4. PROBABILITY DISTRIBUTIONS
دالة توزيع الاحتمال

Probability Distribution is a mathematical function that, stated in simple terms, can be thought of as providing the probabilities of occurrence of different possible outcomes in an experiment.

توزيع الاحتمال هو دالة حسابية يمكن التعبير عنها ببساطة كمجموع احتمالات حدوث نتائج مختلفة في تجربة.

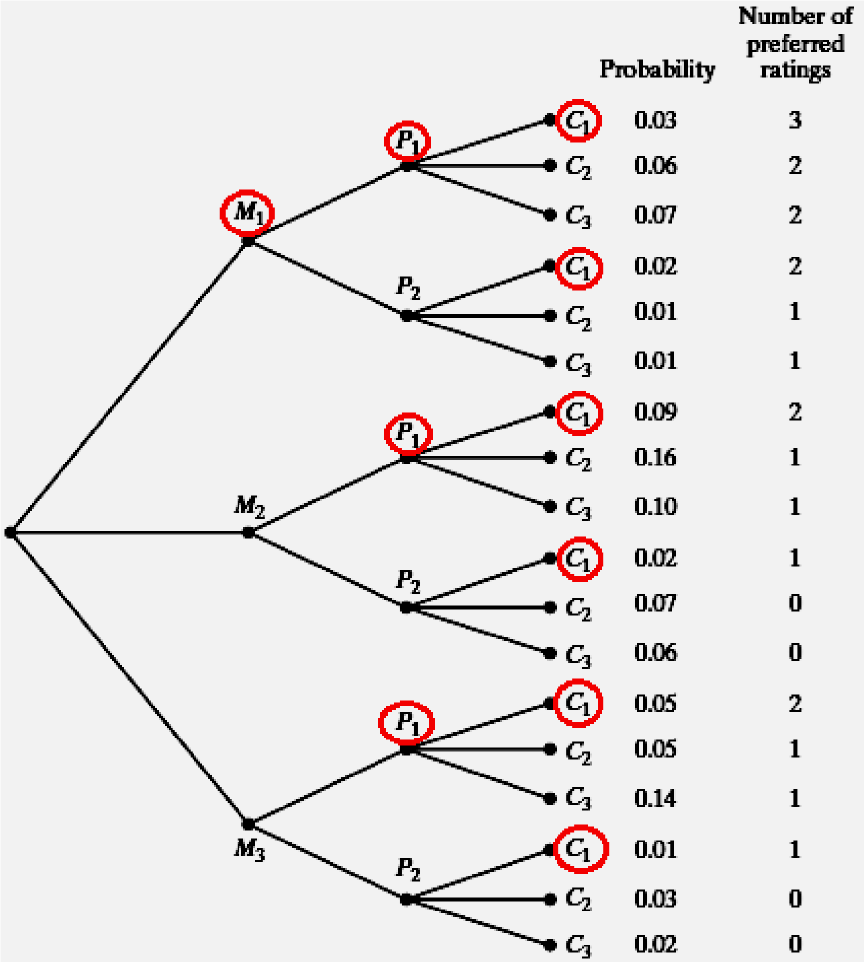
Discrete Random Variables (المتغير العشوائي المتقطع (المنفصل

A **random discrete variable** is obtained by assigning a numerical value to each outcome of a particular experiment **(Preferred Attributes)**.

In other words, **discrete random variables** are those who take discrete values, that is, **countable values** such as 0,1,2,3, .. and so on. Some examples of discrete random variables are **number of Kings** coming while choosing **n** cards from deck of **52 cards**.

Random variables are denoted by capital letters ***X, Y*** and so on, to distinguish them from their possible values given in lowercase ***x, y***.

Example 1: let us refer to the used car example given before and the corresponding probabilities shown in The figure. Now let us refer to $M1$ (low current mileage), $P1$ (moderate price), and $C1$ (inexpensive to operate) **as preferred attributes**. Find the probabilities that a used car will get 0, 1, 2, or 3 preferred attributes (**random variable values)** denoted by **x**.



Solution: The given figure indicates for each outcome the number of preferred attributes. Adding the respective probabilities, we find that for **0** preferred attributes the probability is

$P(X=0)$ =0.07 + 0.06 + 0.03 + 0.02 = 0.18

and for **one** preferred attribute the probability is

$P(X=1)$ = 0.01 + 0.01 + 0.16 + 0.10 + 0.02 +0.05 + 0.14 + 0.01 = 0.50

For **two** preferred attributes, the probability is: $P(X=2)$ = 0.06 + 0.07 + 0.02 + 0.09 + 0.05= 0.29 and for **three** preferred attributes the probability is $P(X=3)$ = 0.03.

These results may be summarized, as in the following table, where x denotes a possible number of preferred attributes and

x	0	1	2	3
Probability	0.18	0.50	0.29	0.03

Probability is the **probability distribution function** denoted by $f(x) = P(X = x)$ which assigns probability to each possible outcome x that is called **the probability distribution**.

Probability distributions Characteristics

- The **probability distribution** of a discrete random variable X is a list of the possible values of $\boldsymbol{X}$ together with their probabilities

$$f(x) = P[X = x]$$

The probability distribution always satisfies the conditions

$$f(x) \geq 0 \qquad \text{and} \qquad \sum_{all\ x} f(x) = 1$$

- Besides the probability $f(x)$ that the value of a random variable is x, there is an important related function. It gives the probability $F(x)$ that the value of a random variable is less than or equal to x. Specifically,

$$F(x) = P[X \leq x] \text{ for all } -\infty < x < \infty$$

and we refer to the function $F(x)$ as the **cumulative distribution function.** Referring to the used car **Example 1** and basing our calculations on the table on, **cumulative distribution function** can be calculate as the following:

x	0	1	2	3
$f(x) = P[X = x]$	0.18	0.50	0.29	0.03
$F(x)$	0.18	0.68	0.97	1.00

Binomial Distribution

Binomial distribution is applicable to data where **one of two mutually exclusive and independent outcomes are possible** as a result of a single experiment. The experiment is called a Bernoulli trial. From the language of games of chance, we might say that in each of these examples we are interested in the probability of getting x **successes** in n **trials**, or, in other words, x **successes** and $n - x$ **failures** in n attempts.

Consider an experiment consisting of

- n trials
- that are independent and
- that each have a constant probability p of success.

Then the total number of successes X is a **random variable** that has a **binomial distribution** with parameters n and p, which is written as $X \sim B(n, p)$.

