

SERVICE SCIENCE

Mark S. Daskin

Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117
U.S.A.

e-mail: msdaskin@umich.edu

©, Copyright, 2010, Mark S. Daskin

Chapter Level Table of Contents

A. Sums of series – basic formulae.....	6
B. Overview of probability	11
C. References.....	69

Detailed Table of Contents

A.	Sums of series – basic formulae.....	6
B.	Overview of probability	11
B.1.	Introduction and basic definitions.....	11
B.2	Axioms of probability	13
B.3	Joint, marginal and conditional probabilities and Bayes’ theorem.....	13
B.4	Counting, ordered pairs, permutations and combinations	16
B.5	Random variables.....	17
B.6	Discrete random variables.....	17
B.7	Continuous random variables	32
B.8	Moment and probability generating functions.....	53
B.9	Generating random variables	60
B.10	Random variables in Excel	67
C.	References.....	69

List of Figures

Figure B.1 -- Venn diagram of graduate school admission events	12
Figure B.2 -- Binomial distribution for various values of n and p	23
Figure B.3 -- Geometric distribution for different values of p	25
Figure B.4 -- Negative binomial distribution for different values of r	26
Figure B.5 -- Poisson distribution for different values of λ	31
Figure B.6 -- Uniform distribution for different values of a and b	35
Figure B.7 -- Example service region and direction of travel within the region	37
Figure B.8 -- Service region with a subregion in which distances are less than or equal to x	38
Figure B.9 -- Distribution of distance between the center of a unit diamond and a randomly selected point within the diamond.....	39
Figure B.10 -- Triangular distribution: the density function of the maximum of two standard uniform random variables	40
Figure B.11 -- Standard normal distribution.....	42
Figure B.12 -- The exponential distribution	47
Figure B.13 -- Shapes of the Erlang- k distribution for different values of k	51
Figure B.14 -- Summary of key probability distributions and their relationships	52
Figure B.15 -- Cumulative distribution of X	60
Figure B.16 -- Finding the value of a random variable X from a draw of a uniform random variable using the cumulative distribution of X	61
Figure B.17 -- True versus sampled distributions for a Poisson random variable with a true mean of 10	64
Figure B.18 -- Sampling from the exponential distribution.....	65
Figure B.19 -- Sampling from the normal distribution.....	66

List of Tables

Table A.1 -- Summary of key summations	8
Table B.1 -- Graduate school admission example probabilities	14
Table B.2 -- Example of joint and marginal probabilities	15
Table B.3 -- Approximating the binomial by the Poisson	32
Table B.4 -- Relationships between discrete and continuous random variables	33
Table B.5 -- Selected values of the cumulative normal distribution	43
Table B.6 -- Approximating the binomial distribution by the normal distribution	44
Table B.7 -- Approximating the Poisson distribution by the normal distribution	44
Table B.8 -- EXCEL information for the same distribution	62
Table B.9 -- Using the LOOKUP function to sample from the distribution given in Table B.8	63
Table B.10 -- Some probability functions in EXCEL	67

A. Sums of series – basic formulae

In this appendix, we briefly review how to compute the value of an infinite (or sometimes finite) sum. This is useful in the analysis of many queueing problems. The general approach is shown below:

$$S = \sum_{n=0}^{\infty} \alpha^n \quad (\text{A.1})$$

$$S = 1 + \alpha + \alpha^2 + \dots \quad (\text{A.2})$$

$$\alpha S = \alpha + \alpha^2 + \dots \quad (\text{A.3})$$

$$S(1 - \alpha) = 1 \quad (\text{A.4})$$

$$S = \frac{1}{1 - \alpha} \quad (\text{A.5})$$

The infinite sum whose value we wish to determine is shown in (A.1). In (A.2) we write out the infinite sum in expanded form. In (A.3) we multiply the left and right hand sides of (A.2) by α . In (A.4) we subtract (A.3) from (A.2). The key here is that all of the terms on the right hand side cancel except for the 1 in equation (A.2). Finally, in (A.5), we simply divide each side of (A.4) by $(1 - \alpha)$.

For this approach to work, we clearly require $\alpha < 1$. Otherwise, the sum will diverge as each term is then larger than the preceding term. This restriction is okay, as we often are summing probabilities all of which satisfy this condition. Also, if $\alpha = 1$, the sum would diverge.

This approach may have to be used repeatedly to get more complicated results. The finite sum results for $\alpha = 1$ are obtained by induction. The list (and Table A.1) below summarize some key sums that are of use to us in this text.

$$1. \quad \sum_{n=0}^{\infty} \alpha^n = \frac{1}{1 - \alpha} \quad \alpha < 1 \quad (\text{A.6})$$

$$2. \quad \sum_{n=1}^{\infty} \alpha^n = \frac{\alpha}{1 - \alpha} \quad \alpha < 1 \quad (\text{A.7})$$

$$3. \sum_{n=0}^{\infty} n\alpha^n = \sum_{n=1}^{\infty} n\alpha^n = \frac{\alpha}{(1-\alpha)^2} \quad \alpha < 1 \quad (\text{A.8})$$

$$4. \sum_{n=0}^{\infty} n^2\alpha^n = \sum_{n=1}^{\infty} n^2\alpha^n = 2\frac{\alpha}{(1-\alpha)^3} - \frac{\alpha}{(1-\alpha)^2} = \frac{2\alpha - \alpha(1-\alpha)}{(1-\alpha)^3}$$

$$= \frac{\alpha + \alpha^2}{(1-\alpha)^3} \quad \alpha < 1 \quad (\text{A.9})$$

$$5. \sum_{n=0}^M \alpha^n = \frac{1 - \alpha^{M+1}}{(1-\alpha)} \quad \alpha \neq 1 \quad (\text{A.10})$$

$$6. \sum_{n=1}^M \alpha^n = \frac{\alpha - \alpha^{M+1}}{(1-\alpha)} \quad \alpha \neq 1 \quad (\text{A.11})$$

$$7. \sum_{n=0}^M n\alpha^n = \sum_{n=1}^M n\alpha^n = \frac{\alpha - (M+1)\alpha^{M+1} + M\alpha^{M+2}}{(1-\alpha)^2} \quad \alpha \neq 1 \quad (\text{A.12})$$

$$8. \sum_{n=0}^M n^2\alpha^n = \sum_{n=1}^M n^2\alpha^n$$

$$= 2 \left\{ \frac{\alpha - \alpha^{M+1}}{(1-\alpha)^3} \right\} - \frac{\alpha}{(1-\alpha)^2} - \frac{(2M-1)\alpha^{M+1}}{(1-\alpha)^2} - \frac{M^2\alpha^{M+1}}{(1-\alpha)} \quad \alpha \neq 1 \quad (\text{A.13})$$

Table A.1 -- Summary of key summations

Form	Infinite ($\alpha < 1$)	Finite ($\alpha \neq 1$) <i>upper limit = M</i>	Finite ($\alpha = 1$) <i>upper limit = M</i>
$\sum \alpha^n$	$\sum_{n=0}^{\infty} \alpha^n = \frac{1}{1-\alpha}$ $\sum_{n=1}^{\infty} \alpha^n = \frac{\alpha}{1-\alpha}$	$\sum_{n=0}^M \alpha^n = \frac{1-\alpha^{M+1}}{(1-\alpha)}$ $\sum_{n=1}^M \alpha^n = \frac{\alpha-\alpha^{M+1}}{(1-\alpha)}$	$\sum_{n=0}^M \alpha^n = M+1$ $\sum_{n=1}^M \alpha^n = M$
$\sum n\alpha^n$	$\frac{\alpha}{(1-\alpha)^2}$	$\frac{\alpha-(M+1)\alpha^{M+1}+M\alpha^{M+2}}{(1-\alpha)^2}$	$\frac{M(M+1)}{2}$
$\sum n^2\alpha^n$	$2\frac{\alpha}{(1-\alpha)^3}-\frac{\alpha}{(1-\alpha)^2}$	$2\left\{\frac{\alpha-\alpha^{M+1}}{(1-\alpha)^3}\right\}-\frac{\alpha}{(1-\alpha)^2}$ $-\frac{(2M-1)\alpha^{M+1}}{(1-\alpha)^2}-\frac{M^2\alpha^{M+1}}{(1-\alpha)}$	$\frac{M(M+1)(2M+1)}{6}$

The blocks below summarize the derivation of the first 8 sums.

$$\begin{aligned}
 S &= \sum_{n=0}^{\infty} \alpha^n \\
 &= 1 + \alpha + \alpha^2 + \dots \\
 \alpha S &= \alpha + \alpha^2 + \dots \\
 S(1-\alpha) &= 1 \\
 S &= \frac{1}{1-\alpha}
 \end{aligned}$$

$$\begin{aligned}
 S &= \sum_{n=1}^{\infty} \alpha^n \\
 &= \alpha + \alpha^2 + \alpha^3 + \dots \\
 \alpha S &= \alpha^2 + \alpha^3 + \dots \\
 S(1-\alpha) &= \alpha \\
 S &= \frac{\alpha}{1-\alpha}
 \end{aligned}$$

$$\begin{aligned}
 \sum_{n=0}^{\infty} n\alpha^n &= \sum_{n=1}^{\infty} n\alpha^n \\
 S &= \alpha + 2\alpha^2 + 3\alpha^3 + \dots \\
 \alpha S &= \alpha^2 + 2\alpha^3 + \dots \\
 S(1-\alpha) &= \alpha + \alpha^2 + \alpha^3 + \dots \\
 &= \frac{\alpha}{1-\alpha} \\
 S &= \frac{\alpha}{(1-\alpha)^2}
 \end{aligned}$$

$$\sum_{n=0}^{\infty} n^2 \alpha^n = \sum_{n=1}^{\infty} n^2 \alpha^n$$

$$S = \alpha + 4\alpha^2 + 9\alpha^3 + \dots$$

$$\alpha S = \alpha^2 + 4\alpha^3 + 9\alpha^4 + \dots$$

$$S(1-\alpha) = \alpha + 3\alpha^2 + 5\alpha^3 + 7\alpha^4 + \dots$$

$$T = \alpha + 3\alpha^2 + 5\alpha^3 + 7\alpha^4 + \dots$$

$$\alpha T = \alpha^2 + 3\alpha^3 + 5\alpha^4 + 7\alpha^5 + \dots$$

$$T(1-\alpha) = \alpha + 2\alpha^2 + 2\alpha^3 + 2\alpha^4 + \dots$$

$$= 2(\alpha + \alpha^2 + \alpha^3 + \alpha^4 + \dots) - \alpha$$

$$= 2\frac{\alpha}{1-\alpha} - \alpha$$

$$T = 2\frac{\alpha}{(1-\alpha)^2} - \frac{\alpha}{1-\alpha}$$

$$S = 2\frac{\alpha}{(1-\alpha)^3} - \frac{\alpha}{(1-\alpha)^2}$$

$$\sum_{n=1}^M \alpha^n = \alpha + \alpha^2 + \dots + \alpha^M$$

$$S = \alpha + \alpha^2 + \dots + \alpha^M$$

$$\alpha S = \alpha^2 + \dots + \alpha^M + \alpha^{M+1}$$

$$S(1-\alpha) = \alpha - \alpha^{M+1}$$

$$S = \frac{\alpha - \alpha^{M+1}}{(1-\alpha)}$$

$$\sum_{n=0}^M n\alpha^n = \sum_{n=1}^M n\alpha^n$$

$$S = \alpha + 2\alpha^2 + 3\alpha^3 + \dots + M\alpha^M$$

$$\alpha S = \alpha^2 + 2\alpha^3 + \dots + (M-1)\alpha^M + M\alpha^{M+1}$$

$$S(1-\alpha) = \alpha + \alpha^2 + \alpha^3 + \dots + \alpha^M - M\alpha^{M+1}$$

$$= \frac{\alpha - \alpha^{M+1}}{1-\alpha} - M\alpha^{M+1}$$

$$S = \frac{\alpha - \alpha^{M+1}}{(1-\alpha)^2} - \frac{M\alpha^{M+1}}{(1-\alpha)}$$

$$= \frac{\alpha - \alpha^{M+1}}{(1-\alpha)^2} - \frac{M(1-\alpha)\alpha^{M+1}}{(1-\alpha)^2}$$

$$= \frac{\alpha - (M+1)\alpha^{M+1} + M\alpha^{M+2}}{(1-\alpha)^2}$$

$$\sum_{n=0}^M n^2 \alpha^n = \sum_{n=1}^M n^2 \alpha^n$$

$$S = \alpha + 4\alpha^2 + 9\alpha^3 + \dots + M^2 \alpha^M$$

$$\alpha S = \alpha^2 + 4\alpha^3 + 9\alpha^4 + \dots + (M-1)^2 \alpha^M + M^2 \alpha^{M+1}$$

$$S(1-\alpha) = \alpha + 3\alpha^2 + 5\alpha^3 + 7\alpha^4 + \dots + (2M-1)\alpha^M - M^2 \alpha^{M+1}$$

$$T = \alpha + 3\alpha^2 + 5\alpha^3 + 7\alpha^4 + \dots + (2M-1)\alpha^M \quad \text{Note: Not same as above...}$$

$$\alpha T = \alpha^2 + 3\alpha^3 + 5\alpha^4 + 7\alpha^5 + \dots + (2M-2)\alpha^M + (2M-1)\alpha^{M+1}$$

$$T(1-\alpha) = \alpha + 2\alpha^2 + 2\alpha^3 + 2\alpha^4 + \dots + 2\alpha^M - (2M-1)\alpha^{M+1}$$

$$= 2 \left\{ \frac{\alpha - \alpha^{M+1}}{(1-\alpha)} \right\} - \alpha - (2M-1)\alpha^{M+1}$$

$$T = 2 \left\{ \frac{\alpha - \alpha^{M+1}}{(1-\alpha)^2} \right\} - \frac{\alpha}{(1-\alpha)} - \frac{(2M-1)\alpha^{M+1}}{(1-\alpha)}$$

$$S = 2 \left\{ \frac{\alpha - \alpha^{M+1}}{(1-\alpha)^3} \right\} - \frac{\alpha}{(1-\alpha)^2} - \frac{(2M-1)\alpha^{M+1}}{(1-\alpha)^2} - \frac{M^2 \alpha^{M+1}}{(1-\alpha)}$$

Finally, the following special forms are of use to us.

$$9. \quad e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad (\text{A.14})$$

$$10. \quad e^{ax} = \sum_{n=0}^{\infty} \frac{(ax)^n}{n!} = 1 + ax + \frac{(ax)^2}{2!} + \frac{(ax)^3}{3!} + \dots \quad (\text{A.15})$$

$$11. \quad (x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \quad (\text{A.16})$$

$$12. \quad \sum_{k=0}^n \binom{n}{k} = 2^n \quad (\text{A.17})$$

B. Overview of probability

B.1. Introduction and basic definitions

Probability is a branch of mathematics that studies the likelihood of different things happening. The probability of something happening measures the likelihood that the “something” will happen. Probabilities are numbers between 0 and 1. Values close to 1 indicate that something is likely to occur while values close to 0 indicate that something is unlikely to happen. A value of 0 indicates that something can not occur and a value of 1 indicates that it will definitely occur.

More formally, we think of an *experiment* in probability as any action or process whose outcome is uncertain. For example, we might think about a student applying to five graduate schools. This is an experiment. Similarly, waiting to be seated at a crowded restaurant is another experiment. The *sample space* of an experiment is the set of all possible outcomes of the experiment. In the case of graduate school applications, the sample space is the set of all possible combinations of acceptances and rejections at the five schools. There are 2^5 or 32 such combinations. In the case of waiting for a table the sample space is the set of all non-negative real numbers representing the time until the party is seated.

An *event* is a collection of outcomes of an experiment. The event “accepted by at least one graduate school” is represented by the set $\{1,2,3,4,5\}$ of the number of possible acceptances. There are $2^5 - 1$ or 31 outcomes in the sample space that comprise the event “accepted by at least one graduate school.” The event “have to wait” is the set represented by $w > 0$, where w is the time that a patron waits at the restaurant. A *simple* event is one with a single outcome (e.g., accepted at the University of Michigan and rejected at the other four schools), while a *compound* event has more than one outcome (e.g., accepted at 4 or more schools – with six outcomes comprising the event – or wait 10 or fewer minutes).

The *union* of two events, A and B, is denoted by $A \cup B$ and is the set of all outcomes in *either* A or B. The *intersection* of two events, A and B, is denoted by $A \cap B$ and is the set of all events in *both* A and B. The complement of an event, A, is denoted by A^c , and is the set of all events that are not in A. This concepts are illustrated for the graduate school application experiment in the Venn diagram shown in Figure B.1. Let us define the following three events: $A = \{0,1,2,3\}$, $B = \{1,2,3,4\}$ and $C = \{3,4,5\}$. Be sure you can identify regions of the figure that correspond to A^c , $A \cup B$, $A \cap B$, $B \cap C$ and $A \cap B \cap C$.

Two events are *mutually exclusive* if their intersection is the empty set. Two (or more) events are *collectively exhaustive* if their union represents the entire sample space. In the Venn diagram of Figure B.1, no two events are mutually exclusive. Events A and C are collectively exhaustive.

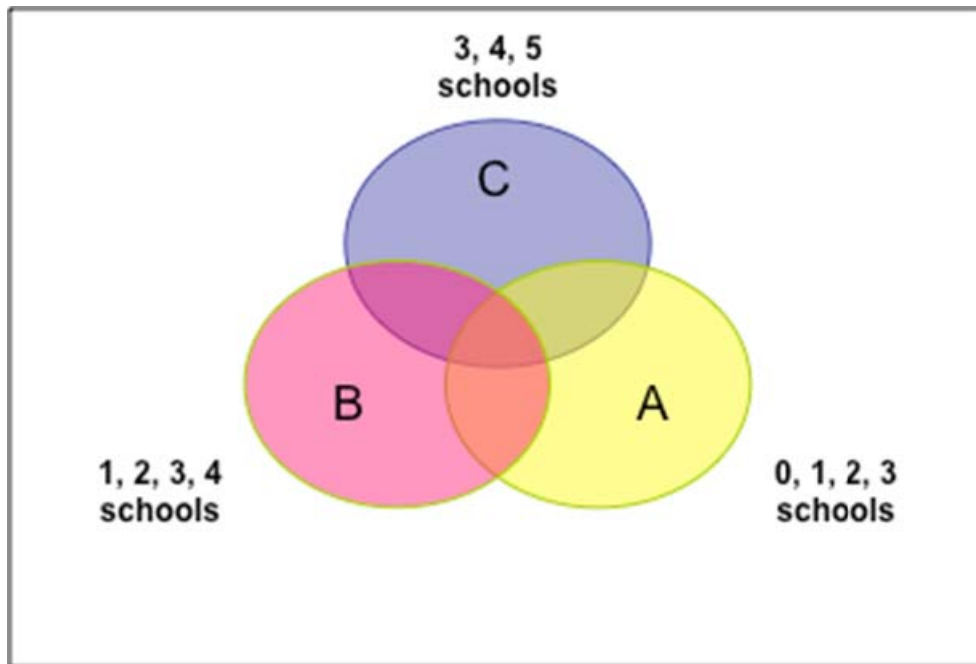


Figure B.1 -- Venn diagram of graduate school admission events

B.2 Axioms of probability

We now come to three key axioms of probability:

1. For any event A , $P(A) \geq 0$, where $P(A)$ denotes the probability that event A occurs.
2. $P(S) = 1$, where S represents the entire sample space.
3. If A_1, A_2, \dots, A_n are mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n P(A_k).$$

Also, if A_1, A_2, \dots is an infinite collection of

mutually exclusive events, then
$$P(A_1 \cup A_2 \cup \dots) = \sum_{k=1}^{\infty} P(A_k).$$

These axioms have a number of implications. For example, they imply $P(A^c) = 1 - P(A)$. They also imply that if two events A and B are mutually exclusive, then $P(A \cap B) = 0$. Importantly, they imply that for *any* two events A and B ,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (\text{B.1})$$

You should be sure you can verify these implications using a Venn diagram.

