

# Generalized linear model

In *statistics*, the **generalized linear model** (**GLM**) is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a *link function* and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

Generalized linear models were formulated by John Nelder and Robert Wedderburn as a way of unifying various other statistical models, including linear regression, logistic regression and Poisson regression.<sup>[1]</sup> They proposed an iteratively reweighted least squares method for maximum likelihood estimation of the model parameters. Maximum-likelihood estimation remains popular and is the default method on many statistical computing packages. Other approaches, including Bayesian approaches and least squares fits to variance stabilized responses, have been developed.

## Contents

**Intuition**

**Overview**

**Model components**

Probability distribution

Linear predictor

Link function

**Fitting**

Maximum likelihood

Bayesian methods

**Examples**

General linear models

Linear regression

Binary data

Logit link function

Probit link function as popular choice of inverse cumulative distribution function

Complementary log-log (cloglog)

Identity link

Variance function

Multinomial regression

Ordered response

Unordered response

Count data

**Extensions**

Correlated or clustered data

Generalized additive models

**Confusion with general linear models**

**See also**

**Notes**

**References**

**Further reading**

## Intuition

Ordinary linear regression predicts the expected value of a given unknown quantity (the *response variable*, a random variable) as a linear combination of a set of observed values (*predictors*). This implies that a constant change in a predictor leads to a constant change in the response variable (i.e. a *linear-response model*). This is appropriate when the response variable has a normal distribution (intuitively, when a response variable can vary essentially indefinitely in either direction with no fixed "zero value", or more generally for any quantity that only varies by a relatively small amount, e.g. human heights).

However, these assumptions are inappropriate for some types of response variables. For example, in cases where the response variable is expected to be always positive and varying over a wide range, constant input changes lead to geometrically varying, rather than constantly varying, output changes. As an example, a prediction model might predict that 10 degree temperature decrease would lead to 1,000 fewer people visiting the beach is unlikely to generalize well over both small beaches (e.g. those where the expected attendance was 50 at a particular temperature) and large beaches (e.g. those where the expected attendance was 10,000 at a low temperature). The problem with this kind of prediction model would imply a temperature drop of 10 degrees would lead to 1,000 fewer people visiting the beach, a beach whose expected attendance was 50 at a higher temperature would now be predicted to have the impossible attendance value of −950. Logically, a more realistic model would instead predict a constant *rate* of increased beach attendance (e.g. an increase in 10 degrees leads to a doubling in beach attendance, and a drop in 10 degrees leads to a halving in attendance). Such a model is termed an *exponential-response model* (or *log-linear model*, since the logarithm of the response is predicted to vary linearly).

Similarly, a model that predicts a probability of making a yes/no choice (a Bernoulli variable) is even less suitable as a linear-response model, since probabilities are bounded on both ends (they must be between 0 and 1). Imagine, for example, a model that predicts the likelihood of a given person going to the beach as a function of temperature. A reasonable model might predict, for example, that a change in 10 degrees makes a person two times more or less likely to go to the beach. But what does "twice as likely" mean in terms of a probability? It cannot literally mean to double the probability value (e.g. 50% becomes 100%, 75% becomes 150%, etc.). Rather, it is the *odds* that are doubling: from 2:1 odds, to 4:1 odds, to 8:1 odds, etc. Such a model is a *log-odds* or *logistic model*.

Generalized linear models cover all these situations by allowing for response variables that have arbitrary distributions (rather than simply normal distributions), and for an arbitrary function of the response variable (the *link function*) to vary linearly with the predicted values (rather than assuming that the response itself must vary linearly). For example, the case above of predicted number of beach attendees would typically be modeled with a Poisson distribution and a log link, while the case of predicted

probability of beach attendance would typically be modeled with a Bernoulli distribution (or binomial distribution, depending on exactly how the problem is phrased) and a log-odds (or *logit*) link function.

