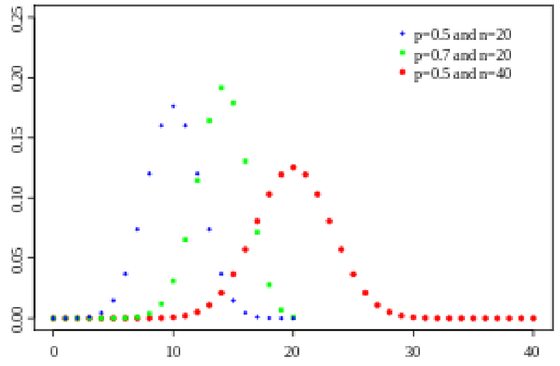
Binomial distribution

In probability theory and statistics, the **binomial distribution** with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking a yes–no question, and each with its own boolean-valued outcome: success/yes/true/one (with probability p) or failure/no/false/zero (with probability $q = 1 - p$). A single success/failure experiment is also called a Bernoulli trial or Bernoulli experiment and a sequence of outcomes is called a Bernoulli process; for a single trial, i.e., $n = 1$, the binomial distribution is a Bernoulli distribution. The binomial distribution is the basis for the popular binomial test of statistical significance.

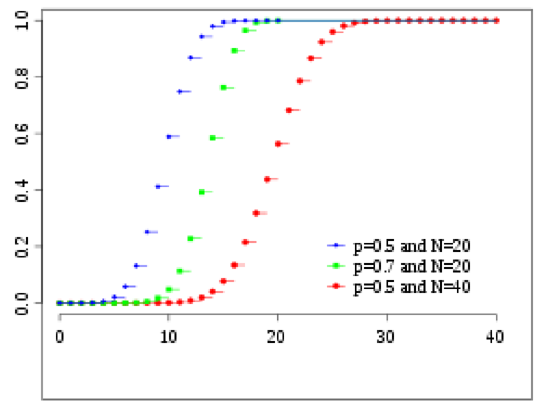
The binomial distribution is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N. If the sampling is carried out without replacement, the draws are not independent and so the resulting distribution is a hypergeometric distribution, not a binomial one. However, for N much larger than n, the binomial distribution remains a good approximation, and is widely used.

Binomial distribution

Probability mass function



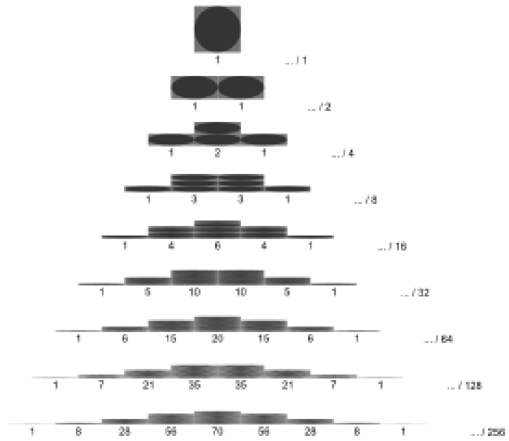
Cumulative distribution function



Notation	$B(n, p)$
Parameters	$n \in \{0, 1, 2, \ldots\}$ – number of trials $p \in [0, 1]$ – success probability for each trial $q = 1 - p$
Support	$k \in \{0, 1, \ldots, n\}$ – number of successes
pmf	$\binom{n}{k} p^k q^{n-k}$
CDF	$I_q(n - k, 1 + k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	npq
Skewness	$\frac{q - p}{\sqrt{npq}}$
Ex. kurtosis	$\frac{1 - 6pq}{npq}$
Entropy	

Contents

	$\frac{1}{2}\log_2(2\pi enpq) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(q + pe^t)^n$
CF	$(q + pe^{it})^n$
PGF	$G(z) = [q + pz]^n$
Fisher information	$g_n(p) = \frac{n}{pq}$ (for fixed n)



Binomial distribution for $p = 0.5$ with *n* and *k* as in Pascal's triangle

The probability that a ball in a Galton box with 8 layers (*n* = 8) ends up in the central bin (*k* = 4) is $70/256$.

Definitions

Probability mass function

In general, if the random variable *X* follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, we write $X \sim B(n, p)$. The probability of getting exactly *k* successes in *n* independent Bernoulli trials is given by the probability mass function:

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for *k* = 0, 1, 2, ..., *n*, where

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

is the binomial coefficient, hence the name of the distribution. The formula can be understood as follows. *k* successes occur with probability p^k and $n - k$ failures occur with probability $(1 - p)^{n - k}$. However, the *k* successes can occur anywhere among the *n* trials, and there are $\binom{n}{k}$ different ways of distributing *k* successes in a sequence of *n* trials.

