

Published in "Random Walks in Biology", 1983, Princeton University Press

PROBABILITIES AND PROBABILITY DISTRIBUTIONS

Howard C. Berg

Table of Contents

-  [PROBABILITIES](#)
-  [PROBABILITY DISTRIBUTIONS](#)
-  [THE BINOMIAL DISTRIBUTION](#)
-  [THE GAUSSIAN DISTRIBUTION](#)
-  [THE POISSON DISTRIBUTION](#)

Next

[Next](#)[Contents](#)

1. PROBABILITIES

Classical or *a priori* probabilities are defined in terms of the possible outcomes of a trial, recognized in advance as equally probable. In the toss of a coin, the probability of getting a head is $1/2$: the number of outcomes that give a head, 1, divided, by the total number of possible outcomes, head or tail, 2. In the toss of a die, the probability of getting one dot is $1/6$: the number of outcomes that give one dot, 1, divided by the total number of possible outcomes, one through six dots, 6. In general, the probability of event a is

$$p(a) = \frac{a}{n}. \quad (\text{A.1})$$

where a is the number of equally probable outcomes that satisfy criteria a , and n is the total number of equally probable outcomes.

In the examples just given, the outcomes are *mutually exclusive*; i.e., only one outcome is possible at a time. If events a and b are mutually exclusive, then

$$p(a \text{ or } b) = p(a) + p(b), \quad (\text{A.2})$$

and

$$p(\text{not } a) = 1 - p(a). \quad (\text{A.3})$$

For the toss of a coin, $p(\text{head or tail}) = p(\text{head}) + p(\text{tail}) = 1$, and $p(\text{not head}) = 1 - p(\text{head}) = 1/2$. For the toss of a die, $p(1 \text{ dot or } 2 \text{ dots}) = p(1 \text{ dot}) + p(2 \text{ dots}) = 1/3$, and $p(\text{not } 1 \text{ dot}) = 1 - p(1 \text{ dot}) = 5/6$.

In these examples, the outcomes also are *statistically independent*; i.e., the occurrence of one event does not affect that of another. If events a and b are statistically independent, then

$$p(a \text{ and } b) = p(a)p(b). \quad (\text{A.4})$$

The probability of obtaining heads in each of two tosses of a coin is $(1/2)(1/2) = 1/4$. The probability of obtaining a single dot in each of two tosses of a die is $(1/6)(1/6) = 1/36$.

Events are *conditional* if the probability of one event depends on the occurrence of another. If the probability that b will occur, given that a has occurred, is $p(b/a)$, then the probability that both will occur

is

$$p(a \text{ and } b) = p(a)p(b/a). \quad (\text{A.5})$$

For example, the probability of drawing two aces from a deck of cards is $(4/52)(3/51) = 1/221$. If the first card were put back into the deck and the deck reshuffled, then the probability of drawing the second ace would not be conditioned on the drawing of the first, and the probability would be $(4/52)(4/52) = 1/169$, in accord with Eq. A.4.

Here are some rules of thumb for dealing with *compound events*, when the events are both mutually exclusive and statistically independent:

$$\begin{aligned} p(\text{neither } a \text{ nor } b) &= p(\text{not } a \text{ and not } b) \\ &= [1 - p(a)][1 - p(b)] \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} p(\text{either } a \text{ or } b \text{ but not both}) &= p(a \text{ and not } b, \text{ or not } a \text{ and } b) \\ &= p(a)[1 - p(b)] + [1 - p(a)]p(b) \end{aligned} \quad (\text{A.7})$$

$$p(\text{either } a \text{ or } b \text{ or both}) = 1 - p(\text{neither } a \text{ nor } b). \quad (\text{A.8})$$

[Next](#)
[Contents](#)

[Next](#)[Contents](#)[Previous](#)

2. PROBABILITY DISTRIBUTIONS

Suppose we were to toss an unbiased coin 4 times in succession. What is the probability, $P(k)$, of obtaining k heads? There are 16 different ways the coins might land; each is equally probable. Let's write down all 16 but group them according to how many heads appear, using the binary notation 1 = heads, 0 = tails:

<i>Sequence</i>	<i>No. heads</i>	<i>$P(k)$</i>
0000	0	$P(0) = 1/16$
1000	1	
0100	1	
0010	1	
0001	1	$P(1) = 4/16 = 1/4$
1100	2	
1010	2	
1001	2	
0110	2	$P(2) = 6/16 = 3/8$
0101	2	
0011	2	
1110	3	
1101	3	$P(3) = 4/16 = 1/4$
1011	3	
0111	3	
1111	4	$P(4) = 1/16$

From Eq. A. 1, the probability of a given sequence is $1/16$. From Eq. A.2, the probability of obtaining the sequence 1000, 0100, 0010, or 0001, i.e., a sequence in which 1 head appears, is $1/16 + 1/16 + 1/16 +$

$1/16 = 1/4$. The third column lists these probabilities, $P(k)$.

The results of these calculations can be summarized by plotting $P(k)$ as a function of k , as shown in [Fig. A.1](#). Such a plot is called a theoretical probability distribution. The peak value, $k = 2$, is the *most probable* value. Since the curve is symmetric about $k = 2$, this value also must be the mean value. One can compute the *mean* or *expectation* value of k by adding the number of heads obtained for each of the sequences shown in the table and dividing by 16, or - and this is the same thing - by weighting each possible value of k by the probability of obtaining that value, $P(k)$, and computing the sum:

$$\langle k \rangle = \sum_{k=0}^4 kP(k). \quad (\text{A.9})$$

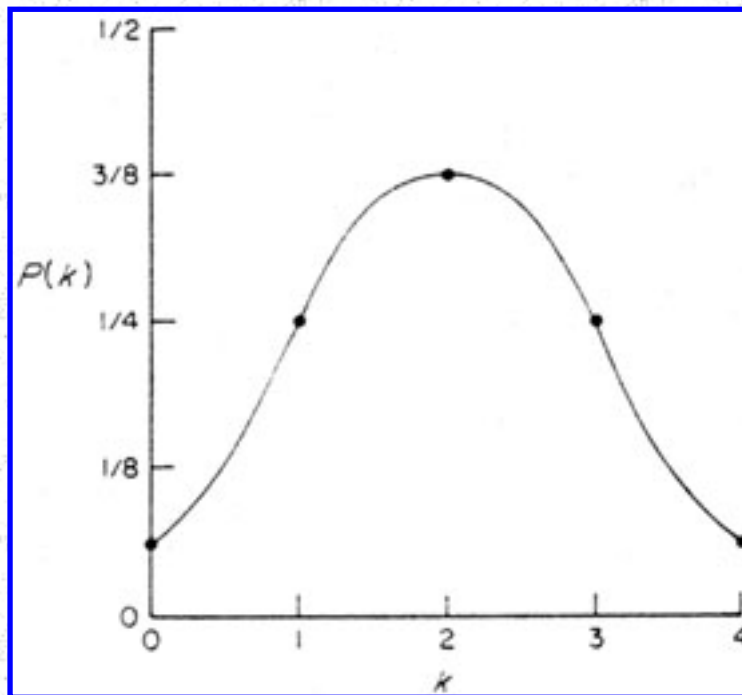


Figure A.1. The probability, $P(k)$, of obtaining k heads in 4 tosses of an unbiased coin. This theoretical probability distribution is *discrete*; it is defined only for the integer values $k = 0, 1, 2, 3$, or 4 . A smooth line is drawn through the points only to make it easier to visualize the trend. The distribution has a mean value 2 and a standard deviation 1.

The value of this sum is $(0)(1/16) + (1)(1/4) + (2)(3/8) + (3)(1/4) + (4)(1/16) = 0 + 1/4 + 3/4 + 3/4 + 1/4 = 2$, as expected. Note that

$$\sum_{k=0}^4 P(k) = 1. \quad (\text{A.10})$$

The probability of obtaining 0, 1, 2, 3, or 4 heads is 1. The distribution is properly *normalized*.