The **probability distribution** function of a $B(n, p)$ random variable is:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad for \quad x = 0, 1, 2, 3$$

Sometimes $P(X = x)$ can be denoted by $b\,(x; n, p)$

Note: $\binom{n}{x} = \dfrac{n!}{x!\,.\,(n-x)!}$

Example 2: Binomial probability distribution n =3

When a relay tower for wireless phone service breaks down, it quickly becomes an expensive proposition for the phone company, and the cost increases with the time it is inoperable. From company records, it is postulated that the probability is **0.90** that the breakdown can be repaired within one hour. For the next three breakdowns, on different days and different towers,

(a) List all possible outcomes in terms of success, *S*, repaired within one hour, and failure, *F*, not repaired within one hour.

(b) Find the probability distribution of the number of successes, ***X*,** among the **3 repairs**.

Solution (a) The possible outcmes are 2^3=2 × 2 × 2 = 8 can be arranged as follows:

X	Possiblities
0	**FFF**
1	**FFS,FSF,SFF**
2	**FSS,SFS,SSF**
3	**SSS**

Sample space = **{ FFF,FFS,FSF,SFF,FSS,SFS,SSF, SSS},** as shown in the table where the number of successes *X* is the same for each outcome in a row.

(b) Not that the results of repairs on different days and different towers should be independent. Also, the probability of success **0.90** is the same for each repair. Therefore the probability distribution according to the given formula can be calculated as the following for each x:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad for\ x = 0, 1, 2, 3 \ \textit{Note:} \binom{n}{x} = \frac{n!}{x! \, . \, (n-x)!}$$

$$P(X = 0) = \binom{3}{0} 0.9^0 (1-0.9)^{3-0} = 0.001$$

$$P(X = 1) = \binom{3}{1} 0.9^1 (1-0.9)^{3-1} = 0.027$$

$$P(X = 2) = \binom{3}{2} 0.9^2 (1-0.9)^{3-2} = 0.243$$

$$P(X = 3) = \binom{3}{3} 0.9^3 (1-0.9)^{3-3} = 0.729$$

$$\sum_{all\ x} f(x) = 0.001+0.027+0.243+0.729=1.0$$

EXAMPLE 3: It has been claimed that in 60% of all solar-heat installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one-third in

(a) four of five installations;

(b) at least four of five installations?

Solution:

(a) Substituting $\boldsymbol{x}$ = 4, ***n*** = 5, and $\boldsymbol{p}$ = 0.60 into the formula for the binomial distribution:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \textit{Note:} \binom{n}{x} = \frac{n!}{x!\,.\,(n-x)!}$$

$$P(X = 4) = \binom{5}{4} 0.6^4 (1-0.6)^{5-4} = 0.2592$$

(b) Substituting x = 5, n = 5, and p = 0.60 into the formula for the binomial distribution:

$$P(X = 5) = \binom{5}{5} 0.6^5 (1-0.6)^{5-5} = 0.07776$$

Therefore, the probability of at least four of five installations is :

$$P\ (x \geq 4) = 0.2592 + 0.07776 = 0.33696 \approx 0.337$$

If n is large, the calculation of **binomial probabilities** can become quite boring. It is convenient to refer to special tables. **Table 1** (given in separate file) gives the values of **cumulative probabilities** rather than the values of $P(X = x)$ or $b(x; n, p)$:

$$B(x; n, p) = \sum_{k=0}^{x} \binom{n}{x} p^x (1-p)^{n-x} \quad for \quad x = 0, 1, 2, 3$$

$$Note: \binom{n}{x} = \frac{n!}{x!\,.\,(n-x)!}$$

Where $B(x; n, p)$ is **cumulative probabilities . Table 1** is design for n **= 2** to n **= 20** and p **= 0.05, 0.10, 0.15, . . . , 0.90, 0.95**.

The table gives the values of **cumulative probabilities,** $B(x; n, p)$**,** rather than the values of $b(x; n, p)$, because the values of $B(x; n, p)$ are the ones needed more often in statistical applications. Note, however, the values of $b(x; n, p)$ can be obtained by subtracting adjacent entries in **Table 1** since the two cumulative probabilities $B(x; n, p)$ and $B(x-1; n, p)$ differ by the single term $b(x; n, p)$

$$P(X = x) = b(x; n, p) = B(x; n, p) - B(x-1; n, p)$$

Example 4: If the probability is 0.05 that a certain wide-flange column will fail under a given axial load, what are the probabilities that among 16 such columns

(a) at most two will fail;

(b) at least four will fail?

Solution

(a) Table 1 shows that

$$\boldsymbol{B(x; n, p) = B(2; 16, 0.05) = \sum_{k=0}^{2}} b\,(\,x\,;\,16,\,0.05)$$

From Table 1 $\quad \boldsymbol{B(2; 16, 0.05) = 0.9571}$

(b)

$$\sum_{k=4}^{16} b\,(\,x\,;\,16,\,0.05) = 1 - B\,(3;\,16,\,0.05)$$

From Table 1 $\quad = 1 - 0.9930 \quad = 0.0070.$

The Hypergeometric Distribution

The hypergeometric distribution is a discrete probability distribution that describes the probability of $\boldsymbol{x}$ **successes** in $\boldsymbol{n}$ draws, without replacement, from a finite population of size $\boldsymbol{N}$ that contains exactly $\boldsymbol{a}$ objects with that feature, wherein each draw is either a success or a failure.

To solve the problem the $\boldsymbol{x}$ successes (defectives) can be chosen in $\binom{a}{x}$ ways, the $\boldsymbol{n-x}$ failures (nondefectives) can be chosen in $\binom{N-a}{n-x}$ ways, and hence, $\boldsymbol{x}$ successes and $\boldsymbol{n-x}$ failures can be chosen in $\binom{a}{x}\binom{N-a}{n-x}$ ways. Also, $\boldsymbol{n}$ objects can be chosen from a set of $\boldsymbol{N}$ objects in $\binom{N}{n}$ ways, and if we consider all the possibilities as equally likely, it follows that for sampling without replacement **the probability of getting "$\boldsymbol{x}$ successes in $\boldsymbol{n}$ trials"** is

Hypergeometric Distribution Probabilities

$$P(X = x) = h(x; n, a, N) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} \; for \; x = 0, 1, 2, \ldots\ldots n$$

where $\boldsymbol{x}$ cannot exceed $\boldsymbol{a}$ and $\boldsymbol{n-x}$ cannot exceed $\boldsymbol{N-a}$. This equation defines the **hypergeometric distribution**, whose parameters are the sample size $\boldsymbol{n}$, the lot size(or population size) $\boldsymbol{N}$, and the number of "**successes**" in the lot $\boldsymbol{a}$.