## Overview

In a generalized linear model (GLM), each outcome **Y** of the dependent variables is assumed to be generated from a particular distribution in an exponential family, a large class of probability distributions that includes the normal, binomial, Poisson and gamma distributions, among others. The mean, **μ**, of the distribution depends on the independent variables, **X**, through:

$$\mathbf{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$$

where E(Y) is the expected value of **Y**; **Xβ** is the *linear predictor*, a linear combination of unknown parameters **β**; *g* is the link function.

In this framework, the variance is typically a function, **V**, of the mean:

$$\mathbf{Var}(\mathbf{Y}) = \mathbf{V}(\boldsymbol{\mu}) = \mathbf{V}(g^{-1}(\mathbf{X}\boldsymbol{\beta})).$$

It is convenient if **V** follows from an exponential family of distributions, but it may simply be that the variance is a function of the predicted value.

The unknown parameters, **β**, are typically estimated with maximum likelihood, maximum quasi-likelihood, or Bayesian techniques.

## Model components

The GLM consists of three elements<sup>[2]</sup>:

1. An exponential family of probability distributions.
2. A linear predictor  $\eta = \mathbf{X}\boldsymbol{\beta}$ .
3. A link function *g* such that  $\mathbf{E}(\mathbf{Y}|\mathbf{X}) = \boldsymbol{\mu} = g^{-1}(\eta)$ .

### Probability distribution

An **overdispersed exponential family** of distributions is a generalization of an exponential family and the exponential dispersion model of distributions and includes those families of probability distributions, parameterized by **θ** and **τ**, whose density functions *f* (or probability mass function, for the case of a discrete distribution) can be expressed in the form

$$f_Y(\mathbf{y} \mid \boldsymbol{\theta}, \tau) = h(\mathbf{y}, \tau) \exp \left( \frac{\mathbf{b}(\boldsymbol{\theta})^T \mathbf{T}(\mathbf{y}) - A(\boldsymbol{\theta})}{d(\tau)} \right).$$

The *dispersion parameter*, **τ**, typically is known and is usually related to the variance of the distribution. The functions **h(y, τ)**, **b(θ)**, **T(y)**, **A(θ)**, and **d(τ)** are known. Many common distributions are in this family, including the normal, exponential, gamma, Poisson, Bernoulli, and (for fixed number of trials) binomial, multinomial, and negative binomial.

For scalar **Y** and **θ**, this reduces to

$$f_Y(y \mid \theta, \tau) = h(y, \tau) \exp \left( \frac{b(\theta)T(y) - A(\theta)}{d(\tau)} \right).$$

**θ** is related to the mean of the distribution. If **b(θ)** is the identity function, then the distribution is said to be in canonical form (or *natural form*). Note that any distribution can be converted to canonical form by rewriting **θ** as **θ'** and then applying the transformation **θ = b(θ')**. It is always possible to convert **A(θ)** in terms of the new parametrization, even if **b(θ')** is not a one-to-one function; see comments in the page on exponential families. If, in addition, **T(y)** is the identity and **τ** is known, then **θ** is called the *canonical parameter* (or *natural parameter*) and is related to the mean through

$$\boldsymbol{\mu} = \mathbf{E}(\mathbf{Y}) = \nabla A(\boldsymbol{\theta}).$$

For scalar **Y** and **θ**, this reduces to

$$\mu = \mathbf{E}(Y) = A'(\theta).$$

Under this scenario, the variance of the distribution can be shown to be<sup>[3]</sup>

$$\mathbf{Var}(\mathbf{Y}) = \nabla^2 A(\boldsymbol{\theta})d(\tau).$$

For scalar **Y** and **θ**, this reduces to

$$\mathbf{Var}(Y) = A''(\theta)d(\tau).$$

### Linear predictor

The linear predictor is the quantity which incorporates the information about the independent variables into the model. The symbol *η* (Greek "eta") denotes a linear predictor. It is related to the expected value of the data through the link function.

*η* is expressed as linear combinations (thus, "linear") of unknown parameters **β**. The coefficients of the linear combination are represented as the matrix of independent variables **X**. *η* can thus be expressed as

$$\eta = \mathbf{X}\boldsymbol{\beta}.$$

Link function

The link function provides the relationship between the linear predictor and the mean of the distribution function. There are many commonly used link functions, and their choice is informed by several considerations. There is always a well-defined *canonical* link function which is derived from the exponential of the response's density function. However, in some cases it makes sense to try to match the domain of the link function to the range of the distribution function's mean, or use a non-canonical link function for algorithmic purposes, for example Bayesian probit regression.

