Basic notions of probability theory are applied to problems of performance analysis of on-line real-time systems.

Frequently used is APL as both an analytical tool and as a scratch pad in working out the examples.

Elements of probability for system design

by A. O. Allen

Modern on-line computer systems are complex and, therefore, often difficult to analyze. This paper aims at providing a directed review of probability theory and problems that are necessary for further study of queuing theory and queuing models. The approach here is to introduce the basic concepts through practical examples.

It is important to have access to a computer when designing and analyzing computer systems. An APL terminal system was used in writing this paper. Therefore, suggestions, examples, and calculations are given in terms of that system. Of course, any programming language with computational capability, such as FORTRAN, ALGOL, BASIC or PL/1, may be used to write programs for expressing the algorithms or models of this paper.

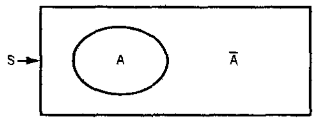
Basic concepts of probability theory

The most basic concept in probability theory is that of a *sample space*. The phenomenon under study is assumed to be described as an experiment that can be repeated under the same conditions. A *sample space* of such an experiment is a set S of *sample points* or *elementary events* so chosen that the outcome of each event of the experiment corresponds to exactly one element of the set S and vice versa. The concept of sample space is illustrated by the following example.

Example 1. We are interested in polling a communication line that is not busy, and on which there are six terminals. A sample space S for the experiment of polling terminals 1 through 6 might be the set of six-tuples $(x_1, x_2, x_3, x_4, x_5, x_6)$. Here each x_i is either 1 or 0 indicating that terminal number i is (1) or is not (0) ready to send a message to the computer. Thus the sample point (1, 0, 0, 1, 0, 1) corresponds to the outcome that terminals numbered 1, 4, and 6 are ready to transmit a message, while terminals 2, 3, and 5 are not ready. The six-terminal sample space consists of $2^6 = 64$ sample points. Composite events, such as the event that five of the terminals are ready to transmit, may also be of interest. This event consists of the six sample points (0, 1, 1, 1, 1, 1), (1, 0, 1, 1, 1, 1), (1, 1, 0, 1, 1, 1), (1, 1, 1, 0, 1, 1), (1, 1, 1, 1, 0, 1), and (1, 1, 1, 1, 0).

In general, an *event* is any subset of the sample space S, that is, any collection of sample points from S. An event where A consists of no sample points is written $A = \emptyset$, in which $\emptyset$ is the *empty set*. Sometimes called the *impossible event* in probability theory, $\emptyset$ is paradoxically considered to be a perfectly acceptable event, as is S itself.

Figure 1 Event A and its complement $\bar{A}$



To every event A there corresponds the complementary event $\bar{A}$, also called the *complement* of A, which consists of all points of S not contained in A and is defined by the condition "A does not occur." Thus, in particular, $\bar{S} = \emptyset$ and $\bar{\emptyset} = S$. This concept is illustrated in Figure 1.

Figure 2 Event $A \cup B$

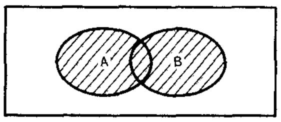
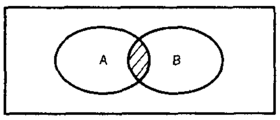
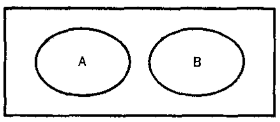


Figure 3 Event $A \cap B$



With each two events A and B, there can be associated two new events that correspond to the intuitive ideas that either A or B occur and that both A and B occur. The first of these events is designated $A \cup B$, (that is, A or B), and consists of all sample points that belong to A or to B (or possibly both). The second event, designated $A \cap B$, (that is, A and B), consists of those sample points that belong both to A and to B. If A and B have no points in common, that is, if $A \cap B = \emptyset$, then A and B are *mutually exclusive*. These concepts are illustrated in Figures 2, 3, and 4. In Figure 2, $A \cup B$ is represented by the shaded area. $A \cap B$ is shaded in Figure 3. In Figure 4, A and B are mutually exclusive events.

Figure 4 Mutually exclusive events A and B



Example 2. Suppose the sample space is that of Example 1. Let A be an event in which at least two terminals are in the ready state, and let B be an event in which not more than four terminals are in the ready state. Then $A \cup B$ is the whole space S, and $A \cap B$ is the collection of elements of S in which two, three, or four of the components x_i are 1. If C is the event in which exactly one of the terminals is in the ready state, then A and C are mutually exclusive events.

Combinatorial analysis

To be able to discuss examples of more practical value, we now proceed to some results from combinatorial analysis. Two of the basic concepts in combinatorial analysis are those of permutation and combination. A *permutation* is an ordered selection of objects from a set S. A *combination* is an unordered selection of objects from a set S. Repetitions may or may not be allowed. Unless a statement is made to the contrary, it is assumed in this paper that repetitions are not allowed in permutations and combinations.

Example 3. Let S consist of the three letters x, y, and z. There are the following nine two-letter permutations in S, with repetitions permitted: xx, xy, xz, yy, yx, yz, zz, zx, zy. And there are the following six permutations without repetitions: xy, xz, yx, yz, zx, zy. There are also the following six combinations with repetitions permitted: xx, yy, zz, xy, xz, yz. In these combinations the permutations xy and yx, for example, are not distinguished because combinations are unordered. There are three combinations without repetitions: xy, xz, yz.

multiplication principle

Many combinatorial formulas can be derived from the *multiplication principle*. Suppose an operation can be broken up into two phases in a particular way. The first phase can be performed in m different ways. After the operation has been performed in any one of these ways, the second phase can be performed in n different ways. Then the whole operation can be performed in $m \times n$ different ways.

The product of all positive integers from 1 through n is called "n factorial," and denoted by $n!$ Thus $n!$ is expressed as follows:

$$n! = n \times (n-1) \times (n-2) \cdots 3 \times 2 \times 1$$

Zero factorial 0! is defined as 1.