In creating reference tables for binomial distribution probability, usually the table is filled in up to *n*/2 values. This is because for $k > n/2$, the probability can be calculated by its complement as

$$f(k,n,p) = f(n-k,n,1-p).$$

Looking at the expression *f*(*k*, *n*, *p*) as a function of *k*, there is a *k* value that maximizes it. This *k* value can be found by calculating

$$\frac{f(k+1,n,p)}{f(k,n,p)}=\frac{(n-k)p}{(k+1)(1-p)}$$

and comparing it to 1. There is always an integer M that satisfies[1]

$$(n+1)p-1\leq M<(n+1)p.$$

$f(k, n, p)$ is monotone increasing for $k < M$ and monotone decreasing for $k > M$, with the exception of the case where $(n + 1)p$ is an integer. In this case, there are two values for which f is maximal: $(n + 1)p$ and $(n + 1)p - 1$. M is the *most probable* outcome (that is, the most likely, although this can still be unlikely overall) of the Bernoulli trials and is called the mode.

Cumulative distribution function

The cumulative distribution function can be expressed as:

$$F(k;n,p)=\Pr(X\leq k)=\sum_{i=0}^{\lfloor k\rfloor}\binom{n}{i}p^i(1-p)^{n-i}$$

where $\lfloor k\rfloor$ is the "floor" under k, i.e. the greatest integer less than or equal to k.

It can also be represented in terms of the regularized incomplete beta function, as follows:[2]

$$\begin{aligned}F(k;n,p)&=\Pr(X\leq k)\\&=I_{1-p}(n-k,k+1)\\&=(n-k)\binom{n}{k}\int_0^{1-p}t^{n-k-1}(1-t)^k\,dt.\end{aligned}$$

which is equivalent to the cumulative distribution function of the F-distribution[3]

$$F(k;n,p)=F_{F\text{-}distribution}\left(x=\frac{1-p}{p}\frac{k+1}{n-k};d_1=2(n-k),d_2=2(k+1)\right).$$

Some closed-form bounds for the cumulative distribution function are given below.

Examples

Suppose a biased coin comes up heads with probability 0.3 when tossed. What is the probability of achieving 0, 1,..., 6 heads after six tosses?

$$\Pr(0\text{ heads})=f(0)=\Pr(X=0)=\binom{6}{0}0.3^0(1-0.3)^{6-0}=0.117649$$

$$\Pr(1\text{ heads})=f(1)=\Pr(X=1)=\binom{6}{1}0.3^1(1-0.3)^{6-1}=0.302526$$

$$\Pr(2\text{ heads})=f(2)=\Pr(X=2)=\binom{6}{2}0.3^2(1-0.3)^{6-2}=0.324135$$

$$\Pr(3\text{ heads})=f(3)=\Pr(X=3)=\binom{6}{3}0.3^3(1-0.3)^{6-3}=0.18522$$

$$\Pr(4\text{ heads})=f(4)=\Pr(X=4)=\binom{6}{4}0.3^4(1-0.3)^{6-4}=0.059535$$

$$\Pr(5\text{ heads}) = f(5) = \Pr(X=5) = \binom{6}{5}0.3^5(1-0.3)^{6-5} = 0.010206$$

$$\Pr(6\text{ heads}) = f(6) = \Pr(X=6) = \binom{6}{6}0.3^6(1-0.3)^{6-6} = 0.000729$$[4]

Properties

Expected value and variance

If $X \sim B(n, p)$, that is, X is a binomially distributed random variable, n being the total number of experiments and p the probability of each experiment yielding a successful result, then the expected value of X is:[5]

$$\operatorname{E}[X] = np.$$

This follows from the linearity of the expected value along with fact that X is the sum of n identical Bernoulli random variables, each with expected value p. In other words, if $X_1, \ldots, X_n$ are identical (and independent) Bernoulli random variables with parameter p, then $X = X_1 + \cdots + X_n$ and

$$\operatorname{E}[X] = \operatorname{E}[X_1 + \cdots + X_n] = \operatorname{E}[X_1] + \cdots + \operatorname{E}[X_n] = p + \cdots + p = np.$$

The variance is:

$$\operatorname{Var}(X) = np(1-p).$$

This similarly follows from the fact that the variance of a sum of independent random variables is the sum of the variances.

Higher moments

The first 6 central moments are given by

$$\begin{aligned}
\mu_1 &= 0,\\
\mu_2 &= np(1-p),\\
\mu_3 &= np(1-p)(1-2p),\\
\mu_4 &= np(1-p)(1+(3n-6)p(1-p)),\\
\mu_5 &= np(1-p)(1-2p)(1+(10n-12)p(1-p)),\\
\mu_6 &= np(1-p)(1-30p(1-p)(1-4p(1-p))+5np(1-p)(5-26p(1-p))+15n^2p^2(1-p)^2).
\end{aligned}$$

