WIKIPEDIA

Bernoulli distribution

In probability theory and statistics, the **Bernoulli distribution**, named after Swiss mathematician Jacob Bernoulli,[1] is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$. Less formally, it can be thought of as a model for the set of possible outcomes of any single experiment that asks a yes–no question. Such questions lead to outcomes that are boolean-valued: a single bit whose value is success/yes/true/one with probability p and failure/no/false/zero with probability q. It can be used to represent a (possibly biased) coin toss where 1 and 0 would represent "heads" and "tails" (or vice versa), respectively, and p would be the probability of the coin landing on heads or tails, respectively. In particular, unfair coins would have $p \neq 1/2$.

The Bernoulli distribution is a special case of the binomial distribution where a single trial is conducted (so n would be 1 for such a binomial distribution). It is also a special case of the **two-point distribution**, for which the possible outcomes need not be 0 and 1.

Bernoulli

Parameters	$0 \leq p \leq 1$ $q = 1 - p$
Support	$k \in \{0, 1\}$
PMF	$\begin{cases} q = 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$ $p^k (1-p)^{1-k}$
CDF	$\begin{cases} 0 & \text{if } k < 0 \\ 1 - p & \text{if } 0 \leq k < 1 \\ 1 & \text{if } k \geq 1 \end{cases}$
Mean	p
Median	$\begin{cases} 0 & \text{if } p < 1/2 \\ [0, 1] & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$
Mode	$\begin{cases} 0 & \text{if } p < 1/2 \\ 0, 1 & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$
Variance	$p(1-p) = pq$
Skewness	$\frac{q - p}{\sqrt{pq}}$
Ex. kurtosis	$\frac{1 - 6pq}{pq}$
Entropy	$-q \ln q - p \ln p$
MGF	$q + pe^t$
CF	$q + pe^{it}$
PGF	$q + pz$
Fisher information	$\frac{1}{pq}$

Contents

Properties

If X is a random variable with this distribution, then:

$$\Pr(X = 1) = p = 1 - \Pr(X = 0) = 1 - q.$$

The probability mass function f of this distribution, over possible outcomes k, is

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ q = 1-p & \text{if } k = 0. \end{cases}$$ [2]

This can also be expressed as

$$f(k;p) = p^k (1-p)^{1-k} \quad \text{for } k \in \{0,1\}$$

or as

$$f(k;p) = pk + (1-p)(1-k) \quad \text{for } k \in \{0,1\}.$$

The Bernoulli distribution is a special case of the binomial distribution with $n = 1$.[3]

The kurtosis goes to infinity for high and low values of p, but for $p = 1/2$ the two-point distributions including the Bernoulli distribution have a lower excess kurtosis than any other probability distribution, namely −2.

The Bernoulli distributions for $0 \leq p \leq 1$ form an exponential family.

The maximum likelihood estimator of p based on a random sample is the sample mean.

Mean

The expected value of a Bernoulli random variable X is

$$\operatorname{E}(X) = p$$

This is due to the fact that for a Bernoulli distributed random variable X with $\Pr(X = 1) = p$ and $\Pr(X = 0) = q$ we find

$$\operatorname{E}[X] = \Pr(X=1) \cdot 1 + \Pr(X=0) \cdot 0 = p \cdot 1 + q \cdot 0 = p.$$ [2]

Variance

The variance of a Bernoulli distributed X is

$$\operatorname{Var}[X] = pq = p(1-p)$$

We first find

$$\operatorname{E}[X^2] = \Pr(X=1) \cdot 1^2 + \Pr(X=0) \cdot 0^2 = p \cdot 1^2 + q \cdot 0^2 = p$$

From this follows

$$\operatorname{Var}[X] = \operatorname{E}[X^2] - \operatorname{E}[X]^2 = p - p^2 = p(1-p) = pq$$ [2]