Example 5: Calculating a probability using the hypergeometric distribution

An Internet-based company that sells discount accessories for cell phones often ships an excessive number of defective products. The company needs better control of quality. Suppose it has **20** identical car chargers on hand but that **5** are defective. If the company decides to randomly select **10** of these items, <u>what is the probability that 2 of the 10 will be defective?</u>

Solution Substituting $\boldsymbol{x}$ = 2, $\boldsymbol{n}$ = 10, $\boldsymbol{a}$ = 5, and $\boldsymbol{N}$ = 20 into the formula for the hypergeometric distribution, then

$$\boldsymbol{P(X=x)=h(x;n,a,N)=h(2;10,5,20)=\frac{\binom{5}{2}\binom{20-5}{10-2}}{\binom{20}{10}}=\frac{10\text{x}6435}{184756}=0.348}$$

In the preceding example, $\boldsymbol{n}$ was not small compared to $\boldsymbol{N}$, and if we had made the mistake of using the binomial distribution with $\boldsymbol{n}$ **= 10** and $\boldsymbol{p}$ **=** $\frac{\mathbf{5}}{\mathbf{20}}$**= 0.25** to calculate the probability of two defectives, the result would have been 0.282, which is much too small. However, when $\boldsymbol{n}$ is small compared to $\boldsymbol{N}$, less than $\frac{\boldsymbol{N}}{\mathbf{10}}$, the composition of the lot is not seriously affected by drawing the sample without replacement, and the binomial distribution with the parameters $\boldsymbol{n}$ and $\boldsymbol{p}$ **=** $\frac{\boldsymbol{a}}{\boldsymbol{N}}$ will yield a good approximation

The Mean and the Variance of a Probability Distribution

The most important statistical measures, describing the location and the variation of the probability distribution are the **mean** and the **variance**.

Mean of Probability Distribution (Mathematical Expectation)

If a random variable $\boldsymbol{X}$ takes on the values $\boldsymbol{x_1, x_2, \ldots, or\ x_k}$, with the probabilities $\boldsymbol{f}\ (\boldsymbol{x_1})$, $\boldsymbol{f}\ (\boldsymbol{x_2}), \ldots, and\ \boldsymbol{f}\ (\boldsymbol{x_k})$, its mathematical expectation or expected value is:

$$\boldsymbol{x_1 \, . \, f(x_1) + x_2 \, . \, f(x_2) + \ . \ . \ . + x_k \, . \, f(x_k) = \sum (value)}\ \textbf{\textit{x (probability)}}$$

The **mean of a probability distribution** is denoted by the Greek letter $\boldsymbol{\mu}$. Alternatively, the **mean** of a random variable $\boldsymbol{X}$, or its probability distribution that is called its **expected value** and is denoted by $\boldsymbol{E}\ (\ \boldsymbol{X}\ )$. It has to be mentioned that both $\boldsymbol{\mu}$ and $\boldsymbol{E}\ (\ \boldsymbol{X}\ )$ refer to the same quantity. Finally, the **mean of discrete probability distribution** can be described as:

$$\boldsymbol{\mu = E(X) = \sum_{all\ x} x \, . \, f(x)}$$

Example 6: Find the mean of the probability distribution of the **number of heads** obtained in **3 flips** of a balanced coin.

Solution: The probabilities can be easily be verified by counting equally likely possibilities as the following:

P(0)= (TTT) $= \frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2} = \frac{1}{8}$

P(1)=(HTT,THT,TTH) $= \left(\frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2}\right) + \left(\frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2}\right) + \left(\frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2}\right) = \frac{3}{8}$

P(2)=(HHT,HTH,THH) $= \left(\frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2}\right) + \left(\frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2}\right) + \left(\frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2}\right) = \frac{3}{8}$

P(3)=(HHH) $= \frac{1}{2} \text{ x } \frac{1}{2} \text{ x } \frac{1}{2} = \frac{1}{8}$

Or by using the formula for the binomial distribution with ***n*** **= 3** and ***p*** $= \frac{1}{2}$

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$$P(X = 0) = \binom{3}{0} 0.5^0 (1 - 0.5)^{3-0} = 0.125 = \frac{1}{8}$$

$$P(X = 0) = \binom{3}{1} 0.5^1 (1 - 0.5)^{3-1} = 0.375 = \frac{3}{8}$$

$$P(X = 0) = \binom{3}{2} 0.5^2 (1 - 0.5)^{3-2} = 0.375 = \frac{3}{8}$$

$$P(X = 0) = \binom{3}{3} 0.5^3 (1 - 0.5)^{3-3} = 0.125 = \frac{1}{8}$$

Accordingly $\mu = 0\text{x}\frac{1}{8} + 1\text{x}\frac{3}{8} + 2\text{x}\frac{3}{8} + 3\text{x}\frac{1}{8} = \frac{3}{2}$

It has to be mentioned that the mean of Binomial Distribution can be calculated from the very simple equation:

$$\mu = n \cdot p$$

Also, the mean of hypergeometric distribution can be calculated from the very simple equation:

$$\mu = n \,.\, \frac{a}{N}$$

Example 7: With reference to Example 8 in which 5 of 20 cell phone chargers are defective, find the mean of the probability distribution of the number of defectives in a sample of 10 randomly chosen for inspection.

Solution: Substituting **x = 2, n = 10, a = 5, and N = 20** into the above formula for μ, as the following:

$$\mu = n \,.\, \frac{a}{N} = 10 \text{ x } \frac{5}{20} = 2.5$$

Definition and Interpretation of Variance and Stanard Deviation

Another important summary measure of the distribution of a random variable is the **variance**, which measures the ***spread* or *variability*** in the values taken by the random variable. Whereas the mean or expectation measures the central or average value of the random variable, the variance measures the spread or deviation of the random variable about its mean value. The **variance** is defined of a probability distribution $\boldsymbol{f(x)}$, or that of the random variable $\boldsymbol{X}$ which has that probability distribution, as

$$\boldsymbol{\sigma^2 = \sum_{all\ x} (x-\mu)^2 . f(x)}$$

This measure is not in the same units as the values of the random variable, however it can be adjusted taking the square root. That is expressed in the same units in which the random variable is expressed and it is called The **standard deviation** as:

$$\boldsymbol{\sigma = \sqrt{\sum_{all\ x} (x-\mu)^2 . f(x)}}$$

An alternative formula for variance is:

$$\boldsymbol{\sigma^2 = \sum_{all\ x} x^2 . f(x) - \mu^2}$$

$$\boldsymbol{= E\left[X^2\right] - \mu^2}$$

where $\boldsymbol{E\left[X^2\right]}$ is defined as $\boldsymbol{\sum_{all\ x} x^2 . f(x)}$

Example 8 Determine the variance of the probability distribution of the number of points rolled with a balanced die.

