

Probability Review

Bolker 2008, Chapter 4

Gavin Fay

09/05/2019

Objectives

- ▶ Review probability laws
- ▶ Review definitions of expected value and variance of random variables
- ▶ Present common probability distributions

Why does variability matter?

Variability affects any ecological system.

Noise affects ecological data in two ways:

- measurement error
- process noise

Measurement error is variability in our measurements.

- leads to large confidence intervals and low power

Process noise (process error), variability in the system.

- demographic stochasticity
- environmental stochasticity

We are interested in understanding patterns in our data.

- use probability to describe relationships between processes and data.

Often assume that our data is generated by some stochastic process whose expected value is a function of covariates we are interested in.

Basic probability theory

The *sample space* is the set of all possible outcomes that could occur.

e.g. for a regular six-sided die

$$s\{1, 2, 3, 4, 5, 6\}$$

Probability of an event A is the frequency with which that event occurs.

e.g.

$$P(1) = 1/6$$

Laws of Probability

1. Law of total probability

The probabilities of all possible outcomes of an observation or experiment add to 1.0

$$P(\text{heads}) + P(\text{tails}) = 1.0$$

2. Probability of A or B , or $P(A \cup B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

3. Mutually exclusive vs. independent events

- ▶ two mutually exclusive events cannot be independent
- ▶ mutually exclusive $\implies P(A \cap B) = 0$
- ▶ independence $\implies P(A \cap B) = P(A) \cdot P(B) \neq 0$

Laws of Probability

4. General multiplication rule

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1, A_2) \dots$$

5. Conditional probability

$P(A|B)$, is the probability that A happens if we know or assume B happens.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional probability leads to Bayes' rule

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B \cap A) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

This is mostly termed with A being the model (hypothesis) and B being the data.

i.e. what is the probability of a hypothesis given the data.

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)}$$

with $P(D) = \sum P(D|H) \cdot P(H)$

Random Variables

A random variable is a numerical valued function defined over a sample space.

The probability distribution describes how the frequency of occurrence varies across the sample space.

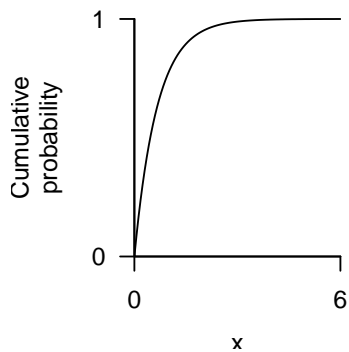
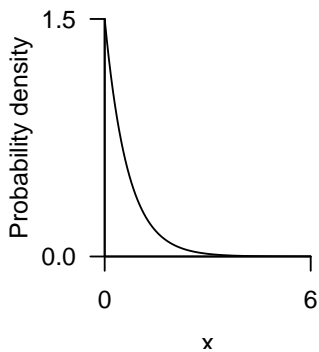
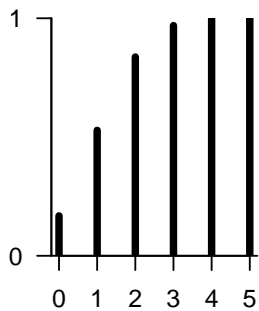
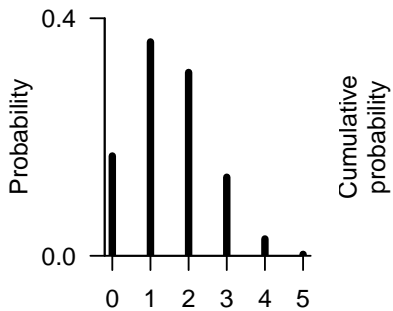
For discrete variables, characterized by $f(x)$,
- the **probability distribution function** (discrete variables)

$$f(x) = \text{Prob}(X = x)$$

(for continuous variables, $f(x)$ is the **probability density function**)

Both types of variables are also described by the **cumulative distribution function**, $F(x)$

$$F(x) = P(X \leq x)$$



Expected Value of Random Variable X

Discrete random variables

$$\mu = E(X) = \sum_{i=0}^{\infty} x_i P(X = x_i)$$

Continuous random variables

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Variance of a Random Variable X , $E[(X - \mu)^2]$

Discrete random variables

$$\text{Var}(X) = \sum_{i=0}^{\infty} (x_i - E(x_i))^2 P(X = x_i)$$

Continuous random variables

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x_i - E(x_i))^2 f(x) dx$$

In general

$$\text{Var}(X) = E(X^2) - (E(X))^2 = E((X - \mu)^2)$$

Variances are additive.