If n is a positive integer and r is a nonnegative integer r, with $r \leq n$, a *binomial coefficient* $\binom{n}{\mathrm{r}}$ is defined as follows:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

The name binomial coefficient derives from the binomial formula

$$(x + y(^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 \cdots + \binom{n}{n}y^n$$

results of combinatorial theory

The following are the summarized results of combinatorial theory that are needed for this discussion of probability:

- The number of permutations of n objects, taken r at a time, without repetitions is

$$P(n, r) = \frac{n!}{(n-r)!}$$

- The number of permutations of n objects taken r at a time, with repetitions allowed, is n^r.
- The number of combinations of n objects taken r at a time without repetition is

$$C(n, r) = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

- The number of combinations of n objects taken r at a time with repetition permitted is

$$\binom{n+r-1}{n-1} = \binom{n+r-1}{r}$$

In early or classical probability theory, all sample spaces were considered to be finite, and each sample point was considered to be equally likely to occur. The definition of the probability P of an event A was defined by the following expression:

$$\mathrm{P}[A] = \frac{n_A}{n}$$

Here n_A is the number of sample points in A (called points favorable to event A) and n is the total number of sample points. This definition is still valid for the case in Example 4.

Example 4. There are $2^6 = 64$ sample points by the combinatorial theorem if no assumption is made about the number of terminals ready. Assume now that exactly three of the six terminals are ready to transmit. Thus the sample space consists of all six-tuples $(x_1, x_2, x_3, x_4, x_5, x_6,)$ for which exactly three of the x_i are 1 (and three of the x_i are 0). The number of sample points in the new sample space is the number of ways of choosing the three components that are 1, that is, the number of ways of choosing three integers from the integers 1, 2, 3, 4, 5, 6, which is computed as follows:

$$\binom{6}{3} = \frac{6!}{3!3!} = 20$$

Assume that each sample point is equally likely, i.e., that each terminal is as likely to be ready to transmit as any other. Let A_1, A_2, A_3, and A_4 be the events that the 1st, 2nd, 3rd, or 4th terminal polled is the first terminal polled that is ready to transmit. (The terminals are polled in the sequence 1, 2, 3, 4, 5, 6.) The probabilities of these events are calculated as follows: A_1 occurs only if $x_1 = 1$, which means that the remaining two 1's must be distributed in the five positions 2 through 6. Hence, the number of sample points in A_1 and the probability of occurence of A_1 are computed as follows:

$$n_{A_1} = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

and

$$\text{P}[A_1] = \frac{n_{A_1}}{n} = \frac{10}{20} = 0.5$$

A_2 occurs only if $x_1 = 0$, $x_2 = 1$, and the remaining two 1's are distributed in the remaining four positions 3 through 6. Hence,

$$\text{P}[A_2] = \frac{n_{A_2}}{n} = \frac{\binom{4}{2}}{20} = \frac{6}{20} = 0.30$$

Proceeding in this way, the probabilities of events A_3 and A_4 are computed as follows:

$$\text{P}[A_3] = \frac{\binom{3}{2}}{20} = \frac{3}{20} = 0.15$$

and

$$\text{P}[A_4] = \frac{\binom{2}{2}}{20} = \frac{1}{20} = 0.05$$

probability measure

For a general sample space, it is no longer assumed that each point is equally likely. In Example 4, the first terminal might be used much more than the others, so that the event "terminal 1 is ready to transmit" should have a probability greater than the event "terminal 2 is ready to transmit." In modern probability theory, a *probability measure* P, which assigns a number P[A] to each event A in a sample space S, is assumed to exist and to satisfy the following axioms or rules:

P1. $\text{P}[A] \geq 0$ for all events A

P2. $\text{P}[S] = 1$

P3. $\text{P}[A \cup B] = \text{P}[A] + \text{P}[B]$ if A and B are mutually exclusive events

Axioms P1, P2, and P3 are certainly satisfied for the classical sample spaces. The added generality provided by probability measures beyond the simple classical theory allows many more problems to be studied using probability theory. The following are immediate consequences of the previous axioms:

R1. $\text{P}[\emptyset] = 0$

R2. $\text{P}[A] = 1 - \text{P}[\overline{A}]$

R3. $\text{P}[A \cup B] = \text{P}[A] + \text{P}[B] - \text{P}[A \cap B]$

conditional probability

Conditional probability is another useful concept, wherein, if $P[A] > 0$, the probability of event B, given that the event A has occurred, is defined as follows:

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

This definition agrees with the intuitive notion of the re-evaluation of the probability of B in the light of the information that A has occurred.

This relationship is sometimes written as follows:

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

and is called the *multiplication rule*. The multiplication rule is stated formally as the probability of the joint occurence of events A and B, and is expressed as follows:

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

Example 5. The supervisor of a group of programmers decides to randomly select two of the five programs that were run one day for his analysis. Three of the five programs were written by programmer Smith and two by programmer Jones. What is the probability that the second program selected was written by Smith, given that Smith wrote the first program selected?

Since, after the first program is selected, half the remaining programs were written by Smith, the probability is 0.5. The previously given definition of conditional probability leads to the same result. Let A be the event that the first program selected was written by Smith, and B the event that the second program selected is also Smith's. Then, by the multiplication rule, the probability of event B, given event A, is computed as follows:

$$P[A \cap B] = \frac{3}{5} \times \frac{2}{4} = \frac{3}{10}$$

Hence,

$$P[B|A] = \frac{P[A \cap B]}{P[A]} = \frac{3/10}{3/5} = 0.5$$

theorem of total probabilities

It is usually much easier to compute conditional probabilities than ordinary probabilities because more information is available about conditional probabilities. The main use of conditional probability is to unravel a set of conditional probabilities to calculate the unconditional probability of an event by the theorem of total probabilities.

Theorem of Total Probabilities. Let $A_1, A_2, \cdots, A_n$ be events such that $A_i \cap A_j = \emptyset$, if $i \neq j$

$P[A_i] > 0$ for all i

and

$$A_1 \cup A_2 \cdots \cup A_n = S$$

Such a family of events is called a *partition of S*. Then for any event A we have the following probability:

$$P[A] = P[A_1]P[A|A_1] + P[A_2]P[A|A_2] + \cdots P[A_n]P[A|A_n]$$

Example 6. Suppose messages arrive at a computer from three different sources with probabilities 0.2, 0.5, and 0.3, respectively. Also the probabilities that messages from sources 1, 2, or 3 will exceed 100 characters in length are 0.2, 0.8, and 0.6, respectively. What is the probability that the next message received will exceed 100 characters in length?