Solution Since $f(x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5,$ and 6, Then

$$\mu = E(X) = \sum_{all\ x} x \, . f(x)$$

$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= \frac{7}{2}$$

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$$

$$= \frac{91}{6}$$

As $\sigma^2 = \sum_{all\ x} x^2 . f(x) - \mu^2$

$= E[X^2] - \mu^2$ then

$$\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Example 9 As part of **a quality-improvement project** focused on **the delivery of mail** at a department office within a large company, data were gathered on **the number of different addresses that had to be changed** so the mail could be redirected to the correct mail stop. The **probability distribution**, given in the given table, describes **the number of redirects per delivery.** <u>Compute the mean and variance</u>.

x	$f(x)$
0	.05
1	.20
2	.45
3	.20
4	.10

Solution $x\,f(x)$ and $x^2\,f(x)$ are determine d in the following:

x	$f(x)$	$x f(x)$	$x^2 f(x)$
0	.05	.0	0.0
1	.20	.2	0.2
2	.45	.9	1.8
3	.20	.6	1.8
4	.10	.4	1.6
Total		2.1	5.4

so $\mu = 2.1$ and $\sigma^2 = 5.4 - (2.1)^2 = 0.990.$

The variance of the binomial distribution with the parameters $\boldsymbol{n}$ and $\boldsymbol{p}$ is given by the following formula:

$$\boldsymbol{\sigma^2 = n\,.\;(1-p)}$$

Example 10 With reference to **Example 4** in It has been claimed that in **60%** of all **solar-heat installations** the utility bill is reduced by at least one-third. Find the standard deviation of the probability distribution of **five** solar-heat installations .

Solution Substituting $\boldsymbol{n}$ = 5, and $\boldsymbol{p}$ = 0.60 into the formula for the variance of a hypergeometric distribution, then

$$\boldsymbol{\sigma^2 = n\,.\;(1-p)\; = 5\,.\;(1-0.6) = 2}$$

Then standard deviation $\boldsymbol{\sigma = \sqrt{2} = 1.41}$

The variance of the hypergeometric distribution with the parameters $\boldsymbol{n}$, $\boldsymbol{a}$, and $\boldsymbol{N}$ is given by the following formula: $\boldsymbol{\sigma^2 = n\,\frac{a}{N}\left(1-\frac{a}{N}\right)\left(\frac{N-n}{N-1}\right)}$

The factor $\boldsymbol{\left(\frac{N-n}{N-1}\right)}$ adjusts for the finite population.

Example 11 With reference to **Example 5** in which 5 of 20 cell phone chargers are defective, find the standard deviation of the probability distribution of the number of defectives in a sample of **10** randomly chosen for inspection.

Solution Substituting $\boldsymbol{n}$ = 10, $\boldsymbol{a}$ = 5, and $\boldsymbol{N}$ = 20 into the formula for the variance of a hypergeometric distribution, then

$$\boldsymbol{\sigma^2 = n\,\frac{a}{N}\left(1-\frac{a}{N}\right)\left(\frac{N-n}{N-1}\right) = 10\,\frac{5}{20}\left(1-\frac{5}{20}\right)\left(\frac{20-10}{20-1}\right) = \frac{75}{76}}$$

$$\boldsymbol{\sigma = \sqrt{75/76} = 0.99}$$

The Poisson Distribution and Rare Events

The Poisson Distribution is often useful to define a random variable that counts the number of events which do not have a natural upper bound and occur within certain specified boundaries. For example, an experimenter may be interested in the number of defects in an item, the number of radioactive particles emitted by a substance, or the number of telephone calls received by an operator within a certain time limit. The **Poisson distribution** is often appropriate to model such situations.

The **Poisson distribution**, with mean λ (lambda), has probabilities given by

$$f(x;\lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!} \qquad \text{for} \quad x = 0, 1, 2, \dots \dots \dots \dots \qquad\qquad \lambda > 0$$

It can be shown that **the mean** and **the variance** of the Poisson distribution with the parameter λ are given by

$$\mu = \lambda \qquad and \qquad \sigma^2 = \lambda$$

Since **the Poisson distribution** has many important applications, it has been extensively tabulated. **Table 2** gives the values of the probabilities:

$$F(x;\lambda) = \sum_{k=0}^{x} f(k;\lambda) = \sum_{k=0}^{x} e^{-\lambda}\,\frac{\lambda^k}{k!}$$

for values of λ in varying increments **from 0.02 to 15**, and its use is very similar to that of Table 1.

Example 12 For health reasons, homes need to be inspected for radon gas which decays and produces alpha particles. One device counts the number of alpha particles that hit its detector. To a good approximation, in one area, the count for the next week follows a Poisson distribution with mean 1.3. Determine

(a) the **probability** of exactly **one** particle next week.

(b) the **probability** of **one or more** particles next week.

(c) the **probability** of at **least two but no more than four** particles next week.

(d) The **variance** of the Poisson distribution.

Solution Unlike the binomial case, there is no choice of a fixed Bernoulli trial here because

one can always work with smaller intervals

(a) $f(x;\lambda) = f(1;1.3) = P(X=1) = e^{-\lambda}\,\dfrac{\lambda^x}{x!}$

$$= e^{-1.3}\,\frac{1.3^1}{1!} \qquad = 0.3543$$

Alternatively, **using Table 2,**

$F(1, 1.3) - F(0, 1.3) = 0.627 - 0.273 = 0.354$

(b) using Table 2

$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1.3} = 1 - 0.273 = 0.727$

(c) using Table 2

$P(2 \leq X \leq 4) = F(4, 1.3) - F(1, 1.3)$

$= 0.989 - 0.627 = 0.362$

λ \ k	0	1	2	3	4
.02	.980	1.000			
.04	.961	.999	1.000		
.06	.942	.998	1.000		
.08	.923	.997	1.000		
.10	.905	.995	1.000		
.15	.861	.990	.999	1.000	
.20	.819	.982	.999	1.000	
.25	.779	.974	.998	1.000	
.30	.741	.963	.996	1.000	
.35	.705	.951	.994	1.000	
.40	.670	.938	.992	.999	1.000
.45	.638	.925	.989	.999	1.000
.50	.607	.910	.986	.998	1.000
.55	.577	.894	.982	.998	1.000
.60	.549	.878	.977	.997	1.000
.65	.522	.861	.972	.996	.999
.70	.497	.844	.966	.994	.999
.75	.472	.827	.959	.993	.999
.80	.449	.809	.953	.991	.999
.85	.427	.791	.945	.989	.998
.90	.407	.772	.937	.987	.998
.95	.387	.754	.929	.981	.997
1.00	.368	.736	.920	.981	.996
1.1	.333	.699	.900	.974	.995
1.2	.301	.663	.879	.966	.992
1.3	.273	.627	.857	.957	.989
1.4	.247	.592	.833	.946	.986
1.5	.223	.558	.809	.934	.981

Example 13 Comparing Poisson and binomial probabilities

It is known that 5% of the books bound at a certain bindery have defective bindings. Find the probability that 2 of 100 books bound by this bindery will have defective bindings using

(a) the formula for the binomial distribution;

(b) the Poisson approximation to the binomial distribution.