B.3 Joint, marginal and conditional probabilities and Bayes' theorem

Now suppose that we know that some particular event B has happened. We are sometimes interested in the probability that some other event A has happened knowing or *given* that event B has happened. For example, suppose the second column of Table B.1a gives the probabilities associated with the 6 simple events in the graduate school application problem outlined above. (We will return to the cumulative probabilities in the third column a bit later.) You should be able to verify that the compound event probabilities listed in Table B.1b are correct. Now suppose event B has occurred, meaning the student has been admitted to 1, 2, 3, or 4 schools. We want to find the probability that event A occurred *given* that event B occurred. Clearly, the only way that event A could happen given that event B has happened is if the student is admitted to 1,

2, or 3 schools. This occurs with probability 0.8352. However, we need to *renormalize* this by the probability that event B occurred. In other words the *conditional* probability that event A occurred *given* that event B occurred is $\frac{0.8352}{0.912} = 0.915789$. In general, we have

$$P(A \text{ occurs given } B \text{ has happened}) = P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (\text{B.2})$$

Table B.1 -- Graduate school admission example probabilities

Number of acceptances	Probability	Cumulative Probability
0	0.07776	0.07776
1	0.2592	0.33696
2	0.3456	0.68256
3	0.2304	0.91296
4	0.0768	0.98976
5	0.01024	1

(a) simple event probabilities

Event	Simple events	Probability
A	0,1,2,3	0.91296
B	1,2,3,4	0.912
C	3,4,5	0.31744

(b) compound event probabilities

The *law of total probability* states that if A_1, A_2, \dots, A_n are mutually exclusive collectively exhaustive events, then

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n) \\ &= \sum_{k=1}^n P(B|A_k)P(A_k) \end{aligned} \quad (\text{B.3})$$

Also, *Bayes' theorem* states that if A_1, A_2, \dots, A_n are mutually exclusive collectively exhaustive events and B is another event with $P(B) > 0$, then

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{k=1}^n P(B|A_k)P(A_k)}. \quad (\text{B.4})$$

Two events A and B are independent if knowing that A has occurred does not change the probability that B has occurred and vice versa. More formally, A and B are independent

if $P(B|A) = P(B)$ and $P(A|B) = P(A)$. You should be able to explain why two mutually exclusive events can not be independent.

To illustrate the importance of conditional probability, and to introduce the concepts of *joint probabilities* and *marginal probabilities*, consider the problem of being tested for some disease such as tuberculosis. There are two possible outcomes of such a test: either the test indicates you have the disease (meaning you test positive) or it indicates you do not have the disease (meaning you test negative). However, the test results are sometimes wrong. Thus, we must also consider the true state of nature: either you have the disease or you do not. This is illustrated in Table B.2. The four probabilities shaded light grey show all four possible combinations of the results of the test and one's true health condition. These indicate that 75% of those tested test negative and truly do not have tuberculosis. Another 5% of those who test negative actually do have the disease, and so on. Note that the light grey shaded numbers sum to 1.0. These are the *joint probabilities* of the test results and your true health. On the right hand side, the table shows the *marginal probabilities* of your true health condition, shown with white lettering on a black background. 83 percent of those tested are truly okay, while 17 percent have tuberculosis. Similarly, the marginal probabilities of the test results are shown at the bottom of the table shaded in dark grey. 20 percent of those tested, test positive for tuberculosis with this test.

Table B.2 -- Example of joint and marginal probabilities

		Test results		Marginal
		Negative	Positive	
True health	OK	0.75	0.08	0.83
	Have TB	0.05	0.12	0.17
Marginal		0.80	0.20	1.00

Now, suppose the test comes back positive indicating that you have tuberculosis. What is the probability that you have tuberculosis *given* that the test is positive? This is simply the joint probability that you have tuberculosis and the test is positive (0.12) divided by the probability that the test is positive (0.20) or 0.6. Thus, there is a 60% chance that you actually have tuberculosis (based on these numbers) if the test is positive.

Similarly, we might ask, if you truly have tuberculosis (which occurs with probability 0.17 for those tested), what are the chances that the test will show that you have tuberculosis? This is simply 0.12 divided by 0.17 or about 0.71.

B.4 Counting, ordered pairs, permutations and combinations

We now turn to the issue of counting. Most of us know how to count, but in probability we need to define three sorts of ways of counting. By *ordered pairs* we mean the number of ways of selecting 1 item from each of K sets of items when the j^{th} set has n_j items. This is simply $\prod_{j=1}^K n_j = n_1 \cdot n_2 \cdot \dots \cdot n_K$. For example, if we have 4 tenured faculty members in Civil Engineering, 5 in Mechanical Engineering and 3 in Industrial Engineering, there are exactly 60 ways of selecting a committee consisting of exactly one tenured faculty member from each of these three departments.

A *permutation* gives the number of ways of selecting k items out of a total of n items when the order of selection matters and items are *not* replaced after each selection. We have n ways to select the first item, $n-1$ ways to select the second, and so on until we have only $n-k+1$ ways to select the k^{th} item. Thus the number of permutations of k items out of n is

$${}_n P_k = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}. \quad (\text{B.5})$$

For example, suppose we have 10 cities and we want to visit 4 of them. We will travel to the first picked city first, the second picked city next and so on. Therefore, the order of selection matters. There are $10 \cdot 9 \cdot 8 \cdot 7 = 5040$ possible routes.

Finally, the number of *combinations* of k items out of a total of n is the number of ways of selecting k items from a total of n when the order of selection does not matter. This is given by

$${}_n C_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}. \quad (\text{B.6})$$

For example, suppose we have 150 candidate ambulance sites in a city and we wish to select 10 of them for ambulance bases. We have

$$\binom{150}{10} = \frac{150!}{10!(150-10)!} = 1,169,554,298,222,310 \approx 1.17 \times 10^{15}. \text{ If we wanted to select 20}$$

bases, we would have about 3.63×10^{24} possible combinations of 20 sites out of 150 candidate sites.

B.5 Random variables

Formally, a *random variable* is any rule that associates a number with every outcome of the sample space. For example, if a tuberculosis test is the experiment of interest, there are two outcomes: positive and negative. If we associate 1 with the event positive and 0 with the event negative, then we have defined a random variable associated with this experiment. There are two types of random variables: *discrete random variables* and *continuous random variables*. Loosely speaking we can think of discrete random variables as those that take on values corresponding to integers while continuous random variables take on values corresponding to any real number. (There are some cases in which a random variable has characteristics of both a discrete and a continuous random variable. For example, the waiting time in a queue has a probability mass associated with no waiting. A probability mass is associated with a discrete random variable. If the waiting time is strictly positive, the random variable behaves like a continuous random variable with a density function.)

B.6 Discrete random variables

We begin by focusing on discrete random variables. If X is a random variable, then we associate a *probability mass function* with X . The probability mass function or pmf allocates the total probability of 1 across all possible outcomes of the random variable X . We denote the probability that the random variable X takes on a particular

value x by either $p(X = x)$ or more simply $p(x)$. Clearly we require $p(x) \geq 0$ for all values of x and

$$\sum_{\text{all } x} p(x) = 1. \quad (\text{B.7})$$

We also are interested in the probability that the random variable X takes on a value less than or equal to x . This is the *cumulative distribution* of X and is given by

$$P(X \leq x) = \sum_{y \leq x} p(y). \quad (\text{B.8})$$

(Note that y in (B.8) is just a dummy variable of summation.) For example, Table B.1a gives both the probabilities of the simple events (in the second column) and the cumulative probabilities in the third column.

The *expected value* of a (discrete) random variable is the mean or average value that the random variable takes on. The expected value of a discrete random variable X is given by

$$E(X) = \sum_{\text{all } x} x \cdot p(x). \quad (\text{B.9})$$

For example, the expected number of graduate schools to which a student is admitted based on the probabilities in Table B.1a is given by:

$$\begin{aligned} E(X) &= 0(0.07776) + 1(0.2592) + 2(0.3456) + 3(0.2304) + 4(0.0768) + 5(0.01024) \\ &= 2 \end{aligned}$$

If $h(X)$ is a function of the random variable X , then the expected value of $h(X)$ is given by

$$E[h(X)] = \sum_{\text{all } x} h(x)p(x). \quad (\text{B.10})$$

It is worth noting that, in general, *the expected value of a function is not the same as the function evaluated at the mean or expected value of the random variable used to define the function*. In other words, in general, $E[h(X)] \neq h[E(X)]$. If $h(X)$ is linear in X , however, then $E[h(X)] = h[E(X)]$.

One important non-linear function of a random variable is the *variance*. The variance is defined as

$$\text{Var}(X) = E(X - E[X])^2. \quad (\text{B.11})$$

Clearly, in this expression we have $h(X) = (X - E[X])^2$. Thus, the variance is given by

$$\sigma_X^2 = \text{Var}(X) = V(X) = E\left\{(X - E[X])^2\right\} = \sum_{\text{all } x} (x - E[X])^2 \cdot p(x) \quad (\text{B.12})$$

We can show that the variance is also equal to

$$\text{Var}(X) = E\left\{(X - E[X])^2\right\} = E(X^2) - E^2(X). \quad (\text{B.13})$$

For example, for the graduate school admission problem we have

$$\begin{aligned} \text{Var}(X) &= E\left\{(X - E[X])^2\right\} \\ &= (-2)^2(0.07776) + (-1)^2(0.2592) + 0(0.3456) + (1)^2(0.2304) + (2)^2(0.0768) + (3)^2(0.01024) \\ &= 1.2 \end{aligned}$$

We can also calculate this by first computing

$$\begin{aligned} E(X^2) &= 0(0.07776) + 1(0.2592) + 4(0.3456) + 9(0.2304) + 16(0.0768) + 25(0.01024) \\ &= 5.2 \end{aligned}$$

and then computing

$$\text{Var}(X) = E(X^2) - E^2(X) = 5.2 - 2^2 = 1.2.$$

The variance measures the dispersion of the distribution. Clearly, if all values of the random variable were the same, the variance would be 0 since $X - E[X] = 0$ in that case for all realizations of the random variable X . The larger the variance, the more spread out the distribution is.

Note that the units of the variance are rather strange. For example, in the graduate school admission case, the variance is measured in terms of schools². It is not clear what a squared school is! We therefore also use the standard deviation as a measure of the dispersion of a distribution. The *standard deviation* is simply the square root of the variance. In the graduate school admission case, the standard deviation is $\sqrt{1.2} \cong 1.095$.

So, why are we interested in the variance at all? The reason is that there are certain things we can do with the variance that we can not do with the standard deviation directly. For example, suppose we have a linear function of a random variable. In other words we are interested in $aX + b$, where a and b are constants. This might represent the profit associated with running a small clinic as a function of the number of patients that we see. The random variable X is the number of patients seen on a day and a is the marginal profit (revenue minus variable cost) associated with treating a patient and b is the fixed cost of operating the clinic even if we see no patients. The variance of the profit is then given by

$$\begin{aligned} \text{Var}(aX + b) &= E\left\{\left[aX + b - (aE(X) + b)\right]^2\right\} \\ &= E\left\{\left[aX - aE(X)\right]^2\right\} \\ &= a^2\text{Var}(X) \end{aligned} \tag{B.14}$$

More generally, if $h(X)$ is a function of a discrete random variable X , then the variance of $h(X)$ is given by

$$\text{Var}[h(X)] = \sum_{\text{all } x} \{h(x) - E[h(X)]\}^2 p(x). \tag{B.15}$$

Also, if X and Y are independent random variables, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y). \tag{B.16}$$

We can also show that

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) \tag{B.17}$$

when X and Y are independent. *Note that we add the variances even when we are looking at the difference between two random variables.*

When they are not independent, there is a *covariance* term that must be added to this. Qualitatively, the covariance and correlation coefficient measure the degree to which X and Y vary together. If X generally increases as Y increases, then the covariance between X and Y is positive and the variance of the sum of X and Y is greater than the sum of the variances of X and Y . On the other hand, if X and Y tend to move in opposite directions, meaning that as one increases the other decreases, then the variance of the sum

of the random variables is less than the sum of the variances of the random variables. Note that none of these equations apply directly to the standard deviations. *You should never add standard deviations!*

More formally, if X and Y are two random variables with a joint probability distribution given by $p(x, y)$, then the *covariance* of X and Y is defined as

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \quad (\text{B.18})$$

To illustrate this, consider again the tuberculosis testing example of Table B.2. Let us associate the number 0 with a negative test result and 1 with a positive test result. Similarly, let 0 represent a true health condition of not having tuberculosis, while 1 represents having the disease. With these associations, we can compute the following values:

$$\begin{aligned} E(\text{test}) &= 0.2 & E(\text{test}^2) &= 0.2 & \text{Var}(\text{test}) &= 0.2 - (0.2)^2 = 0.16 \\ E(\text{health}) &= 0.17 & E(\text{health}^2) &= 0.17 & \text{Var}(\text{health}) &= 0.17 - (0.17)^2 = 0.1411 \\ E(\text{test} \cdot \text{health}) &= 0.12 \\ \text{Cov}(\text{test}, \text{health}) &= E(\text{test} \cdot \text{health}) - E(\text{test})E(\text{health}) = 0.12 - (0.2)(0.17) = 0.086 \end{aligned}$$

The first line of results above gives the mean and variance of the test results. The second gives the mean and variance of the true health results. The third line computes the expected value of the product of health and the test. Since there is only one case in which the health random variable and the test random variable are both positive (the case in which you test positive and truly have the disease), and since the product of the two random variables in this case is 1, the expected value of the product is simply equal to the joint probability of both events occurring, or 0.12. The covariance is computed in the next and final line above.

The covariance is positive meaning that in general a negative result corresponds with a disease-free condition and a positive result corresponds with an unhealthy condition. How strong is this trend, however? What does a covariance of 0.086 mean? To understand how strong this correlation is, we need to compute the correlation coefficient. The *correlation coefficient* of X and Y is defined as

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \quad (\text{B.19})$$

In this example, the correlation coefficient is equal to $\frac{0.086}{\sqrt{(0.16)(0.1411)}} \cong 0.572$. The correlation coefficient is always between -1 and 1. A value close to -1 or 1 indicates a very strong (linear) correlation. A value close to 0 means little or no correlation. This value indicates a moderately strong positive correlation between the test results and a person's true health condition.

We now turn to specific discrete random variables that are of interest to us. We begin with the notion of a *Bernoulli trial*. A Bernoulli trial is any experiment that has exactly two outcomes. We denote these outcomes by *success* and *failure*. In the graduate school admission problem success might mean the student is admitted to a particular school and failure would mean she is not admitted. In testing for tuberculosis, success might mean that the patient is positive (it is a success in the sense that we have discovered one source of the patient's ailment) and failure means the test is negative. In general, choosing which outcome to call a success and which to call a failure is arbitrary, though we clearly have to be consistent in the definition of success or failure in any particular analysis.

Suppose we now have a sequence of n independent Bernoulli trials each with the same probability of success, p . The total number of successes in n independent Bernoulli trials is given by the *binomial* distribution. In particular, the probability that we have exactly x successes in n trials is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n. \quad (\text{B.20})$$

The expected number of success is given by

$$E(X) = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} = np. \quad (\text{B.21})$$

Similarly, we can show that

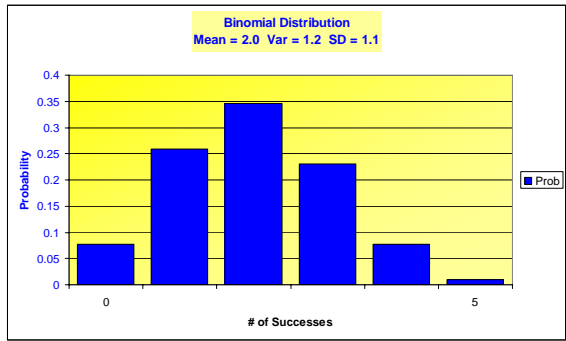
$$E(X^2) = n^2 p^2 - np^2 + np \quad (\text{B.22})$$

and

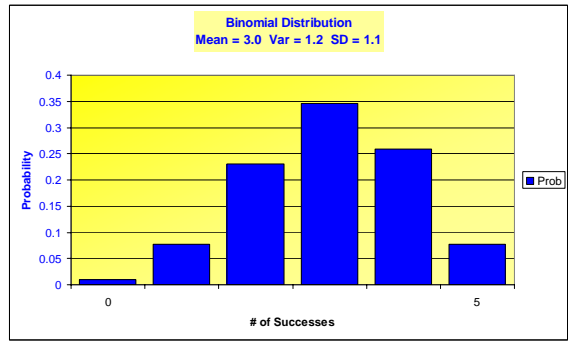
$$\begin{aligned} \text{Var}(X) &= E(X^2) - \{E(X)\}^2 \\ &= n^2 p^2 - np^2 + np - (np)^2 \\ &= np - np^2 \\ &= np(1-p) \end{aligned} \tag{B.23}$$

Figure B.2 shows the binomial distribution for different values of n and p .

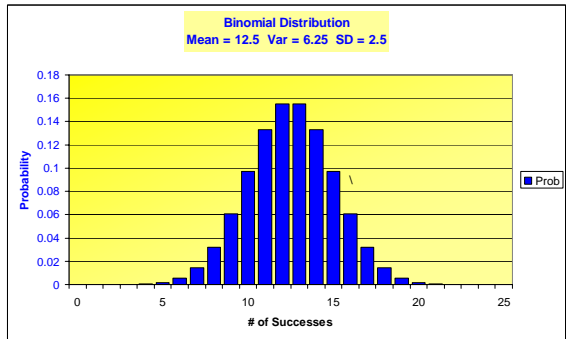
See [What do they look like.xls](#) to visualize different probability mass functions and probability density functions



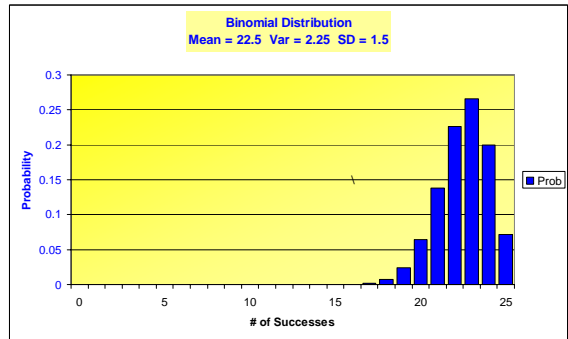
(a) $n=5, p=0.4$



(b) $n=5, p=0.6$



(c) $n=25, p=0.5$



(d) $n=25, p=0.9$

Figure B.2 -- Binomial distribution for various values of n and p

The *geometric distribution* is closely related to the binomial distribution and to Bernoulli trials. The geometric distribution gives the number of trials until the first success in a series of independent Bernoulli trials each with probability of success of p . Clearly, if the first success is on trial m , we must have had $m-1$ failures before that. Thus, the probability that the first success is on trial m is

$$P(\text{first success on trial } m) = P(M = m) = (1-p)^{m-1}p \text{ for } m = 1, 2, \dots \quad (\text{B.24})$$

Note that while this is a discrete distribution, the random variable – the trial number of the first success – can take on an infinite number of values. Our first task is to ensure that this is indeed a probability mass function. Clearly, all terms are positive for $0 < p < 1$.