Let A_1, A_2, and A_3 be the events that a message is received from source 1, 2, or 3, respectively. Let A be the event that the next message exceeds 100 characters in length. Then, by the theorem of total probabilities, we have the following computation of the probability of A:

$$P[A] = P[A_1]PA|A_1] + P[A_2]P[A|A_2] + P[A_3]P[A|A_3]$$

$$= 0.2 \times 0.2 + 0.5 \times 0.8 + 0.3 \times 0.6 = 0.62$$

Two events A and B are *independent* if $P[A \cap B] = P[A]P[B]$. This definition is equivalent to the requirement that both $P[B|A] = P[B]$ and $P[A|B] = P[A]$. Two independent events have the property that neither event influences the occurrence of the other.

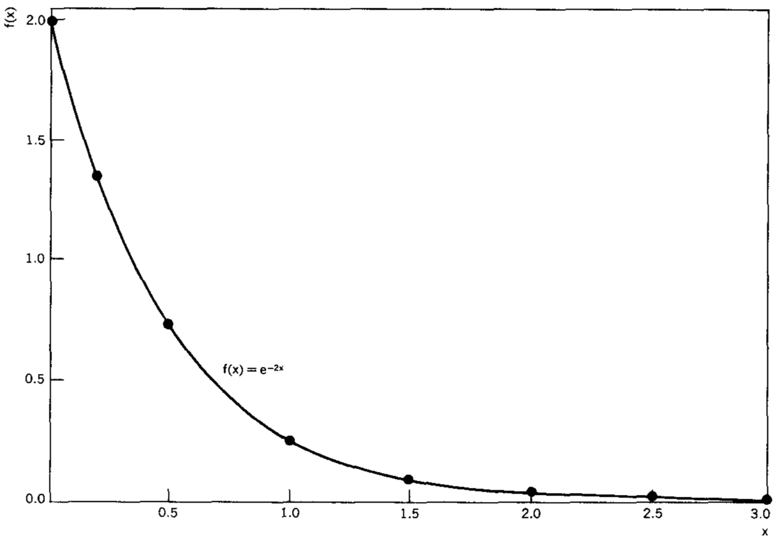
random variable

A *random variable X* is a real-valued function defined on a sample space. A capital letter (such as X) generally indicates a random variable, and a small letter (such as x) indicates a possible value of the random variable. Random variables are usually classified as being either discrete or continuous, although, occasionally a random variable of interest may be of a mixed type. Typical examples of random variables that occur in an online computer system are the following, where 1, 2, and 3 are discrete and 4 and 5 are continuous:

1. Number of inquiries received at the central computer during the last hour.
2. Number of inquiries received, but not processed by time t.
3. Number of I/O buffers in use at time t.
4. Time to process an inquiry that arrives at time t.
5. Time interval between the arrivals of the last two inquiries.

A random variable X is *discrete* if: (a) it assumes only a finite number of values, say $x_1, x_2, \cdots x_n$; or (b) it assumes countably many values $x_1, x_2, x_3. \ . \ . \ .$ The random variables 1, 2, and 3 just

Figure 5 Density function for the exponential random variable with parameter 2



given are discrete. Associated with a discrete random variable X is its *probability mass function* p, which is defined for each x_i in the range of X by $p(x_i) = \mathrm{P}[X = x_i]$. Here "$X = x_i$" is a shorthand notation for the event that is the collection of all points s in the same space S for which $X(s) = x_i$. Probability mass function is abbreviated in this paper as "p.m.f."

Example 7. Implicitly defined in Example 4 is a discrete random variable X that counts the number of polls made until the first ready terminal is found. X assumes only the values 1, 2, 3, and 4. Its p.m.f. p is defined by $p(1) = 0.5, p(2) = 0.3, p(3) = 0.15$, and $p(4) = 0.05$.

A random variable X is said to be *continuous* if it can assume all the values of some interval of the real line, and if there exists a *density function* f for X with the following properties:

$f(x) \geq 0$ for all x in the range of X and

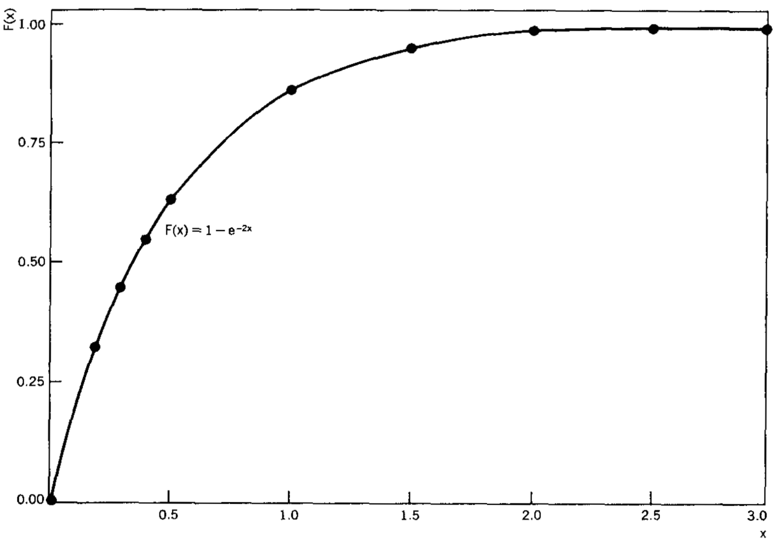
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\mathrm{P}[a \leq X \leq b] = \int_a^b f(x)dx$$

The probability that X assumes a value in any interval from a to b is the area under the density curve over the interval.

The *distribution function* F of a random variable X is defined by $F(x) = \mathrm{P}[X \leq x]$. Thus, by extension the distribution function of a continuous random variable X is expressed as follows:

Figure 6 Distribution function for the exponential random variable with parameter 2



$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

The distribution function of X as a discrete random variable with p.m.f. p is given as follows:

$$F(x) = \sum_{x_i \le x} p(x_i)$$

The sum on the right is evaluated for all the x_i such that $x_i \le x$. Thus, if F is the distribution function for the random variable of Example 7, then

$$F(0) = 0,\ F(1) = p(1) = 0.5,\ F(2.5) = p(1) + p(2) = 0.8$$

$$F(3) = p(1) + p(2) + p(3) = 0.95$$

$$F(5) = p(1) + p(2) + p(3) + p(4) = 1.0$$

Example 8. Let $\lambda > 0$. A random variable X is said to be exponentially distributed with parameter λ if X has a density function f defined as follows:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \ge 0 \end{cases}$$

Such a random variable also has a distribution function F defined as follows:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \ge 0 \end{cases}$$