Solution

(a) Substituting x **= 2,** n **= 100, and** p **= 0.05** into the formula for the binomial distribution:

$$b(x; n, p) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$b(2; 100, 0.05) = P(X = 2) = \binom{100}{2} 0.05^2 (1 - 0.05)^{100-2} = 0.0812$$

(b) Substituting x **= 2** and $\lambda = \mu = n \,.\, p$ **= 100(0.05) = 5**

into the formula for the Poisson Distribution:

$$f(x; \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$f(2\,; 5) = e^{-5} \frac{5^2}{2!} = 0.0842$$

λ \ k	0	1	2
2.8	.061	.231	.469
3.0	.050	.199	.423
3.2	.041	.171	.380
3.4	.033	.147	.340
3.6	.027	.126	.303
3.8	.022	.107	.269
4.0	.018	.092	.238
4.2	.015	.078	.210
4.4	.012	.066	.185
4.6	.010	.056	.163
4.8	.008	.048	.143
5.0	.007	.040	.125

The error if it would be calculated by using the Poisson approximation is only 0.003. If **Table 2** is used instead of using a calculator, the following will be obtained:

$$f(2; 5) = F(2; 5) - F(1; 5) = 0.125 - 0.040 = 0.085$$

Poisson Processes

In this section. It will be concerned with **processes taking place over continuous intervals of time** or space, such as the occurrence of imperfections on a continuously produced of cloth. It will be shown that the best model that can be used to describe these situations is the Poisson distribution.

To find the probability of $\boldsymbol{x}$ successes during a time interval of length $\boldsymbol{T}$, the interval is divided into $\boldsymbol{n}$ equal parts of length $\Delta \boldsymbol{t}$, so that $\boldsymbol{T} = \boldsymbol{n} \cdot \Delta \boldsymbol{t}$. This means that the probability of $\boldsymbol{x}$ successes in the time interval $\boldsymbol{T}$ is given by the binomial probability $\boldsymbol{b}(\boldsymbol{x}; \boldsymbol{n}, \boldsymbol{p})$ with:

$$\boldsymbol{n} = \frac{\boldsymbol{T}}{\boldsymbol{\Delta t}} \qquad \text{and} \qquad \boldsymbol{p} = \boldsymbol{\alpha} \cdot \boldsymbol{\Delta t}$$

Where $\boldsymbol{\alpha}$ is the average (mean) number of successes per unit time. As $\boldsymbol{n} \to \infty$ then the probability of $\boldsymbol{x}$ successes during the time interval $\boldsymbol{T}$ is given by the corresponding **Poisson probability** with the parameter:

$$\boldsymbol{\lambda} = \boldsymbol{n} \cdot \boldsymbol{p} = \frac{\boldsymbol{T}}{\boldsymbol{\Delta t}} \cdot (\boldsymbol{\alpha} \,.\, \boldsymbol{\Delta t}) = \boldsymbol{\alpha} \cdot \boldsymbol{T}$$

Note that $\boldsymbol{\lambda}$ is the mean of this Poisson distribution.

Example 14 If a bank receives on the average α **= 6** bad checks per day, what are the probabilities that it will receive

(a) 4 bad checks on any given day?

(b) 10 bad checks over any **2 consecutive days**?

Solution

(a) Substituting x **= 4** bad checks, T **=1** day and $\lambda = \alpha \cdot T$ **= 6 · 1 = 6** into the formula for the Poisson distribution, then

$$f(x;\lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!}$$

$$f(4;6) = e^{-6}\,\frac{6^4}{4!} = 0.1339$$

Or from Table 2 $f(4;6) = F(4;6) - F(3;6) = 0.285 - 0.151 = 0.134$

(b) Here T **= 2** days then $\lambda = \alpha \cdot T$ **= 6x2 = 12** so it will be found f**(10; 12)**:

$$\begin{aligned} f(10;12) &= F(10;12) - F(9;12) \\ &= 0.347 - 0.242 \\ &= 0.105 \end{aligned}$$

Example 15 A computing system manager states that the rate of interruptions to the internet service is **0.2 per week**. Use **the Poisson distribution** to find the probability of

(a) **one** interruption in **3 weeks**

(b) at least two interruptions in **5 weeks**

(c) at **most one** interruption in **15 weeks**.

Solution as α is the average number of successes per unit time that is the interruption per week in the given example, then α **= 0.2 interruption/week**, the following solution are done from Table 2:

a) with x **= 1** and $\lambda = \alpha \cdot T$ **= (0.2) · 3 = 0.6**, then

$$F(x;\lambda) = F(1;0.6) - F(0;0.6) = 0.878 - 0.549 = 0.329$$

(b) with $x \geq 2$ and $\lambda = \alpha \cdot T$ **= (0.2) · 5 = 1.0**, then

$$1 - F(x;\lambda) = 1 - F(1;1.0) = 1 - 0.736 = 0.264$$

(c) with x **= 1** and $\lambda = \alpha \cdot T$ **= (0.2) · 15 = 3.0**, then

$$F(x;\lambda) = F(1;3.0) = 0.199$$

Continuous Random Variables

There is also exist random variables whose set of possible values is uncountable. Two examples are the time that a train arrives at a specified stop and the lifetime of a transistor. Let $\boldsymbol{X}$ be such a random variable. We say that $\boldsymbol{X}$ is a continuous random variable if there exists a nonnegative function $\boldsymbol{f}$, defined for all real $\boldsymbol{x} \in (-\infty, \infty)$, having the property that, for any set $\boldsymbol{B}$ of real numbers

$$\boldsymbol{P\{X \in B\} = \int_B f(x)\, dx}$$

In general, it is written as $\boldsymbol{P(a \leq X \leq b)}$ for the probability with the points of the sample space for which the value of a random variable falls on the interval from $\boldsymbol{a}$ to $\boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are constants with $\boldsymbol{a \leq b}$. Then the probability that the random variable with which it is concerned on a value in the interval from $\boldsymbol{a}$ to $\boldsymbol{b}$ is given by

$$\boldsymbol{P(a \leq X \leq b) = \int_a^b f(x)\, dx}$$

The function $\boldsymbol{f}$ is called **the probability density function** of the random variable $\boldsymbol{X}$.

