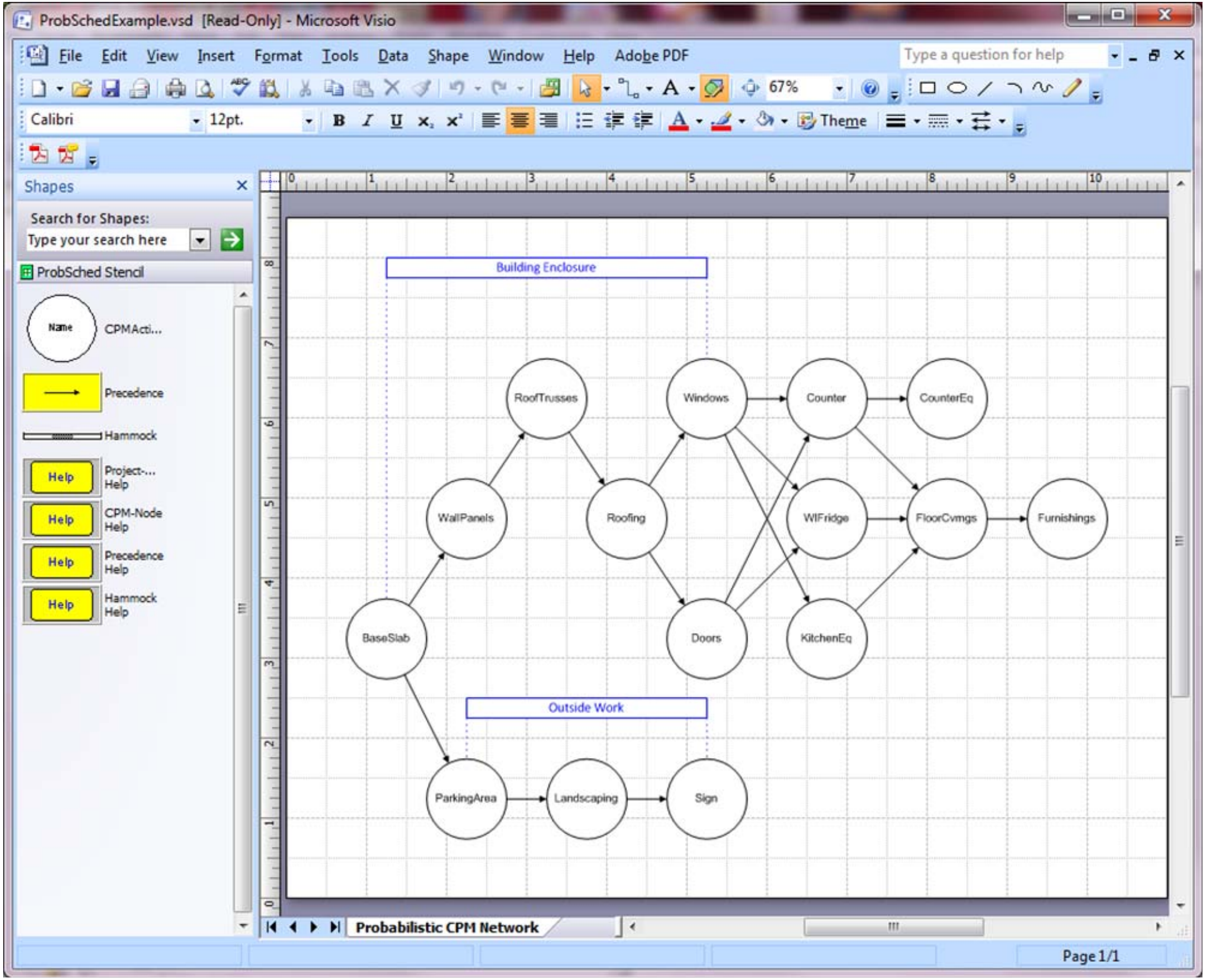


Detailed statistics collected by TE_8

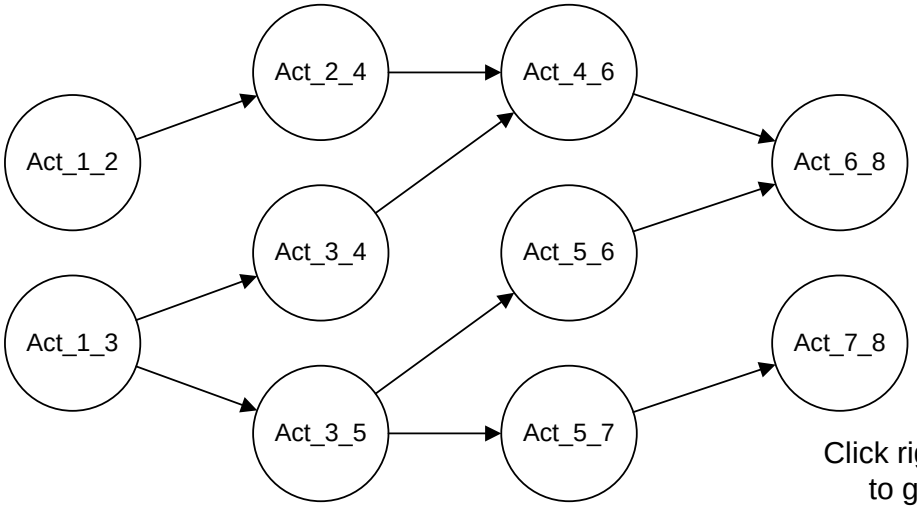
Range		Hits	%Hits
<=	35	17	1.7
<=	36	39	3.9
<=	37	92	9.2
<=	38	177	17.7
<=	39	293	29.3
<=	40	414	41.4
<=	41	562	56.2
<=	42	717	71.7
<=	43	825	82.5
<=	44	898	89.8
<=	45	938	93.8
>	45	62	6.2



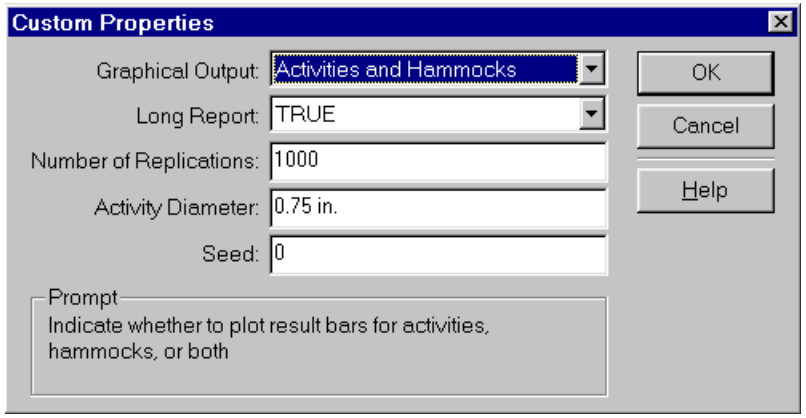

Act_1_3
0.00
0.47
16.00
16.48
Act_1_2
0.00
6.39
12.08
18.47
Act_2_4
12.08
18.47
24.11
30.51
Act_3_4
16.00
16.50
30.01
30.51
Act_3_5
16.00
22.54
25.07
31.61
Act_5_7
25.07
31.63
32.07
38.63
Act_5_6
25.07
34.51
27.07
36.51
Act_4_6
30.49
30.51
36.50
36.51
Act_7_8
32.07
38.63
34.05
40.60
Act_6_8
36.50
36.51
40.59
40.60
Activity description
Early Start
Late Start
Early time frame
Late time frame
Early Finish
Late Finish
Free Float
Total Float
0%
Criticality Color Coding
Hammock Color
100%
Project Statistics (1000 repl.)
Mean
SD
Duration: 40.60 2.83
Cost: 0.00 0.00
STROBOSCOPE ProbSched
PERT (P&G)
Output generated by
STROBOSCOPE
ProbSched


ProbSchedExample.vsd [Read-Only] - Microsoft Visio
File Edit View Insert Format Tools Data Shape Window Help Adobe PDF
Type a question for help
Calibri
12pt.
Theme
Shapes
Search for Shapes:
Type your search here
ProbSched Stencil
Name
CPMActi...
Precedence
Hammock
Help
Project-... Help
CPM-Node Help
Precedence Help
Hammock Help
Building Enclosure
RoofTrusses
Windows
Counter
CounterEq
WallPanels
Roofing
WIFridge
FloorCvrngs
Furnishings
BaseSlab
Doors
KitchenEq
Outside Work
ParkingArea
Landscaping
Sign
Probabilistic CPM Network
Page 1/1