Mode

Usually the mode of a binomial $B(n, p)$ distribution is equal to $\lfloor (n+1)p \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. However, when $(n + 1)p$ is an integer and p is neither 0 nor 1, then the distribution has two modes: $(n + 1)p$ and $(n + 1)p - 1$. When p is equal to 0 or 1, the mode will be 0 and n correspondingly. These cases can be summarized as follows:

$$\text{mode} = \begin{cases} \lfloor (n+1)\,p \rfloor & \text{if } (n+1)p \text{ is 0 or a noninteger},\\ (n+1)\,p\ \text{ and }\ (n+1)\,p-1 & \text{if } (n+1)p \in \{1,\ldots,n\},\\ n & \text{if } (n+1)p = n+1. \end{cases}$$

Proof: Let

$$f(k) = \binom{n}{k} p^k q^{n-k}.$$

For $p = 0$ only $f(0)$ has a nonzero value with $f(0) = 1$. For $p = 1$ we find $f(n) = 1$ and $f(k) = 0$ for $k \neq n$. This proves that the mode is 0 for $p = 0$ and n for $p = 1$.

Let $0 < p < 1$. We find

$$\frac{f(k+1)}{f(k)} = \frac{(n-k)p}{(k+1)(1-p)}.$$

From this follows

$$\begin{aligned} k > (n+1)p - 1 &\Rightarrow f(k+1) < f(k) \\ k = (n+1)p - 1 &\Rightarrow f(k+1) = f(k) \\ k < (n+1)p - 1 &\Rightarrow f(k+1) > f(k) \end{aligned}$$

So when $(n+1)p - 1$ is an integer, then $(n+1)p - 1$ and $(n+1)p$ is a mode. In the case that $(n+1)p - 1 \notin \mathbb{Z}$, then only $\lfloor (n+1)p - 1 \rfloor + 1 = \lfloor (n+1)p \rfloor$ is a mode.[6]

Median

In general, there is no single formula to find the median for a binomial distribution, and it may even be non-unique. However several special results have been established:

- If np is an integer, then the mean, median, and mode coincide and equal np.[7][8]
- Any median m must lie within the interval $\lfloor np \rfloor \le m \le \lceil np \rceil$.[9]
- A median m cannot lie too far away from the mean: $|m - np| \le \min\{\ln 2, \max\{p, 1 - p\}\}$.[10]
- The median is unique and equal to m = round(np) when $|m - np| \le \min\{p, 1 - p\}$ (except for the case when $p = \frac{1}{2}$ and n is odd).[9]
- When $p = 1/2$ and n is odd, any number m in the interval $\frac{1}{2}(n - 1) \le m \le \frac{1}{2}(n + 1)$ is a median of the binomial distribution. If $p = 1/2$ and n is even, then $m = n/2$ is the unique median.

Tail bounds

For $k \le np$, upper bounds for the lower tail of the distribution function can be derived. Recall that $F(k; n, p) = \Pr(X \le k)$, the probability that there are at most k successes.

Hoeffding's inequality yields the bound

$$F(k; n, p) \le \exp\left(-2\frac{(np-k)^2}{n}\right),$$

and Chernoff's inequality can be used to derive the bound

$$F(k; n, p) \le \exp\left(-\frac{1}{2\,p}\frac{(np-k)^2}{n}\right).$$

Moreover, these bounds are reasonably tight when $p = 1/2$, since the following expression holds for all $k \ge 3n/8$[11]

$$F(k;n,\tfrac{1}{2}) \leq \frac{14}{15} \exp\left(-\frac{16(\frac{n}{2}-k)^2}{n}\right).$$

However, the bounds do not work well for extreme values of p. In particular, as $p \to 1$, value $F(k;n,p)$ goes to zero (for fixed k, n with $k < n$) while the upper bound above goes to a positive constant. In this case a better bound is given by [12]

$$F(k;n,p) \leq \exp\left(-nD\left(\frac{k}{n} \parallel p\right)\right) \qquad \text{if } 0 < \frac{k}{n} < p$$

where $D(a \parallel p)$ is the relative entropy between an a-coin and a p-coin (i.e. between the Bernoulli(a) and Bernoulli(p) distribution):

$$D(a \parallel p) = (a)\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p}.$$

Asymptotically, this bound is reasonably tight; see [12] for details. An equivalent formulation of the bound is

$$\Pr(X \geq k) = F(n-k;n,1-p) \leq \exp\left(-nD\left(\frac{k}{n} \parallel p\right)\right) \qquad \text{if } p < \frac{k}{n} < 1.$$

Both these bounds are derived directly from the Chernoff bound. It can also be shown that,

$$\Pr(X \geq k) = F(n-k;n,1-p) \geq \frac{1}{(n+1)^2} \exp\left(-nD\left(\frac{k}{n} \parallel p\right)\right) \qquad \text{if } p < \frac{k}{n} < 1.$$

This is proved using the method of types (see for example chapter 11 of *Elements of Information Theory* by Cover and Thomas [13]).

We can also change the $(n+1)^2$ in the denominator to $\sqrt{2n}$, by approximating the binomial coefficient with Stirling's formula.[14]

Statistical Inference

Estimation of parameters

When n is known, the parameter p can be estimated using the proportion of successes: $\widehat{p} = \frac{x}{n}$. This estimator is found using maximum likelihood estimator and also the method of moments. This estimator is unbiased and uniformly with minimum variance, proven using Lehmann–Scheffé theorem, since it is based on a minimal sufficient and complete statistic (i.e.: x). It is also consistent both in probability and in MSE.