Note also that if a is a constant (does not depend on the variable k),

$$\langle a \rangle = \sum_{k=0}^4 a P(k) = a \sum_{k=0}^4 P(k) = a, \quad (\text{A.11})$$

and

$$\langle ak \rangle = \sum_{k=0}^4 ak P(k) = a \sum_{k=0}^4 k P(k) = a \langle k \rangle. \quad (\text{A.12})$$

It is useful to have some measure of the width or spread of a distribution about its mean. One might compute $\langle k - \mu \rangle$, the expectation value of the deviation of k from the mean $\mu = \langle k \rangle$, but the answer always comes out 0. It makes more sense to compute the expectation value of the square of the deviation, namely

$$\sigma^2 = \langle (k - \mu)^2 \rangle = \langle k^2 - 2\mu k + \mu^2 \rangle. \quad (\text{A.13})$$

This quantity is called the *variance*. Its square root, σ , is called the *standard deviation*. Since $\mu = \langle k \rangle$ and $\mu^2 = \langle k \rangle^2$ are constants, Eq. A.13 can be simplified. It follows from Eqs. A.11 and A.12 that

$$\sigma^2 = \langle k^2 \rangle - 2\mu \langle k \rangle + \mu^2 = \langle k^2 \rangle - \langle k \rangle^2, \quad (\text{A.14})$$

where

$$\langle k^2 \rangle = \sum_{k=0}^4 k^2 P(k). \quad (\text{A.15})$$

For the distribution of [Fig. A.1](#), $\langle k^2 \rangle = (0)(1/16) + (1)(1/4) + (4)(3/8) + (9)(1/4) + (16)(1/16) = 0 + 1/4 + 6/4 + 9/4 + 1 = 5$, $\langle k \rangle^2 = 4$, and $\sigma^2 = 5 - 4 = 1$.

It is instructive to sit down and actually flip a coin 4 times in succession, count the number of heads, and then repeat the experiment a large number of times. One can then construct an *experimental* probability distribution, with $P(0)$ equal to the number of experiments that give 0 heads divided by the total number of experiments, $P(1)$ equal to the number of experiments that give 1 head divided by the total number of experiments, etc. Two such distributions are shown in [Fig. A.2](#). In the first (x), the total number of

experiments was 10. In the second (o), the total number was 100.

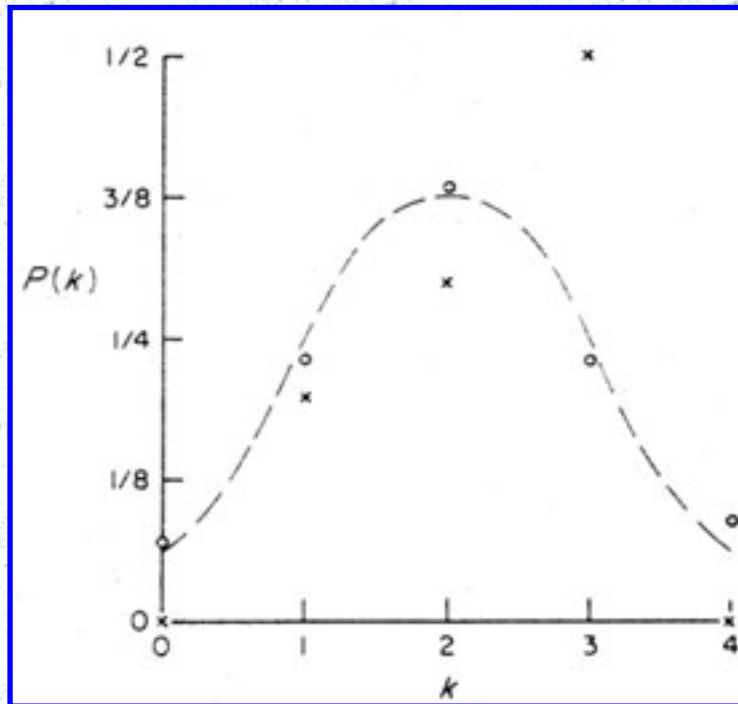


Figure A.2. Two experimental probability distributions. In one (x), a coin was flipped 4 times in 10 successive experiments; the mean value is 2.30, the standard deviation is 1.22. In the other (o), a coin was flipped 4 times in 100 successive experiments; the mean value is 2.04, the standard deviation 1.05. The dashed curve is the theoretical probability distribution of [Fig. A.1](#).