Also, we have
$$\sum_{m=1}^{\infty} (1-p)^{m-1}p = p \sum_{m=0}^{\infty} (1-p)^m = p \sum_{m=0}^{\infty} q^m = \frac{p}{1-q} = \frac{p}{1-(1-p)} = \frac{p}{p} = 1,$$

where we have let $q = 1 - p$. (See Appendix A for a review of infinite sums.) Also, we have

$$\begin{aligned} E(M) &= \sum_{m=1}^{\infty} m(1-p)^{m-1}p \\ &= \frac{p}{1-p} \sum_{m=1}^{\infty} m(1-p)^m \\ &= \frac{p}{1-p} \sum_{m=0}^{\infty} mq^m \quad . \\ &= \frac{p}{1-p} \left(\frac{1-p}{p^2} \right) \\ &= \frac{1}{p} \end{aligned} \quad (\text{B.25})$$

We can also show that

$$\text{Var}(M) = \frac{1-p}{p^2}. \quad (\text{B.26})$$

We sometimes write the geometric distribution in terms of the number of *failures* before the first success. If K is the random number representing the number of failures before the first success, then we have

$$P(K = k) = (1-p)^k p \text{ for } k=0, 1, 2, \dots \quad (\text{B.27})$$

We also have

$$E(K) = \frac{1}{p} - 1 = \frac{1-p}{p} \quad (\text{B.28})$$

and

$$\text{Var}(K) = \frac{1-p}{p^2}. \quad (\text{B.29})$$

Note that this form of the geometric is the same as the earlier form except that we have subtracted 1 from every value of the random variable defined as the number of trials until the first success. Therefore, the expected value of this form is 1 less than that of the first form and the variance is the same.

Figure B.3 shows the geometric distribution for two different values of p . This is using the second form of the distribution in which the random variable is the number of failures until the first success.

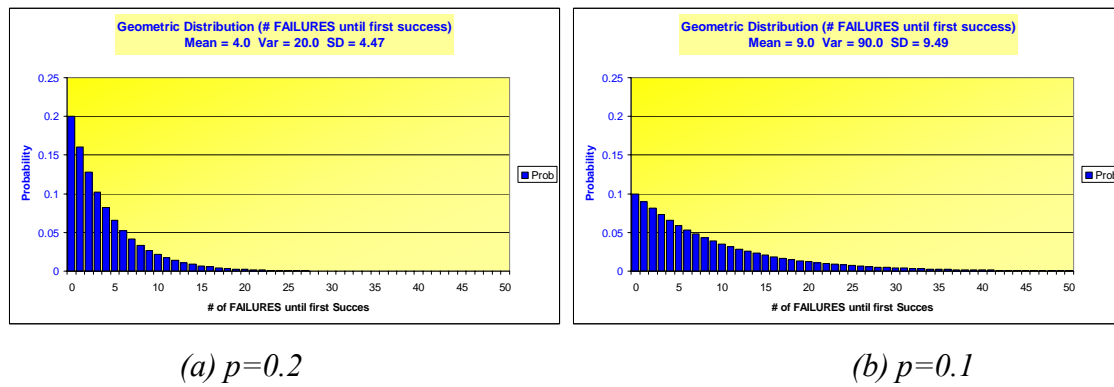


Figure B.3 -- Geometric distribution for different values of p

The *negative binomial* distribution is closely related to the geometric distribution. It gives the distribution of the number of failures until the r^{th} success. If X is the random variable representing this number, we have

$$P(X = x) = \binom{x+r-1}{r-1} p^r (1-p)^x \text{ for } x=0,1,2,\dots \quad (\text{B.30})$$

We also can show that

$$E(X) = r \frac{1-p}{p} \tag{B.31}$$

and

$$Var(X) = r \frac{1-p}{p^2} . \tag{B.32}$$

You should be able to verify these by noting that the negative binomial is simply the sum of r independent identically distributed geometric random variables (each representing the distribution of the number of failures until the first success). Figure B.4 shows the negative binomial distribution when $p=0.25$ for four values of r .

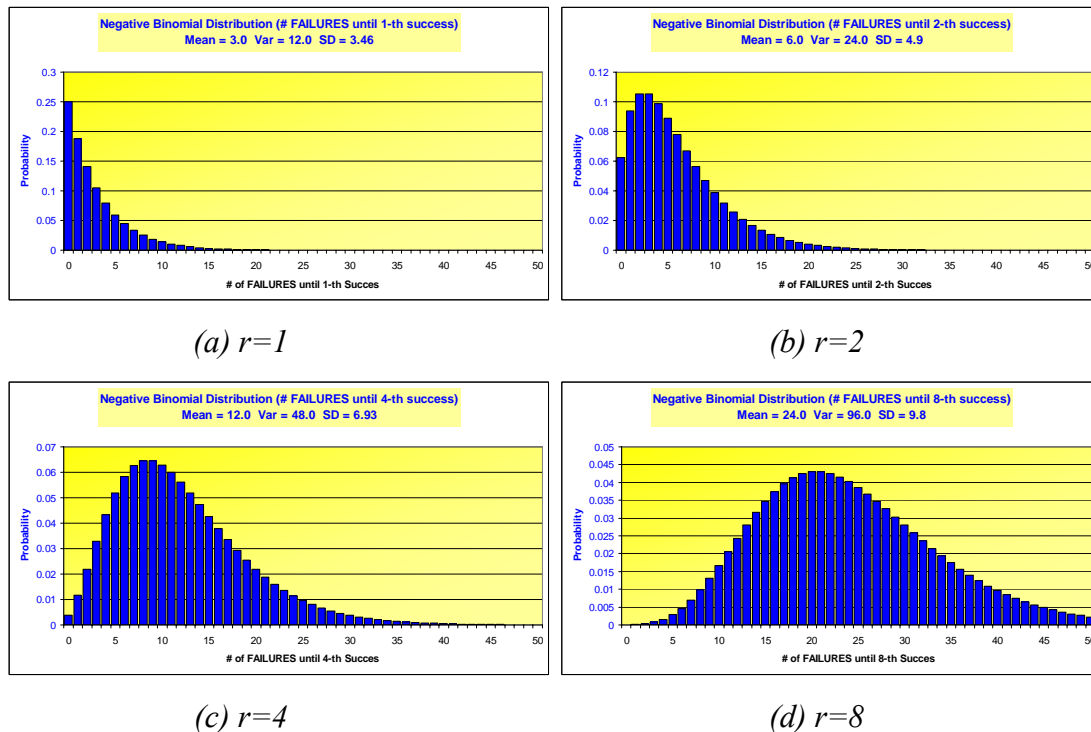


Figure B.4 -- Negative binomial distribution for different values of r

(Note that the vertical scale changes with each graph)

The *hypergeometric distribution* is derived as follows. We are given N individuals of whom M have a particular characteristic which we will consider a success. For example, we may have a class of 50 students of whom 23 are seniors (which we will call success). A sample of n is taken from the population without replacement. The

hypergeometric distribution gives the distribution of the number of successes in a sample of size n . If X is this random variable, we have

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \text{ for } \max(0, n - N + M) \leq x \leq \min(n, M). \quad (\text{B.33})$$

We also have

$$E(X) = n \frac{M}{N} = np, \quad (\text{B.34})$$

where $p = \frac{M}{N}$ is the fraction of the entire population that has the characteristic we are labeling a success. Note the similarity to the expected value of the binomial distribution. Also,

$$\text{Var}(X) = \left(\frac{N-n}{N-1} \right) n \left(\frac{M}{N} \right) \left(1 - \frac{M}{N} \right) = \left(\frac{N-n}{N-1} \right) np(1-p). \quad (\text{B.35})$$

Note that this is similar to the variance of the binomial distribution, except that we have a correction term $\left(\frac{N-n}{N-1} \right)$ included in the variance. Also note that this correction term is always less than 1 as long as the sample has 2 or more elements in it.

As the final discrete distribution of interest, we consider the *Poisson distribution*. One way to derive the Poisson distribution is to begin with the binomial distribution and to let $n \rightarrow \infty$ and $p \rightarrow 0$ while holding their product $np = \lambda$. Recall that for the binomial distribution, we have

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n \quad (\text{B.36})$$

For $x=0$, we have

$$P(X = 0) = (1-p)^n \quad (\text{B.37})$$

Now, let $np = \lambda$ or $p = \lambda/n$. We now have

$$P(X=0) = \left(1 - \frac{\lambda}{n}\right)^n \quad (\text{B.38})$$

Now if we take the natural log of this, we get

$$\ln\{P(X=0)\} = \ln\left\{\left(1 - \frac{\lambda}{n}\right)^n\right\} = n \ln\left(1 - \frac{\lambda}{n}\right) \quad (\text{B.39})$$

But we know that

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad (\text{B.40})$$

If we let $x = -\lambda/n$, we get

$$\ln\left(1 - \frac{\lambda}{n}\right) = -\frac{\lambda}{n} - \frac{\lambda^2}{2n^2} - \frac{\lambda^3}{3n^3} - \frac{\lambda^4}{4n^4} - \dots \quad (\text{B.41})$$

And

$$\begin{aligned} \ln\{P(X=0)\} &= \ln\left\{\left(1 - \frac{\lambda}{n}\right)^n\right\} = n \ln\left(1 - \frac{\lambda}{n}\right) \\ &= n \left\{-\frac{\lambda}{n} - \frac{\lambda^2}{2n^2} - \frac{\lambda^3}{3n^3} - \frac{\lambda^4}{4n^4} - \dots\right\} \\ &= -\lambda - \frac{\lambda^2}{2n} - \frac{\lambda^3}{3n^2} - \frac{\lambda^4}{4n^3} - \dots \end{aligned} \quad (\text{B.42})$$

Now suppose that n gets very large and p gets very small while we hold the product $np = \lambda$ constant. For very large values of n , we have

$$\ln\{P(X = 0)\} \approx -\lambda \quad (\text{B.43})$$

So

$$P(X = 0) \approx e^{-\lambda} \quad (\text{B.44})$$

Now let us consider the subsequent terms. Consider the ratio of two successive terms in the binomial distribution.

$$\begin{aligned} \frac{P(X = k)}{P(X = k - 1)} &= \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} \\ &= \frac{(n-k+1)p}{k(1-p)} \\ &= \frac{np - (k-1)p}{k(1-p)} \\ &= \frac{\lambda - (k-1)p}{k(1-p)} \end{aligned} \quad (\text{B.45})$$

Again, as p gets very small, this is approximately

$$\frac{P(X = k)}{P(X = k - 1)} \approx \frac{\lambda}{k} \quad (\text{B.46})$$

Thus,

$$\begin{aligned}
P(X = 1) &\approx \frac{\lambda}{1} P(X = 0) \approx \lambda e^{-\lambda} \\
P(X = 2) &\approx \frac{\lambda}{2} P(X = 1) \approx \frac{\lambda^2}{2} e^{-\lambda} \\
P(X = 3) &\approx \frac{\lambda}{3} P(X = 2) \approx \frac{\lambda^3}{3!} e^{-\lambda}
\end{aligned}
\tag{B.47}$$

And so on. These are simply the terms of the *Poisson distribution*. In summary, the Poisson probability that $X = x$ is given

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x = 0, 1, 2, \dots \tag{B.48}$$

We can also show that for the Poisson distribution,

$$E(X) = \lambda, \tag{B.49}$$

$$E(X^2) = \lambda^2 + \lambda, \tag{B.50}$$

and

$$\text{Var}(X) = E(X^2) - \{E(X)\}^2 = \lambda^2 + \lambda - \{\lambda\}^2 = \lambda. \tag{B.51}$$

It is often useful to think about λ as a rate per unit time. In that case, the number of Poisson events in a time period of duration t is given by

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \text{ for } n = 0, 1, 2, \dots \tag{B.52}$$

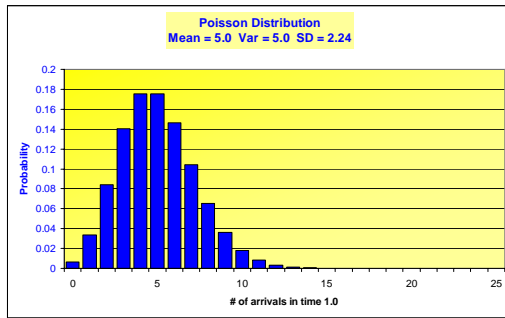
and

$$E(N(t)) = \lambda t \tag{B.53}$$

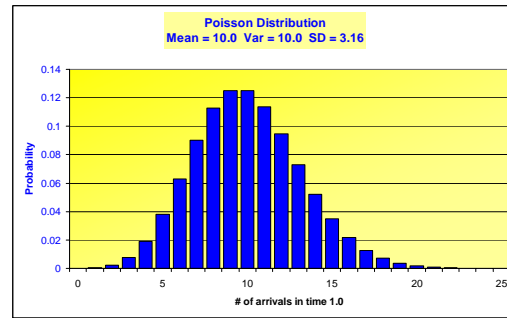
and

$$\text{Var}(N(t)) = \lambda t. \tag{B.54}$$

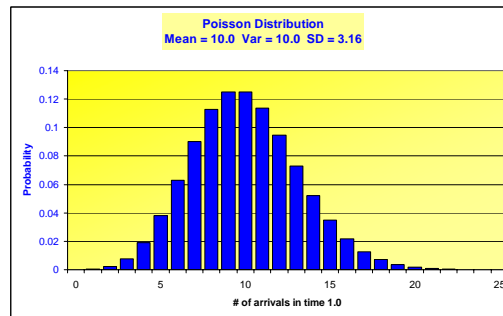
Figure B.5 shows the Poisson distribution for different values of λ for $t=1$.



(a) $\lambda = 5$



(b) $\lambda = 10$



(c) $\lambda = 15$

Figure B.5 -- Poisson distribution for different values of λ

We should note that we can approximate the binomial distribution by the Poisson distribution for large values of n and small values of p . Table B.3 shows the accuracy of this approximation in estimating the value of the probability that $X=5$ for the binomial distribution with different values of n and p . Generally for $n \geq 100$ and $p \leq 0.01$ and $np \leq 20$, the approximation works quite well.

Table B.3 -- Approximating the binomial by the Poisson

n	p	Prob(X=5)		Relative % Error
		Binomial	Poisson	
100	0.050	0.180018	0.175467	2.5%
100	0.025	0.066353	0.066801	0.7%
100	0.010	0.002898	0.003066	5.8%
200	0.050	0.035896	0.037833	5.4%
200	0.025	0.177701	0.175467	1.3%
200	0.010	0.035723	0.036089	1.0%
300	0.050	0.001641	0.001936	17.9%
300	0.025	0.109128	0.109375	0.2%
300	0.010	0.100985	0.100819	0.2%

B.7 Continuous random variables

We now turn our attention to continuous random variables. Table B.4 below summarizes key similarities between probability density functions which are used for continuous random variables and probability mass functions used for discrete random variables. Note that for a continuous random variable the density function must be non-negative (see the second row of the table) but it may be greater than 1 as long as it integrates to 1 (as shown in the third column). We will see distributions like this.

Table B.4 -- Relationships between discrete and continuous random variables

	Discrete	Continuous
Mass or density function	$p(x)$	$f(x)$
Probability is non-negative	$0 \leq p(x) \leq 1$	$0 \leq f(x)$
Total probability is 1	$\sum_{\text{all } x} p(x) = 1$	$\int_{-\infty}^{\infty} f(x) dx = 1$
Cumulative distribution	$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$	$F(x) = P(X \leq x) = \int_{-\infty}^x f(y) dy$
Expected value	$E(X) = \sum_{\text{all } x} xp(x)$	$E(X) = \int_{-\infty}^{\infty} xf(x) dx$
Expected value of X^2	$E(X^2) = \sum_{\text{all } x} x^2p(x)$	$E(X^2) = \int_{-\infty}^{\infty} x^2f(x) dx$
Variance	$V(X) = E(X^2) - (E(X))^2$	$V(X) = E(X^2) - (E(X))^2$
Expected value of a function of X	$E(h(X)) = \sum_{\text{all } x} h(x)p(x)$	$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x) dx$

Perhaps the simplest continuous distribution is the *uniform distribution*. This distribution is non-negative over an interval between a and b and has a constant height of $\frac{1}{b-a}$ in this interval and 0 everywhere else. Specifically, if X is a uniform random variable, we have

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases} \quad (\text{B.55})$$

The cumulative uniform distribution is given by:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (\text{B.56})$$

and the expected value is

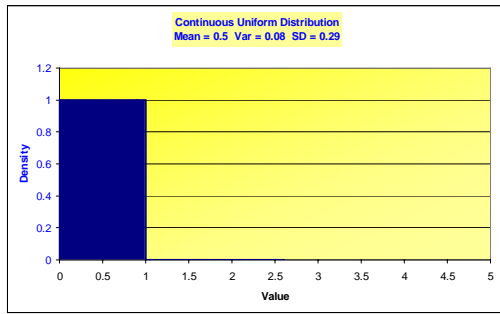
$$E(X) = \frac{a+b}{2} \quad (\text{B.57})$$

Finally, the variance is

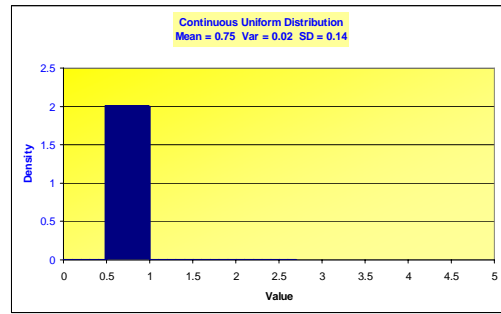
$$V(X) = \frac{(b-a)^2}{12}. \quad (\text{B.58})$$

The *standard uniform distribution* has $a=0$ and $b=1$ so that $E(X) = \frac{1}{2}$ and $V(X) = \frac{1}{12}$.

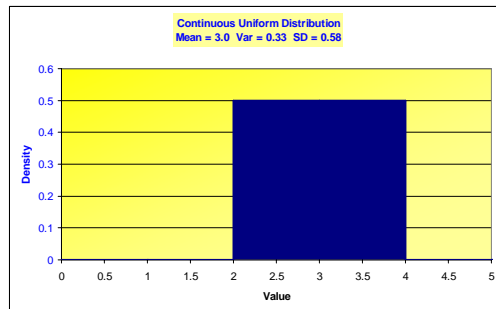
Figure B.6 shows both the standard uniform distribution as well as a few other variants of the distribution.



(a) Standard uniform; $a=0$, $b=1$



(b) $a=0.5$, $b=1$



(c) $a=2$, $b=4$

Figure B.6 -- Uniform distribution for different values of a and b

(Note that in (b) the density function takes on values greater than 1, but it still integrates to 1 over the entire range of the random variable.)

At this point, it is useful to introduce another way of finding the expected value of a random variable. We can show that if X is a non-negative random variable (meaning that X can not take on negative values), then

$$E(X) = \int_0^{\infty} [1 - F(x)] dx \quad (\text{B.59})$$

where $F(x)$ is the cumulative distribution of X . To illustrate this, consider the uniform distribution, whose distribution is given above. In this case,

$$\begin{aligned}
E(X) &= \int_0^{\infty} [1 - F(x)] dx \\
&= \int_0^a 1 dx + \int_a^b \left\{ 1 - \frac{x-a}{b-a} \right\} dx \\
&= a + \int_a^b \left\{ \frac{b-x}{b-a} \right\} dx \\
&= a - \frac{(b-x)^2}{2(b-a)} \Big|_a^b \\
&= a + \frac{b-a}{2} \\
&= \frac{a+b}{2}
\end{aligned} \tag{B.60}$$

which is just what we had above. Sometimes it is easier to compute the expected value of a (continuous) random variable in this manner.

For example, consider the problem of finding the expected distance between a facility located in the center of a diamond shaped region and a randomly chosen demand point in the region. Demands are uniformly distributed in the region and travel is at 45 degrees to the sides of the region. This is illustrated in Figure B.7 below.