Figure 5 shows the density function for an exponentially distributed random variable with parameter $\lambda = 2$. Figure 6 shows its distribution function. To calculate the probability that this random variable assumes a value between 1 and 3, find the area under the curve in Figure 5 between $x = 1$ and $x = 3$ by using the following relationship:

$$\begin{aligned} P[1 \leq X \leq 3] &= P[X \leq 3] - P[X \leq 1] = F(3) - F(1) \\ &= 1 - e^{-6} - (1 - e^{-2}) \\ &= e^{-2} - e^{-6} = 0.135335 - 0.002479 = 0.132856 \end{aligned}$$

Two random variables X and Y are *independent*, if, for any a, b, c, d, the following conditions prevail:

$$P[a \leq X \leq b \text{ and } c \leq Y \leq d] = P[a \leq X \leq b]\,P[c \leq Y \leq d]$$

This means that, if X and Y are discrete, then $P[X = x \text{ and } Y = y] = p_X(x)p_Y(y)$, where p_X is the p.m.f. of X and p_Y is the p.m.f. of Y. If X and Y are continuous, the probabilities of X and Y are computed as follows:

$$P[a \leq X \leq b \text{ and } c \leq Y \leq d] = \int_a^b f_X(x)\,dx \int_c^d f_Y(y)\,dy$$

Functions of random variables

Some parameters of random variables are important in summarizing the properties of that variable in a way that makes it easy to understand how to use it to make probability estimates. Let $m(X)$ be a function of a random variable X such as $2X$, X^2 etc. Then the *expected value* $E[m(X)]$ where X is a discrete random variable is defined as

$$E[m(X)] = \sum_i m(x_i)p(x_i) = m(x_1)p(x_1) + m(x_2)p(x_2) + \cdots$$

If X is a continuous random variable, the expected value is defined as

$$E[m(X)] = \int_{-\infty}^{\infty} m(x)f(x)\,dx$$

The two most important parameters used to describe a random variable are the *mean* (or expected value) $\mu = E[X]$, and the *standard deviation* σ, where σ^2 is the *variance* of X defined by

$$\sigma^2 = \mathrm{Var}[X] = E[X - \mu)^2].$$

The sequence of *moments* of X defined by $E[X^k]$, where $k = 1, 2, 3, \cdots$, is sometimes of interest. The first moment is the expected value or *mean*. From the definition, we see that, if X is discrete, then for each $k = 1, 2, 3, \cdots$

$$E[X^k] = x_1{}^k p(x_1) + x_2{}^k p(x_2) + \cdots + x_n{}^k p(x_n) + \cdots$$

and, if X is continuous, then

$$E[X^k] = \int_{-\infty}^{\infty} x^k f(x)\,dx$$

It can be shown that, if all the moments for X exist, they uniquely determine X. That is if X and Y have the same sequence of moments, then $X = Y$.

Example 9. For Example 7, the mean or expected value of X is computed as follows:

$$\mu = \mu_X = E[X] = \Sigma x_i p(x_i)$$
$$= 1 \times 0.5 + 2 \times 0.3 + 3 \times 0.15 + 4 \times 0.05 = 1.75$$

and the *variance* and *standard deviation* of X are computed as follows:

$$\sigma^2 = \sigma_X{}^2 = \mathrm{Var}[X] = \Sigma(x_i - \mu)^2 p(x_i)$$
$$= (1 - 1.75)^2 \times 0.5 + (2 - 1.75)^2 \times 0.30$$
$$+ (3 - 1.75)^2 \times 0.15 + (4 - 1.75)^2$$
$$\times 0.05 = 0.7875$$

so that $\sigma = \mathrm{Var}[X]^{\frac{1}{2}} = 0.8874$.

This example can be generalized. Consider a communication line with m terminals attached of which n are ready to transmit. Let X be the number of polls made until the first terminal that is ready is found. Then X can assume only the values $i = 1, 2, 3, \cdots, m - n + 1$ with

$$p(i) = \mathrm{P}[X = i] = \binom{m-i}{n-1} \Big/ \binom{m}{n}$$

$E[X]$ and $\mathrm{Var}[X]$ can then be calculated by the formulas

$$\mu = E[X] = \sum_{i=1}^{m-n+1} ip(i)$$

$$\mathrm{Var}[X] = \sum_{i=1}^{m-n+1} (i - \mu)^2 p(i)$$

A vector **P**, whose components are the probabilities $p(i)$, $i = 1, 2, \cdots, m - n + 1$, can be generated by the APL statement

```
P←((N-1)!M-⍳M-N-1)÷N!M
```

The vector **X** composed of all values assumed by the random variable X, is produced by

```
X←⍳M-N-1
```

The mean and variance of X can then be calculated by the formulas

```
EX← +/P×X
```

and

```
VARX←+/P×(X-EX)*2
```

Consider an installation consisting of 20 terminals with 7 ready to transmit. Then the mean and the variance from the formulas just given are the following.

$EX = 2.625$

$\text{Var}X = 3.17708333$

Hence, the expected value of X is $\mu = 2.625$ with standard deviation

$\sigma = [\text{Var}X]^{\frac{1}{2}} = 1.82145775$

The mean value μ of a random variable X can be thought of as a typical or average value of X. The standard deviation is a measure of how typical this value is. That is, σ has the same units as μ and measures the tendency of the values of X to cluster close to μ. Chebychev's inequality, to be discussed later in this paper, brings this point out more clearly.

Listed in the following, are properties M1 – M4 of the mean, and properties V1 – V4 of the variance for later reference. X and Y are assumed to be arbitrary random variables (except for the requirements stated below) and c is an arbitrary constant.

M1: $E[c] = c$ (The expectation of a constant random variable is the value of the constant.)

M2: $E[cX] = cE[X]$

M3: $E[X + Y] = E[X] + E[Y]$ (X and Y need not be independent.)

M4: $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ if X and Y are independent

V1: $\text{Var}[c] = 0$

V2: $\text{Var}[cX] = c^2\text{Var}[X]$

V3: $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ if X and Y are independent

V4: $\text{Var}[X] = E[X^2] - (E[X])^2$

Bernoulli random variable

Several important discrete random variables are derived from the concept of a Bernoulli sequence of trials. A *Bernoulli trial* or *experiment* is one in which there are only two possible outcomes, called "success" or "failure" with respective probabilities P and Q, where $P + Q = 1$. A sequence of such trials is a *Bernoulli sequence* if the probability of success or failure is constant from trial to trial. A *Bernoulli random variable* is thus one that assumes only two values: 1 (for success) with probability P and 0 (for failure) with probability Q, where $Q = 1 - P$.