When using a distribution function with a canonical parameter  $\theta$ , the canonical link function is the function that expresses  $\theta$  in terms of  $\mu$ , i.e.  $\theta = b(\mu)$ . For the most common distributions, the mean  $\mu$  is one of the parameters in the standard form of the distribution's density function, and then  $b(\mu)$  is the function as defined above that maps the density function into its canonical form. When using the canonical link function,  $b(\mu) = \theta = \mathbf{X}\beta$ , which allows  $\mathbf{X}^T \mathbf{Y}$  to be a sufficient statistic for  $\beta$ .

Following is a table of several exponential-family distributions in common use and the data they are typically used for, along with the canonical link functions and their inverses (sometimes referred to as the mean function, as done here).

Common distributions with typical uses and canonical link functions					
Distribution	Support of distribution	Typical uses	Link name	Link function, $\mathbf{X}\beta = g(\mu)$	Mean function
<u>Normal</u>	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\beta = \mu$	$\mu = \mathbf{X}\beta$
<u>Exponential</u>	real: $(0, +\infty)$	Exponential-response data, scale parameters	<u>Negative inverse</u>	$\mathbf{X}\beta = -\mu^{-1}$	$\mu = -(\mathbf{X}\beta)^{-1}$
<u>Gamma</u>					
<u>Inverse Gaussian</u>	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\beta = \mu^{-2}$	$\mu = (\mathbf{X}\beta)^{-1/2}$
<u>Poisson</u>	integer: $0, 1, 2, \dots$	count of occurrences in fixed amount of time/space	<u>Log</u>	$\mathbf{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbf{X}\beta)$
<u>Bernoulli</u>	integer: $\{0, 1\}$	outcome of single yes/no occurrence	<u>Logit</u>	$\mathbf{X}\beta = \ln\left(\frac{\mu}{1 - \mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)} = \frac{1}{1 + \exp(-\mathbf{X}\beta)}$
<u>Binomial</u>	integer: $0, 1, \dots, N$	count of # of "yes" occurrences out of N yes/no occurrences			
<u>Categorical</u>	integer: $[0, K)$	outcome of single K-way occurrence			
	K-vector of integer: $[0, 1]$ , where exactly one element in the vector has the value 1				
<u>Multinomial</u>	K-vector of integer: $[0, N]$	count of occurrences of different types (1 .. K) out of N total K-way occurrences			

In the cases of the exponential and gamma distributions, the domain of the canonical link function is not the same as the permitted range of the mean. In particular, the linear predictor may be positive, which would give an impossible negative mean. When maximizing the likelihood, precautions must be taken to avoid this. An alternative is to use a noncanonical link function.

Note also that in the case of the Bernoulli, binomial, categorical and multinomial distributions, the support of the distributions is not the same type of data as the parameter being predicted. In all of these cases, the predicted parameter is one or more probabilities, i.e. real numbers in the range  $[0, 1]$ . The resulting model is known as *logistic regression* (or *multinomial logistic regression* in the case that K-way rather than binary values are being predicted).

For the Bernoulli and binomial distributions, the parameter is a single probability, indicating the likelihood of occurrence of a single event. The Bernoulli still satisfies the basic condition of the generalized linear model in that, even though a single outcome will always be either 0 or 1, the *expected value* will nonetheless be a real-valued probability, i.e. the probability of occurrence of a "yes" (or 1) outcome. Similarly, in a binomial distribution, the expected value is  $Np$ , i.e. the expected proportion of "yes" outcomes will be the probability to be predicted.

For categorical and multinomial distributions, the parameter to be predicted is a K-vector of probabilities, with the further restriction that all probabilities must add up to 1. Each probability indicates the likelihood of occurrence of one of the K possible values. For the multinomial distribution, and for the vector form of the categorical distribution, the expected values of the elements of the vector can be related to the predicted probabilities similarly to the binomial and Bernoulli distributions.