It has to be mentioned that

$$\boldsymbol{f(x) \geq 0} \quad \text{for all} \quad \boldsymbol{x}$$

and

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

As in the discrete case, let $\boldsymbol{F(x)}$ be the probability that a random variable with the probability density $\boldsymbol{f\ (x)}$ takes on a value less than or equal to $\boldsymbol{x}$. Again refer to the corresponding function $\boldsymbol{F}$ as the **cumulative distribution function** or just the **distribution function** of the random variable. Thus, for any value $\boldsymbol{x}$, $\boldsymbol{F(x) = P(\,X \leq x\,)}$ is the area under **the probability density function** over the interval $\boldsymbol{-\infty}$ to $\boldsymbol{x}$. In the usual calculus notation for the integral

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

The probability that the random variable value on the interval from $\boldsymbol{a}$ to $\boldsymbol{b}$ is $\boldsymbol{F(\,b) - F(\,a)}$, and accordingly it follows that

$$\boldsymbol{\frac{dF(x)}{dx} = f\ (x)}$$

Example 16 Calculating probabilities from the probability density function

If a random variable has the probability density

$$f(x) = \begin{cases} 2e^{-2x} & for\ \ x > 0 \\ 0 & for\ \ x \le 0 \end{cases}$$

find the probabilities that it will take on a value

(a) between 1 and 3;

(b) greater than 0.5.

Solution Evaluating the necessary integrals, then

(a)

$$\int_{1}^{3} 2e^{-2x}\,dx = -\left|e^{-2x}\right| = -\left(e^{-2\text{x}3} - e^{-2\text{x}1}\right) = e^{-2} - e^{-6} = 0.133$$

(b)

$$\int_{0.5}^{\infty} 2e^{-2x}\,dx = -\left|e^{-2x}\right| = -\left(e^{-2\text{x}\infty} - e^{-2\text{x}0.5}\right) = e^{-1} - e^{-\infty} = e^{-1} - 0 = 0.368$$

Note $e^{-\infty} = \frac{1}{e^{\infty}} = \frac{1}{\infty}$=0

The Normal Distribution

The **normal** or **Gaussian distribution** is the most important of all continuous probability distributions and is used extensively as the basis for many statistical inference methods. Its importance stems from the fact that it is a natural probability distribution for directly modeling error distributions and many other naturally occurring phenomena. In addition, by virtue of the **central limit theorem**, which is discussed in Section 5.3, the normal distribution provides a useful, simple, and accurate approximation to the distribution of general sample averages.

يعتبر التوزيع الطبيعي (جاوسن) و احد من أهم توزيعات الاحتمالات المستمرة ويستخدم على نطاق واسع كأساس للعديد من طرق الاستدلال الإحصائي و تنبع أهميته من حقيقة أنه توزيع احتمالي طبيعي لتوزيعات العديد من الظواهر المختلفة لتي تحدث بشكل طبيعي بالإضافة إلى أنه يوفر التوزيع الطبيعي و الدقيق لتوزيع متوسطات العينات العامة.

Probabilities relating to normal distributions are usually obtained from special tables, **(Table 3)**. This table relates to the **standard normal distribution**, namely, **the normal distribution** with $\mu = 0$ and its entries are the values of

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt = P(Z \leq z)$$

To find the probability that a random variable having the standard normal distribution will take on a value between a and b, we use the equation

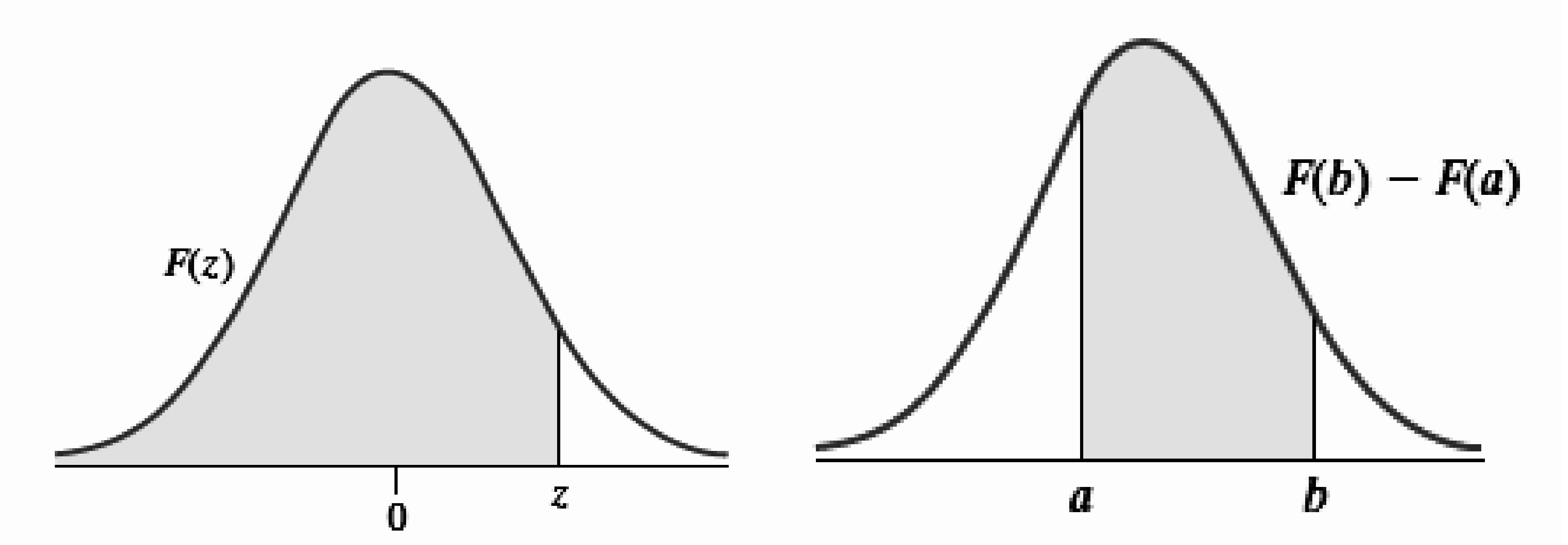
$$P(a < Z \leq b) = F(b) - F(a)$$

Probabilities relating to normal distributions are usually obtained from special tables, **(Table 3)**. This table relates to the **standard normal distribution**, namely, **the normal distribution** with $\boldsymbol{\mu = 0}$ and its entries are the values of

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\,dt = P(Z \le z)$$

To find the probability that a random variable having the standard normal distribution will take on a value between $\boldsymbol{a}$ and $\boldsymbol{b}$, we use the equation

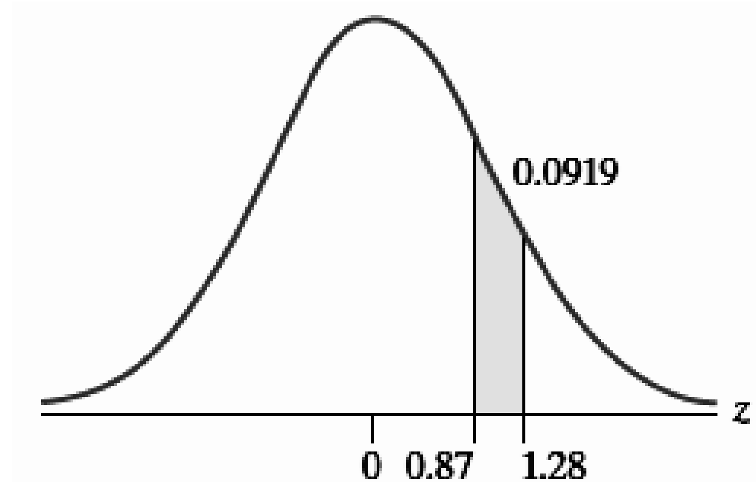
$$\boldsymbol{P(\,a < Z \le b\,) = F(b) - F(a)}$$



Example 17 Calculating some standard normal probabilities

Find the probabilities that a random variable having the standard normal distribution will take on a value

(a) between 0.87 and 1.28;
(b) between −0.34 and 0.62;
(c) greater than 0.85;
(d) greater than −0.65.