If the service region has an area A , then each side of the service region has a length \sqrt{A} and the maximum distance between the facility and any point is $\sqrt{A/2}$. This distance is realized for any point on the boundary of the service region.

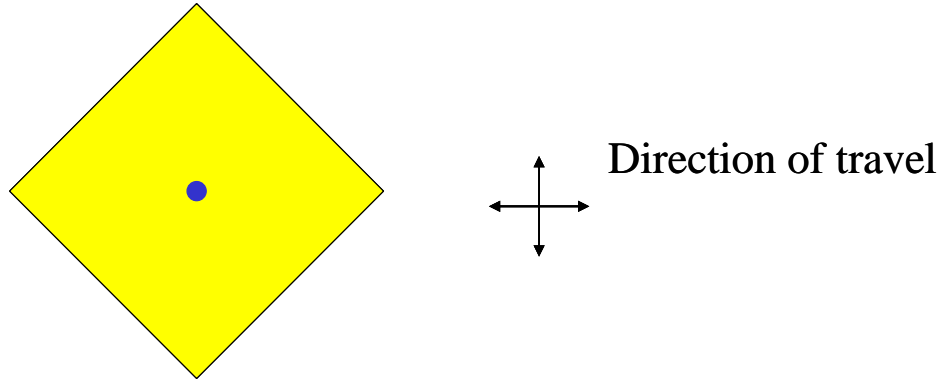


Figure B.7 -- Example service region and direction of travel within the region

Now let us find the cumulative distribution of the distance between a randomly selected point in the diamond and the facility at the center of the diamond. In Figure B.8, we show the original service region as well as a subregion, or smaller diamond. Any point inside (or on the boundary) of the smaller diamond is at a distance of x or less from the center of the service region. Therefore, the probability that a randomly selected point is at a distance x or less is given by the ratio of the area of the subregion to the area of the entire region. The area of the subregion is $2x^2$ and so the cumulative probability of the distance between the center of the service region and a randomly selected point is $\frac{2x^2}{A}$.

We can now use the formula above for finding the expected distance between the center and a randomly selected demand. This distance is given by:

$$\begin{aligned}
 E(Dist) &= \int_0^{\infty} \{1 - F(x)\} dx \\
 &= \int_0^{\sqrt{A/2}} \left\{1 - \frac{2x^2}{A}\right\} dx \\
 &= x - \frac{2x^3}{3A} \Bigg|_0^{\sqrt{A/2}} \\
 &= \frac{2}{3} \sqrt{\frac{A}{2}}
 \end{aligned}
 \tag{B.61}$$

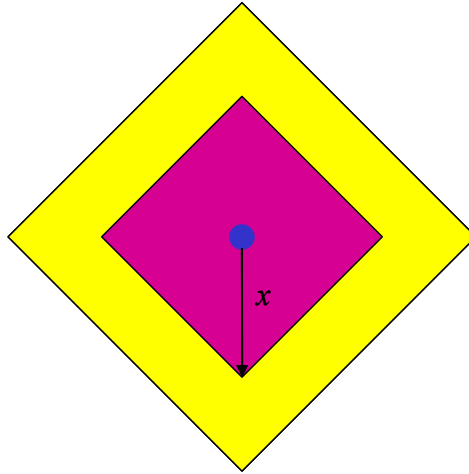


Figure B.8 -- Service region with a subregion in which distances are less than or equal to x

Thus, the expected distance is $\frac{2}{3}$ of the maximum distance. Knowing that the cumulative distribution is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{2x^2}{A} & 0 \leq x \leq \sqrt{A/2} \\ 1 & x > \sqrt{A/2} \end{cases} \quad (\text{B.62})$$

we can compute the density function of the distance by taking the derivative (appropriately) of the cumulative distribution. We find that the density function is simply

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{4x}{A} & 0 \leq x \leq \sqrt{A/2} \\ 0 & x > \sqrt{A/2} \end{cases} \quad (\text{B.63})$$

This is simply a *triangular density* function as shown in Figure B.9.

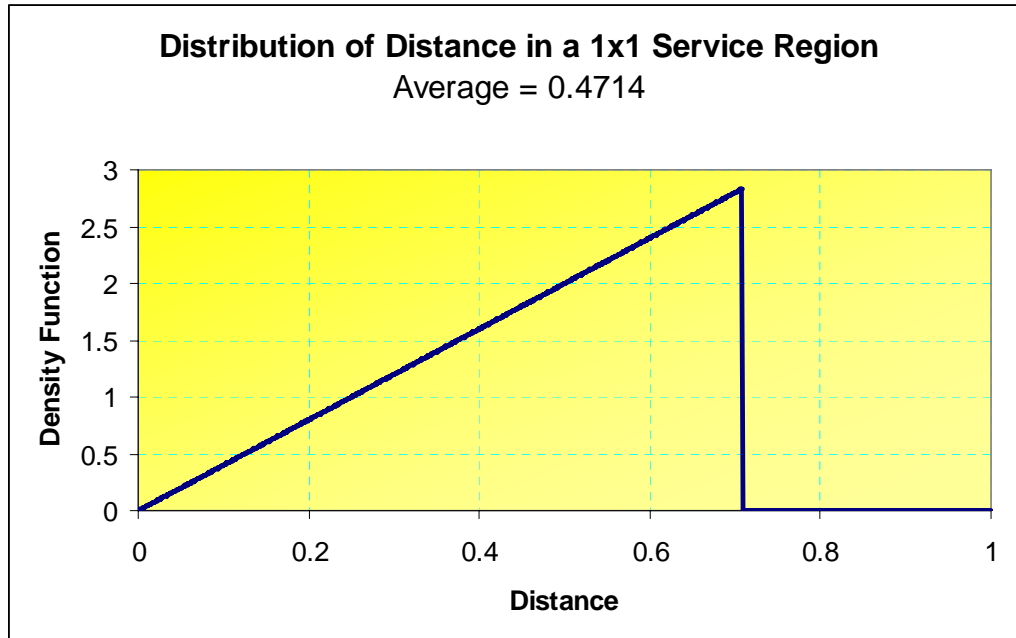


Figure B.9 -- Distribution of distance between the center of a unit diamond and a randomly selected point within the diamond

Because of its simplicity, the uniform distribution is also a good distribution to use to illustrate certain other functions of random variables. Sometimes we are interested in the minimum or maximum of two independent random variables. Suppose, for example, we want the distribution of the maximum of two independent standard uniform random variables. In finding the distribution of the minimum or the maximum, it is often easiest to work with the cumulative distribution. Thus, let X_1 and X_2 be the two independent uniform random variables, and let $Y = \max(X_1, X_2)$. The probability that $Y \leq y$ is the probability that $X_1 \leq y$ and $X_2 \leq y$. Since X_1 and X_2 are independent, the probability that $X_1 \leq y$ and $X_2 \leq y$ is simply the product of the probability that $X_1 \leq y$ and the probability that $X_2 \leq y$. In other words, we have

$$\begin{aligned}
P(Y \leq y) &= P(X_1 \leq y \text{ and } X_2 \leq y) \\
&= P(X_1 \leq y)P(X_2 \leq y) \\
&= y \cdot y \\
&= y^2
\end{aligned}
\tag{B.64}$$

Since the density function is just the derivative of the cumulative distribution we

$$\text{have } f(y) = \begin{cases} 2y & 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}
\tag{B.65}$$

This is just a *triangular* distribution. This distribution is shown in Figure B.10.

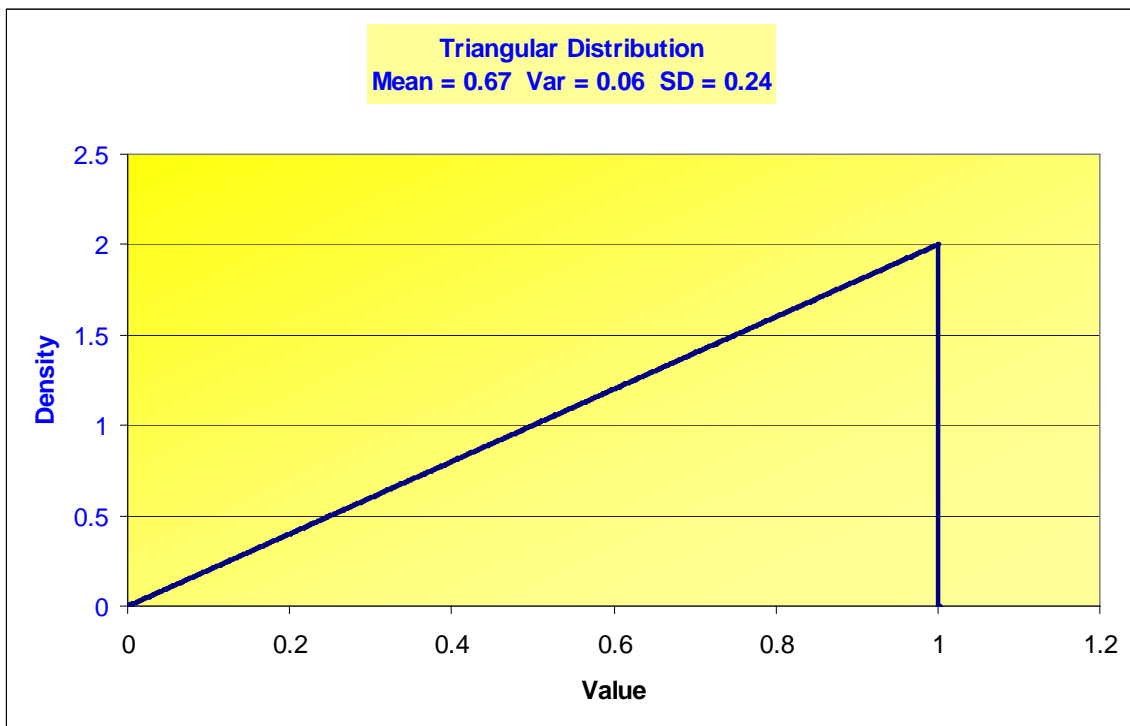


Figure B.10 -- Triangular distribution: the density function of the maximum of two standard uniform random variables

We can also find the distribution of $Z = \min(X_1, X_2)$. Using similar logic we have

$$\begin{aligned}
P(Z \geq z) &= P(X_1 \geq z \text{ and } X_2 \geq z) \\
&= [1 - P(X_1 \leq z)] \cdot [1 - P(X_2 \leq z)] \\
&= (1 - z)(1 - z) \\
&= 1 - 2z + z^2
\end{aligned}
\tag{B.66}$$

So

$P(Z \leq z) = 1 - P(Z \geq z) = 1 - (1 - 2z + z^2) = 2z - z^2$. Therefore, the density function of the minimum of two standard uniform random variables is given by

$$f(z) = \begin{cases} 2 - 2z & 0 \leq z \leq 1 \\ 0 & \text{elsewhere} \end{cases}
\tag{B.67}$$

This too is a triangular distribution, but it slopes down from 2 (when $z=0$) to 0 (when $z=1$). It has a mean of $1/3$ and, like the maximum of two standard uniform random variables, it has a variance of $1/18$.

The *normal distribution* may also be one of most recognized and common of probability distributions. If X is normally distributed with mean μ and variance σ^2 , then the density function of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}
\tag{B.68}$$

This distribution is defined for all values of the random variable X . In other words, X can take on any value from $-\infty$ to ∞ . Figure B.11 shows the *standard normal* distribution; i.e., the normal distribution with a mean of 0 and a variance of 1.

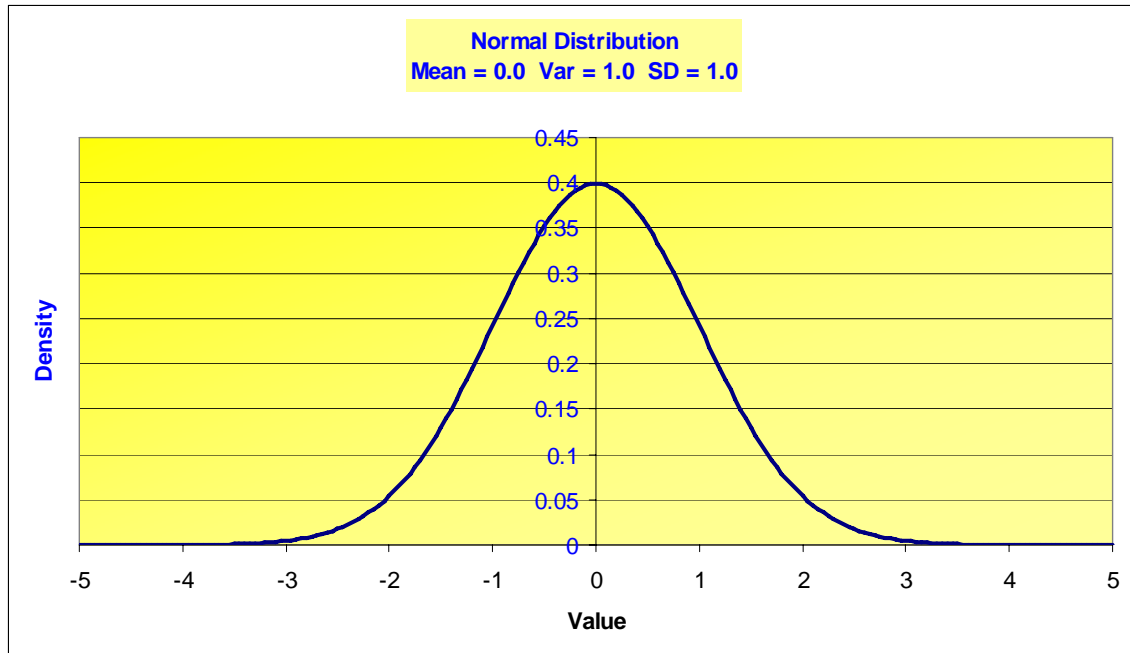


Figure B.11 -- Standard normal distribution

There is no closed form formula for the cumulative distribution of X , though the cumulative of the standard normal distribution is widely tabulated and there are a number of very good numerical approximations for the cumulative distribution. Table B.5 tabulates key values of the standard normal cumulative distribution. The third column in the table gives the probability that the random variable is within the given number of standard deviations of the mean.

Table B.5 -- Selected values of the cumulative normal distribution

Standard deviations, k, from the mean	P(X < k*sd)	P(-k*sd < x < k*sd)
0.000	0.500	0.0000
0.500	0.691	0.3829
1.000	0.841	0.6827
1.282	0.900	0.8002
1.645	0.950	0.9000
1.960	0.975	0.9500
2.326	0.990	0.9800
2.576	0.995	0.9900
3.000	0.999	0.9973

Note that if Y is normally distributed with mean μ_Y and variance σ_Y^2 , then $X = \frac{Y - \mu_Y}{\sigma_Y}$

is a standard normal random variable with mean 0 and standard deviation 1.

We can also use the normal distribution to approximate the binomial distribution when $np > 10$ and $n(1-p) > 10$. In this case, we are approximating a discrete distribution by a continuous distribution and so it is appropriate to include a continuity correction of 0.5. Thus, if Z has a binomial distribution with $n=50$ and $p=0.6$, then we can approximate the cumulative binomial distribution by a normal distribution with mean $np=30$ and variance $np(1-p)=12$. In other words, we would approximate

$$P(Z \leq z) = \sum_{k=0}^z \binom{n}{k} p^k (1-p)^{n-k} \text{ by } P(Y \leq k + 0.5) = P\left(X \leq \frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right),$$

where Y now has a normal distribution with mean $np=30$ and variance $np(1-p)=12$, and X is a standard normal random variable. Table B.6 shows how well this approximation does.

Table B.6 -- Approximating the binomial distribution by the normal distribution

x	P(X≤x)		Absolute Difference
	Binomial	Normal	
23	0.0314	0.0303	0.0011
25	0.0978	0.0970	0.0008
27	0.2340	0.2352	0.0013
29	0.4390	0.4426	0.0037
31	0.6644	0.6675	0.0031
33	0.8439	0.8438	0.0001
35	0.9460	0.9438	0.0022
37	0.9867	0.9848	0.0019

Similarly, we can approximate the Poisson distribution by the normal distribution if $\lambda \geq 20$. For example, suppose that W is a Poisson random variable with $\lambda = 30$. Table B.7 shows how well the normal distribution (including the continuity correction) approximates the Poisson in this case. Note that the variance of the normal distribution is $\lambda = 30$ in this case.

Table B.7 -- Approximating the Poisson distribution by the normal distribution

w	P(W≤w)		Absolute Difference
	Poisson	Normal	
23	0.1146	0.1177	0.0030
25	0.2084	0.2057	0.0027
27	0.3329	0.3240	0.0088
29	0.4757	0.4636	0.0121
31	0.6186	0.6079	0.0107
33	0.7444	0.7386	0.0059
35	0.8426	0.8423	0.0003
37	0.9110	0.9145	0.0036

The normal distribution arises frequently in statistics, due in large part to the Central Limit Theorem which states that, if X_1, X_2, \dots, X_n is a random sample from some distribution which is *not* necessarily normal, but has a mean μ and variance σ^2

and if n is sufficiently large, then (a) the sample average $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ is approximately normally distributed with mean $E(\bar{X}) = \mu$ and variance $Var(\bar{X}) = \frac{\sigma^2}{n}$ and (b) the total $T = \sum_{i=1}^n X_i$ is approximately normally distributed with mean $E(T) = n\mu$ and variance $Var(T) = n\sigma^2$. This is a key theorem in statistics as it allows us to derive confidence intervals and to perform hypothesis testing related to the true (but unknown) mean of a distribution as long as the sample size is large enough. In general, we need $n \geq 30$ for this theorem to hold reasonably well. We will see this again in a later section of this appendix where we discuss one procedure for generating (approximately) normally distributed random variables.

Finally, we turn our attention to two key distributions used extensively in queueing theory: the *exponential* and *Erlang- k* distributions. As we will see, these two distributions are intimately linked to each other and to the *Poisson* distribution. Specifically, if the number of arrivals or events in time t , $N(t)$, follows a Poisson distribution or Poisson process, then we have

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad n = 0, 1, 2, \dots \quad (\text{B.69})$$

and in particular,

$$P(N(t) = 0) = e^{-\lambda t}. \quad (\text{B.70})$$

Also, if the number of arrivals of one type of customer is Poisson with parameter λ_1 (e.g., the number of non-critical patients at an emergency room) and the number of arrivals of another type of customer is also Poisson with parameter λ_2 (e.g., the number

of critical patients arriving at an emergency room), then the sum of the number of arrivals of the first and second types of customers is also Poisson with parameter, $\lambda_1 + \lambda_2$. This is known as the *superposition principle of the Poisson process*.

If the time between arrivals is exponential, we have

$$P(\text{interarrival time} \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t} \quad t \geq 0 \quad (\text{B.71})$$

and

$$P(\text{interarrival time} > t) = \int_t^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda t} \quad t \geq 0 \quad (\text{B.72})$$

where in (B.69)-(B.73), λ is the rate of customer arrivals per unit time.

From equations (B.70) and (B.72), we see that **Poisson arrivals imply that the time between arrivals is exponentially distributed and vice versa**. The *exponential* probability density function can be obtained by taking the derivative of (B.71) with respect to t . This results in

$$f(t) = \lambda e^{-\lambda t} \quad t \geq 0 \quad (\text{B.73})$$

The mean and variance of the exponential distribution are computed as follows (using integration by parts):

$$E(T) = \int_0^{\infty} t \lambda e^{-\lambda t} dt = -t e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda} \int_0^{\infty} \lambda e^{-\lambda t} dt = \frac{1}{\lambda} \quad (\text{B.74})$$

$$E(T^2) = \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt = -t^2 e^{-\lambda t} \Big|_0^{\infty} + 2 \int_0^{\infty} t e^{-\lambda t} dt = \frac{2}{\lambda} \int_0^{\infty} t \lambda e^{-\lambda t} dt = \frac{2}{\lambda^2} \quad (\text{B.75})$$

$$\text{Var}(T) = E(T^2) - \{E(T)\}^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad (\text{B.76})$$

The exponential distribution (in blue) along with the cumulative distribution for the exponential are shown below, in Figure B.12, for the case of $\lambda = 0.2$.