If you do not like flipping coins, you can build a probability machine of the sort shown in [Fig. A.3](#). If the machine is level and well made, the probability that a ball bounces to the right or the left on striking the next pin is $1/2$. The number of successive trials is equal to the number of rows of pins. If you drop 100 balls through this machine, they will pile up in the bins at the bottom, forming a distribution like the one shown in [Fig. A.2](#). Or do the experiments on a computer: ask for a random number uniformly distributed between 0 and 1; if the number is less than or equal to $1/2$, call it a head; if it is greater than $1/2$, call it a tail. The data shown in [Fig. A.2](#) were generated in this way.

The theoretical expectations are more closely met the larger the number of samples. What is the likelihood that the deviations between the experimental distributions and the theoretical distribution, evident in [Fig. A.2](#), occur by chance? By how much are the mean values of the experimental distributions likely to differ from the mean value predicted by the theory? Questions of this kind often are encountered in the laboratory. They are dealt with in books on data reduction and error analysis and will not be pursued here.

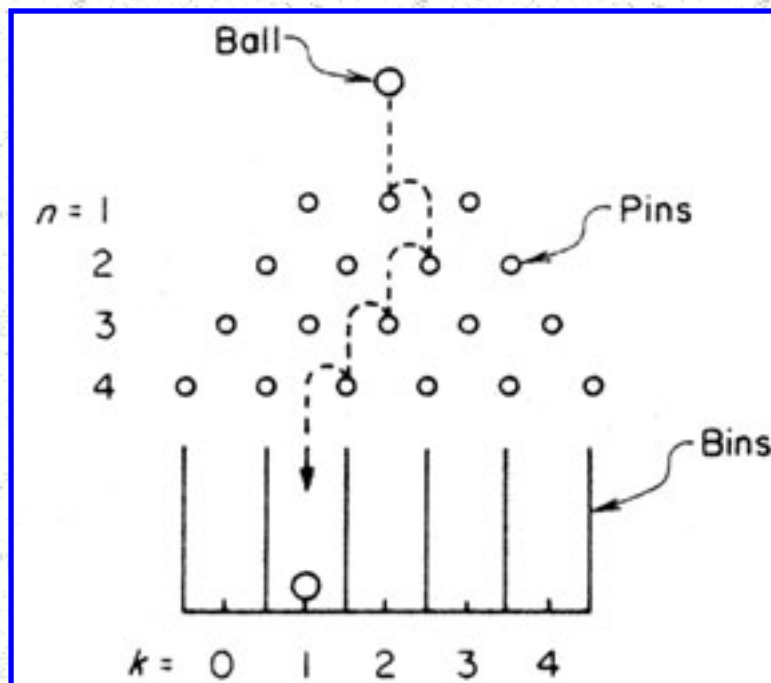


Figure A.3. A probability machine of 4 rows. A ball dropped on the center pin of the top row bounces to the right or left and executes a random walk as it moves from row to row, arriving finally at one of the bins at the bottom. The rows and bins are numbered $n = 1, 2, 3, 4$ and $k = 0, 1, 2, 3, 4$, respectively; n is the bounce number and k is the number of bounces made to the right.

[Next](#)
[Contents](#)
[Previous](#)

[Next](#)[Contents](#)[Previous](#)

3. THE BINOMIAL DISTRIBUTION

What if we flip a biased coin, with the probability of a head p and the probability of a tail $q = 1 - p$? The probability of a given sequence, e.g., 100010 ..., in which k heads appear in n flips is, by Eq. A.4, $pqqqpq$..., or

$$p^k q^{n-k}. \quad (\text{A.16})$$

There are a total of 2^n possible sequences. Only some of these give k heads and $n - k$ tails. Their number is

$$\frac{n!}{k!(n-k)!} \equiv \binom{n}{k}, \quad (\text{A.17})$$

where $0!$, whenever it appears in the denominator, is understood to be 1.