A closed form Bayes estimator for p also exists when using the Beta distribution as a conjugate prior distribution. When using a general $Beta(\alpha,\beta)$ as a prior, the posterior mean estimator is: $\widehat{p_b} = \frac{x+\alpha}{n+\alpha+\beta}$. The Bayes estimator is asymptoticly efficient and as the sample size approaches infinity ($n \to \infty$), it approaches the MLE solution. The Bayes estimator is biased (how much depends on the priors), admissible and consistent in probability.

For the special case of using the standard uniform distribution as a non-informative prior ($Beta(\alpha=1,\beta=1)=U(0,1)$), the posterior mean estimator becomes $\widehat{p_b} = \frac{x+1}{n+2}$ (a posterior mode should just lead to the standard estimator). This method is called the rule of succession, which was introduced in the 18th century by Pierre-Simon Laplace.

When estimating p with very rare events and a small n (e.g.: if x=0), then using the standard estimator leads to $\widehat{p}=0$, which sometimes is unrealistic and undesirable. In such cases there are various alternative estimators.[15] One way is to use the Bayes estimator, leading to: $\widehat{p_b}=\frac{1}{n+2}$). Another method is to use the upper bound of the confidence interval obtained using the rule of three: $\widehat{p_{rule\ of\ 3}}=\frac{3}{n}$)

Confidence intervals

Even for quite large values of n, the actual distribution of the mean is significantly nonnormal.[16] Because of this problem several methods to estimate confidence intervals have been proposed.

In the equations for confidence intervals below, the variables have the following meaning:

- n_1 is the number of successes out of n, the total number of trials
- $\widehat{p}=\frac{n_1}{n}$ is the proportion of successes
- z is the $1-\frac{1}{2}\alpha$ quantile of a standard normal distribution (i.e., probit) corresponding to the target error rate α. For example, for a 95% confidence level the error α = 0.05, so $1-\frac{1}{2}\alpha$ = 0.975 and z = 1.96.

Wald method

$$\widehat{p}\pm z\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

A continuity correction of 0.5/n may be added.

Agresti–Coull method

[17]

$$\tilde{p}\pm z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+z^2}}.$$

Here the estimate of p is modified to

$$\tilde{p}=\frac{n_1+\frac{1}{2}z^2}{n+z^2}$$

Arcsine method

[18]

$$\sin^2\left(\arcsin\left(\sqrt{\widehat{p}}\right)\pm\frac{z}{2\sqrt{n}}\right).$$

Wilson (score) method

The notation in the formula below differs from the previous formulas in two respects:[19]

- Firstly, z_x has a slightly different interpretation in the formula below: it has its ordinary meaning of 'the xth quantile of the standard normal distribution', rather than being a shorthand for 'the $(1 - x)$-th quantile'.
- Secondly, this formula does not use a plus-minus to define the two bounds. Instead, one may use $z = z_{\alpha/2}$ to get the lower bound, or use $z = z_{1-\alpha/2}$ to get the upper bound. For example: for a 95% confidence level the error α = 0.05, so one gets the lower bound by using $z = z_{\alpha/2} = z_{0.025} = -1.96$, and one gets the upper bound by using $z = z_{1-\alpha/2} = z_{0.975} = 1.96$.

$$\frac{\widehat{p} + \frac{z^2}{2n} + z\sqrt{\frac{\widehat{p}\,(1-\widehat{p}\,)}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}$$ [20]

Comparison

The exact (Clopper–Pearson) method is the most conservative.[16]

The Wald method, although commonly recommended in textbooks, is the most biased.

Related distributions

Sums of binomials

If $X \sim B(n, p)$ and $Y \sim B(m, p)$ are independent binomial variables with the same probability p, then $X + Y$ is again a binomial variable; its distribution is $Z=X+Y \sim B(n+m, p)$:

$$\begin{aligned}\operatorname{P}(Z=k) &= \sum_{i=0}^{k}\left[\binom{n}{i}p^{i}(1-p)^{n-i}\right]\left[\binom{m}{k-i}p^{k-i}(1-p)^{m-k+i}\right]\\ &= \binom{n+m}{k}p^{k}(1-p)^{n+m-k}\end{aligned}$$

However, if X and Y do not have the same probability p, then the variance of the sum will be smaller than the variance of a binomial variable distributed as $B(n+m, \bar{p})$.

Ratio of two binomial distributions

This result was first derived by Katz et al. in 1978.[21]

Let $X \sim B(n,p_1)$ and $Y \sim B(m,p_2)$ be independent. Let $T = (X/n)/(Y/m)$.

Then $\log(T)$ is approximately normally distributed with mean $\log(p_1/p_2)$ and variance $((1/p_1) - 1)/n + ((1/p_2) - 1)/m$.