Suppose a Bernoulli trial with probability P of success is repeated exactly n times. The random variable X, which counts the number of successes in the n trials is called a *binomial random variable* with parameters n and P. It can be shown that $E[X] = nP$ and $\text{Var}[X] = nQP$. The random variable X can assume only the values $0, 1, 2, \cdots, n$. The p.m.f. p of a binomial random variable X is expressed as follows:

binomial random variable

$$p(k) = b(k; n, P) = \text{P}[X = k] = \binom{n}{k} P^k Q^{n-k}$$

where

$$k = 0, 1, \cdots, n$$

Suppose a sequence of Bernoulli trials is continued until the first success is recorded. The random variable X, which counts the number of failures preceding the first success, is called a *geometric random variable* with parameter P. It can be shown that the p.m.f. p of a geometric random variable, the expected value, and the variance can be expressed as follows:

geometric random variable

$$P(k) = \text{P}[X = k] = (1 - P)^k P = Q^k P$$

where

$$k = 0, 1, 2, 3, \cdots,$$

and

$$E[X] = \frac{Q}{P}, \ \text{Var}[X] = \frac{Q}{P^2}.$$

X is a *Poisson random variable* with parameter $\lambda > 0$, if $\text{P}[X = k] = e^{-\lambda} \lambda^k / k!$, for $k = 0, 1, 2, 3, \cdots$, and X assumes no other values.

Poisson random variable

For a Poisson random variable

$$E[X] = \text{Var}[X] = \lambda$$

The Poisson random variable is frequently used in queueing theory and many other areas of applied probability theory. Such diverse phenomena as the number of raisins per cubic inch of raisin bread; the number of typographical errors per page of a book; the number of customers entering a bank during the noon hour; the number of inquiries per minutes in an on-line computer system; or the number of seeks per second on a disk drive may be described by a Poisson random variable.

The information on the discrete random variables we have discussed is summarized in Table 1.

Table 1 Facts about discrete random variables discussed in this paper

Random Variable	Parameters	Probability Mass Function $p(.)$	Mean $E[X]$	Variance $\sigma^2 = E[X^2] - E^2[X]$
Bernoulli	$0 \leq P \leq 1$	$p(1) = P$ $p(0) = Q = 1 - P$	P	PQ
Binomial	$n = 1, 2, \cdots$ $0 \leq P \leq 1$	$p(k) = \binom{n}{k} P^k Q^{n-k}$ $k = 0, 1, \cdots, n$	nP	nPQ
Geometric	$0 \leq P \leq 1$	$p(k) = Q^k P$ $k = 0, 1, 2, \cdots$	$\frac{Q}{P}$	$\frac{Q}{P^2}$
Poisson	$\lambda > 0$	$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, 2, \cdots$	λ	λ

Example 10. There is a master file of 120,000 records stored as a sequential file on a disk pack in blocks of six records each. Each day the master file is run against a transaction file, and approximately five percent of the master records are updated. The records to be updated are assumed to be distributed uniformly over the master file. An entire block of records on the disk pack must be updated if any records in a block have to be updated. Approximately how many blocks must be updated?

Let X be the random variable that counts the number of records in a block that are to be updated. Thus X can assume only the values 0, 1, 2, 3, 4, 5, and 6. It is reasonable to assume that X has a binomial distribution with parameters $n = 6$ and $P = 0.05$. (A Bernoulli trial consists of checking a record to determine whether it must be updated, that is, to determine whether it is listed in the transaction file.) A given block must be updated if $X \geq 1$, that is with probability $\mathrm{P}[X \geq 1] = 1 - \mathrm{P}[X = 0]$. Hence the probability that any given block must be updated is

$$1 - b(0; 6, 0.05) = 1 - (0.95)^6 = 1 - 0.735 = 0.265$$

and the approximate number of blocks that must be updated is $20{,}000 \times 0.265 = 5{,}300$.

law of large numbers

The *law of large numbers* was used in making the previous calculation. The law of large numbers states that as the number of times an experiment is repeated increases (i.e., as n increases) the proportion of the outcomes in which a given event A occurs n_A / n approaches $\mathrm{P}[A]$. In symbols the law of large numbers may be written as follows:

$$\lim_{n \to \infty} \frac{n_A}{n} = \mathrm{P}[A]$$

In the above example P[A] is assumed to be approximately $n_A \div 20{,}000$, where A is the event $[X \geq 1]$, so that

$$n_A = 20{,}000 \times \mathrm{P}[A] = 20{,}000 \times 0.265 = 5300$$

Another approach is to let Y be the random variable that counts the number of blocks that must be updated. Then Y has a binomial distribution with parameters $n = 20{,}000$ and $P = 0.265$. Hence the expected number of blocks to be updated $E[Y]$ is $nP = 20{,}000 \times 0.265 = 5300$, with standard deviation

$$\sigma = \sqrt{nQP} = 62.414.$$

Let Z be the random variable that counts the number of blocks of the master file that are read before the first block to be updated is found. Then Z has a geometric distribution. Therefore, the probability that the first block must be updated is $p(0) = P = 0.265$, and the expected value of Z is

$$\frac{Q}{P} = \frac{0.735}{0.265} = 2.774$$

Example 11. In a teleprocessing network with 20 communication lines, it has been found that the probability that any given line is in use is 0.6 and that the lines operate independently. What is the probability that ten or more lines are in operation?

Let X be the number of lines in operation, and assume that X has a binomial distribution with parameters $n = 20$ and $P = 0.6$. The required probability is then

$$\sum_{k=10}^{20} \binom{20}{k} (0.6)^k (0.4)^{20-k} = 0.872479$$

This is a tedious calculation to carry out manually, but it can be accomplished by means of the following APL statement:

```
+/(K!20)×(0.6*K)×(0.4*20-K←9+⍳11)
```

The required probability can also be approximated by using the normal distribution, as shown later in this paper.

Continuous random variables

uniform distribution

Summarized in Table 2 are facts about some of the most common theoretical continuous random variables that are used in computer system design and analysis. If X has a *uniform distribution* its values are restricted to a finite interval, and the probability that the value of X falls in any particular subinterval is the ratio of the length of that subinterval to the length of the whole interval. Thus, if X is uniformly distributed on the interval 10 to 20, then the probability that X lies between 15 and 20 is $5/10 = 0.5$. Also, by Table 2, E[X] = 15 and Var[X] = $10^2 \div 12 = 8.33$.