Since any one or another of these sequences will do, the probability that exactly k heads occur in n flips is, by Eq. A.2.

$$P(k; n, p) = \binom{n}{k} p^k q^{n-k}. \quad (\text{A.18})$$

This is the binomial distribution. The coefficient $\binom{n}{k}$ is the binomial coefficient, the number of combinations of n things taken k and $n - k$ at a time. You have seen it before in algebra in the binomial theorem:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (\text{A.19})$$

We can use the binomial theorem to show that the binomial distribution is normalized:

$$\sum_{k=0}^n P(k; n, p) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1^n = 1. \quad (\text{A.20})$$

As an example, let's work out the case of 4 flips of an unbiased coin. If $p = q = 1/2$, then $p^k q^{n-k} = (1/2)^n = (1/2)^4 = 1/16$ for all values of k , and the probabilities $P(0;4,1/2)$, ..., $P(4;4,1/2)$ are equal to the binomial coefficients $\binom{4}{0}$, ..., $\binom{4}{4}$ times this factor. Since

$$\binom{4}{0} = \binom{4}{4} = \frac{4!}{0!4!} = 1,$$

$$\binom{4}{1} = \binom{4}{3} = \frac{4!}{1!3!} = 4$$

and

$$\binom{4}{2} = \frac{4!}{2!2!} = 6,$$

we obtain the probabilities 1/16, 1/4, 3/8, 1/4, and 1/16, as before.

The expectation value of k is

$$\langle k \rangle = \sum_{k=0}^n k P(k; n, p) = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k}. \quad (\text{A.21})$$

To evaluate this, note that the $k = 0$ term is 0 and that $k/k! = 1/(k-1)!$, so that

$$\langle k \rangle = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k}.$$

Next, factor out np :

$$\langle k \rangle = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k}.$$

Finally, change variables by substituting $m = k - 1$ and $s = n - 1$:

$$\begin{aligned}
 \langle k \rangle &= np \sum_{m=0}^s \frac{s!}{m!(s-m)!} p^m q^{s-m} \\
 &= np \sum_{m=0}^s \binom{s}{m} p^m q^{s-m}.
 \end{aligned}$$

The sum in this expression is the same as the one in Eq. A.20; only the labels have been changed. Thus,

$$\langle k \rangle = np. \quad (\text{A.22})$$

One can evaluate the expectation value of k^2 in a similar fashion by two successive changes in variables and show that

$$\langle k^2 \rangle = (np)^2 + npq. \quad (\text{A.23})$$

The variance of k , Eq. A.14, is

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2 = npq, \quad (\text{A.24})$$

and its standard deviation is

$$\sigma_k = (npq)^{1/2}. \quad (\text{A.25})$$

An example of the binomial distribution is given in [Fig. A.4](#), which shows the theoretical distribution $P(k;10,1/6)$. This is the probability of obtaining a given side k times in 10 throws of a die.