Conditional binomials

If $X \sim B(n, p)$ and $Y \mid X \sim B(X, q)$ (the conditional distribution of Y, given X), then Y is a simple binomial random variable with distribution $Y \sim B(n, pq)$.

For example, imagine throwing n balls to a basket U_X and taking the balls that hit and throwing them to another basket U_Y. If p is the probability to hit U_X then $X \sim B(n, p)$ is the number of balls that hit U_X. If q is the probability to hit U_Y then the number of balls that hit U_Y is $Y \sim B(X, q)$ and therefore $Y \sim B(n, pq)$.

[Proof]

Since $X \sim B(n,p)$ and $Y \sim B(X,q)$, by the law of total probability,

$$\begin{aligned}\Pr[Y=m] &= \sum_{k=m}^{n} \Pr[Y=m \mid X=k] \Pr[X=k] \\ &= \sum_{k=m}^{n} \binom{n}{k}\binom{k}{m} p^k q^m (1-p)^{n-k}(1-q)^{k-m}\end{aligned}$$

Since $\binom{n}{k}\binom{k}{m} = \binom{n}{m}\binom{n-m}{k-m}$, the equation above can be expressed as

$$\Pr[Y=m] = \sum_{k=m}^{n} \binom{n}{m}\binom{n-m}{k-m} p^k q^m (1-p)^{n-k}(1-q)^{k-m}$$

Factoring $p^k = p^m p^{k-m}$ and pulling all the terms that don't depend on k out of the sum now yields

$$\begin{aligned}\Pr[Y=m] &= \binom{n}{m} p^m q^m \left(\sum_{k=m}^{n} \binom{n-m}{k-m} p^{k-m}(1-p)^{n-k}(1-q)^{k-m}\right) \\ &= \binom{n}{m}(pq)^m \left(\sum_{k=m}^{n} \binom{n-m}{k-m} (p(1-q))^{k-m}(1-p)^{n-k}\right)\end{aligned}$$

After substituting $i = k-m$ in the expression above, we get

$$\Pr[Y=m] = \binom{n}{m}(pq)^m \left(\sum_{i=0}^{n-m} \binom{n-m}{i} (p-pq)^i (1-p)^{n-m-i}\right)$$

Notice that the sum (in the parentheses) above equals $(p-pq+1-p)^{n-m}$ by the binomial theorem. Substituting this in finally yields

$$\begin{aligned}\Pr[Y=m] &= \binom{n}{m}(pq)^m (p-pq+1-p)^{n-m} \\ &= \binom{n}{m}(pq)^m (1-pq)^{n-m}\end{aligned}$$

and thus $Y \sim B(n,pq)$ as desired.

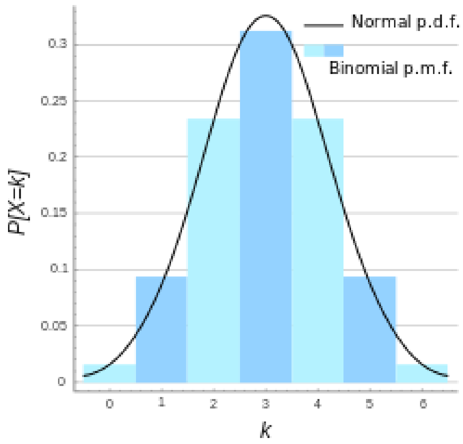
Bernoulli distribution

The Bernoulli distribution is a special case of the binomial distribution, where $n = 1$. Symbolically, $X \sim B(1, p)$ has the same meaning as $X \sim \text{Bernoulli}(p)$. Conversely, any binomial distribution, $B(n, p)$, is the distribution of the sum of n Bernoulli trials, Bernoulli(p), each with the same probability p.[22]

Poisson binomial distribution

The binomial distribution is a special case of the Poisson binomial distribution, or general binomial distribution, which is the distribution of a sum of n independent non-identical Bernoulli trials $B(p_i)$.[23]

Normal approximation



Binomial probability mass function and normal probability density function approximation for $n = 6$ and $p = 0.5$

If n is large enough, then the skew of the distribution is not too great. In this case a reasonable approximation to $B(n, p)$ is given by the normal distribution

$$\mathcal{N}(np,\, np(1-p)),$$

and this basic approximation can be improved in a simple way by using a suitable continuity correction. The basic approximation generally improves as n increases (at least 20) and is better when p is not near to 0 or 1.[24] Various rules of thumb may be used to decide whether n is large enough, and p is far enough from the extremes of zero or one:

- One rule[24] is that for $n > 5$ the normal approximation is adequate if the absolute value of the skewness is strictly less than 1/3; that is, if

$$\frac{|1-2p|}{\sqrt{np(1-p)}} = \frac{1}{\sqrt{n}} \left| \sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} \right| < \frac{1}{3}.$$