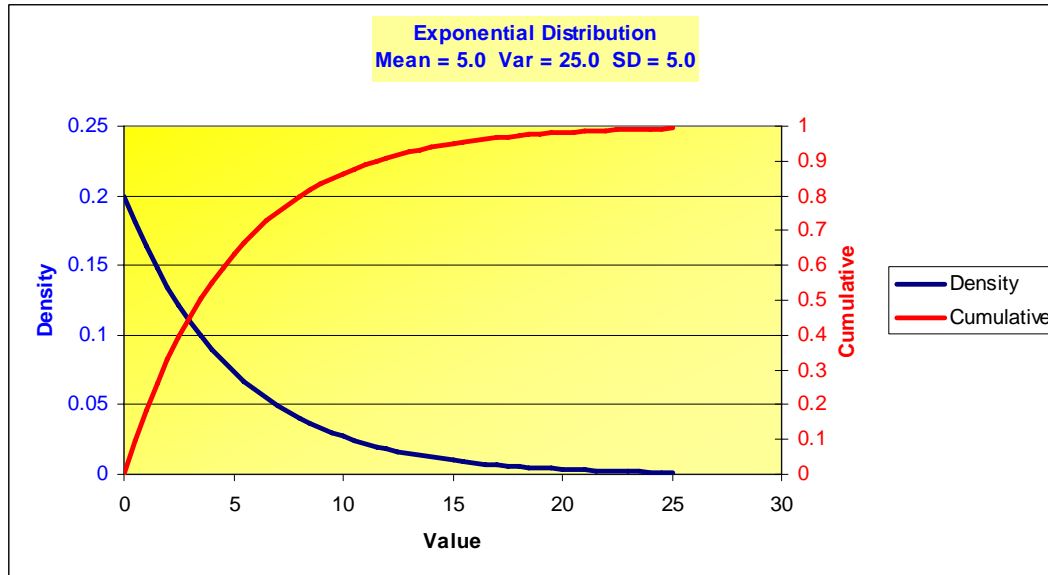


Figure B.12 -- The exponential distribution

The exponential distribution has a key property, the *memoryless property*, that makes it particularly useful in queuing theory. In words the property states that if the interarrival times (for example) are exponential, then the probability distribution of the *remaining* time until an arrival *given* that we have already waited T_0 minutes is also exponential with the same (original) parameter. To see this, we note that if t =interarrival time, for $K > T_0$,

$$\begin{aligned}
 P(t > K | t > T_0) &= \frac{P(t > K \text{ AND } t > T_0)}{P(t > T_0)} \\
 &= \frac{P(t > K)}{P(t > T_0)} \\
 &= \frac{e^{-\lambda K}}{e^{-\lambda T_0}} \\
 &= e^{-\lambda(K-T_0)}
 \end{aligned}
 \tag{B.77}$$

but $K - T_0$ is the *remaining* time. So the

$$P(\text{remaining time} > K - T_0 = R | t > T_0) = e^{-\lambda(K-T_0)} = e^{-\lambda R} \quad (\text{B.78})$$

which is exponential with the original parameter λ . In other words, if bus arrivals follow a Poisson process with a mean of 6 per hour (one every 10 minutes on average), the expected *additional* waiting time *given* we have been waiting 8 minutes is *10 more minutes, not 2 minutes*. Note that an estimate of 2 more minutes would be wrong for virtually *all* distributions of interarrival times.

(This property means that, in modeling a queue with Poisson arrivals, we do not need to know when the last person arrived to characterize the state of the system. Similarly, we do not need to know how long the current customers have been in service if service times are exponentially distributed. Qualitatively speaking this means that the state space described by the number of people in the system is *Markovian*, meaning we do not have to worry about *how* we got to the state in order to fully describe the probability distribution of the state space at some future point in time.)

If X_1, X_2, \dots, X_k are k independent identically distributed random variables, each with an exponential distribution given by $f_{X_i}(x_i) = \lambda e^{-\lambda x_i} \quad x_i \geq 0, \quad i = 1, 2, \dots, k$, then $S_k = \sum_{i=1}^k X_i$ is a random variable with an *Erlang- k* distribution:

$$f_{S_k}(s) = \frac{\lambda (\lambda s)^{k-1} e^{-\lambda s}}{(k-1)!} \quad s \geq 0 \quad k = 1, 2, \dots \quad (\text{B.79})$$

This equation is derived below based on the cumulative distribution, which we also derive. Sometimes we write

$$f_{S_k}(s) = \frac{k \nu (k \nu s)^{k-1} e^{-k \nu s}}{(k-1)!} \quad s \geq 0 \quad k = 1, 2, \dots \quad (\text{B.80})$$

where $\lambda = k\nu$. We have

$$E(S_k) = \frac{k}{\lambda} = \frac{1}{\nu} \quad (\text{B.81})$$

$$\text{Var}(S_k) = \frac{k}{\lambda^2} = \frac{1}{k\nu^2} \quad (\text{B.82})$$

So the Erlang- k distribution is the distribution of the sum of k i.i.d. (independent, identically distributed) exponential random variables. The cumulative Erlang- k may be found by noting that

$$\begin{aligned} P(S_k > s) &= P(\text{sum of } k \text{ i.i.d. Exponential random variables} > s) \\ &= P(k-1 \text{ or fewer Poisson arrivals in time } s) \\ &= \sum_{n=0}^{k-1} \frac{(\lambda s)^n e^{-\lambda s}}{n!} \end{aligned} \quad (\text{B.83})$$

So, the Poisson and Erlang- k are related. In particular,

$$\begin{aligned} P(k \text{ or more Poisson arrivals in time } s) &= P(S_k < s) \\ &= \text{cumulative Erlang-}k \end{aligned} \quad (\text{B.84})$$

This relationship between the cumulative Erlang- k distribution and the Poisson distribution can be used to derive the probability mass function for the Erlang- k

distribution as follows. Note that $P(S_k > s) = \sum_{n=0}^{k-1} \frac{(\lambda s)^n e^{-\lambda s}}{n!}$, so

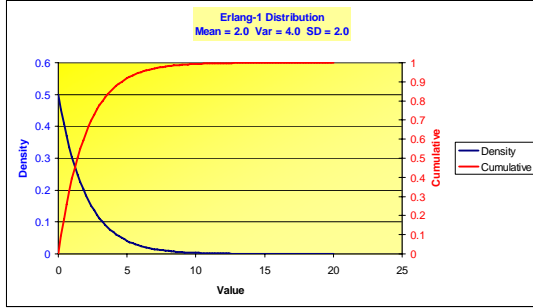
$$P(S_k \leq s) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda s)^n e^{-\lambda s}}{n!}. \quad (\text{B.85})$$

This is the cumulative Erlang- k distribution. For any cumulative distribution, if we take the derivative, we get the density function. Thus, we have:

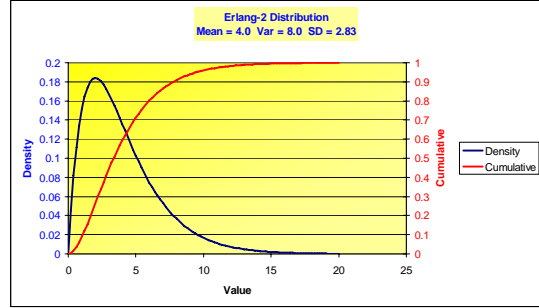
$$\begin{aligned}
f_{S_k}(s) &= \frac{dP(S_k \leq s)}{ds} \\
&= \frac{d \left\{ 1 - \sum_{n=0}^{k-1} \frac{(\lambda s)^n e^{-\lambda s}}{n!} \right\}}{ds} \\
&= - \sum_{n=0}^{k-1} \frac{n \lambda (\lambda s)^{n-1} e^{-\lambda s}}{n!} + \lambda \sum_{n=0}^{k-1} \frac{(\lambda s)^n e^{-\lambda s}}{n!} \\
&= - \sum_{n=1}^{k-1} \frac{n \lambda (\lambda s)^{n-1} e^{-\lambda s}}{n!} + \lambda \sum_{n=0}^{k-1} \frac{(\lambda s)^n e^{-\lambda s}}{n!} \\
&= - \sum_{n=1}^{k-1} \frac{\lambda (\lambda s)^{n-1} e^{-\lambda s}}{(n-1)!} + \sum_{n=0}^{k-1} \frac{\lambda (\lambda s)^n e^{-\lambda s}}{n!} \\
&= - \sum_{n=0}^{k-2} \frac{\lambda (\lambda s)^n e^{-\lambda s}}{n!} + \sum_{n=0}^{k-1} \frac{\lambda (\lambda s)^n e^{-\lambda s}}{n!} \\
&= \frac{\lambda (\lambda s)^{k-1} e^{-\lambda s}}{(k-1)!}
\end{aligned} \tag{B.86}$$

Figure B.13 shows the Erlang- k distribution for various values of k , where the distribution is written in the form of (B.79) so that the mean increases with k . Note that as k increases, the distribution looks more and more like a normal distribution. This is as expected from the Central Limit Theorem (CLT), since the Erlang- k distribution is nothing more than the sum of k i.i.d. exponential random variables and the CLT tells us that the distribution of the sum of i.i.d. random variables approaches the normal distribution as the number of random variables being added together increases.

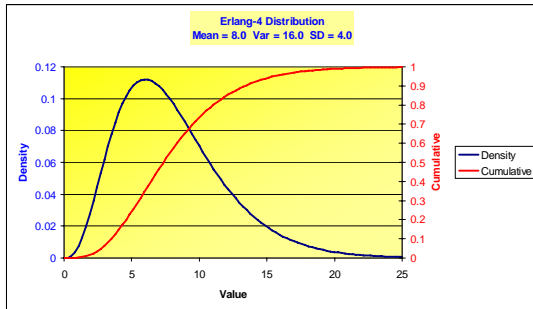
Figure B.14 summarizes the key distributions we have discussed above.



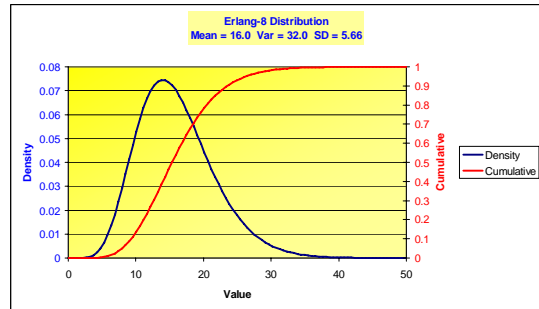
(a) $\lambda = 0.5; k=1$



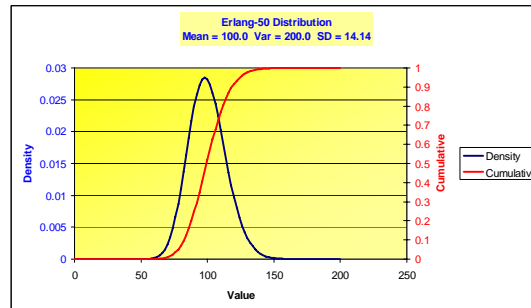
(b) $\lambda = 0.5; k=2$



(c) $\lambda = 0.5; k=4$



(d) $\lambda = 0.5; k=8$



(e) $\lambda = 0.5; k=50$

Figure B.13 -- Shapes of the Erlang- k distribution for different values of k

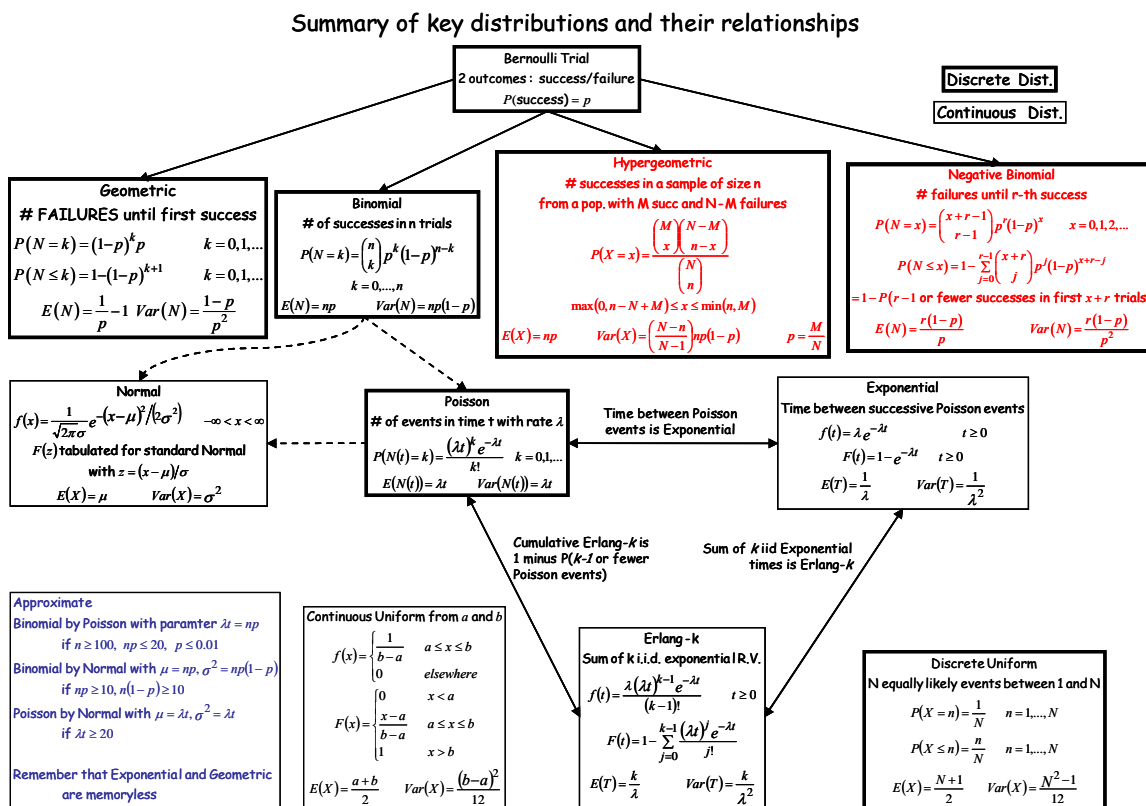


Figure B.14 -- Summary of key probability distributions and their relationships

B.8 Moment and probability generating functions¹

At times, it is easier to deal with probability distributions using what is called the transform domain. In this section, we outline moment and probability generating functions. We begin with moment generating functions.

The *moment generating function* of a random variable X is defined as

$$MGF(\theta) = E(e^{\theta X}) \quad (\text{B.87})$$

where θ is a parameter of the MGF. The following key properties of the moment generating function are important:

1. There is a one-to-one relationship between a moment generating function and a probability mass function. Knowing one automatically gives you the other.
2. If $T = X + Y$ and X and Y are independent random variables then

$$E(e^{\theta T}) = E(e^{\theta(X+Y)}) = E(e^{\theta X})E(e^{\theta Y}). \quad (\text{B.88})$$

So the MGF of the sum of independent random variables is the product of the MGFs of the individual random variables. Thus, the MGF of the Erlang- k distribution is the product of k MGFs of exponential distributions as discussed below.

3. Finally, we note that since $e^{\theta X} = 1 + (\theta X) + \frac{(\theta X)^2}{2!} + \frac{(\theta X)^3}{3!} + \frac{(\theta X)^4}{4!} + \dots$, we

have

$$\left. \frac{d^n MGF(\theta)}{d\theta^n} \right|_{\theta=0} = E(X^n), \quad (\text{B.89})$$

hence the name *moment generating function*.

¹ Note that this section is advanced material that can be skipped by most readers. It will be used to a limited extent in chapter 3 on queueing.

To illustrate the use of moment generating functions, let us consider again the exponential and Erlang- k distributions. The MGF of the exponential distribution is given by

$$E(e^{\theta X}) = \int_0^{\infty} e^{\theta x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - \theta} \int_0^{\infty} (\lambda - \theta) e^{-(\lambda - \theta)x} dx = \frac{\lambda}{\lambda - \theta}. \quad (\text{B.90})$$

Since the Erlang- k distribution is the sum of k i.i.d. exponential random variables, the MGF of the Erlang- k distribution is given by

$$\left(\frac{\lambda}{\lambda - \theta} \right)^k \quad (\text{B.91})$$

using the second property above. We can now use property (3) above to find the mean and variance of the Erlang- k distribution. The mean is given by

$$\begin{aligned} E(X) &= \left. \frac{dMGF(\theta)}{d\theta} \right|_{\theta=0} \\ &= \left. \frac{d \left(\frac{\lambda}{\lambda - \theta} \right)^k}{d\theta} \right|_{\theta=0} \\ &= k \left(\frac{\lambda}{\lambda - \theta} \right)^k \frac{\lambda}{(\lambda - \theta)^2} \Big|_{\theta=0} \\ &= \frac{k}{\lambda} \end{aligned} \quad (\text{B.92})$$

We also have

$$\begin{aligned} E(X^2) &= \left. \frac{d^2 MGF(\theta)}{d\theta^2} \right|_{\theta=0} \\ &= \left. \frac{d \frac{k}{\lambda} \left(\frac{\lambda}{\lambda - \theta} \right)^{k+1}}{d\theta} \right|_{\theta=0} \\ &= \frac{k}{\lambda} (k+1) \left(\frac{\lambda}{\lambda - \theta} \right)^k \frac{\lambda}{(\lambda - \theta)^2} \Big|_{\theta=0} \\ &= \frac{k(k+1)}{\lambda^2} \end{aligned} \quad (\text{B.93})$$

So

$$\begin{aligned}
\text{Var}(X) &= E(X^2) - E^2(X) \\
&= \frac{k(k+1)}{\lambda^2} - \frac{k^2}{\lambda^2} \\
&= \frac{k}{\lambda^2}
\end{aligned}
\tag{B.94}$$

As expected (B.92) and (B.94) agree with (B.81) and (B.82) respectively.

The *probability generating function* is the second type of generating function we want to consider. The probability generating function (PGF) of a discrete non-negative random variable X is defined as:

$$PGF(z) = \sum_{x=0}^{\infty} z^x p_x \tag{B.95}$$

Just as the MGF is a function of θ , so the PGF is a function of z . There are several key properties of the PGF:

1. $PGF(z) \Big|_{z=1} = 1$. This is clear from substituting $z=1$ into the definition (B.95)

of the PGF.

2. $\frac{d^n PGF(z)}{dz^n} \Big|_{z=0} = n! p_n$. Thus, by taking successive derivatives of the PGF and

evaluating each at $z=0$, we can obtain back the individual probabilities (multiplied by $n!$).

3. $\frac{dPGF(z)}{dz} \Big|_{z=1} = \sum_{n=0}^{\infty} z^{n-1} n p_n \Big|_{z=1} = E(X)$. In other words, taking the first

derivative of the PGF with respect to z and evaluating it at $z=1$ gives the expected value of the random variable.

4. $\frac{d^2 PGF(z)}{dz^2} \Big|_{z=1} = \sum_{n=0}^{\infty} z^{n-2} n(n-1) p_n \Big|_{z=1}$
 $= \sum_{n=0}^{\infty} z^{n-2} n^2 p_n \Big|_{z=1} - \sum_{n=0}^{\infty} z^{n-2} n p_n \Big|_{z=1}$. Thus, we can obtain
 $= E(X^2) - E(X)$

the $E(X^2)$ by taking the second derivative of the PGF with respect to z and

evaluating it at $z=1$ and then adding the mean or expected value back into the expression.

