Some Useful Distributions in Bayesian Analysis with Models from Educational Measurement

Robert J. Mislevy
University of Maryland
November 25, 2001

Introduction

This 'cheat sheet' is a short list of some statistical distributions that appear in Bayesian analyses of educational measurement models. It gives the name of the distribution, some ways it is used, and conjugate priors if they are available. More distributions are discussed, and each is discussed more deeply, in Appendix A of Gelman, Carlin, Stern, & Rubin (1995). The following distributions are included here:

- Normal
- Gamma
- Bernoulli
- Binomial
- Beta
- Categorical
- Multinomial
- Dirichlet
- Log-normal
- Logit-normal
- Poisson
- Exponential

Measurement Models

Some key ideas in choosing probability distributions in Bayesian analyses of measurement-model data are these:

Observations of a particular nature are made—e.g., right/wrong responses, response categories, elapsed time to an event, counts of some phenomenon. E.g.

A probability distribution is chosen that produces that kind of data, with probabilities that depend on one or more parameters of the distribution. Those parameters imply the shape, central tendency, spread, etc. of the modeled distribution.

The parameters of the measurement model are expressed as functions of one or more parameters for a person (characterizing some characteristic(s) of people in which we are interested), and one or more parameters for each observational setting (e.g., item, test) that indicates characteristics such as its difficulty, sensitivity, etc.

Higher-level distributions—hyperpriors—are specified for the parameters of people and of items. Their form is based on the character of the person and of the item parameters—e.g., continuous on the real line, continuous and positive, categorical, etc.

If we have collateral information about persons, we will often take it into account with a linear model for the location of the hyperdistribution for the person parameter(s). If we have collateral information about items, we will often take it into account with a linear model for the location of the hyperdistribution for the item or test difficulty parameter(s).

The Normal Distribution

For continuous data on the real line.

Parameters are mean and precision (precision=1/variance). Mean can be any real number. Precision must be positive; smaller values indicate less precision/more variance, larger values indicate more precision, tighter variance.

Formula

BUGS: `x ~ dnorm(mu,tau)`

Sufficient statistics from *n* independent observations are sample mean and sum of squares.

Pictures

Conjugate priors: Normal for the mean, if precision is known. Gamma for precision.

Minimally informative prior: Precision -> 0

The Gamma Distribution

For continuous data on the positive half line.

Conjugate prior for precision in the normal distribution.

Formula

Parameters: a & b.

BUGS: `x ~ dgamma(r, mu)`

Sufficient statistics

Relationships among (a,b), mean and variance of the gamma(a,b), and sample-size and sum-of-squares of a hypothetical sample from a normal distribution with known mean:

parameters of gamma(a,b)	moments of gamma(a,b)	correspondence to sample of observations from a normal distribution with known mean
a	mean = a/b	sample size ~ $2a$
b	variance = a/b^2	sum of squares around mean ~ $2b$

The Bernoulli Distribution

For one observation of a variable that can take only two values.

The parameter is *p*, the probability of getting a designated one of the values, arbitrarily called a ‘success’ (e.g., probability of getting Heads on a coin flip). 1-*p* is the probability of the other value. *p* is in [0,1].

Formula

BUGS: `r ~ dbern(p)` [BUGS assigns the categories values 0 and 1]

Sufficient statistics from *n* independent observations are number of trials and count (or, equivalently, proportion) of successes.

Conjugate prior is the Beta distribution (see below).

The Binomial Distribution

Count of successes in *n* independent observations of a Bernoulli process; e.g., number of heads observed in 12 coin flips.

The parameters are *n*, the number of trials, and *p*, the probability of success on each trial.

Formula

BUGS: `r ~ dbin(p, n)` [BUGS assigns categories the values 0 and 1]

Sufficient statistics from *n* independent observations are count of trials and number (or, equivalently, proportion) of successes.

Conjugate prior is the Beta distribution (see below).

The Beta Distribution

Distribution over [0,1].

Parameters are (*a,b*), both positive numbers.

Formula

Pictures

BUGS: `p ~ dbeta(a, b)`

Is conjugate prior for the *p* parameter in the Bernoulli and in the binomial.

As prior for p, can think of a-1 successes in (a+b-2) trials of the Bernoulli process at issue.

The Categorical Distribution

Generalization of the Bernoulli to more than 2 categories: Think of one observation of a variable that can be in one of K categories.

The parameters in the categorical distribution are the probabilities for the categories: p_k, for k=1,...,K.