Skewness

The skewness is $\frac{q-p}{\sqrt{pq}} = \frac{1-2p}{\sqrt{pq}}$. When we take the standardized Bernoulli distributed random variable $\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}}$ we find that this random variable attains $\frac{q}{\sqrt{pq}}$ with probability p and attains $-\frac{p}{\sqrt{pq}}$ with probability q. Thus we get

$$\begin{aligned}
\gamma_1 &= \mathrm{E}\left[\left(\frac{X-\mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}}\right)^3\right] \\
&= p \cdot \left(\frac{q}{\sqrt{pq}}\right)^3 + q \cdot \left(-\frac{p}{\sqrt{pq}}\right)^3 \\
&= \frac{1}{\sqrt{pq}^3}\left(pq^3 - qp^3\right) \\
&= \frac{pq}{\sqrt{pq}^3}(q-p) \\
&= \frac{q-p}{\sqrt{pq}}
\end{aligned}$$

Higher moments and cumulants

The central moment of order k is given by

$$\mu_k = (1-p)(-p)^k + p(1-p)^k.$$

The first six central moments are

$$\begin{aligned}
\mu_1 &= 0, \\
\mu_2 &= p(1-p), \\
\mu_3 &= p(1-p)(1-2p), \\
\mu_4 &= p(1-p)(1-3p(1-p)), \\
\mu_5 &= p(1-p)(1-2p)(1-2p(1-p)), \\
\mu_6 &= p(1-p)(1-5p(1-p)(1-p(1-p))).
\end{aligned}$$

The higher central moments can be expressed more compactly in terms of μ_2 and μ_3

$$\begin{aligned}
\mu_4 &= \mu_2(1-3\mu_2), \\
\mu_5 &= \mu_3(1-2\mu_2), \\
\mu_6 &= \mu_2(1-5\mu_2(1-\mu_2)).
\end{aligned}$$

The first six cumulants are

$$\begin{aligned}\kappa_1 &= p,\\ \kappa_2 &= \mu_2,\\ \kappa_3 &= \mu_3,\\ \kappa_4 &= \mu_2(1-6\mu_2),\\ \kappa_5 &= \mu_3(1-12\mu_2),\\ \kappa_6 &= \mu_2(1-30\mu_2(1-4\mu_2)).\end{aligned}$$

Related distributions

- If $X_1, \dots, X_n$ are independent, identically distributed (i.i.d.) random variables, all Bernoulli trials with success probability *p*, then their sum is distributed according to a binomial distribution with parameters *n* and *p*:

  $$\sum_{k=1}^{n} X_k \sim \mathrm{B}(n,p)$$ (binomial distribution).[2]

  The Bernoulli distribution is simply $\mathrm{B}(1,p)$, also written as $\mathrm{Bernoulli}(p)$.

- The categorical distribution is the generalization of the Bernoulli distribution for variables with any constant number of discrete values.
- The Beta distribution is the conjugate prior of the Bernoulli distribution.
- The geometric distribution models the number of independent and identical Bernoulli trials needed to get one success.
- If $Y \sim \mathrm{Bernoulli}\left(\frac{1}{2}\right)$, then $2Y-1$ has a Rademacher distribution.

See also

- Bernoulli process, a random process consisting of a sequence of independent Bernoulli trials
- Bernoulli sampling
- Binary entropy function
- Binary decision diagram

References

1. James Victor Uspensky: *Introduction to Mathematical Probability*, McGraw-Hill, New York 1937, page 45
2. Bertsekas, Dimitri P. (2002). *Introduction to Probability*. Tsitsiklis, John N., Τσιτσικλής, Γιάννης Ν. Belmont, Mass.: Athena Scientific. ISBN 188652940X. OCLC 51441829 (https://www.worldcat.org/oclc/51441829).
3. McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models, Second Edition*. Boca Raton: Chapman and Hall/CRC. Section 4.2.2. ISBN 0-412-31760-5.

Further reading

- Johnson, N. L.; Kotz, S.; Kemp, A. (1993). *Univariate Discrete Distributions* (2nd ed.). Wiley. ISBN 0-471-54897-9.
- Peatman, John G. (1963). *Introduction to Applied Statistics*. New York: Harper & Row. pp. 162–171.

External links

- "Binomial distribution" (https://www.encyclopediaofmath.org/index.php?title=Binomial_distribution), *Encyclopedia of Mathematics*, EMS Press, 2001 [1994]
- Weisstein, Eric W. "Bernoulli Distribution" (https://mathworld.wolfram.com/BernoulliDistribution.html). *MathWorld*.
- Interactive graphic: Univariate Distribution Relationships (http://www.math.wm.edu/~leemis/chart/UDR/UDR.html)