5. Finally, there is a one-to-one relationship between a PGF and a probability mass function.

To illustrate the use of the PGF, consider the binomial distribution. The PGF for this distribution is

$$\sum_{x=0}^n z^x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^n \binom{n}{x} (pz)^x (1-p)^{n-x} = (1-p+pz)^n \quad (\text{B.96})$$

Note that when $z=1$, this evaluates to 1 as expected by property (1) above. If we take the first derivative of this with respect to z , and evaluate the expression at $z=1$, we obtain

$$\left. \frac{d(1-p+pz)^n}{dz} \right|_{z=1} = np(1-p+pz)^{n-1} \Big|_{z=1} = np \quad (\text{B.97})$$

which agrees with the mean of the binomial distribution in equation (B.21). Similarly, if we take the second derivative with respect to z and evaluate the expression at $z=1$, we obtain

$$\begin{aligned} \left. \frac{d^2(1-p+pz)^n}{dz^2} \right|_{z=1} &= \left. \frac{dn p(1-p+pz)^{n-1}}{dz} \right|_{z=1} \\ &= n(n-1)p^2 \\ &= E(N^2) - E(N) \end{aligned} \quad (\text{B.98})$$

If we now use this to obtain the variance we find

$$\begin{aligned} \text{Var}(N) &= E(N^2) - E^2(N) \\ &= \{E(N^2) - E(N)\} + E(N) - E^2(N) \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np - np^2 \\ &= np(1-p) \end{aligned} \quad (\text{B.99})$$

which agrees with the variance of the binomial distribution in equation (B.23). Note that the term in braces in the second line of (B.99) is the second derivative of the PGF evaluated at $z = 1$ found in (B.98).

As a second example, consider the PGF of the Poisson distribution. The PGF is

$$\begin{aligned} \sum_{x=0}^{\infty} z^x \frac{\lambda^x e^{-\lambda}}{x!} &= \sum_{x=0}^{\infty} \frac{(\lambda z)^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} e^{\lambda z} \sum_{x=0}^{\infty} \frac{(\lambda z)^x e^{-\lambda z}}{x!} \\ &= e^{-\lambda} e^{\lambda z} \end{aligned} \tag{B.100}$$

Again, note that when we evaluate this at $z = 1$, we obtain 1 as expected. If we use this to evaluate the mean, we get

$$\left. \frac{d e^{-\lambda} e^{\lambda z}}{dz} \right|_{z=1} = \lambda e^{-\lambda} e^{\lambda z} \Big|_{z=1} = \lambda \tag{B.101}$$

which agrees with the mean of the Poisson distribution in (B.49). As before, we can use this to evaluate the variance by first computing the second derivative with respect to z and evaluating that at $z = 1$ as follows

$$\begin{aligned} \left. \frac{d^2 e^{-\lambda} e^{\lambda z}}{dz^2} \right|_{z=1} &= \left. \frac{d \lambda e^{-\lambda} e^{\lambda z}}{dz} \right|_{z=1} \\ &= \left. \lambda^2 e^{-\lambda} e^{\lambda z} \right|_{z=1} \\ &= \lambda^2 \\ &= E(N^2) - E(N) \end{aligned} \tag{B.102}$$

To compute the variance we now have

$$\begin{aligned} \text{Var}(N) &= E(N^2) - E^2(N) \\ &= \{E(N^2) - E(N)\} + E(N) - E^2(N) \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda \end{aligned} \tag{B.103}$$

which agrees with the variance of the Poisson distribution in (B.51).

Finally, probability generating functions are one way of solving an infinite series of equations. Consider a process (which we describe in chapter 3 on queueing) for which the probabilities are governed by the following infinite system of equations.

$$\rho P_n = P_{n+1} \quad n = 0, 1, \dots \quad (\text{B.104})$$

If we now multiply each side of this equation by z^n and sum over all values of n , we obtain

$$\sum_{n=0}^{\infty} \rho z^n P_n = \sum_{n=0}^{\infty} z^n P_{n+1} = \frac{1}{z} \sum_{n=0}^{\infty} z^{n+1} P_{n+1} = \frac{1}{z} \sum_{n=1}^{\infty} z^n P_n \quad (\text{B.105})$$

The left hand side of (B.105) is simply $\rho \text{PGF}(z)$ for some (unknown) distribution. The

right hand side is simply $\frac{1}{z} \sum_{n=1}^{\infty} z^n P_n = \frac{1}{z} \left[\sum_{n=0}^{\infty} z^n P_n - P_0 \right] = \frac{1}{z} [\text{PGF}(z) - P_0]$. Thus we

have

$$\begin{aligned} \rho \text{PGF}(z) &= \frac{1}{z} [\text{PGF}(z) - P_0] \\ \rho z \text{PGF}(z) &= \text{PGF}(z) - P_0 \\ \text{PGF}(z)[1 - \rho z] &= P_0 \end{aligned} \quad (\text{B.106})$$

Evaluating this at $z = 1$, we find

$$\begin{aligned} \text{PGF}(1)[1 - \rho] &= P_0 \\ 1 - \rho &= P_0 \end{aligned} \quad (\text{B.107})$$

So, substituting (B.107) back into (B.106), we have

$$\begin{aligned} \text{PGF}(z)[1 - \rho z] &= 1 - \rho \\ \text{PGF}(z) &= \frac{1 - \rho}{1 - \rho z} = (1 - \rho)(1 - \rho z)^{-1} \end{aligned} \quad (\text{B.108})$$

We can now evaluate the mean and variance of the probability mass function governed by (B.104). The mean is simply

$$E(N) = \left. \frac{d(1 - \rho)(1 - \rho z)^{-1}}{dz} \right|_{z=1} = \rho(1 - \rho)(1 - \rho z)^{-2} \Big|_{z=1} = \frac{\rho}{1 - \rho} \quad (\text{B.109})$$

Taking the second derivative with respect to z and evaluating that at $z = 1$, we obtain

$$\begin{aligned}
E(N^2) - E(N) &= \left. \frac{d^2(1-\rho)(1-\rho z)^{-1}}{dz^2} \right|_{z=1} \\
&= \left. \frac{d\rho(1-\rho)(1-\rho z)^{-2}}{dz} \right|_{z=1} \\
&= \frac{2\rho^2}{(1-\rho)^2}
\end{aligned} \tag{B.110}$$

And finally the variance is given by

$$\begin{aligned}
Var(N) &= E(N^2) - E^2(N) \\
&= \{E(N^2) - E(N)\} + E(N) - E^2(N) \\
&= \frac{2\rho^2}{(1-\rho)^2} + \frac{\rho}{1-\rho} - \left(\frac{\rho}{1-\rho}\right)^2 \\
&= \frac{2\rho^2}{(1-\rho)^2} + \frac{\rho(1-\rho)}{(1-\rho)^2} - \frac{\rho^2}{(1-\rho)^2} \\
&= \frac{\rho}{(1-\rho)^2}
\end{aligned} \tag{B.111}$$

Note that (B.109) agrees with the mean of the geometric distribution shown in (B.28) (after substituting $p = 1 - \rho$) and the variance agrees with the variance of the geometric distribution given in (B.29) after the same substitution. In fact, equations (B.104) together with the normalizing equation $\sum_{n=0}^{\infty} P_n = 1$ describe a geometric distribution. To

see this, observe that

$$\begin{aligned}
\left. \frac{d^n(1-\rho)(1-\rho z)^{-1}}{dz} \right|_{z=0} &= n! \rho^n (1-\rho) = n! P_n \\
P_n &= \rho^n (1-\rho)
\end{aligned} \tag{B.112}$$

where the second equality in the first row of (B.112) comes from property (2) of the probability generating functions outlined above. Note that the probability P_n in (B.112) agrees with the geometric probability terms in (B.27), again after substituting $p = 1 - \rho$ into (B.27).

B.9 Generating random variables

Almost every computer language, as well as Microsoft Excel and other spreadsheet programs, has the ability to generate standard uniform random variables; i.e., uniform random variables between 0 and 1. Using a uniform random number between 0 and 1, we can then generate a random variable from any distribution, provided we have either the probability mass function or the cumulative distribution. To see this, consider first a simple discrete distribution of a random variable that can take on only 3 values: 1, 2 or 3. For this random variable,

$$P(X = 1) = 0.2, P(X = 2) = 0.3, \text{ and } P(X = 3) = 0.5. \quad (\text{B.113})$$

The cumulative distribution function of X is shown in Figure B.15.

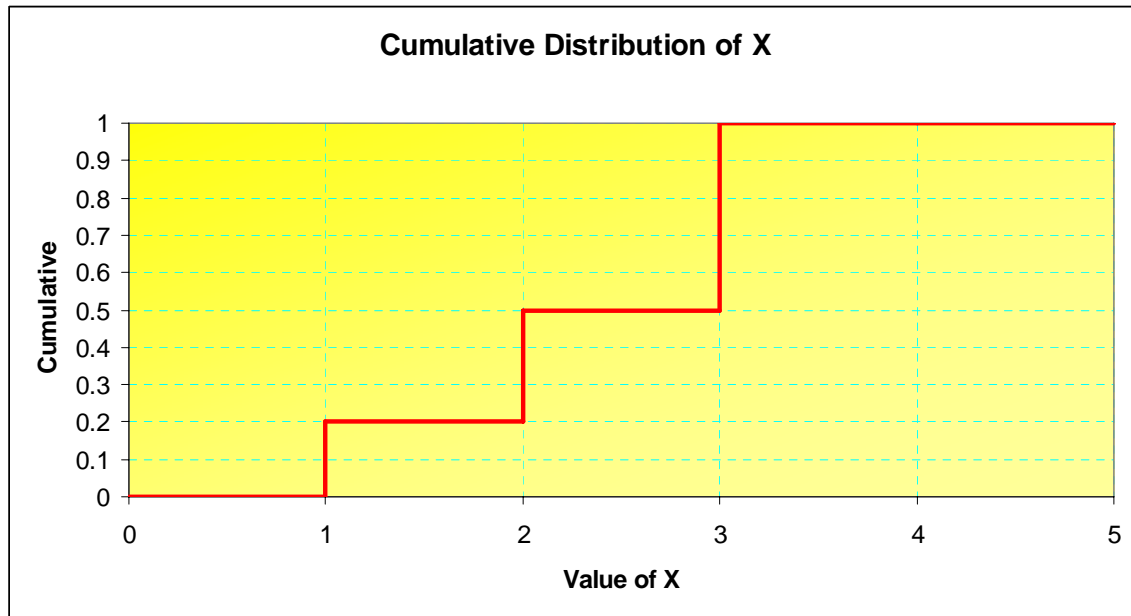


Figure B.15 -- Cumulative distribution of X

We can now generate realizations of the random variable X based on the uniform random variable, U , using the following formula:

$$X = \begin{cases} 1 & U \leq 0.2 \\ 2 & 0.2 < U \leq 0.5 \\ 3 & U > 0.5 \end{cases} \quad (\text{B.114})$$

Note that determining the value of X can be thought of as starting on the vertical axis (the cumulative distribution axis) with the value of U , the uniform random variable, and moving across until we hit the cumulative distribution function. We then read down to the X axis and return the corresponding value. This is shown in Figure B.16 for a value of $U=0.43$.

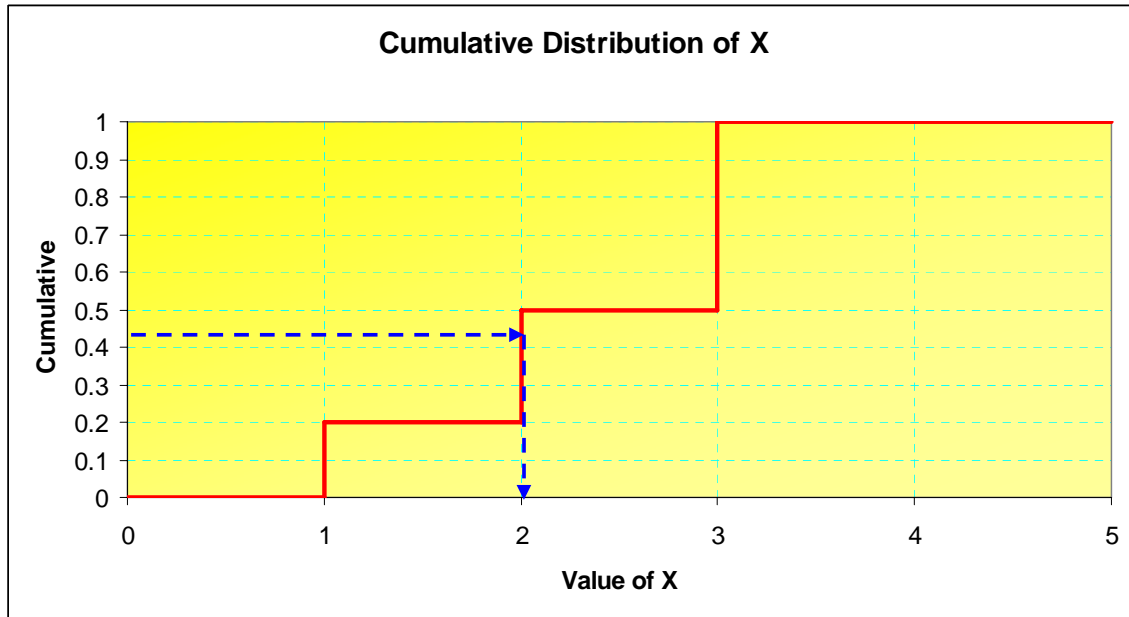


Figure B.16 -- Finding the value of a random variable X from a draw of a uniform random variable using the cumulative distribution of X

To generate such a random variable in EXCEL, all we need is the probability mass function and the LOOKUP function in EXCEL. From the probability mass function, we compute the $P(X < x)$ for every value of x in the distribution. Note that we want the probability that the random variable takes on a value strictly less than x . Table B.8 shows the key information for the distribution above.

Table B.8 -- EXCEL information for the same distribution

	C	D	E
78	x	Prob	P(X<x)
79	1	0.2	0
80	2	0.3	0.2
81	3	0.5	0.5

To compute a random variable from this distribution, we draw a random variable from the standard uniform distribution using the RAND() function in EXCEL. We then use this along with the LOOKUP function to return a value drawn from this distribution. Specifically, we use:

$$\text{LOOKUP}(\text{RAND}(),\text{P}(X<x),x) \quad (\text{B.115})$$

where $\text{P}(X<x)$ points to cells E79:E81 in the third column in Table B.8 and the third parameter of the LOOKUP function, x , points to cells C79:C81 in the first column. Table B.9a shows the results of using this function 10,000 times and tabulating the number of times the function returned, 1, 2, and 3, respectively. We would expect about 1 to appear about 2,000 times, 2 to appear about 3,000 times and 3 to appear about 5,000 times out of the 10,000 times the function in (B.115) was exercised. Table B.9b shows the results of repeating this exercise 10 times and totaling the number of times each value appears. Note how closely the results of both tests conform with our expectations.

Table B.9 -- Using the LOOKUP function to sample from the distribution given in**Table B.8**

Single run	
Value	Results
1	1979
2	3012
3	5009
Total	10000

(a) single run of 10,000 draws

10 runs	
Value	Results
1	20003
2	30097
3	49900
Total	100000

(b) 10 runs of 10,000 draws each

The same technique can be used to sample from other distributions. For example, suppose we want to sample from a Poisson distribution with a mean of $\lambda = 10$. Again, we would set up a table with the values of x (e.g., from 0 to 25). We would then compute the probability that the random variable is strictly less than x . This is then used in a LOOKUP function. Using the LOOKUP function 10,000 times and plotting the frequency of each value (shown in blue) against the true probabilities, we obtain a figure similar to Figure B.17. Note that both the sampled mean (10.004) and the sampled variance (9.940) are remarkably close to the true values of 10. Also note how closely the true and sampled distributions resemble each other.

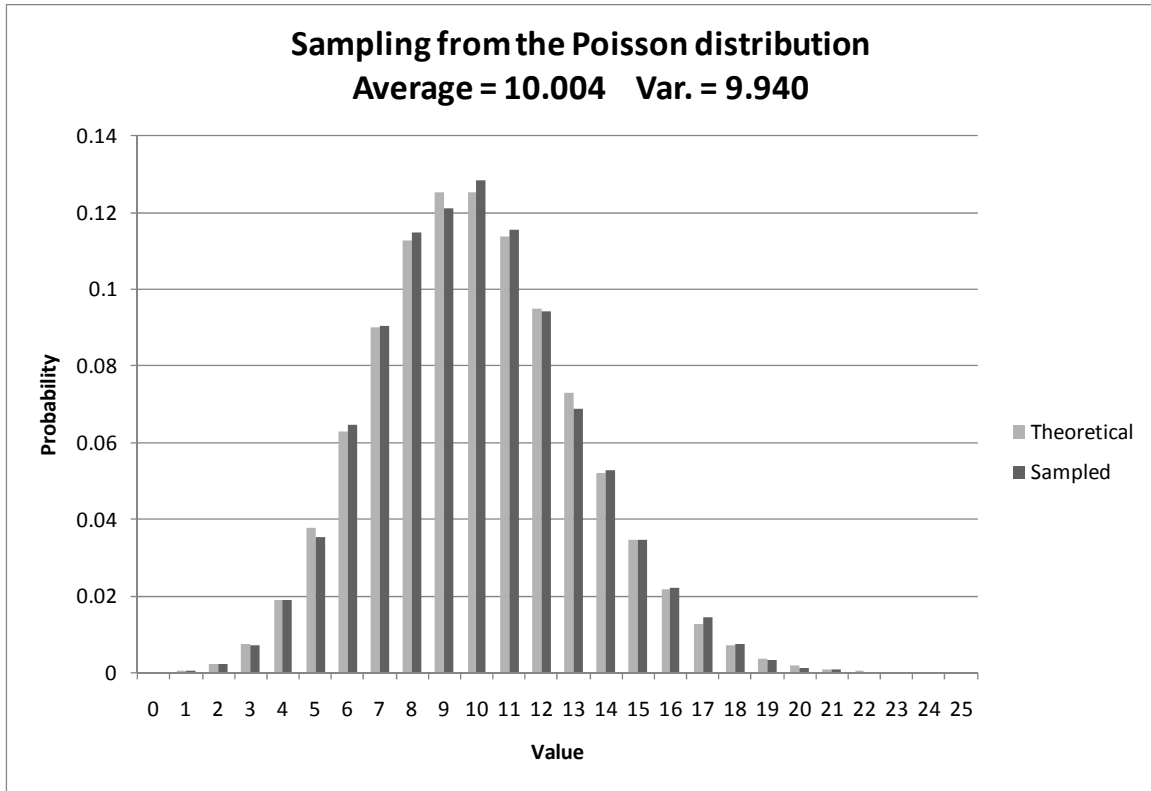


Figure B.17 -- True versus sampled distributions for a Poisson random variable with a true mean of 10

When the cumulative distribution is known in closed form and can readily be inverted, we can use a similar technique. Consider, for example, the exponential distribution, for which $F(x) = 1 - e^{-\mu x}$ for $x \geq 0$. The technique shown in Figure B.16 corresponds to setting a standard uniform random variable, U , equal to this cumulative distribution and then solving for x as shown below

$$\begin{aligned}
 U &= 1 - e^{-\mu x} \\
 e^{-\mu x} &= 1 - U \\
 -\mu x &= \ln(1 - U) \\
 x &= \frac{-\ln(1 - U)}{\mu}
 \end{aligned}
 \tag{B.116}$$

Since $1 - U$ is also uniformly distributed between 0 and 1, we can actually simply use $x = \frac{-\ln(U)}{\mu}$ and avoid one subtraction. Figure B.18 shows the results of sampling from the exponential distribution in this manner 10,000 times with $\mu = 0.1$. Recall that for the exponential distribution we have $E(X) = 1/\mu$ and $Var(X) = 1/\mu^2$, so we would expect the average to be 10 and the variance to be 100. Again, note how closely the sampled values are to the theoretical values and how closely the sampled distribution (dark grey) resembles the expected or true distribution (shown in light grey).