Table 2 Facts about continuous random variables discussed in this paper

Name	Parameters	Density function	Mean $\mu = E[X]$	Variance $\sigma^2 = E[X^2] - E^2[X]$	Distribution function $F(x) = P[X \leq x]$
Uniform over interval a to b	a and b real with $a < b$	$\frac{1}{b-a}$, if $a < x < b$ 0 otherwise	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0 if $x \leq a$ $\frac{x-a}{b-a}$ if $a < x \leq b$ 1 if $b < x$
Normal	$-\infty < \mu < +\infty$ $\sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$	μ	σ^2	$\int_{-\infty}^{x} f(t)dt$
Exponential	$\lambda > 0$	$\lambda e^{-\lambda t}$, if $t \geq 0$, 0 otherwise	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	0, if $x \leq 0$ $1 - e^{-\lambda x}$, if $x > 0$
Erlang-k	$k = 1, 2, 3 \cdots$ $\lambda > 0$	$\frac{(\lambda k)(\lambda k t)^{k-1} e^{-\lambda k t}}{(k-1)!}$ if $t \geq 0$ 0 otherwise	$\frac{1}{\lambda}$	$\frac{1}{k\lambda^2}$	0 if $x \leq 0$ $1 - e^{-\lambda k x} \sum_{j=0}^{k-1} \frac{(\lambda k x)^j}{j!}$ if $x \geq 0$

normal distribution

The *normal distribution* is the most widely used distribution in applied probability theory. This is true not only because many useful random variables are approximately normally distributed, but also because the *normal distribution* can sometimes be used to approximate Poisson and binomial probabilities.

The normal distribution with a mean of 0 and a standard deviation of 1 is called the *standard normal distribution*. A random variable X, normally distributed with mean μ and standard deviation σ is indicated by writing "X is $N(\mu, \sigma^2)$." The standard normal distribution is the only normal distribution that need be tabulated because of the remarkable fact that if X is $N(\mu, \sigma^2)$ then the random variable Z defined by the following expression:

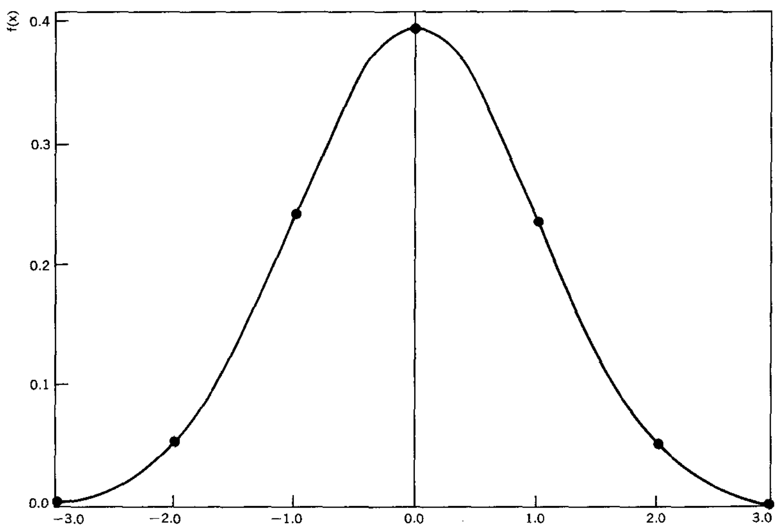
$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution, that is, Z is $N(0, 1)$.

Values of Z that correspond to values of X are called z-values. Figure 7 shows the density function for the standard normal distribution. Values of the standard normal distribution function can be calculated from data given in Table 4 and by formula 26.2.17 in Reference 1.

Example 12. Suppose that it has been found that the response time of an on-line inquiry system has approximately a normal distribution with a mean of 1.5 seconds and standard deviation of 250 milliseconds. What is the probability that the response time for an inquiry will not exceed 1.8 seconds?

Figure 7 Density function for the standard normal distribution $f(x) = 1/\sqrt{2\pi}\, e^{-x^2/2}$



The Z-value corresponding to 1.8 seconds is $(1.8 - 1.5)/0.25 = 1.2$. The required probability is the probability that $Z \leq 1.2$ has been calculated by Reference 1 or obtained from Table 4 as 0.88493.

exponential distribution

The exponential distribution is especially important for queueing theory because of its Markov or "memoryless" property. (The exponential distribution is illustrated in Example 8.) Suppose X is an exponentially distributed random variable that measures the time between successive message arrivals to a message switching system, and t time units have passed since the last message arrived. The Markov property of X is as follows: The probability that more than h additional time units will pass before the next message arrives is exactly the same as the probability that more than h time units will have passed before the next message arrives, given that a message has just arrived. That is, the system "forgets" that t time units have passed since the last message arrived. Symbolically, the Markov property is expressed as follows:

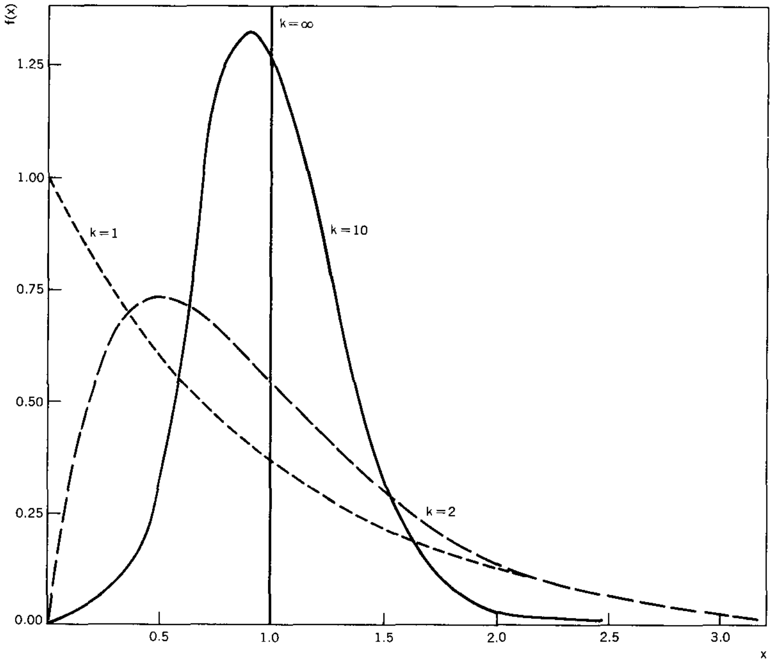
$$P[X > t + h | X > t] = P[X > h]$$

Erlang-k distribution

The *Erlang-k distribution* is important for queueing theory. The Erlang-1 distribution coincides with the exponential distribution. In Figure 8 we show the Erlang-k density function for $k = 1$, 2, 10, and ∞. When $k = \infty$ the random variable is constant, since $\sigma = 0$. The mean and the variance of an Erlang-k random variable are related by the equation

$$\text{Var}[X] = \frac{E[X]^2}{k}$$

Figure 8 Density function for Erlang distributed random variables with average value 1