- A stronger rule states that the normal approximation is appropriate only if everything within 3 standard deviations of its mean is within the range of possible values; that is, only if

$$\mu \pm 3\sigma = np \pm 3\sqrt{np(1-p)} \in (0, n).$$

This 3-standard-deviation rule is equivalent to the following conditions, which also imply the first rule above.

$$n > 9\left(\frac{1-p}{p}\right) \quad \text{and} \quad n > 9\left(\frac{p}{1-p}\right).$$

[Proof]

The rule $np \pm 3\sqrt{np(1-p)} \in (0, n)$ is totally equivalent to request that

$$np - 3\sqrt{np(1-p)} > 0 \quad \text{and} \quad np + 3\sqrt{np(1-p)} < n.$$

Moving terms around yields:

$$np > 3\sqrt{np(1-p)} \quad \text{and} \quad n(1-p) > 3\sqrt{np(1-p)}.$$

Since $0 < p < 1$, we can apply the square power and divide by the respective factors np^2 and $n(1-p)^2$, to obtain the desired conditions:

$$n > 9\left(\frac{1-p}{p}\right) \quad \text{and} \quad n > 9\left(\frac{p}{1-p}\right).$$

Notice that these conditions automatically imply that $n > 9$. On the other hand, apply again the square root and divide by 3,

$$\frac{\sqrt{n}}{3} > \sqrt{\frac{1-p}{p}} > 0 \quad \text{and} \quad \frac{\sqrt{n}}{3} > \sqrt{\frac{p}{1-p}} > 0.$$

Subtracting the second set of inequalities from the first one yields:

$$\frac{\sqrt{n}}{3} > \sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} > -\frac{\sqrt{n}}{3};$$

and so, the desired first rule is satisfied,

$$\left|\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}}\right| < \frac{\sqrt{n}}{3}.$$

- Another commonly used rule is that both values np and $n(1-p)$ must be greater than or equal to 5. However, the specific number varies from source to source, and depends on how good an approximation one wants. In particular, if one uses 9 instead of 5, the rule implies the results stated in the previous paragraphs.

[Proof]

Assume that both values np and $n(1-p)$ are greater than 9. Since $0 < p < 1$, we easily have that

$$np \geq 9 > 9(1-p) \quad \text{and} \quad n(1-p) \geq 9 > 9p.$$

We only have to divide now by the respective factors p and $1-p$, to deduce the alternative form of the 3-standard-deviation rule:

$$n > 9\left(\frac{1-p}{p}\right) \quad \text{and} \quad n > 9\left(\frac{p}{1-p}\right).$$

The following is an example of applying a continuity correction. Suppose one wishes to calculate Pr($X \leq 8$) for a binomial random variable X. If Y has a distribution given by the normal approximation, then Pr($X \leq 8$) is approximated by Pr($Y \leq 8.5$). The addition of 0.5 is the continuity correction; the uncorrected normal approximation gives considerably less accurate results.

This approximation, known as de Moivre–Laplace theorem, is a huge time-saver when undertaking calculations by hand (exact calculations with large n are very onerous); historically, it was the first use of the normal distribution, introduced in Abraham de Moivre's book *The Doctrine of Chances* in 1738. Nowadays, it can be seen as a consequence of the central limit theorem since B(n, p) is a sum of n independent, identically distributed Bernoulli variables with parameter p. This fact is the basis of a hypothesis test, a "proportion z-test", for the value of p using x/n, the sample proportion and estimator of p, in a common test statistic.[25]

For example, suppose one randomly samples n people out of a large population and ask them whether they agree with a certain statement. The proportion of people who agree will of course depend on the sample. If groups of n people were sampled repeatedly and truly randomly, the proportions would follow an approximate normal distribution with mean equal to the true proportion p of agreement in the population and with standard deviation

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

Poisson approximation

The binomial distribution converges towards the Poisson distribution as the number of trials goes to infinity while the product *np* remains fixed or at least *p* tends to zero. Therefore, the Poisson distribution with parameter $\lambda = np$ can be used as an approximation to B(*n*, *p*) of the binomial distribution if *n* is sufficiently large and *p* is sufficiently small. According to two rules of thumb, this approximation is good if $n \geq 20$ and $p \leq 0.05$, or if $n \geq 100$ and $np \leq 10$.[26]

Concerning the accuracy of Poisson approximation, see Novak,[27] ch. 4, and references therein.

Limiting distributions

- *Poisson limit theorem*: As *n* approaches ∞ and *p* approaches 0 with the product *np* held fixed, the Binomial(*n*, *p*) distribution approaches the Poisson distribution with expected value $\lambda = np$.[26]
- *de Moivre–Laplace theorem*: As *n* approaches ∞ while *p* remains fixed, the distribution of

$$\frac{X - np}{\sqrt{np(1-p)}}$$

approaches the normal distribution with expected value 0 and variance 1. This result is sometimes loosely stated by saying that the distribution of *X* is asymptotically normal with expected value *np* and variance $np(1 - p)$. This result is a specific case of the central limit theorem.