SIMULATING PERT NETWORKS USING STROBOSCOPE
Example Using the ProbSched GUI and Perry & Greig Formulas



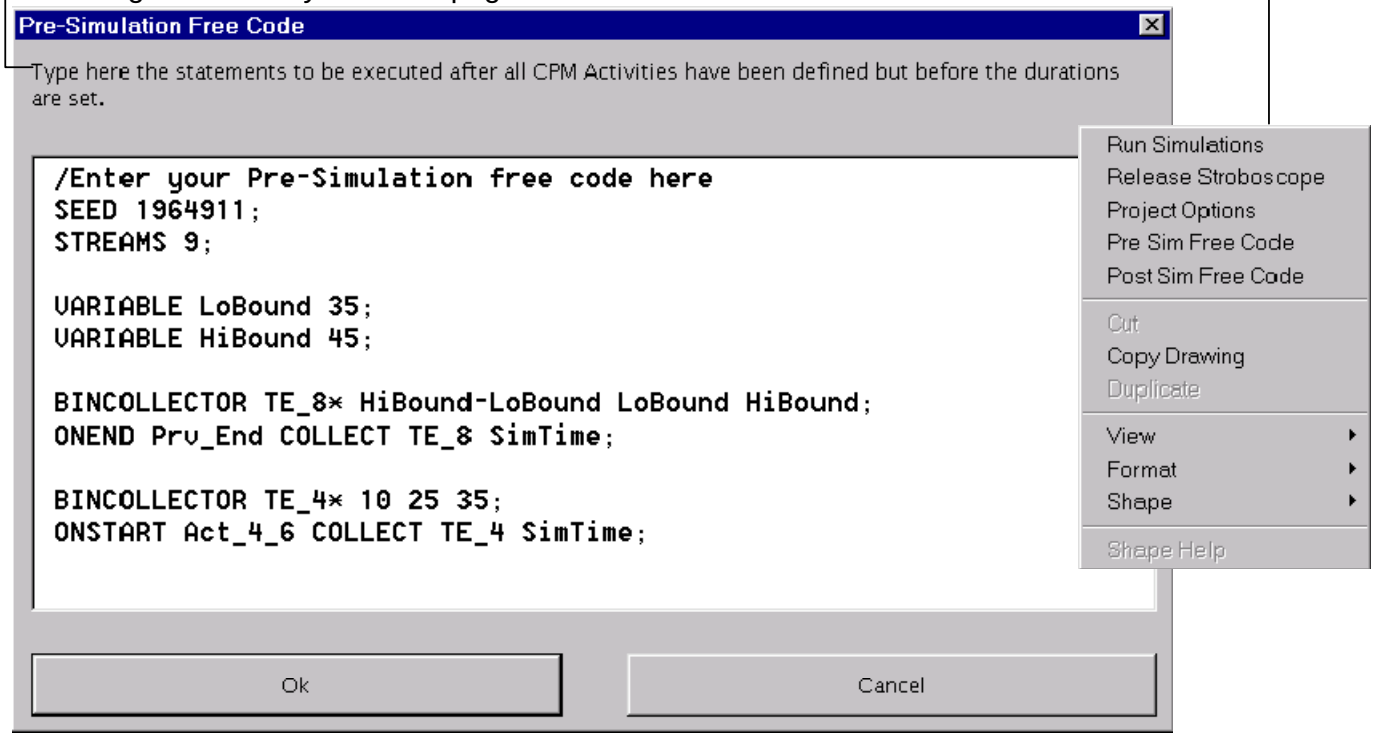
Click right button anywhere on page and select Project Options to set