Example 13. The time between the arrival of any two consecutive inquiries in an on-line system has been found to have an Erlang distribution of order 5 with average value of 50 milliseconds. What is the probability that the time interval between the next two arrivals does not exceed 75 milliseconds? What is the probability that the time interval does not exceed 50 milliseconds?

Let X be the random variable. Then $E[X] = 0.05$ sec so that $\lambda = 1/0.05$. Hence the probability that X does not exceed 75 milliseconds is given by $F_X(0.075)$ which by Table 2 has the value

$$1 - e^{-\lambda k x} \sum_{j=0}^{k-1} \frac{(\lambda k x)^j}{j!} = 1 - e^{-7.5} \sum_{j=0}^{4} \frac{(7.5)^j}{j!} = 0.8679$$

Similarly, the probability that the value of X does not exceed 50 milliseconds is

$$1 - e^{-5} \sum_{j=0}^{4} \frac{5^j}{j!} = 0.5595$$

These calculations can also be made by using the APL statement

```
1-(*-Z)×+/((Z←L×K×X)*U)÷!U←¯1+⍳K
```

and executing it with the values of $\lambda = L$, $k = K$, and $x = X$.

Combining random variables

Suppose a random variable is the sum of n independent random variables, that is, $t = t_1 + t_2 + \cdots + t_n$. Then by properties M3 and V3 previously listed, the expected value and variance of t are given as follows:

$$E[t] = E[t_1] + E[t_2] + \cdots + E[t_n]$$

and

$$\mathrm{Var}[t] = \mathrm{Var}[t_1] + \mathrm{Var}[t_2] + \cdots + \mathrm{Var}[t_n].$$

Now suppose that ten different types of inquiries arrive at the computer, each with a different length distribution. The method for calculating the mean and the standard deviation of the message length of all messages arriving at the computer is given by the Combination Theorem.

Combination Theorem. Suppose the random variable Z assumes the value of the random variable X_1 or X_2 or . . . or X_n with the respective probabilities P_1, P_2, $P_3 \cdots$ or P_n. Then the moments and variance of Z are calculated as follows:

$$E[Z^k] = P_1 E[X_1{}^k] + P_2 E[X_2{}^k] + \cdots + P_n E[X_n{}^k]$$

for $k = 1, 2, 3, \cdots$,
and

$$\mathrm{Var}[Z] = E[Z^2] - (E[Z])^2$$

Note that the theorem does not say that

$$\mathrm{Var}[Z] = P_1 \mathrm{Var}[X_1] + \cdots + P_n \mathrm{Var}[X_n]$$

This formula is not valid.

Example 14. Ten different types of messages arrive at a central computer. The fraction of each type, their means and standard deviations of message length in bytes, are shown in Table 3. Find the mean and the standard deviation of the message length of the mix of all messages that arrive at the computer.

The expected value $E[Z]$ is calculated by the formula $E[Z] = \sum_{i=1}^{10} P_i E[X_i] = 164.25$ bytes. Also calculate the $E[X_i{}^2]$, for $i = 1, 2, \cdots 10$, by using the formula $E[X_i{}^2] = \mathrm{Var}[X_i] + (E[X_i])^2$ to obtain the respective values 10,100; 14,544; 40,400; 5,650; 90,625; 27,200; 130,896; 2,516; 3,609; and 17,000. Calculated next is

$$E[Z^2] = \sum_{i=1}^{10} P_i E[X_i{}^2] = 37{,}256.875$$

Finally,

$$\sigma_Z = \{E[Z^2] - E^2[Z]\}^{\frac{1}{2}} = 101.38 \text{ bytes}$$

Table 3 Fraction, mean length, and standard deviation of ten message types

Message type	*Fraction* P_i	*Mean length*	*Standard deviation*
1	0.100	100	10
2	0.050	120	12
3	0.200	200	20
4	0.050	75	5
5	0.025	300	25
6	0.075	160	40
7	0.150	360	36
8	0.050	50	4
9	0.150	60	3
10	0.150	130	10

These calculations can also be made by using APL/360. If $\text{Var}[Z]$ were calculated using the incorrect formula

$$\text{Var}[Z] = P_1 \text{ Var}[X_1] + P_2 \text{ Var}[X_2] + \cdots + P_n \text{ Var}[X_n]$$

a value of 445.625 would be obtained, although the correct value is 10,278.8125. This yields 21.11 for σ_Z, although the true value is 101.38.

three inequalities

Three inequalities are now stated that have both theoretical and practical value.

Markov's inequality. Let X be a random variable that assumes only nonnegative values. Then, for any $t > 0$,

$$\text{P}[X \geq t] \leq E[X]/t$$

Chebychev's inequality. Let X be a random variable with expected value μ and standard deviation σ. Then for every number $k > 0$

$$\text{P}[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

and

$$\text{P}[|X - \mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}$$

One-sided inequality. Let X be a random variable with mean $E[X]$ and variance $\text{Var}[X]$. Then

$$\text{P}[X \geq t] \leq \frac{\text{Var}[X]}{\text{Var}[X] + (t - E[X])^2}$$

where

$$t \geq E[X]$$

and

$$P[X \leq t] \leq \frac{\text{Var}[X]}{\text{Var}[X] + (t - E[X])^2}$$

where

$$t \leq E[X]$$

Example 15. Assume that you have constructed a mathematical model of a proposed on-line system and find that the mean response time is 400 milliseconds with a standard deviation of 116 milliseconds. The design criterion given is that 90% of all response times must be less than 750 milliseconds. Will the proposed system satisfy this requirement?