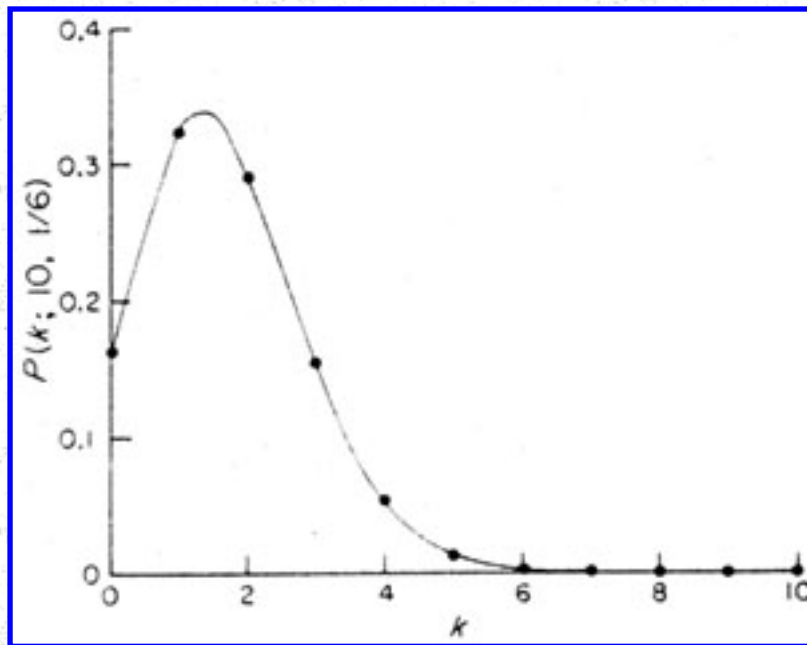


Figure A.4. The binomial distribution for $n = 10$, $p = 1/6$. The mean value is 1.67, the standard deviation 1.18.

[Next](#)[Contents](#)[Previous](#)

[Next](#)[Contents](#)[Previous](#)

4. THE GAUSSIAN DISTRIBUTION

In many of the problems dealt with in this book, the number of trials, n , is very large. A small molecule undergoing diffusion, for example, steps to the right or left millions of times in a microsecond, not 4 times in a few seconds, as the ball in the apparatus of [Fig. A.3](#). There are two asymptotic limits of the binomial distribution. One, the Gaussian, or normal, distribution, is obtained when the probability of a success, p , is finite, i.e., if $np \rightarrow \infty$ as $n \rightarrow \infty$. The other, the Poisson distribution, is obtained if p is very small, so small that np remains finite as $n \rightarrow \infty$.

The derivation of the Gaussian distribution involves the use of Stirling's approximation for the factorials of the binomial coefficients:

$$n! \simeq (2\pi n)^{1/2} (n/e)^n, \quad (\text{A.26})$$

where e is the base of the natural logarithms. The result is

$$P(k; n, p) \rightarrow P(k; \mu, \sigma) dk = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(k-\mu)^2/2\sigma^2} dk, \quad (\text{A.27})$$

where $\mu = \langle k \rangle = np$ and $\sigma = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = (npq)^{1/2}$, as before. $P(k; \mu, \sigma) dk$ is the probability that k will be found between k and $k + dk$, where dk is infinitesimal. The distribution is continuous rather than discrete. Expectation values are found by taking integrals rather than sums. The distribution is symmetric about the mean, μ , and its width is determined by σ . The area of the distribution is 1, so its height is inversely proportional to σ .

If we define $u = (k - \mu) / \sigma$, i.e., plot the distribution with the abscissa in units of σ and the origin at μ , then

$$P(k) dk \rightarrow P(u) du = \frac{1}{(2\pi)^{1/2}} e^{-u^2/2} du. \quad (\text{A.28})$$

$P(u)$ is called the normal curve of error; it is shown in [Fig. A.5](#). As an exercise, use your tables of definite integrals and show that

$$\int_{-\infty}^{\infty} P(u) du = 1, \quad (\text{A.29})$$

$$\int_{-\infty}^{\infty} u P(u) du = 0, \quad (\text{A.30})$$

and

$$\int_{-\infty}^{\infty} u^2 P(u) du = 1. \quad (\text{A.31})$$

Eq. A.30 can be done by inspection: $P(u)$ is an even function of u , so $uP(u)$ must be an odd function of u . The distribution $P(u)$ is normalized, its mean value is 0, and its variance and standard deviation are 1.