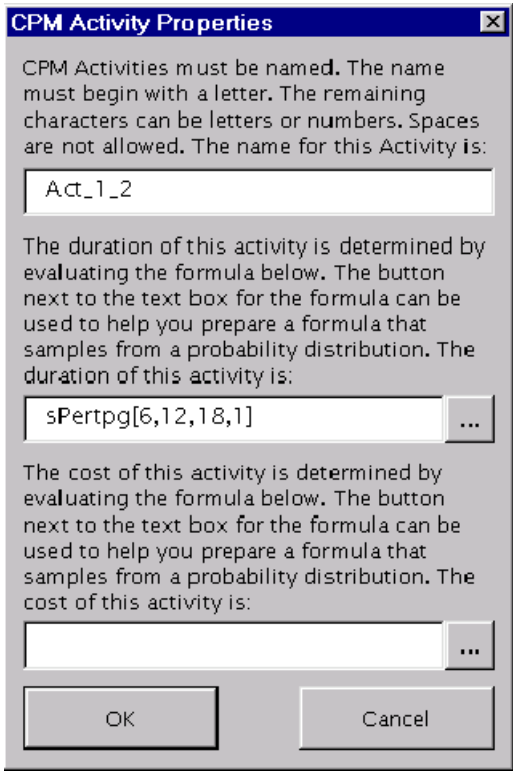


Click right button anywhere on page to get to simulation control menu

Drop an activity on the page or double-click an exisitng activity to set its attributes

Click right button anywhere on page and select Pre-Sim Free Code to...





PERT Simulation Example (Harris, Fig. 12.8) using EXCEL and @RISK

E[W], SD[W] are given by the orginal PERT formulas

W = Act. Dur. ~ BETA(a1,a2;a:b)

Activity on Arrow Table X = Standardized Act. Dur. ~ BETA(a1,a2;0:1)

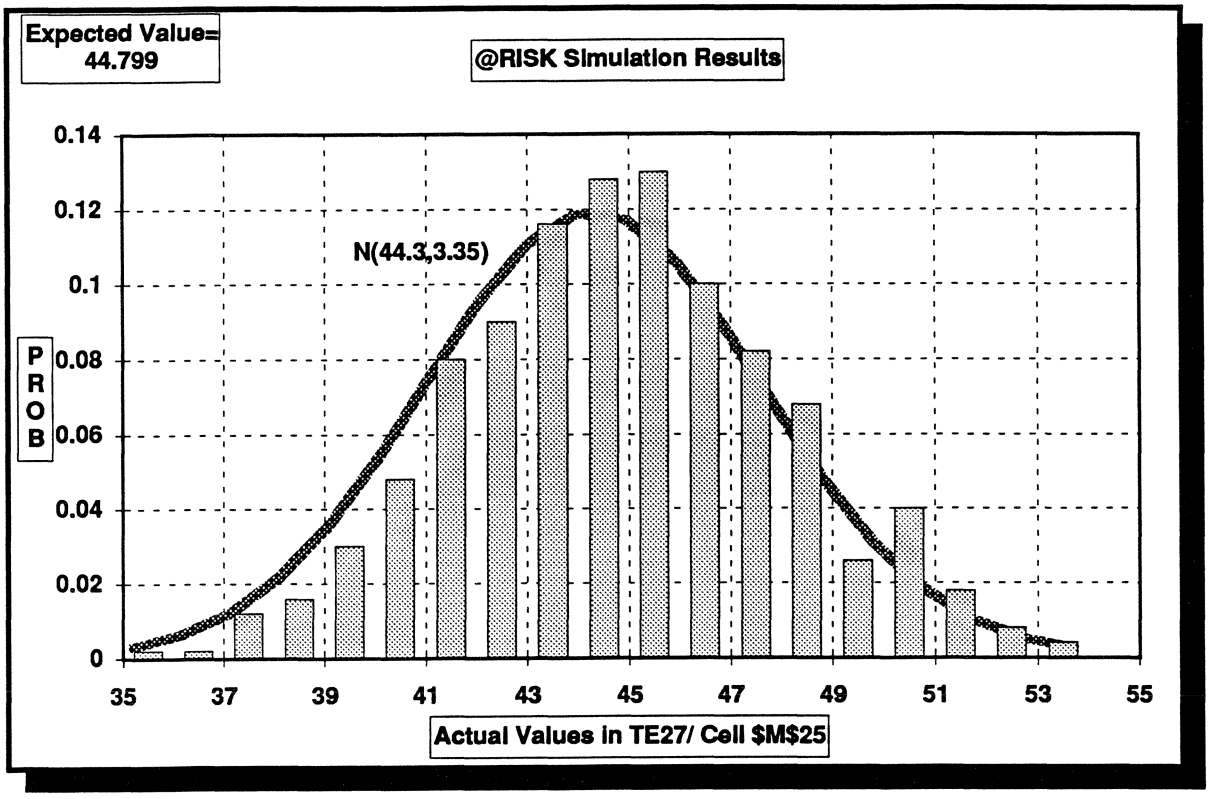
i	j	a	M	b	E[W]	SD[W]	E[X]	a1	a2	W	ESDij	EFDij
1	3	3	6	12	6.5	1.5	0.389	2.938	4.617	6.5	0	6.5
3	5	4	6	9	6.167	0.833	0.433	3.397	4.443	6.167	6.5	12.67
3	7	6	15	20	14.33	2.333	0.595	4.568	3.106	14.33	6.5	20.83