Formula

BUGS: `r ~ dcat(p[])` [BUGS assigns categories the values 1, 2, ..., K]

Sufficient statistics from n independent observations are number of trials and count (or, equivalently, proportion) of observations in each category.

Conjugate prior is the Dirichlet distribution (see below).

The Multinomial Distribution

Generalization of the binomial to more than 2 categories: Think of n observations of a variable that can be in one of K categories. The observation is a vector of the number of observations in each category.

The parameters in the categorical distribution are the number of trials, n, and the probabilities for the categories: p_k, for k=1,...,K.

Formula

BUGS: `r[] ~ dmulti(p[], N)`

Sufficient statistics from n independent observations are number of trials and count (or, equivalently, proportion) of observations in each category.

Conjugate prior is the Dirichlet distribution (see below).

The Dirichlet Distribution

Generalization of the beta to K variables, each on [0,1].

Parameters are (a_1,...,a_K), all positive numbers.

Formula

BUGS: `p[] ~ ddirch(alpha[])`

Is conjugate prior for the p_k parameters in the categorical.

As prior for p_ks, can think of $\Sigma a_k – k$ trials of the categorical process at issue, with the number of observations in category k being a_k-1. That is, $\Sigma a_k – k$ is like the sample size of a hypothetical experiment, and (a_k-1)/($\Sigma a_k – k$) is the proportion for category k.

Aside: The Exponential and Logit *Functions* (not *distributions*)

The exponential function and the logit function are often used in IRT, to move back and forth between the real line and the (0,1) range of probabilities. They are inverses of one another. The logit function takes *p* on (0,1) to log(p/(1-p)). The exponential function takes a real valued variable *x* to (0,1) via exp(*x*)/(1+ exp(*x*)).

The exponential function takes numbers on the real line to (0,1). In many IRT models, a linear model on the real line is used to combine one or more parameters each for people and items. The result gets mapped onto (0,1) via the logit function to represent a probability of a correct response. This is then used as the *p* parameter of a Bernoulli distribution, to characterize the probability of a correct response by that person to that item.

The Log-Normal Distribution

For continuous data on the positive half line.

It's obtained by taking the natural logs of variables from a normal distribution, so it is easy to think about the parameters using normal-distribution intuition. Its parameters are the mean and precision of the normal distribution before the log transformation.

Formula

BUGS: `r ~ dlnorm(mu, tau)`

Easy-to-think-about hyper-priors for the parameters of the log-normal are normal for its mean and gamma for its precision.

Can be used as the prior for item slope parameters in the 3-parameter IRT BILOG and Parscale, because it is easy to take derivatives of.

The Logit-Normal Distribution

For continuous data on (0,1). Similar unimodal shape and range as Beta distributions, for Beta distributions that have both parameters greater than 1.

The exponential function takes a real valued variable *x* to (0,1) via exp(*x*)/(1+ exp(*x*)).

The Logit-Normal Distribution is obtained by taking the exponential function of variables from a normal distribution, so it is easy to think about the parameters using normal-distribution intuition. Its parameters are the mean and precision of the normal distribution on the real line before the transformation.

Formula

BUGS: `logit(r) ~ dnorm(mu, tau)`

Easy-to-think-about hyper-priors for the parameters of the logit-normal are normal for its mean and gamma for its precision.

Can be used as the prior for item asymptote (*c*) parameters in the 3-parameter logistic IRT model.

The Poisson Distribution

Count of occurrences of some event in a given period of time. It is assumed that the probability of occurrence at any instant is the same. The possible values are integers, 0, 1, 2, ...

The parameter is λ, and higher values imply more occurrences.

Formula

BUGS: `r ~ dpois(lambda)`

Both the mean and variance of the Poisson distribution are λ.

Sufficient statistics from n independent observations are counts of occurrences.

Rasch developed a Poisson measurement model for counts of reading errors (Rasch, 1960/1980). The λ for the count of errors for a given person reading a given text is the quotient of a text-difficulty parameter over a person-ability parameter.

Conjugate prior for λ is the gamma distribution.

The Exponential Distribution

Distribution for the waiting time of an occurrence of an event that follows a Poisson process. The possible values are positive real numbers. The parameter is λ, and higher numbers imply shorter waiting times.

The exponential distribution is fundamental in engineering in reliability analysis; $1/\lambda$ is the Mean Time To Failure.

Formula

BUGS: `x ~ dexp(lambda)`

The Exponential distribution is a special case of the Gamma distribution, with a=1. Can use gamma formulas to figure out its mean and variance.

The gamma distribution is the conjugate prior for λ.

References

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research/Chicago: University of Chicago Press (reprint).