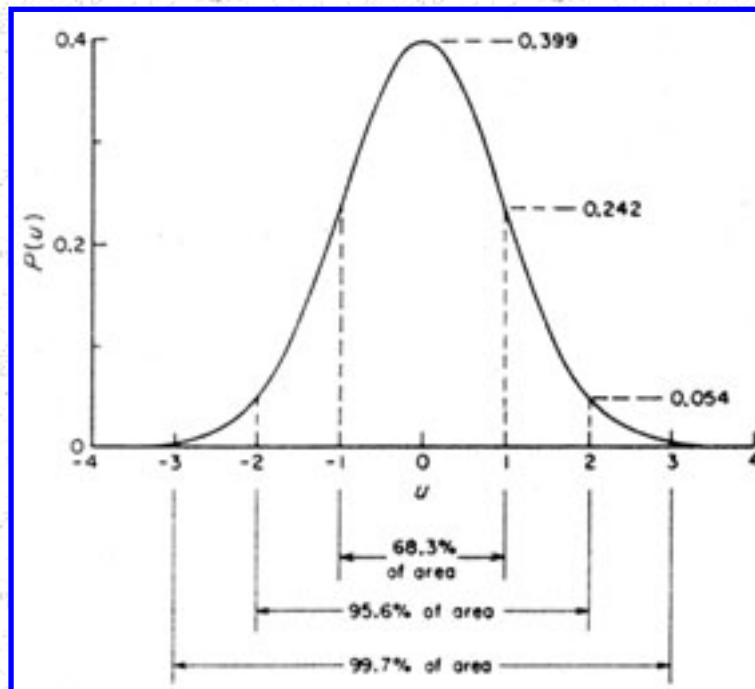


Figure A.5. The normal curve of error: the Gaussian distribution plotted in units of the standard deviation σ with its origin at the mean value μ . The area under the curve is 1. Half the area falls between $u = \pm 0.67$.

[Next](#)
[Contents](#)
[Previous](#)

[Contents](#)
[Previous](#)

5. THE POISSON DISTRIBUTION

As noted above, the Poisson distribution is obtained as an asymptotic limit of the binomial distribution when p is very small. The result is

$$P(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}, \quad (\text{A.32})$$

where $\mu = np$, as before, and $0!$ is understood to be 1. This distribution is determined by one rather than two constants: $\sigma = (npq)^{1/2}$, but $q = 1 - p \approx 1$, so $\sigma = (np)^{1/2} = \mu^{1/2}$. The standard deviation is equal to the square-root of the mean. The Poisson distribution is discrete: $P(0; \mu) = e^{-\mu}$ is the probability of 0 successes, given that the mean number of successes is μ , etc. The probability of 1 or more successes is $1 - P(0; \mu) = 1 - e^{-\mu}$. The distribution $P(k; 1.67)$ is shown in [Fig. A.6](#).

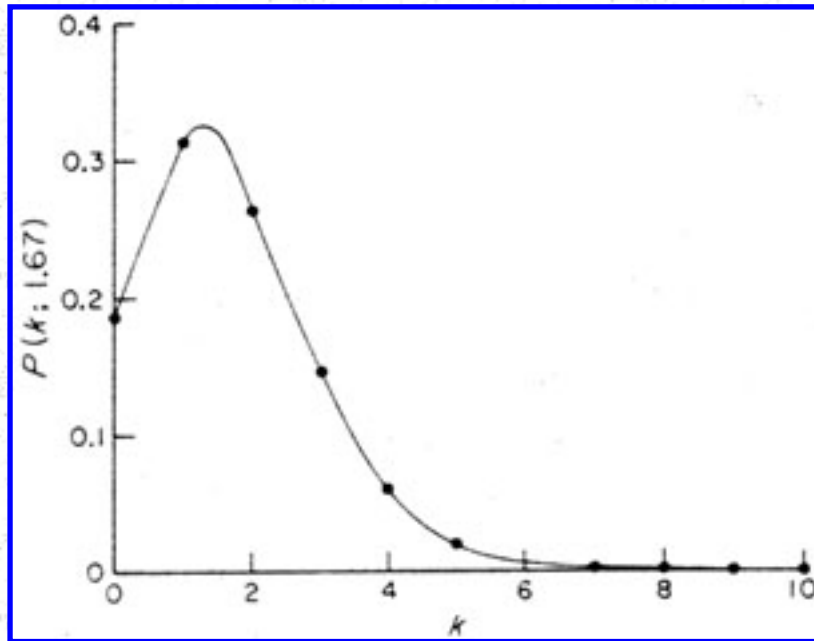


Figure A.6. The Poisson distribution $P(k; 1.67)$. The mean value is 1.67, the standard deviation 1.29. The curve is similar to the binomial distribution shown in [Fig. A.4](#), but it is defined for values of $k > 10$. For example, $P(20; 1.67) = 2.2 \times 10^{-15}$.

[Contents](#)
[Previous](#)