Solution It is helpful to first indicate the area of interest in a graph as in the figures

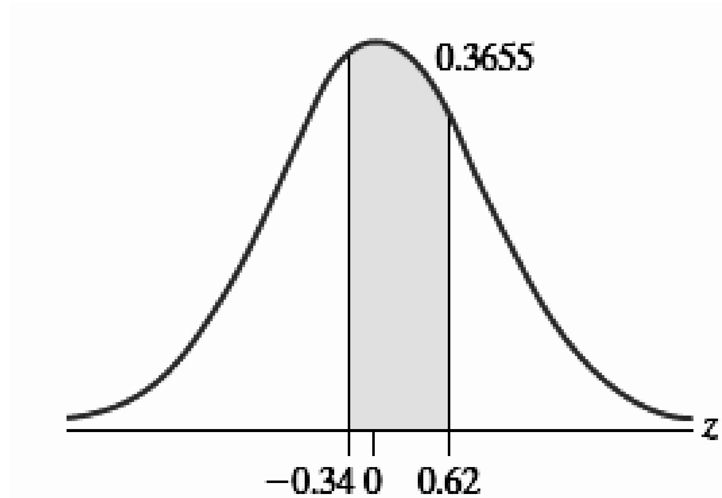
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5973	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

(a) Looking up the necessary values in Table 3,),

$$\boldsymbol{F\ (1.28) - F\ (0.87) = 0.8997 - 0.8078}$$
$$\boldsymbol{= 0.0919}$$

(b) between −0.34 and 0.62
As indicated in the figure

$$F\,(0.62) - F\,(-0.34) = 0.7324 - 0.3669$$
$$= 0.3655$$

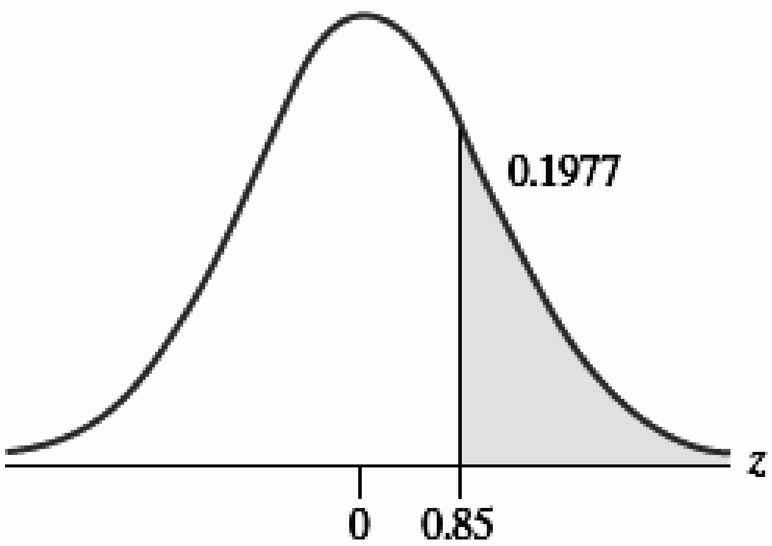


z	0.00	0.01	0.02	0.03	0.04	
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5973	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264

(c) greater than 0.85;
As indicated in the figure,

1 − F (0.85) = 1 − 0.8023
= 0.1977

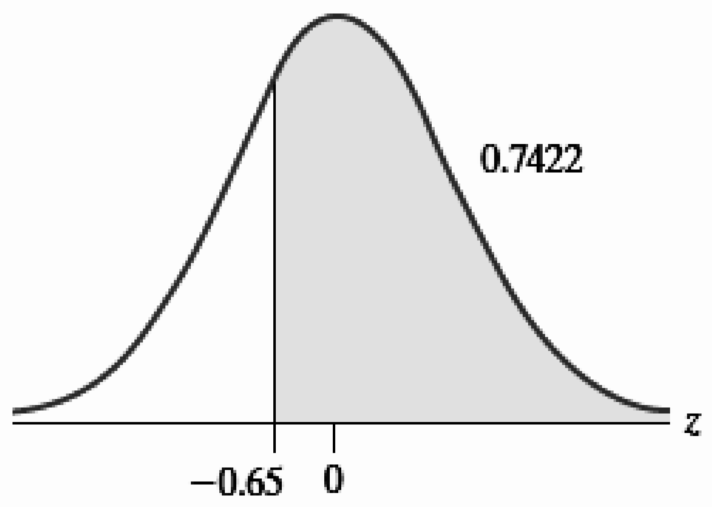


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5973	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315

(d) greater than −0.65

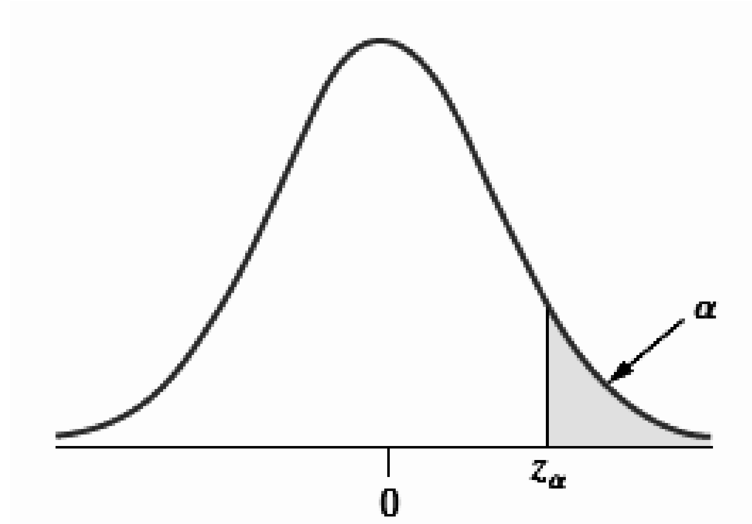
As indicated in Figure 5.9 for part (d)