Beta distribution

The binomial distribution and beta distribution are different views of the same model of repeated Bernoulli trials. The binomial distribution is the PMF of *k* successes given *n* independent events each with a probability *p* of success. Mathematically, when $\alpha = k+1$ and $\beta = n-k+1$, the beta distribution and the binomial distribution are related by a factor of $n+1$:

$$Beta(p;\alpha;\beta) = (n+1)Binom(k;n;p)$$

Beta distributions also provide a family of prior probability distributions for binomial distributions in Bayesian inference:[28]

$$P(p;\alpha,\beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}.$$

Given a uniform prior, the posterior distribution for the probability of success *p* given *n* independent events with *k* observed successes is a beta distribution.[29]

Computational methods

Generating binomial random variates

Methods for random number generation where the marginal distribution is a binomial distribution are well-established.[30][31]

One way to generate random samples from a binomial distribution is to use an inversion algorithm. To do so, one must calculate the probability that $\Pr(X = k)$ for all values k from 0 through n. (These probabilities should sum to a value close to one, in order to encompass the entire sample space.) Then by using a pseudorandom number generator to generate samples uniformly between 0 and 1, one can transform the calculated samples into discrete numbers by using the probabilities calculated in the first step.

History

This distribution was derived by James Bernoulli. He considered the case where $p = r/(r + s)$ where p is the probability of success and r and s are positive integers. Blaise Pascal had earlier considered the case where $p = 1/2$.

See also

- Logistic regression
- Multinomial distribution
- Negative binomial distribution
- Beta-binomial distribution
- Binomial measure, an example of a multifractal measure.[32]
- Statistical mechanics

References

1. Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (Third ed.). New York: Wiley. p. 151 (theorem in section VI.3).
2. Wadsworth, G. P. (1960). *Introduction to Probability and Random Variables* (https://books.google.com/books?id=NNtQAAAAMAAJ&pg=PA52). New York: McGraw-Hill. p. 52.
3. Jowett G H (1963), The Relationship Between the Binomial and F Distributions, Journal of the Royal Statistical Society D, 13, 55-57.
4. Hamilton Institute. "The Binomial Distribution" (http://www.hamilton.ie/ollie/EE304/Binom.pdf) October 20, 2010.
5. See Proof Wiki (https://proofwiki.org/wiki/Expectation_of_Binomial_Distribution)
6. See also Nicolas, André (January 7, 2019). "Finding mode in Binomial distribution" (https://math.stackexchange.com/q/117940). *Stack Exchange*.
7. Neumann, P. (1966). "Über den Median der Binomial- and Poissonverteilung". *Wissenschaftliche Zeitschrift der Technischen Universität Dresden* (in German). **19**: 29–33.
8. Lord, Nick. (July 2010). "Binomial averages when the mean is an integer", The Mathematical Gazette 94, 331-332.
9. Kaas, R.; Buhrman, J.M. (1980). "Mean, Median and Mode in Binomial Distributions". *Statistica Neerlandica*. **34** (1): 13–18. doi:10.1111/j.1467-9574.1980.tb00681.x (https://doi.org/10.1111%2Fj.1467-9574.1980.tb00681.x).
10. Hamza, K. (1995). "The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions". *Statistics & Probability Letters*. **23**: 21–25. doi:10.1016/0167-7152(94)00090-U (https://doi.org/10.1016%2F0167-7152%2894%2900090-U).
11. Matoušek, J, Vondrak, J: *The Probabilistic Method* (lecture notes) [1] (https://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15859-f09/www/handouts/matousek-vondrak-prob-ln.pdf).
12. R. Arratia and L. Gordon: *Tutorial on large deviations for the binomial distribution*, Bulletin of Mathematical Biology 51(1) (1989), 125–131 [2] (https://link.springer.com/article/10.1007%2FBF02458840).
13. Theorem 11.1.3 in Cover, T.; Thomas, J. (2006). *Elements of Information Theory* (https://books.google.com/books?id=VWq5GG6ycxMC&pg=PA350) (2nd ed.). Wiley. p. 350. ISBN 9781118585771.
14. "Sharper Lower Bounds for Binomial/Chernoff Tails" (https://math.stackexchange.com/q/1564088). *Stack Exchange*. December 7, 2015.
15. Razzaghi, Mehdi. "On the estimation of binomial success probability with zero occurrence in sample." Journal of Modern Applied Statistical Methods 1.2 (2002): 41. url (https://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1673&context=jmasm)