Fitting

Maximum likelihood

The maximum likelihood estimates can be found using an iteratively reweighted least squares algorithm or a Newton's method with updates of the form:

$$\beta^{(t+1)} = \beta^{(t)} + \mathcal{J}^{-1}(\beta^{(t)})u(\beta^{(t)}),$$

where  $\mathcal{J}(\beta^{(t)})$  is the observed information matrix (the negative of the Hessian matrix) and  $u(\beta^{(t)})$  is the score function; or a Fisher's scoring method:

$$\beta^{(t+1)} = \beta^{(t)} + \mathcal{I}^{-1}(\beta^{(t)})u(\beta^{(t)}),$$

where  $\mathcal{I}(\beta^{(t)})$  is the Fisher information matrix. Note that if the canonical link function is used, then they are the same.<sup>[4]</sup>

Bayesian methods

In general, the posterior distribution cannot be found in closed form and so must be approximated, usually using Laplace approximations or some type of Markov chain Monte Carlo method such as Gibbs sampling.

Examples

## General linear models

A possible point of confusion has to do with the distinction between generalized linear models and the general linear model, two broad statistical models. The general linear model may be viewed as a special case of the generalized linear model with identity link and responses normally distributed. As most exact results of interest are obtained only for the general linear model, the general linear model has undergone a somewhat longer historical development. Results for the generalized linear model with non-identity link are asymptotic (tending to work well with large samples).

## Linear regression

A simple, very important example of a generalized linear model (also an example of a general linear model) is linear regression. In linear regression, the use of the least-squares estimator is justified by the Gauss–Markov theorem, which does not assume that the distribution is normal.

From the perspective of generalized linear models, however, it is useful to suppose that the distribution function is the normal distribution with constant variance and the link function is the identity, which is the canonical link if the variance is known.

For the normal distribution, the generalized linear model has a closed form expression for the maximum-likelihood estimates, which is convenient. Most other GLMs lack closed form estimates.

## Binary data

When the response data, *Y*, are binary (taking on only values 0 and 1), the distribution function is generally chosen to be the Bernoulli distribution and the interpretation of *μ<sub>i</sub>* is then the probability, *p*, of *Y<sub>i</sub>* taking on the value one.

There are several popular link functions for binomial functions.

### Logit link function

The most typical link function is the canonical logit link:

$$g(p) = \ln\left(\frac{p}{1 - p}\right).$$

GLMs with this setup are logistic regression models (or *logit models*).

### Probit link function as popular choice of inverse cumulative distribution function

Alternatively, the inverse of any continuous cumulative distribution function (CDF) can be used for the link since the CDF's range is [0, 1], the range of the binomial mean. The normal CDF Φ is a popular choice and yields the probit model. Its link is

$$g(p) = \Phi^{-1}(p).$$

The reason for the use of the probit model is that a constant scaling of the input variable to a normal CDF (which can be absorbed through equivalent scaling of all of the parameters) yields a function that is practically identical to the logit function, but probit models are more tractable in some situations than logit models. (In a Bayesian setting in which normally distributed prior distributions are placed on the parameters, the relationship between the normal priors and the normal CDF link function means that a probit model can be computed using Gibbs sampling, while a logit model generally cannot.)

### Complementary log-log (cloglog)

The complementary log-log function may also be used:

$$g(p) = \log(-\log(1 - p)).$$

This link function is asymmetric and will often produce different results from the logit and probit link functions.<sup>[5]</sup> The cloglog model corresponds to applications where we observe either zero events (e.g., defects) or one or more, where the number of events is assumed to follow the Poisson distribution<sup>[6]</sup>. The Poisson assumption means that

$$\mathbf{Pr}(0) = \exp(-\mu),$$

where *μ* is a positive number denoting the expected number of events. If *p* represents the proportion of observations with at least one event, its complement