$1 - F(-0.65) = 1 - 0.2578 = 0.7422$



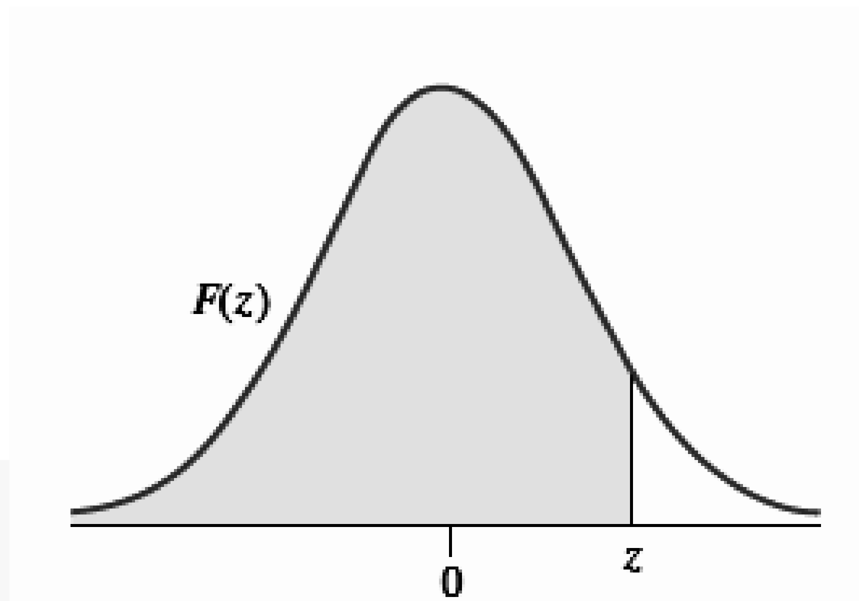
z	0.00	0.01	0.02	0.03	0.04	0.05
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801

There are also problems in which we are given probabilities relating to standard normal distributions and asked to find the corresponding values of $\boldsymbol{Z}$. Let $\boldsymbol{Z_\alpha}$ be such that the probability is $\boldsymbol{\alpha}$ that it will be exceeded by a random variable having the standard normal distribution. That is, $\boldsymbol{\alpha = P(\,Z > z\alpha\,)}$ as illustrated in the figure 5.



Remember that the probabilities relating to normal distributions obtained from **Table 3** is that given as shown in the figure.

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt = P(Z \le z)$$



Example 18 **Two important values for Z_α**

Find (a) z0.01; (b) z0.05.

Solution

(a) Since $F(z_{0.01}) = \mathbf{0.99}$, we look for the entry in **Table 3** which is closest to **0.99** and get **0.9901** corresponding to **z = 2.33**. Thus $z_{0.01} = \mathbf{2.33}$.

z	0.00	0.01	0.02	0.03	0.04	0.05
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929

(b) Since $F(z_{0.05}) = \mathbf{0.95}$, we look for the entry in **Table 3** which is closest to **0.95** and get **0.9495** and **0.9505** corresponding to $z = \mathbf{1.64}\ \boldsymbol{and}\ z = \mathbf{1.65}$. Thus, by interpolation, $z_{0.05} = \mathbf{1.645}$.

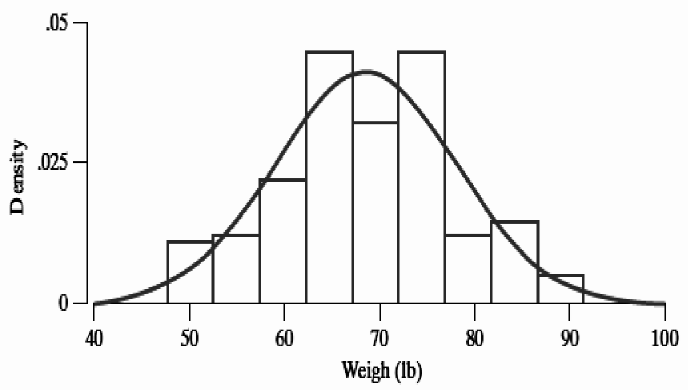
z	0.00	0.01	0.02	0.03	0.04	0.05
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744

To use **Table 3** in connection with a random variable $\boldsymbol{X}$ which has a normal distribution with the mean $\boldsymbol{\mu}$ and the variance $\boldsymbol{\sigma^2}$, we refer to the corresponding **standardized random variable**,

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

Example 19 A major manufacturer of processed meats monitors the amount of each ingredient. The weight(lb) of cheese per run is measured on $\boldsymbol{n} = \mathbf{80}$ occasions.

72.2 67.8 78.0 64.4 76.3 72.3 73.1 71.7 66.2 63.3 85.4 67.4
66.3 76.3 57.7 50.3 77.4 63.1 73.9 67.4 74.7 68.2 87.4 86.4
69.4 58.0 63.3 72.7 73.6 68.8 63.3 63.3 73.0 64.8 73.1 70.9
85.9 74.4 75.9 72.3 84.3 61.8 79.2 64.3 65.4 66.7 77.2 50.0
70.3 90.4 63.9 62.1 68.2 55.1 52.6 68.5 55.2 73.5 53.7 61.7
47.9 72.3 61.1 71.8 83.1 71.2 58.8 61.8 86.8 64.5 52.3 58.3
65.9 80.2 75.1 59.9 62.3 48.8 64.3 75.4



The figure suggests that the histogram, and therefore the population distribution, is well approximated by a normal distribution with mean $\boldsymbol{\mu} = \mathbf{68.4}$ and standard deviation $\boldsymbol{\sigma} = \mathbf{9.6}$ pounds. Using the normal population distribution,

(a) Find the probability of using **80 or more** pounds of cheese.

(b) Set a limit so that only **10 % of production** runs have less than $\boldsymbol{L}$ **pounds** of cheese.

(c) Determine a new mean for the distribution so that only **5 % of the runs** have less than $\boldsymbol{L}$ **pounds**.

Solution (a) $Z = (X - 68.4)/9.6$ and, from Table 3, we get

$$1 - F\left(\frac{80 - 68.4}{9.6}\right) = 1 - F(1.208) = 1 - .8865 = .1135$$

About 1 out of 9 production runs will result in more than 80 pounds of cheese.

(b) From Table 3, the entry with probability closest to .1000 is $z_{0.10} = 1.28$. The limit L is given by

$$L = \mu - \sigma z_{0.10} = 68.4 - 9.6 \times 1.28 = 56.1 \text{ pounds}$$

(c) The new value of the mean μ must satisfy

$$-z_{.05} = \frac{L - \mu}{9.6}$$

where $z_{0.05} = 1.645$ so

$$\mu = L + 9.6 \times z_{0.05} = 56.1 + 9.6 \times 1.645 = 71.9 \text{ pounds}$$

The mean must be increased by 3.5 pounds to decrease the percentage of units below the limit L from 10% to 5%.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	$\boxed{0.8997}$	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	$\boxed{0.9495}$	$\boxed{0.9505}$	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633