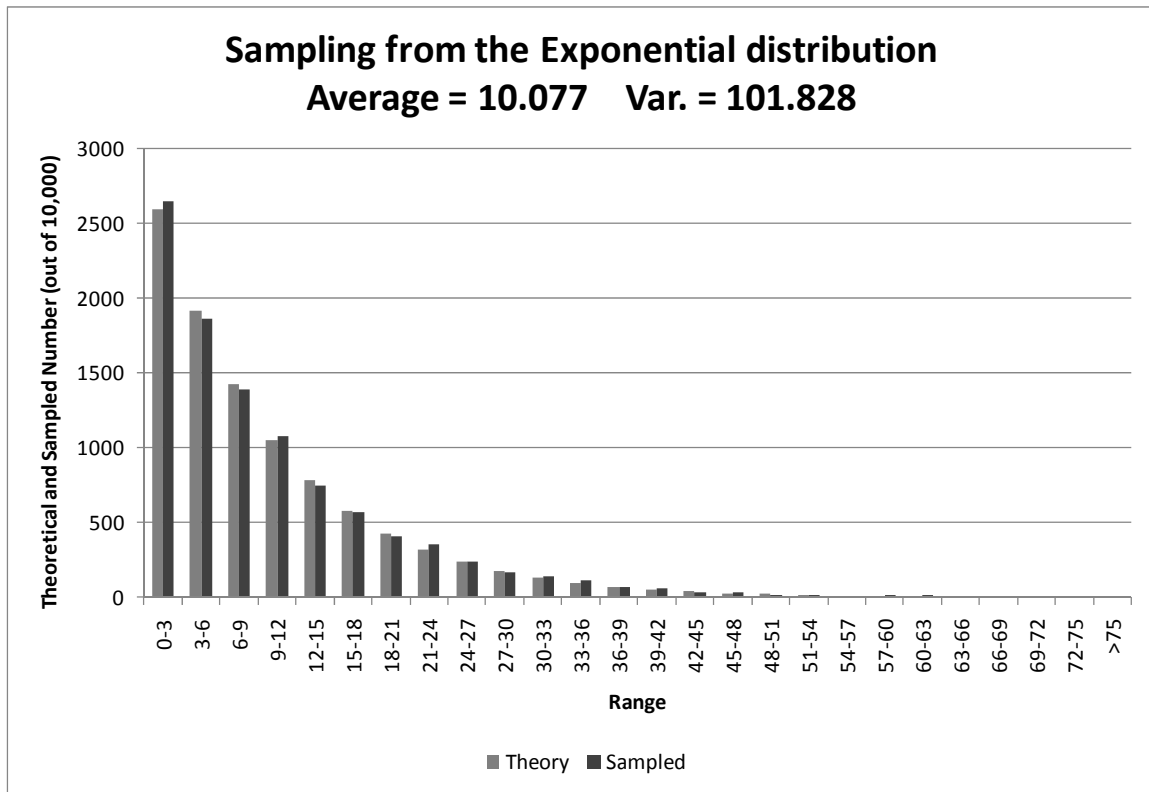


Figure B.18 -- Sampling from the exponential distribution

For some distributions, the cumulative distribution is not known in closed form. This is true for the normal distribution. In this case, however, we can obtain two

samples, X and Y , from the standard normal distribution from two uniform random variables, U and V , using the following equations:

$$\begin{aligned} X &= \sqrt{-2\ln U} \cdot \cos(2\pi V) \\ Y &= \sqrt{-2\ln U} \cdot \sin(2\pi V) \end{aligned} \tag{B.117}$$

Figure B.19 shows the results of sampling 10,000 times from the normal distribution using equations (B.117). As before, note how closely both the average and variance are to their theoretical values of 0 and 1 respectively and how closely the sampled values are to the theoretical values.

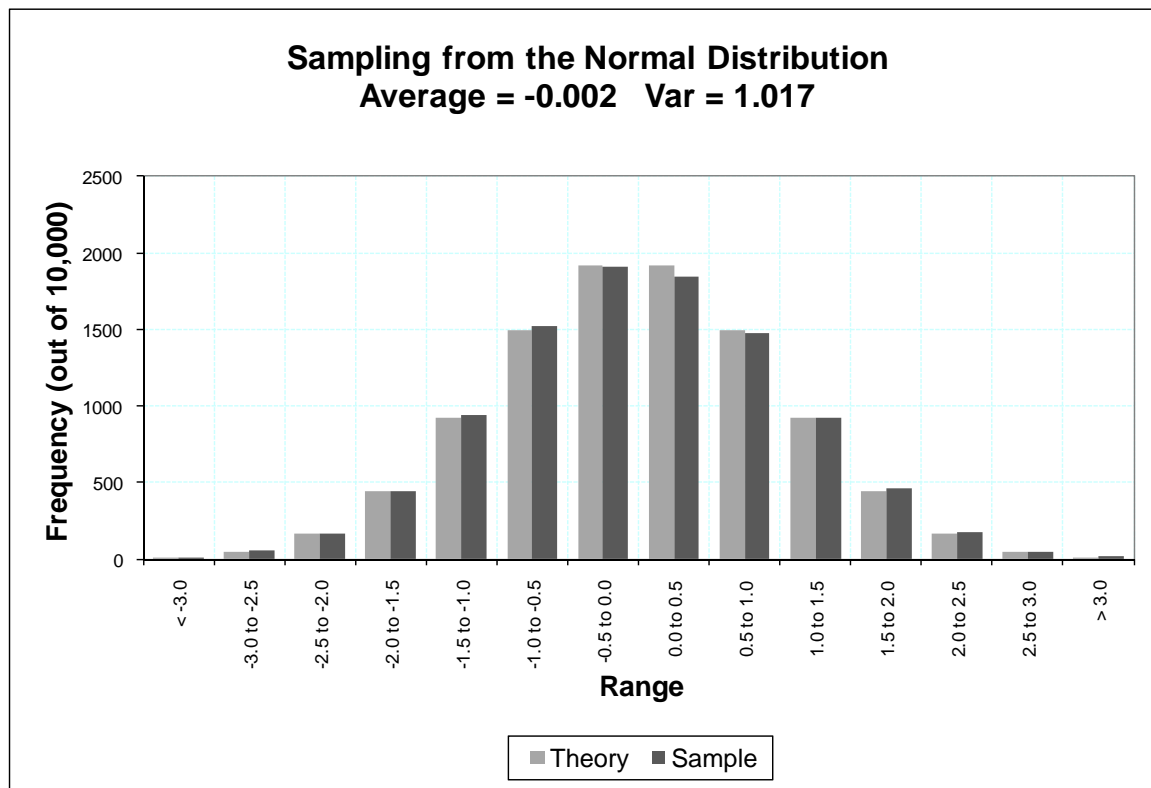


Figure B.19 -- Sampling from the normal distribution

B.10 Random variables in Excel

Microsoft EXCEL incorporates a number of probability functions. Some are shown in Table B.10 below which gives (where appropriate) the form of the function for both the probability mass function (for discrete random variables) or the density function (for continuous random variables) and the cumulative distribution.

Table B.10 -- Some probability functions in EXCEL

Distribution	Functional Form	Probability Mass or Density	Cumulative Distribution
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	BinomDist(x,n,p,false)	BinomDist(x,n,p,true)
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	Poisson(x,λ,false)	Poisson(x,λ,true)
Hypergeometric	$\frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$	HypGeom(x,n,M,N)	
Negative Binomial	$\binom{x+r-1}{r-1} p^r (1-p)^x$	NegBinom(x,r,p)	
Exponential	$\lambda e^{-\lambda x}$	ExponDist(x,λ,false)	ExponDist(x,λ,true)
Erlang-k	$\frac{\lambda (\lambda s)^{k-1} e^{-\lambda s}}{(k-1)!}$	GammaDist(s,k,1/λ,false)	GammaDist(s,k,1/λ,true)
Normal	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$	NormDist(x,μ,σ,false)	NormDist(x,μ,σ,true)
Standard Continuous Uniform	$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$	Rand()	

In addition to these functions, a few other EXCEL functions are of use to us. To compute the number of combinations of k items taken out of n , we can use COMBIN(n,k). To compute the number of permutations of k items selected from n

items, we can use PERMUT(n,k). We can use CRITBINOM(n,p, α) to find the smallest value of x , such that $\sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k} \geq \alpha$.

If we have the cumulative probability, p , associated with an event whose outcomes are normally distributed with mean μ and standard deviation σ , we can use

NORMINV(p, μ, σ) to find the value of x such that. $\int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} dy = p$ If

the normal distribution in question is a standard normal distribution for which $\mu = 0$ and $\sigma = 1$, we can simply use NORMSINV(p).

C. References

- Angéniol, B., G. de La Croix Vaubois and J.-Y. Le Texier, 1988, "Self-Organizing Feature Maps and the Travelling Salesman Problem," *Neural Networks*, **1**, 289-293.
- Ayers, I., 2007, Super Crunchers: Why Thinking by Numbers Is the New Way to Be Smart, Bantam Books, New York.
- Bailey, N. T. J. and J. D. Welch, 1952, "Appointment systems in hospital outpatient departments," *The Lancet*, **259**, 1105-1108.
- Bailyn, L., R. Collins, and Y. Song, 2007, "Self-scheduling for hospital nurses: an attempt and its difficulties," *Journal of Nursing Management*, **15**, 72-77.
- Baker, B. M. and M. A. Ayechev, 2003, "A genetic algorithm for the vehicle routing problem," *Computers and Operations Research*, **30**, 787-800.
- Baker, S. 2008, The Numerati, Houghton Mifflin Co., Boston, MA.
- Banks, J., 1999, "Introduction to simulation," *Proceedings of the 1999 Winter Simulation Conference*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock and G. W. Evans, eds., 7-13.
- Banks, J., J. S. Carson, B. L. Nelson and D. Nicol, 2005, Discrete Event System Simulation. Fourth edition., Prentice Hall, Inc., Upper Saddle River, NJ.
- Bard, J. F. and H. W. Purnomo, 2005a, "Hospital-wide reactive scheduling of nurses with preference considerations," *IIE Transactions*, **37**:7, 589-608.
- Bard, J. F., 1998, Practical Bilevel Optimization: Algorithms and Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Bard, J. F. and H. W. Purnomo, 2005b, "Preference scheduling for nurses using column generation," *European Journal of Operational Research*, **164**:2, 510-534.
- Bard, J. F. and H. W. Purnomo, 2007, "Cyclic preference scheduling of nurses using a Lagrangian-based heuristic," *Journal of Scheduling*, **10**:1, 5-23.
- Bartholdi, J. J. and L. K. Platzman, 1982, "An $O(N \log N)$ Planar Traveling Salesman Heuristic Based on Spacefilling Curves," *Operations Research Letters*, **1**:4, 121-125.
- Bartholdi, J. J. and L. K. Platzman, 1988, "Heuristics Based on Spacefilling Curves for Combinatorial Problems in Euclidean Space," *Management Science*, **34**:3, 291-305

- Bartholdi, J. J., L. K. Platzman, R. L. Collins, and W. H. Warden, 1983, "A Minimal Technology Routing System for Meals on Wheels", *Interfaces*, **3**, June, 1-8.
- Benedict, J. M., 1983, "Three Hierarchical Objective Models Which Incorporate the Concept of Excess Coverage to Locate EMS Vehicles or Hospitals." (M. S. Daskin, advisor), M.S. Thesis, Department of Civil Engineering, Northwestern University, Evanston, IL.
- Bertsimas D. and A. Thiele, 2006, "Robust and data-driven optimization: modern decision-making under uncertainty", in Tutorials on Operations Research, M. Johnson, B. Norman and N. Secomandi, eds., Ch. 4, pp. 95-122, INFORMS, Hanover, MD.
- Birge, J. R. and F. Louveaux, 1999, Introduction to Stochastic Programming, Springer Series in Operations Research, New York.
- Bodin, L. and A. Panken, 2003, "High tech for a higher authority: The placement of graduating rabbis from Hebrew Union College -- Jewish Institute of Religion," *Interfaces*, **33**:3, 1-11.
- Box, G. E. P., and N. R. Draper, 1987, Empirical Model-Building and Response Surfaces, Wiley, New York, NY.
- Bullock, C. S. III., 1982, "The Inexact Science of Congressional Redistricting.", *American Political Science Association*, **15**:Summer, 431-438. Burke, P. J., 1956, "The Output of a Queueing System," *Operations Research*, **4**:6, 699-704.
- Burke, E. K., P. De Causmaecker, S. Petrovic, and G. Vanden Berghe, 2002, "A multi criteria metaheuristic approach to Nurse rostering," *Proceedings of Congress on Evolutionary Computation*, CEC2002, Honolulu, IEEE Press, 1197-1202.
- Burke, E. K., P. De Causmaecker, G. Vanden Berghe, and H. Van Landeghem, 2004, "The State of the Art of Nurse Rostering," *Journal of Scheduling*, **7**:6, 441-99.
- Carter, M. W., S. D. Lapierre, 2001, "Scheduling Emergency Room Physicians," *Health Care Management Science*, **4**: 347-60.
- Cheang, B., H. Li, A. Lim and B. Rodrigues, 2003, "Nurse rostering problems – a bibliographic survey," *European Journal of Operational Research*, **151**, 447-460.
- Chiang, W.-C. and R. A. Russell, 1996, "Simulated annealing metaheuristics for the vehicle routing problem with time windows," *Annals of Operations Research*, **63**, 3-27.
- Chopra, S. and P. Meindl, 2004, Supply Chain Management: Strategy, Planning and Operation, Prentice Hall.

- Christofides, N., 1976, "Worst-case analysis of a new heuristic for the traveling salesman problem," Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA.
- Church, R.L., and M. Meadows, 1979, "Location Modeling Utilizing Maximal Service Distance Criteria." *Geographical Analysis*, **11**, 358-379.
- Church, R.L., and C. ReVelle, 1974, "The Maximal Covering Location Problem." *Papers of the Regional Science Association*, **32**, 101-118. Clarke, A. B. and R. L. Disney, 1970, Probability and random processes for engineers and scientists, Wiley, New York.
- Church, R. L., M. P. Scaparra, and R. S. Middleton, 2004, "Identifying Critical Infrastructure: The Median and Covering Facility Interdiction Problems," *Annals of the Association of American Geographers*, **94**:3, 491-502.
- Clarke, G. and J. W. Wright, 1964, "Scheduling of vehicles from a central depot to a number of delivery points," *Operations Research*, **12**, 568-581.
- Cohon, J. L., 1978, Multiobjective Programming and Planning, Academic Press, New York, NY.
- Current, J., M. S. Daskin, and D. Schilling, 2002, "Discrete Network Location Models," in Facility Location Theory: Applications and Methods, edited by Z. Drezner and H. Hamacher, 81-118, Chapter 3, Springer-Verlag, Berlin.
- Daganzo, C. F., 1991, Logistics Systems Analysis, Lecture Notes in Economics and Mathematical Sciences, Vol. 361, edited by M. Beckmann and W. Krelle, Springer-Verlag, Berlin.
- Daganzo, C. F., 2003, A Theory of Supply Chains, Vol. 526, Lecture Notes in Economics and Mathematical Sciences, Vol 526, edited by W. Trockel G. Fandel, Springer-Verlag, Berlin.
- Dantzig, G. B., 1998, Linear Programming and Extensions, Princeton University Press, Princeton, NJ.
- Daskin, M. S., 1982, "Application of an Expected Covering Model to Ems System Design," *Decision Sciences*, **13**:3, 416-439.
- Daskin, M. S., 1983, "A Maximum Expected Covering Location Model: Formulation, Properties and Heuristic Solution," *Transportation Science*, **17**, 48-70.
- Daskin, M. S., C. Coullard, and Z.-J. M. Shen, 2002, "An Inventory-Location Model: Formulation, Solution Algorithm and Computational Results," *Annals of Operations Research*, **110**, 83-106.

- Daskin, M. S., K. Hogan, and C. ReVelle, 1988, "Integration of Multiple, Excess, Backup and Expected Covering Models," *Environment and Planning B*, **15**, 15-35.
- Daskin, M. S., and E. H. Stern, 1981, "A Hierarchical Objective Set Covering Model for Emergency Medical Service Deployment," *Transportation Science*, **15**:2, 137-152.
- Davenport, T. H., and J. G. Harris, 2007, Competing on Analytics: The New Science of Winning, Harvard Business Press, Boston, MA.
- Davis, A., 2009, "The Nurse Shortage Crisis," IE 490 Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.
- Davis, M. M., and J. Heineke, 2003, Managing Services: Using Technology to Create Value, McGraw Hill/Irwin, Boston, MA.
- De Grano, M. L., D. J. Medeiros, D. J., and D. Eitel, 2009, "Accommodating individual preferences in nurse scheduling via auctions and optimization," *Health Care Management Science*, **12**:3, 228-242.
- Dekle, J., M. S. Laverie, E. Martin, H. Emir-Farinas, and R. L. Francis, 2005, "A Florida County Locates Disaster Recovery Centers" *Interfaces*, **35**:2, 133-139.
- de la Merced, M. J., 2008, "Starbucks Announces It Will Close 600 Stores," The New York Times, July 2, 2008, online at http://www.nytimes.com/2008/07/02/business/02sbux.html?_r=1&scp=1&sq=starbucks%20600%20stores&st=cse.
- Denton, B. T., A. J. Miller, H. J. Balasubramanian, T. R. Huschka, 2010, "Optimal Allocation of Surgery Blocks to Operating Rooms Under Uncertainty," *Operations Research*, forthcoming.
- Denton, B. T., A. S. Rahman, H. Nelson, and A. C. Bailey, 2006, "Simulation of a Multiple Operating Room Surgical Suite," *Proceedings of the 2006 Winter Simulation Conference*, L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.), 414-424.
- Denton, B., J. Viapiano, and A. Vogl, 2007, "Optimization of surgery sequencing and scheduling decisions under uncertainty," *Health Care Management Science*, **10**, 13-24.
- Devore, J. L., 2008, Probability and Statistics for Engineering and the Sciences, Seventh edition, Thomson, Australia.
- Doerner, K., R. F. Hartl, V. Maniezzo, and M. Reimann, 2003, "An Ant System Metaheuristic for the Capacitated Arc Routing Problem," *The Fifth Metaheuristics International Conference*, MIC2003, Kyoto, Japan, 16-1 to 16-6

- Drezner, Z., K. Klamroth, A. Schöbel, and G. O. Wesolowsky, 2002, "The Weber Problem." in Facility Location Theory: Applications and Methods, edited by Z. Drezner and H. Hamacher, 1-36, Chapter 1, Springer-Verlag, Berlin.
- Eaton, D. J. et al, 1979, Location Techniques for Emergency Medical Service Vehicles: Volume I – An Analytical Framework for Austin, Texas, Lyndon B. Johnson School of Public Affairs, The University of Texas at Austin, Austin, TX.
- Eaton, D., M. S. Daskin, D. Simmons, B. Bulloch, B., and G. Jansma, 1985, "Determining Emergency Medical Service Vehicle Deployment in Austin, Texas," *Interfaces*, **15**:1, 96-108.
- Erdogan, S. A. and B. T. Denton, 2009, "Surgery Planning and Scheduling: A Literature Review," working paper, Edward Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC.
- Erkut, E., and B. Bozkaya, 1999, "Analysis of Aggregation Errors for the P-Median Problem," *Computers and Operations Research*, **26**, 1075-1096.
- Erkut, E. and A. Ingolfsson, 2000, "Let's Put the Squares in Least-Squares." INFORMS Transactions on Education, **1**:1, 47-50, <http://archive.itejournal.informs.org/Vol1No1/ErkutIngolfsson/ErkutIngolfsson.pdf>.
- Erkut, E., and S. Neuman, 1989, "Analytical Models for Locating Undesirable Facilities," *European Journal of Operational Research*, **40**, 275-291.
- Fitzpatrick, J. M., A. E. While, A. E., and J. D. Roberts, 1999, "Shift work and its impact upon nurse performance: current knowledge and research issues," *Journal of Advanced Nursing*, **29**:1, 18-27.
- Fitzsimmons, J. A., 1973, "A Methodology for Emergency Ambulance Deployment." *Management Science*, **19**:6, 627-36.
- Fitzsimmons, J. A., and M. J. Fitzsimmons, 2008, Service Management: Operations, Strategy, Information Technology, Sixth edition, McGraw Hill/Irwin, Boston, MA.
- Fourer, R., D. M. Gay and B. W. Kernighan, 2002, AMPL: A Modeling Language for Mathematical Programming, Duxbery Press/Brooks/Cole Publishing Company, South San Francisco, CA.
- Francis, P. M., 2006, "Period Vehicle Routing with Service Choice: Models and Algorithms," PhD dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.
- Francis, R. L., T. J. Lowe, and A. Tamir, 2002, "Demand Point Aggregation for Location Models." in Facility Location Theory: Applications and Methods, edited by Z. Drezner and H. Hamacher, 209-234, Chapter 7, Springer-Verlag, Berlin.

