Bernoulli distribution

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **Bernoulli distribution**, named after Swiss scientist Jacob Bernoulli,[1] is the probability distribution of a random variable which takes the value 1 with success probability of p and the value 0 with failure probability of $q = 1 - p$. It can be used to represent a coin toss where 1 and 0 would represent "head" and "tail" (or vice versa), respectively. In particular, unfair coins would have $p \neq 0.5$.

The Bernoulli distribution is a special case of the **two-point distribution**, for which the two possible outcomes need not be 0 and 1. It is also a special case of the binomial distribution; the Bernoulli distribution is a binomial distribution where n=1.

Bernoulli

Parameters	$0 < p < 1, p \in \mathbb{R}$
Support	$k \in \{0, 1\}$
pmf	$\begin{cases} q = (1-p) & \text{for } k = 0 \\ p & \text{for } k = 1 \end{cases}$
CDF	$\begin{cases} 0 & \text{for } k < 0 \\ 1-p & \text{for } 0 \leq k < 1 \\ 1 & \text{for } k \geq 1 \end{cases}$
Mean	p
Median	$\begin{cases} 0 & \text{if } q > p \\ 0.5 & \text{if } q = p \\ 1 & \text{if } q < p \end{cases}$
Mode	$\begin{cases} 0 & \text{if } q > p \\ 0, 1 & \text{if } q = p \\ 1 & \text{if } q < p \end{cases}$
Variance	$p(1-p) (= pq)$
Skewness	$\frac{1-2p}{\sqrt{pq}}$
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	$\frac{1}{p(1-p)}$

Contents

Properties

If X is a random variable with this distribution, we have:

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

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

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

This can also be expressed as

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

The Bernoulli distribution is a special case of the binomial distribution with $n = 1$.[2]

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

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

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

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

Variance

The variance of a Bernoulli distributed X is

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

We first find

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

From this follows

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

Skewness

The skewness is $\frac{q-p}{\sqrt{pq}} = \frac{1-2p}{\sqrt{pq}}$. When we take the standardized Bernoulli distributed random variable $\frac{X - \mathrm{E}[X]}{\sqrt{\operatorname{Var}[X]}}$ 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

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

Related distributions

- If $X_1, \dots, X_n$ are independent, identically distributed (i.i.d.) random variables, all Bernoulli distributed with success probability *p*, then

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

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

- 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 $Y \sim$ Bernoulli(0.5), then $(2Y\text{-}1)$ has a Rademacher distribution.

See also

- Bernoulli process
- Bernoulli sampling
- Bernoulli trial
- Binary entropy function
- Binomial Distribution

Notes

1. James Victor Uspensky: *Introduction to Mathematical Probability*, McGraw-Hill, New York 1937, page 45
2. McCullagh and Nelder (1989), Section 4.2.2.

References

- McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models, Second Edition*. Boca Raton:

Chapman and Hall/CRC. ISBN 0-412-31760-5.
- Johnson, N.L., Kotz, S., Kemp A. (1993) Univariate Discrete Distributions (2nd Edition). Wiley. ISBN 0-471-54897-9

External links

- Hazewinkel, Michiel, ed. (2001), "Binomial distribution", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Weisstein, Eric W., "Bernoulli Distribution" (http://mathworld.wolfram.com/BernoulliDistribution.html), *MathWorld*.
- Interactive graphic: Univariate Distribution Relationships (http://www.math.wm.edu/~leemis/chart/UDR/UDR.html)

Wikimedia Commons has media related to ***Bernoulli distribution***.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Bernoulli_distribution&oldid=711006923"

Categories: Discrete distributions | Distributions with conjugate priors | Exponential family distributions | Probability distributions