5	9	1	2	5	2.333	0.667	0.333	2.333	4.667	2.333	12.67	15
7	17	3	5	10	5.5	1.167	0.357	2.595	4.671	5.5	20.83	26.33
9	11	1	3	5	3	0.667	0.5	4	4	3	15	18
11	13	0	0	0	0	0				0	18	18
11	15	1	2	4	2.167	0.5	0.389	2.938	4.617	2.167	18	20.17
13	15	1	2	4	2.167	0.5	0.389	2.938	4.617	2.167	18	20.17
15	19	3	8	10	7.5	1.167	0.643	4.671	2.595	7.5	20.17	27.67
15	21	0	0	0	0	0				0	20.17	20.17
15	23	2	5	8	5	1	0.5	4	4	5	20.17	25.17
17	27	3	4	6	4.167	0.5	0.389	2.938	4.617	4.167	26.33	30.5
19	23	0	0	0	0	0				0	27.67	27.67
19	27	1	2	4	2.167	0.5	0.389	2.938	4.617	2.167	27.67	29.83
21	23	4	10	15	9.833	1.833	0.53	4.225	3.742	9.833	20.17	30
23	25	2	4	8	4.333	1	0.389	2.938	4.617	4.333	30	34.33
25	27	5	10	15	10	1.667	0.5	4	4	10	34.33	44.33