The solution procedure is to apply in turn each of the inequalities just given. Let X be the random variable describing the response time in milliseconds. By Markov's inequality

$$P[X \geq 750] \leq \frac{400}{750} = 0.5333$$

so that

$$P[X \leq 750] \geq 1 - 0.5333 = 0.4667$$

Markov's inequality leaves doubt as to whether the design criterion is met.

Since

$$P[X \geq 750] = P[X - 400 \geq 350] \leq P[|X - 400| \geq 350]$$
$$\leq \left(\frac{116}{350}\right)^2 = 0.1098$$

by Chebychev's inequality, there remains uncertainty that the design is good enough.

By the one-sided inequality

$$P[X \geq 750] \leq \frac{116^2}{116^2 + 350^2} = \frac{1}{1 + (350/116)^2} = 0.09897$$

From this, the design criterion is seen to be met, because

$$P[X \leq 750] \geq 1 - 0.09897 = 0.90103$$

Note that each of the inequalities (Markov, Chebychev, and one-sided) is sharp, that is, they cannot be improved without strengthening the hypotheses. However, if the actual probability distribution is known, the probabilities involved may differ considerably from the limiting values given by the inequalities. For example, if X has a normal distribution

$$P[|X - \mu| \leq 2\sigma] = 0.9545$$

although Chebychev's inequality yields only that

$$P[|X - \mu| \leq 2\sigma] \geq 0.75$$

There are two approximation theorems that are sometimes helpful in calculating binomial probabilities.

Poisson approximation theorem

Let X be a binomial random variable with parameters n and P. If P is small and n is large, $P[X = k]$ is approximately

$$P[X = k] = e^{-nP}\left[\frac{(nP)^k}{k!}\right]$$

where

$$k = 0, 1, 2, \cdots, n$$

This approximation is reasonably accurate if $n \geq 100$ and $P \leq 0.05$.

Example 16. A computer installation has a library of 100 subroutines. Each week—on the average—unknown errors (bugs) are discovered and corrected in two of the subroutines. (It is widely believed that, no matter how many bugs are corrected in any large program, roughly the same number of new bugs appear in each fixed time interval.) Assuming that the number of subroutines per week with newly discovered and corrected bugs has a binomial distribution, use the Poisson Approximation Theorem to calculate the probability that errors will be found in exactly 0, 1, 2, or 3 subroutines next week.

The Poisson approximation yields the following $P[X=k]$ values: 0.13534, 0.27067, 0.27067, and 0.18045. The correct values can be calculated by the following APL statement:

```
(K!100)×(0.02*K)×0.98*100-K
```

Executed for $k = 0, 1, 2, 3$, the correct values of $P[X = k]$ are 0.13262, 0.27065, 0.27341, and 0.18228.

normal approximation theorem

Let X be a binomial random variable with parameters n and P. Then, if a and b are integers, it is approximately true that

$$P[a \leq X \leq b] = \Phi\left(\frac{b - nP + \frac{1}{2}}{\sqrt{nPQ}}\right) - \Phi\left(\frac{a - nP - \frac{1}{2}}{\sqrt{nPQ}}\right)$$

where Φ is the distribution function for the standard normal distribution, and $P + Q = 1$.

The normal approximation theorem gives reasonably accurate results when $nPQ \geq 10$.

Table 4 The normal distribution function $\Phi(Z) = \int_{-\infty}^{Z} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	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	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995

Example 17. Use the normal approximation to make the calculation of Example 11.

By the normal approximation theorem, the required probability is

$$\Phi\left(\frac{20 - 12 + 0.5}{\sqrt{4.8}}\right) - \Phi\left(\frac{10 - 12 - 0.5}{\sqrt{4.8}}\right)$$

$$= \Phi(3.88) - \Phi(-1.14) = 0.99995 - (1 - \Phi(1.14))$$

$$= 0.99995 - (1 - 0.87286) = 0.87281$$

The values of $\Phi(1.14)$ are obtained from the formula in Reference 1 and Table 4, and the identity $\Phi(-1.14) = 1 - \Phi(1.14)$ follows from the symmetry of the normal density function.

Example 18. In Example 10, the average number of blocks to be updated is 5300. Also, the number of blocks to be updated has a binomial distribution with parameters n = 20,000 and P = 0.265. Use the Normal Approximation Theorem to estimate the probability that between 5200 and 5400 blocks must be updated.

Since $\sigma^2 = nPQ = 20{,}000 \times 0.265 \times 0.735 = 3{,}895.5$, the approximation should be accurate. If X is the number of blocks to be updated, the theorem yields

$$
\begin{aligned}
P[5200 \leq X \leq 5400]& \\
&= \Phi\left(\frac{5400 - 5300 + 0.5}{\sqrt{3895.5}}\right) \\
&\quad - \Phi\left(\frac{5200 - 5300 - 0.5}{\sqrt{3895.5}}\right) \\
&= \Phi(1.61) - \Phi(-1.61) \\
&= \Phi(1.61) - (1 - \Phi(1.61)) \\
&= 2\Phi(1.61) - 1 \\
&= 2 \times 0.9463 - 1 \\
&= 0.8926.
\end{aligned}
$$

The value of $\Phi(1.61)$ was obtained from Table 4. It may also be calculated from Reference 1. The identity $\Phi(-1.61) = 1 - \Phi(1.61)$ follows from the symmetry of the normal density function.

Concluding remarks

We have selected and presented aspects of probability theory and have emphasized their applications to computer system design and analysis. Some general references for further reading on probability theory are given in the brief bibliographic section that follows.

CITED REFERENCE AND BIBLIOGRAPHY

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards (1964).
2. W. Weaver, *Lady Luck: the Theory of Probability*, Anchor Books, Garden City, New York (1963).
3. F. Mosteller, R. E. K. Rourke, and G. B. Thomas, Jr., *Probability With Statistical Applications*, Addison-Wesley Publishing Company, Reading, Massachusetts (1961).
4. D. P. Gaver and G. L. Thompson, *Programming and Probability Models in Operations Research*, Brooks/Cole Publishing Company, Monterey, California (1973).
5. W. Feller, *An Introduction to Probability Theory and its Applications, 3rd Edition*, John Wiley and Sons, New York, New York (1968).
6. T. C. Fry, *Probability and its Engineering Uses, 3rd Edition*, The Bell Telephone Laboratories Series, D. Van Nostrand Company, Inc. (1965).