$$(1 - p) = \mathbf{Pr}(0) = \exp(-\mu),$$

and then

$$(-\log(1 - p)) = \mu.$$

A linear model requires the response variable to take values over the entire real line. Since *μ* must be positive, we can enforce that by taking the logarithm, and letting log(*μ*) be a linear model. This produces the "cloglog" transformation

$$\log(-\log(1 - p)) = \log(\mu).$$

### Identity link

The identity link *g(p) = p* is also sometimes used for binomial data to yield a linear probability model. However, the identity link can predict nonsense "probabilities" less than zero or greater than one. This can be avoided by using a transformation like cloglog, probit or logit (or any inverse cumulative distribution function). A primary merit of the identity link is that it can be estimated using linear math—and other standard link functions are approximately linear matching the identity link near *p* = 0.5.

### Variance function

The variance function for "quasibinomial" data is:

$$\mathbf{Var}(Y_i) = \tau \mu_i (1 - \mu_i)$$

where the dispersion parameter  $\tau$  is exactly 1 for the binomial distribution. Indeed, the standard binomial likelihood omits  $\tau$ . When it is present, the model is called "quasibinomial", and the modified likelihood is called a quasi-likelihood, since it is not generally the likelihood corresponding to any real family of probability distributions. If  $\tau$  exceeds 1, the model is said to exhibit overdispersion.

### Multinomial regression

The binomial case may be easily extended to allow for a multinomial distribution as the response (also, a Generalized Linear Model for counts, with a constrained total). There are two ways in which this is usually done:

#### Ordered response

If the response variable is an ordinal measurement, then one may fit a model function of the form:

$$g(\mu_m) = \eta_m = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p + \gamma_2 + \cdots + \gamma_m = \eta_1 + \gamma_2 + \cdots + \gamma_m \text{ where } \mu_m = \mathbf{P}(Y \leq m).$$

for  $m > 2$ . Different links  $g$  lead to proportional odds models or ordered probit models.

#### Unordered response

If the response variable is a nominal measurement, or the data do not satisfy the assumptions of an ordered model, one may fit a model of the following form:

$$g(\mu_m) = \eta_m = \beta_{m,0} + X_1\beta_{m,1} + \cdots + X_p\beta_{m,p} \text{ where } \mu_m = \mathbf{P}(Y = m \mid Y \in \{1, m\}).$$

for  $m > 2$ . Different links  $g$  lead to multinomial logit or multinomial probit models. These are more general than the ordered response models, and more parameters are estimated.

- 

### Count data

Another example of generalized linear models includes Poisson regression which models count data using the Poisson distribution. The link is typically the logarithm, the canonical link.

The variance function is proportional to the mean

$$\mathbf{var}(Y_i) = \tau \mu_i,$$

where the dispersion parameter  $\tau$  is typically fixed at exactly one. When it is not, the resulting quasi-likelihood model is often described as Poisson with overdispersion or *quasi-Poisson*.

## Extensions

#### Correlated or clustered data

The standard GLM assumes that the observations are uncorrelated. Extensions have been developed to allow for correlation between observations, as occurs for example in longitudinal studies and clustered designs:

- Generalized estimating equations** (GEEs) allow for the correlation between observations without the use of an explicit probability model for the origin of the correlations, so there is no explicit likelihood. They are suitable when the random effects and their variances are not of inherent interest, as they allow for the correlation without explaining its origin. The focus is on estimating the average response over the population ("population-averaged" effects) rather than the regression parameters that would enable prediction of the effect of changing one or more components of **X** on a given individual. GEEs are usually used in conjunction with Huber–White standard errors.<sup>[7]</sup><sup>[8]</sup>
- Generalized linear mixed models** (GLMMs) are an extension to GLMs that includes random effects in the linear predictor, giving an explicit probability model that explains the origin of the correlations. The resulting "subject-specific" parameter estimates are suitable when the focus is on estimating the effect of changing one or more components of **X** on a given individual. GLMMs are also referred to as multilevel models and as mixed model. In general, fitting GLMMs is more computationally complex and intensive than fitting GEEs.