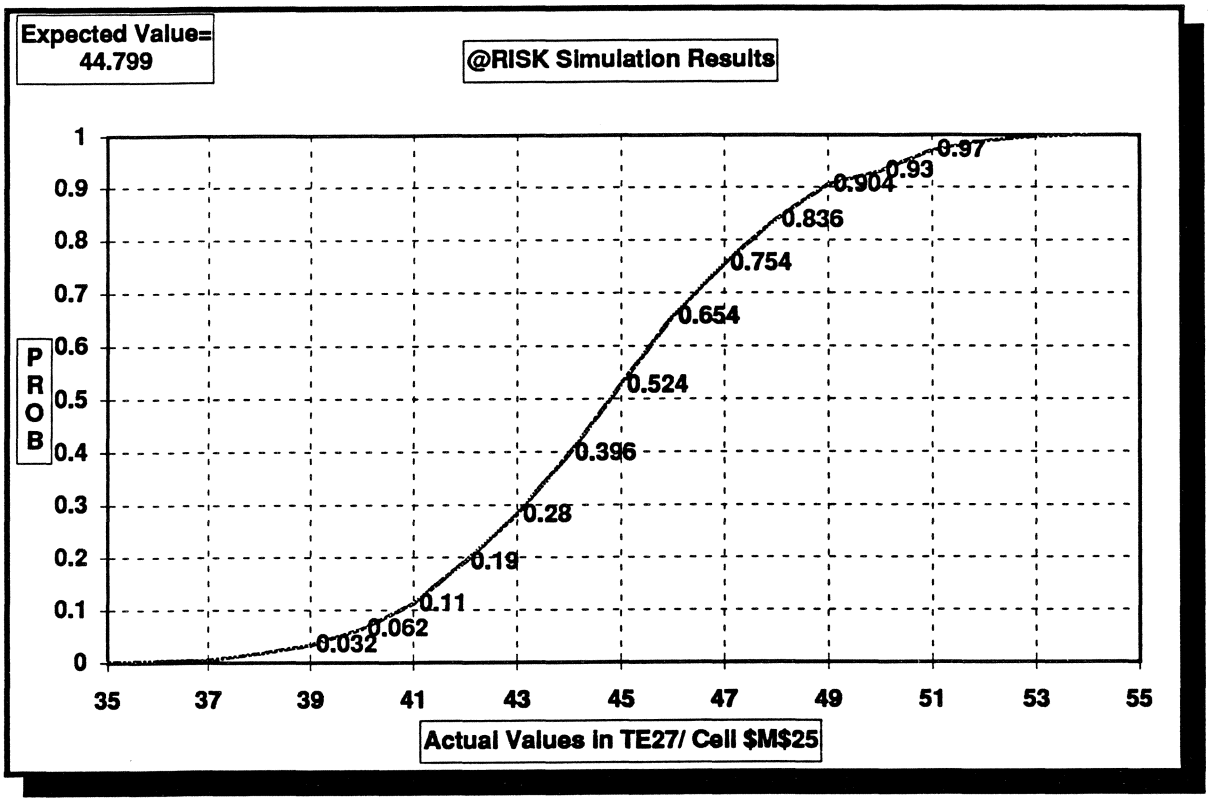
TE27= 44.33

Formulas

E[W] = (a+4*M+b)/6
SD[W] = (b - a) / 6
E[X] = (E[W] - a) / (b-a)
E[X] = (4*M + b -5*a) / {6*(b-a)}
a1 = {36 * E[X] * (1-E[X]) - 1} * E[X]
a2 = {36 * E[X] * (1-E[X]) - 1} * (1 - E[X])
W = a + (b-a) * RiskBeta(a1,a2)

TE27 = MAX(EFD(17,27), EFD(25,27), EFD(19,27))


Expected Value=
44.799
@RISK Simulation Results
N(44.3,3.35)
PROB
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
35
37
39
41
43
45
47
49
51
53
55
Actual Values in TE27/ Cell M25


Expected Value=
44.799
@RISK Simulation Results
PROB
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.032
0.062
0.11
0.19
0.28
0.396
0.524
0.654
0.754
0.836
0.904
0.93
0.97
35
37
39
41
43
45
47
49
51
53
55
Actual Values in TE27/ Cell M25

Simulation Statistics	
Date: 10/6/93 at 17:22	
Iterations: 500	
Simulations: 1	
Worksheet: '3ERTHARR.XLS'	
Output Range: TE27	
Cell: M25	
Minimum=	35.18954086
Maximum=	53.723526
Mean=	44.79873131
Std Deviation=	3.096827805
Variance=	9.590342452
Skewness=	0.044354588
Kurtosis=	2.85043962

Percentile Values	
5Perc=	39.81210938
10Perc=	40.87290192
15Perc=	41.48642941
20Perc=	42.20258179
25Perc=	42.70581913
30Perc=	43.13938751
35Perc=	43.61952267
40Perc=	44.03520203
45Perc=	44.4179533
50Perc=	44.78309822
55Perc=	45.09986973
60Perc=	45.52222824
65Perc=	45.84908485
70Perc=	46.35101128
75Perc=	46.861063
80Perc=	47.29953308
85Perc=	48.16351395
90Perc=	48.76297951
95Perc=	50.11289501

Target Values:

Value#1=	35	45
Prob#1=	0.00%	53.05%
Value#2=	36	46
Prob#2=	0.10%	66.09%
Value#3=	37	47
Prob#3=	0.27%	76.77%
Value#4=	38	48
Prob#4=	0.97%	84.13%
Value#5=	39	49
Prob#5=	2.95%	91.03%
Value#6=	40	50
Prob#6=	6.02%	94.53%
Value#7=	41	51
Prob#7=	10.71%	97.41%
Value#8=	42	52
Prob#8=	18.89%	99.08%
Value#9=	43	53
Prob#9=	28.00%	99.79%
Value#10=	44	54
Prob#10=	39.78%	100.00%