- Francis, P. and K. Smilowitz, 2006, "Modeling techniques for periodic vehicle routing problems," *Transportation Research Part B: Methodological*, **40**, 872-884.
- Francis, P., K. Smilowitz and M. Tzur, 2006, "The Period Vehicle Routing Problem with Service Choice," *Transportation Science*, **40**:4, 439-454.
- Francis P., K. Smilowitz and M. Tzur, 2007, "Flexibility and Complexity in Periodic Distribution Problems," *Naval Research Logistics*, **54**:2, 136-150.
- Friedman, T. L., 2005, The World is Flat: A Brief History of the Twenty-first Century, Farar, Straus and Giroux, New York, NY:.
- Fu, Z., R. Eglese and L.Y.O. Li, 2005, "A new tabu search heuristic for the open vehicle routing problem," *Journal of the Operational Research Society*, **56**, 267-274.
- Garey, M. R., and D. S. Johnson, 1979, Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman and Co., New York.
- Garnett, O., A. Mandelbaum, and M. Reiman, 2002, "Designing a Call Center with Impatient Customers," *Manufacturing & Service Operations Management*, **4**:3, 208-227.
- Gavirneni, S., L. Clark and G. Pataki, 2004, "Schlumberger Optimizes Receiver Location for Automated Meter Reading." *Interfaces*, **34**:3, 208-14.
- Gendreau, M., A. Hertz and G. Laporte, 1994, "A Tabu Search Heuristic for the Vehicle Routing Problem," *Management Science*, **40**:10, 1276-1290.
- Gendreau, M., G. Laporte and R. Séguin, 1996, "Tabu search heuristic for the vehicle routing problem with stochastic demands and customers," *Operations Research*, **44**:3, 469-477.
- Gillett, B., and Miller, L., 1974, "A heuristic algorithm for the vehicle dispatch problem," *Operations Research*, **22**, 340-349.
- Glover, F., 1989, "Tabu Search — Part I," *ORSA Journal on Computing*, 1:3, pp. 190-206.
- Glover, F., 1990a, "Tabu Search — Part II," *ORSA Journal on Computing*, 2:1, pp. 4-32.
- Glover, F. 1990b, "Tabu Search: A Tutorial," *Interfaces*, **20**:4, 74-94.
- Glover, F. and M. Laguna, 1997, *Tabu Search*, Kluwer Academic Publishers, Boston, MA.
- Gold, D. R., S. Rogacz, N. Bock, T.D. Tosteson, T. M. Baum, F. E. Speizer, F. E., and C. A. Czeisler, 1992, "Rotating Shift Work, Sleep, and Accidents Related to

- Sleepiness in Hospital Nurses," *American Journal of Public Health*, **32**:7, 1011-1015.
- Goldberg, D., (1989). Genetic Algorithms in Search Optimization and Machin Learning, Addison-Wesley, Reading, MA.
- Golden, B. L. and A. A. Assad, (eds.), 1988, Vehicle Routing: Methods and Studies, North-Holland, Amsterdam.
- Goldman, A. J., 1971, "Optimal Center Location on Simple Networks," *Transportation Science*, **5**, 212-21.
- Green, L. V., S. Savin, and B. Wang, 2006, "Managing Patient Service in a Diagnostic Medical Facility," *Operations Research*, **54**:1 11-25.
- Gross, D. and C. Harris, 1985, Fundamentals of Queueing Theory, Second Edition, John Wiley & Sons, New York.
- Gupta, D. and B. Denton, 2008, "Appointment scheduling in health care: Challenges and opportunities," *IIE Transactions*, **40**, 800-819.
- Hakimi, S.L., 1964, "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," *Operations Research*, **12**:3, 450-59.
- Hakimi, S. L., 1965, "Optimum Location of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems," *Operations Research*, **13**:3, 462-475.
- Halfin, S. and W. Whitt, 1981, "Heavy-traffic limits for queues with many exponential servers," *Operations Research*, **29**:3, 567-588.
- Hansen, P. and N. Mladenovic, 2001, "Variable neighborhood search: Principles and applications," *European Journal of Operational Research*, **130**, 449-467.
- Heizer, J. and B. Render, 2001, Principles of Operations Management, Prentice Hall, Upper Saddle River, NJ.
- Hillsman, E. L., and R. Rhoda, 1978, "Errors in Measuring Distances from Populations to Service Centers" *Annals of Regional Science*, **12**, 74-88.
- Ho, C. J., and H. S. Lau, 1992, "Minimizing Total-Cost in Scheduling Outpatient Appointments," *Management Science*, **38**:12, 1750-64.
- Hopp, W. J. and M. L. Spearman, 2008, Factory Physics, Third edition, Irwin/McGraw Hill, Boston MA.
- Hopp, W. J. and M. P. van Oyen, 2004, "Agile workforce evaluation: a framework for cross-training and coordination," *IIE Transactions*, **36**, 919-940.

- Huschka, T. R., B. T. Denton, S. Gul and J. W. Fowler, 2007, "Bi-criteria evaluation of an outpatient procedure center via simulation," *Proceedings of the 2007 Winter Simulation Conference*, S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton (eds.), 1510-1518.
- Iser, J. H., B. T. Denton, and R. E. King, 2008, "Heuristics for balancing operating room and post-anesthesia resources under uncertainty," *Proceedings of the 2008 Winter Simulation Conference*, S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler (eds.), 1601-1608.
- Jessop, B., 2008, "Applying Optimization Techniques to Long-Term Workforce Planning," Undergraduate Honors Research Project, Winter, Mark S. Daskin, Advisor, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208.
- Johnson, M. P., and Smilowitz, K. , 2007, "Community-Based Operations Research," in *TutORials in Operations Research*, INFORMS, Institute for Operations Research and the Management Sciences, chapter 6, 102-123.
- Kaandorp, G. C. and G. Koole, 2007, "Optimal outpatient appointment scheduling," *Health Care Management Science*, **10**: 217-229.
- Kane, R. L., T. Shamliyan, C. Mueller, S. Duval, S., and T. J. Wilt, T. J., 2007, "Nurse Staffing and Quality of Patient Care," Evidence Report/Technology Assessment, Number 151, Minnesota Evidence-based Practice Center, Minneapolis, Minnesota.
- Koole, G. and A. Mandelbaum, 2002, "Queueing Models of Call Centers: An Introduction," *Annals of Operations Research*, **113**, 41-59.
- Kruskal, J., 1956, "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem," *Proceedings of the American Mathematical Society*, **17**:1, 48-50.
- Kuby, M., 1987, "Programming Models for Facility Dispersion: The P-Dispersion and Maxisum Dispersion Problems," *Geographical Analysis*, **19**, 315-329.
- Knutsson, A., 2003, "Health disorders of shift workers," *Occupational Medicine*, **53**, 103-108.
- Laporte, G., 1992a, "The Traveling Salesman Problem: An overview of exact and approximate algorithms," *European Journal of Operational Research*, **59**, 231-247.
- Laporte, G., 1992b, "The Vehicle Routing Problem: An overview of exact and approximate algorithms," *European Journal of Operational Research*, **59**, 345-358.

- Laporte, G., 2007, "What You Should Know about the Vehicle Routing Problem," *Naval Research Logistics*, **59**:8, 811-819.
- Laporte, G., Y Nobert, and S. Taillefer, 1988, "Solving a Family of Multi-Depot Vehicle Routing and Location-Routing Problems," *Transportation Science*, **22**:3, 161-172.
- Larson, R. C., 1987, "Perspectives on Queues: Social Justice and the Psychology of Queueing," *Operations Research*, **35**:6, 895-905.
- Lawler, E. L., J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, 1985, *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, John Wiley & Sons, New York.
- Ledyard, J. O., M. Olson, D. Porter, J. A. Swanson and D. P. Torma, 2002, "The First Use of a Combined-Value Auction for Transportation Services," *Interfaces*, **32**:5, 4-12.
- Lee, E. K., and M. Zaider, 2003, "Mixed Integer Programming Approaches to Treatment Planning for Brachytherapy -- Application to Permanent Prostate Implants," *Annals of Operations Research*, **119**, 147-163.
- Lee, E. K., and M. Zaider, 2008, "Operations Research Advances Cancer Therapies," *Interfaces*, **38**:1, 5-25.
- Lien, R., Iravani, S., and K. Smilowitz, 2007, Sequential allocation problems for nonprofit agencies, Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208.
- Lin, S. and B. W. Kernighan, 1973, "An effective heuristic algorithm for the traveling salesman problem." *Operations Research*, **21**, 498-516.
- Malandraki, C. and M. S. Daskin, 1993, "The Maximum Benefit Chinese Postman Problem and the Maximum Benefit Traveling Salesman Problem," *European Journal of Operations Research*, **65**, pp. 218-234.
- Marsh, M. T. and D. A. Schilling, 1994, "Equity measurement in facility location analysis: A review and framework," *European Journal of Operational Research*, **74**, 1-17.
- Meuleau, N. and M Dorigo, 2002, "Ant Colony Optimization and Stochastic Gradient Descent," *Artificial Life*, **8**, 103-121.
- Mehrotra, A., E. J. Johnson, and G. L. Nemhauser, 1998, "An Optimization Based Heuristic for Political Districting," *Management Science*, **44**:8, 1100-1114. Metters, R., K. King-Metters, and M. Pullman, 2003, Successful Service Operations Management, Thomson/South-Western, Mason, OH.

- Min, H., V. Jayaraman and R. Srivastava, 1998, "Combined location-routing problems: A synthesis and future research directions," *European Journal of Operational Research*, **108**, 1-15.
- Muthuraman, K. and M. Lawley, 2008, "A stochastic overbooking model for outpatient clinical scheduling with no-shows," *IIE Transactions*, **40**:9, 820-837.
- Nahmias, S., 2009, Production and Operations Analysis, Sixth edition, Irwin/McGraw Hill, Boston MA.
- National Fire Protection Association, 2001, NFPA 1710: Standard for the Organization and Deployment of Fire Suppression Operations, Emergency Medical Operations and Special Operations to the Public, Quincy, MA 02269.
- National Resident Matching Program, 2009, *Results and Data: 2009 Main Residency Match*, National Resident Matching Program, Washington, DC.
- Nelson, B. L., 1995, Stochastic Modeling: Analysis and Simulation, McGraw-Hill, Inc., New York.
- Nemhauser, G. L. and L. A. Wolsey. 1988. Integer and Combinatorial Optimization, John Wiley & Sons, New York.
- Newell, G.F., 1971, Applications of Queueing Theory. Edited by M. S. Bartlett, Monographs on Applied Probability and Statistics, Chapman and Hall, London, England.
- Orwell, G., 1945, Animal Farm, original title – Animal Farm: A Fairy Story), Secker and Warburg, London.
- Ozsen, L., C. Coullard, and M. S. Daskin, 2008, "Capacitated Facility Location Model with Risk Pooling," *Naval Research Logistics*, **55**:4, 295-312.
- Ozsen, L., C. Coullard, and M. S. Daskin, 2009, "Facility Location Modeling and Inventory Management with Multi-Sourcing," *Transportation Science*, **43**:4, 455-472.
- Perl, J. and M. S. Daskin, 1985, "A Warehouse Location-Routing Model," *Transportation Research*, **19B**, 381-396.
- Prim, R. C., 1957, "Shortest connection networks and some generalizations," *Bell System Technical Journal*, **36**, 1389–1401.
- Purnomo, H. W., and J. F. Bard, 2007, "Cyclic Preference Scheduling for Nurses Using Branch and Price." *Naval Research Logistics*, **54**:2, 200-220.

- Reimann, M., K. Doerner, and R. F. Hartl, 2004, "D-Ants: Savings Based Ants divide and conquer the vehicle routing problem," *Computers and Operations Research*, **31**, 563-591.
- Render, B., R. M. Stair, Jr., M. E. Hanna, 2003, Quantitative Analysis for Management, Eighth edition, Prentice Hall, Upper Saddle River, NJ.
- Rogers, A. E., W. T. Hwang, L. D. Scott, L. H. Aiken, and D. F. Dinges, 2004, "The working hours of hospital staff nurses and patient safety," *Health Affairs*, **23**:4, 202-212.
- Rubin, D. S., and H. M. Wagner, 1990, "Shadow Prices: Tips and Traps for Managers and Instructors," *Interfaces* **20**:4, 150-57.
- Santibáñez, P., G. Bekiou and K. Yip, 2009, "Fraser Health Uses Mathematical Programming to Plan Its Outpatient Hospital Network," *Interfaces*, **39**:3, 196-208.
- Scaparra, M. P. and R. L. Church, 2008, "A bilevel mixed-integer program for critical infrastructure protection planning," *Computers and Operations Research*, **35**, 1905-1923.
- Schernhammer, E. S., F. Laden, F. E. Speizer, W. C. Willett, D. J. Hunter, I. Kawachi, I., C. S. Fuchs and G. A. Colditz, 2003, "Night-Shift Work and Risk of Colorectal Cancer in the Nurses' Health Study," *Journal of the National Cancer Institute*, **95**:11, 825-828.
- Schernhammer, E. S., F. Laden, F. E. Speizer, W. C. Willett, D. J. Hunter, I. Kawachi, and G. A. Colditz, 2001, "Rotating Night Shifts and Risk of Breast Cancer in Women Participating in the Nurses' Health Study," *Journal of the National Cancer Institute*, **93**:20, 1563-1568.
1. Seago, J. A., 2008, "Nurse Staffing, Models of Care Delivery, and Interventions," Accessed June 23, 2010, <http://archive.ahrq.gov/clinic/ptsafety/chap39.htm>
- Seago, J. A., A. Williamson, and C. Atwood, 2006, "Longitudinal Analyses of Nurse Staffing and Patient Outcomes: More About Failure to Rescue," *JONA The Journal of Nursing Administration*, **36**:1, 13-21.
- Sen, S. and J. L. Higle, 1999, "An Introductory Tutorial on Stochastic Linear Programming Models," *Interfaces*, **29**:2, 33-61.
- Sheffi, Y., 2004, "Combinatorial Auctions in the Procurement of Transportation Services," *Interfaces*, **34**:4, 245-252.
- Shen, Z.-J. M., C. Coullard and M. S. Daskin, 2003, "A joint location-inventory model," *Transportation Science*, **37**:1, 40-55.

- Shen, Z-J M. and M. S. Daskin, 2005, "Tradeoffs Between Customer Service and Cost in an Integrated Supply Chain Design Framework," *Manufacturing and Service Operations Management*, **7**:3, 188-207.
- Snyder, L. V. and M. S. Daskin, 2005, "Reliability Models for Facility Location: The Expected Failure Case," *Transportation Science*, **39**:3, 400-416.
- Snyder, L.V. and M. S. Daskin, 2006, "Models for Reliable Supply Chain Network Design," invited paper, chapter in Reliability and Vulnerability in Critical Infrastructure: A Quantitative Geographic Perspective, A. Murray and T. H. Grubestic, editors, Advances in Spatial Science Series, Springer.
- Snyder, L.V., M. P. Scaparra, M. S. Daskin and R. L. Church, 2006, "Planning for Disruptions in Supply Chain Networks," Tutorials on Operations Research, M. Johnson, B. Norman and N. Secomandi, eds., Ch. 9, pp. 234-257, INFORMS, Hanover, MD.
- Stanton, M. W. and M. Rutherford, 2004, "Hospital Nurse Staffing and Quality of Care," Agency for Healthcare Research and Quality, Research in Action Issue 14, AHRQ Pub. No. 04-0029, Rockville, MD.
- Taillard, E., P. Badeau, M. Gendreau, F Guertin, and J.-Y. Potvin, 1997, "A Tabu Search Heuristic for the Vehicle Routing Problem with Soft Time Windows," *Transportation Science*, **31**:2, 170-186.
- Tamhane, A. C. and D. D. Dunlop, 2000, Statistics and Data Analysis: from Elementary to Intermediate, Prentice Hall, Upper Saddle River, NJ.
- Tekin, E., W. J. Hopp, and M. P. Van Oyen, 2009, "Pooling Strategies for Call Center Agent Cross-training," *IIE Transactions*, **41**:6, 546-561.
- Tekin, E. and I Sabuncuoglu, 2004, "Simulation optimization: A comprehensive review on theory and applications," *IIE Transactions*, **36**, 1067-1081.
- Thompson, J. M. and D. Parr, 2007, "Solving the multi-objective nurse scheduling problem with a weighted cost function," *Annals of Operations Research*, **155**:1, 279-288.
- Toregas, C., R. Swain, C. ReVelle, and L. Bergman, 1971, "The Location of Emergency Service Facilities," *Operations Research*, **19**, 1363-1373.
- Toth, P. and D. Vigo, 2003, "The Granular Tabu Search and Its Application to the Vehicle-Routing Problem," *INFORMS Journal on Computing*, **15**:4, 333-346.
- U. S. Census Bureau, 2009, "Monthly & Annual Retail Trade," Retrieved July 23, 2009, from <http://www.census.gov/retail/>.

- Wasner, M. and G. Zäpfel, 2004, "An integrated multi-depot hub-location vehicle routing model for network planning of parcel service," *International Journal of Production Economics*, **90**, 403-419.
- Waxman, H. A. (prepared for), 2001, National Preparedness: Ambulance Diversions Impede Access to Emergency Rooms. Edited by U.S. House of Representatives. Washington, DC.
- What's Best, 2010, Version 10.0, Lindo Systems, Chicago, IL, www.lindo.com.
- Whitt, W., 2007, "What You Should Know About Queueing Models to Set Staffing Requirements in Service Systems," *Naval Research Quarterly*, 54:5, 476-484.
- Winston, W. L. and M. Venkataramanan, 2003, Introduction to Mathematical Programming -- Operations Research: Volume One, fourth edition, Thomson Brooks/Cole, Pacific Grove, CA.
- Wong, W. and B. Rose, 2008, "Starbucks' exit leaves some areas wondering," The Chicago Tribune, July 19, 2008, online at www.chicagotribune.com/business/chisat-starbucksjul19,0,7595492.story.
- Wright, P. D., K. M. Bretthauer, and M. Côté, 2006, "Reexamining the Nurse Scheduling Problem: Staffing Ratios and Nursing Shortages," *Decision Sciences*, **37**:1, 39-70.