### Generalized additive models

Generalized additive models (GAMs) are another extension to GLMs in which the linear predictor  $\eta$  is not restricted to be linear in the covariates **X** but is the sum of smoothing functions applied to the  $x_j$ :

$$\eta = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots$$

The smoothing functions  $f_j$  are estimated from the data. In general this requires a large number of data points and is computationally intensive.<sup>[9]</sup><sup>[10]</sup>

## Confusion with general linear models

The term "generalized linear model", and especially its abbreviation GLM, are sometimes confused with the term "general linear model". Co-originator John Nelder has expressed regret over this terminology.<sup>[11]</sup>

## See also

- Comparison of general and generalized linear models
- Generalized linear array model
- GLIM (software)
- Quasi-variance

- Natural exponential family
- Tweedie distributions
- Variance functions
- Vector generalized linear model (VGLM)

# Notes

1. Nelder, John; Wedderburn, Robert (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society. Series A (General)*. Blackwell Publishing. **135** (3): 370–384. doi:10.2307/2344614 (https://doi.org/10.2307%2F2344614). JSTOR 2344614 (http s://www.jstor.org/stable/2344614).

2. "6.1 - Introduction to Generalized Linear Models | STAT 504" (https://newonlinecou rses.science.psu.edu/stat504/node/216/). *newonlinecourses.science.psu.edu*. Retrieved 2019-03-18.

3. McCullagh and Nelder (1989), chapter 2.

4. McCullagh and Nelder (1989), page 43.

5. "Complementary Log-log Model" (http://www.stat.ualberta.ca/~kcarrier/STAT562/c omp\_log\_log.pdf) (PDF).

6. "Which Link Function — Logit, Probit, or Cloglog?" (https://bayesium.com/which-lin k-function-logit-probit-or-cloglog/). *Bayesium Analytics*. 2015-08-14. Retrieved 2019-03-17.

7. Zeger, Scott L.; Liang, Kung-Yee; Albert, Paul S. (1988). "Models for Longitudinal Data: A Generalized Estimating Equation Approach". *Biometrics*. International Biometric Society. **44** (4): 1049–1060. doi:10.2307/2531734 (https://doi.org/10.230 7%2F2531734). JSTOR 2531734 (https://www.jstor.org/stable/2531734). PMID 3233245 (https://www.ncbi.nlm.nih.gov/pubmed/3233245).

8. Hardin, James; Hilbe, Joseph (2003). *Generalized Estimating Equations*. London: Chapman and Hall/CRC. ISBN 1-58488-307-3.

9. Hastie & Tibshirani 1990.

10. Wood 2006.

11. Senn, Stephen (2003). "A conversation with John Nelder". *Statistical Science*. **18** (1): 118–131. doi:10.1214/ss/1056397489 (https://doi.org/10.1214%2Fss%2F105639748 9). "I suspect we should have found some more fancy name for it that would have stuck and not been confused with the general linear model, although general and generalized are not quite the same. I can see why it might have been better to have thought of something else."

# References

- Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC. ISBN 978-0-412-34390-2.
- Madsen, Henrik; Thyregod, Poul (2011). *Introduction to General and Generalized Linear Models*. Chapman & Hall/CRC. ISBN 978-1-4200-9155-7.
- McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models, Second Edition*. Boca Raton: Chapman and Hall/CRC. ISBN 0-412-31760-5.
- Wood, Simon (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC. ISBN 1-58488-474-6.

# Further reading

- Dunn, P.K.; Smyth, G.K. (2018). *Generalized Linear Models With Examples in R*. New York: Springer. doi:10.1007/978-1-4419-0118-7 (https://doi.org/10.1007%2F978-1-4419-0118 -7). ISBN 978-1-4419-0118-7.
- Dobson, A.J.; Barnett, A.G. (2008). *Introduction to Generalized Linear Models* (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC. ISBN 1-58488-165-8.
- Hardin, James; Hilbe, Joseph (2007). *Generalized Linear Models and Extensions* (2nd ed.). College Station: Stata Press. ISBN 1-59718-014-9.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Generalized\_linear\_model&oldid=896764202"

**This page was last edited on 12 May 2019, at 18:52 (UTC).**

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.