16. Brown, Lawrence D.; Cai, T. Tony; DasGupta, Anirban (2001), "Interval Estimation for a Binomial Proportion" (http://www-stat.wharton.upenn.edu/~tcai/paper/html/Binomial-StatSci.html), *Statistical Science*, **16** (2): 101–133, CiteSeerX 10.1.1.323.7752 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.323.7752), doi:10.1214/ss/1009213286 (https://doi.org/10.1214%2Fss%2F1009213286), retrieved 2015-01-05
17. Agresti, Alan; Coull, Brent A. (May 1998), "Approximate is better than 'exact' for interval estimation of binomial proportions" (http://www.stat.ufl.edu/~aa/articles/agresti_coull_1998.pdf) (PDF), *The American Statistician*, **52** (2): 119–126, doi:10.2307/2685469 (https://doi.org/10.2307%2F2685469), JSTOR 2685469 (https://www.jstor.org/stable/2685469), retrieved 2015-01-05
18. Pires M. A. Confidence intervals for a binomial proportion: comparison of methods and software evaluation (http://www.math.ist.utl.pt/~apires/PDFs/AP_COMPSTAT02.pdf).
19. Wilson, Edwin B. (June 1927), "Probable inference, the law of succession, and statistical inference" (https://web.archive.org/web/20150113082307/http://psych.stanford.edu/~jlm/pdfs/Wison27SingleProportion.pdf) (PDF), *Journal of the American Statistical Association*, **22** (158): 209–212, doi:10.2307/2276774 (https://doi.org/10.2307%2F2276774), JSTOR 2276774 (https://www.jstor.org/stable/2276774), archived from the original (http://psych.stanford.edu/~jlm/pdfs/Wison27SingleProportion.pdf) (PDF) on 2015-01-13, retrieved 2015-01-05
20. "Confidence intervals" (http://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm). *Engineering Statistics Handbook*. NIST/Sematech. 2012. Retrieved 2017-07-23.
21. Katz D. *et al*.(1978) "Obtaining confidence intervals for the risk ratio in cohort studies". *Biometrics* 34:469–474
22. Taboga, Marco. "Lectures on Probability Theory and Mathematical Statistics" (https://www.statlect.com/probability-distributions/binomial-distribution#hid3). *statlect.com*. Retrieved 18 December 2017.
23. Wang, Y. H. (1993). "On the number of successes in independent trials" (https://web.archive.org/web/20160303182353/http://www3.stat.sinica.edu.tw/statistica/oldpdf/A3n23.pdf) (PDF). *Statistica Sinica*. **3** (2): 295–312. Archived from the original (http://www3.stat.sinica.edu.tw/statistica/oldpdf/A3n23.pdf) (PDF) on 2016-03-03.
24. Box, Hunter and Hunter (1978). *Statistics for experimenters* (https://archive.org/details/statisticsforexp00geor). Wiley. p. 130 (https://archive.org/details/statisticsforexp00geor/page/130).
25. NIST/SEMATECH, "7.2.4. Does the proportion of defectives meet requirements?" (http://www.itl.nist.gov/div898/handbook/prc/section2/prc24.htm) *e-Handbook of Statistical Methods.*
26. NIST/SEMATECH, "6.3.3.1. Counts Control Charts" (http://www.itl.nist.gov/div898/handbook/pmc/section3/pmc331.htm), *e-Handbook of Statistical Methods.*
27. Novak S.Y. (2011) Extreme value methods with applications to finance. London: CRC/ Chapman & Hall/Taylor & Francis. ISBN 9781-43983-5746.
28. MacKay, David (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press; First Edition. ISBN 978-0521642989.
29. https://www.statlect.com/probability-distributions/beta-distribution
30. Devroye, Luc (1986) *Non-Uniform Random Variate Generation*, New York: Springer-Verlag. (See especially Chapter X, Discrete Univariate Distributions (http://luc.devroye.org/chapter_ten.pdf))
31. Kachitvichyanukul, V.; Schmeiser, B. W. (1988). "Binomial random variate generation". *Communications of the ACM*. **31** (2): 216–222. doi:10.1145/42372.42381 (https://doi.org/10.1145%2F42372.42381).
32. Mandelbrot, B. B., Fisher, A. J., & Calvet, L. E. (1997). A multifractal model of asset returns. *3.2 The Binomial Measure is the Simplest Example of a Multifractal*

Further reading

- Hirsch, Werner Z. (1957). "Binomial Distribution—Success or Failure, How Likely Are They?" (https://books.google.com/books?id=KostAAAAIAAJ&pg=PA140). *Introduction to Modern Statistics*. New York: MacMillan. pp. 140–153.
- Neter, John; Wasserman, William; Whitmore, G. A. (1988). *Applied Statistics* (Third ed.). Boston: Allyn & Bacon. pp. 185–192. ISBN 0-205-10328-6.

External links

- Interactive graphic: Univariate Distribution Relationships (http://www.math.wm.edu/~leemis/chart/UDR/UDR.html)
- Binomial distribution formula calculator (http://www.fxsolver.com/browse/formulas/Binomial+distribution)
- Difference of two binomial variables: X-Y (https://math.stackexchange.com/q/1065487) or |X-Y| (https://math.stackexchange.com/q/562119)
- Querying the binomial probability distribution in WolframAlpha (http://www.wolframalpha.com/input/?i=Prob+x+%3E+19+if+x+is+binomial+with+n+%3D+36++and+p+%3D+.6)