Binomial distribution

From Wikipedia, the free encyclopedia

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: a random variable containing single bit of information: 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.

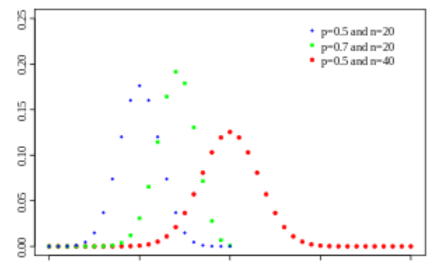
The binomial distribution is frequently used to model the number of successes in a sample of size *n* drawn with replacement from a population of size *N*. If the sampling is carried out without replacement, the draws are not independent and so the resulting distribution is a hypergeometric distribution,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

Contents

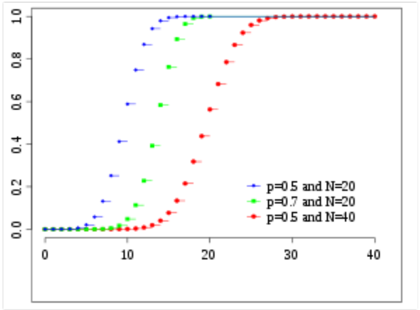
- 1 Specification
  - 1.1 Probability mass function
  - 1.2 Cumulative distribution function
- 2 Example
- 3 Mean
- 4 Variance
- 5 Mode
- 6 Median
- 7 Covariance between two binomials
- 8 Related distributions
  - 8.1 Sums of binomials
  - 8.2 Conditional binomials
  - 8.3 Bernoulli distribution
  - 8.4 Poisson binomial distribution
  - 8.5 Normal approximation
  - 8.6 Poisson approximation
  - 8.7 Limiting distributions
  - 8.8 Beta distribution
- 9 Confidence intervals
- 10 Generating binomial random variates
- 11 Tail bounds
- 12 See also
- 13 References
- 14 External links

binomial

Probability mass function



Cumulative distribution function

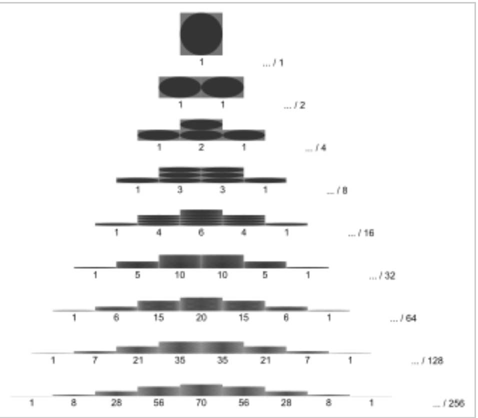


Notation	$B(n, p)$
Parameters	$n \in \mathbf{N}_0$ — number of trials $p \in [0,1]$ — success probability in each trial
Support	$k \in \{0, \ldots, n\}$ — number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2\left(2\pi e\, np(1-p)\right) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.
MGF	$(1-p+pe^t)^n$
CF	$(1-p+pe^{it})^n$
PGF	$G(z) = [(1-p)+pz]^n$.
Fisher information	$g_n(p) = \frac{n}{p(1-p)}$ (for fixed n)

Specification

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 $X \sim \mathrm{B}(n, p)$. The probability of getting exactly *k* successes in *n* trials is given by the probability mass function:



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 $70/256$.

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

for $k = 0, 1, 2, \ldots, n$, where