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$$

The *standard deviation* of a distribution is $\sqrt{\text{Var}}$

The *coefficient of variation* (CV) is $\sqrt{\text{Var}}/\mu$

Summary of probability distributions

Binomial

Describes the number of successes from a fixed number of trials.

Two possible outcomes on each trial, success or failure.

Probability of success is the same in each trial.

Range: discrete, $0 \leq x \leq N$

Distribution:

$$\binom{N}{x} p^x (1 - p)^{N-x}$$

R: `dbinom` `pbinom` `qbinom` `rbinom`

Parameters:

- p [real, 0-1], probability of success [prob]
- N [positive integer], number of trials [size]

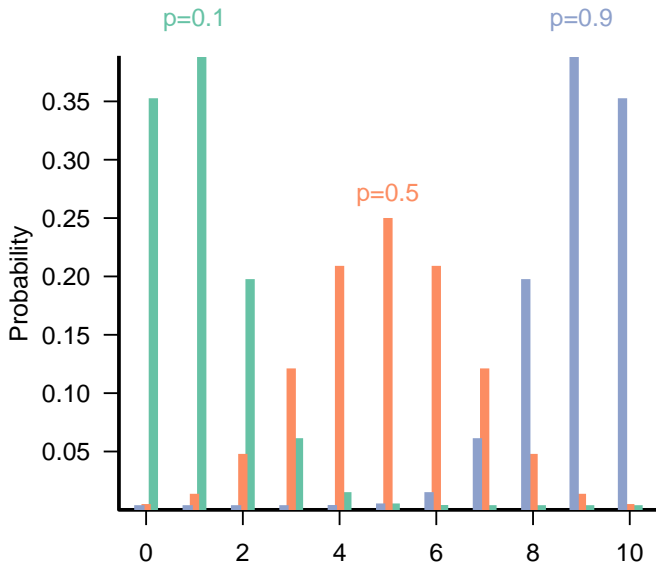
Mean: Np

Variance: $Np(1 - p)$

CV: $\sqrt{(1 - p)/(Np)}$

Conjugate prior: Beta

Binomial



Multinomial

Extension of binomial trials to three or more possible outcomes.

$$X = (X_1, X_2, \dots, X_k)$$

Range: discrete, $0 \leq x_i \leq N$

Distribution:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \binom{N}{x_1, x_2, \dots, x_k} \prod_{i=1}^k p_i^{x_i}$$

R: `dbinom` `pbinom` `qbinom` `rbinom`

Parameters:

- p_i [real, 0-1], $\sum_{i=1}^k p_i = 1$
- N [positive integer], number of samples

$$E(X_i) = Np_i$$

$$\text{Var}(X_i) = Np_i(1 - p_i)$$

$$\text{Cov}(X_i, X_j) = -Np_i p_j, \quad i \neq j$$

Poisson

Describes events which occur randomly and independently in time.

Limit of a binomial distribution in which:

$N \rightarrow \infty, p \rightarrow 0$ while $Np = \mu$ is fixed.

Distribution of “rare events” (i.e., $p \rightarrow 0$).

Range: discrete ($0 \leq x$)

Distribution:

$$\frac{e^{-\lambda} \lambda^n}{n!} \text{ or } \frac{e^{-rt} (rt)^n}{n!}$$

R: dpois, ppois, qpois, rpois

Parameters: λ (real, positive), expected number per sample
[lambda] or r (real, positive), expected number per unit effort,
area, time, etc. (*arrival rate*)

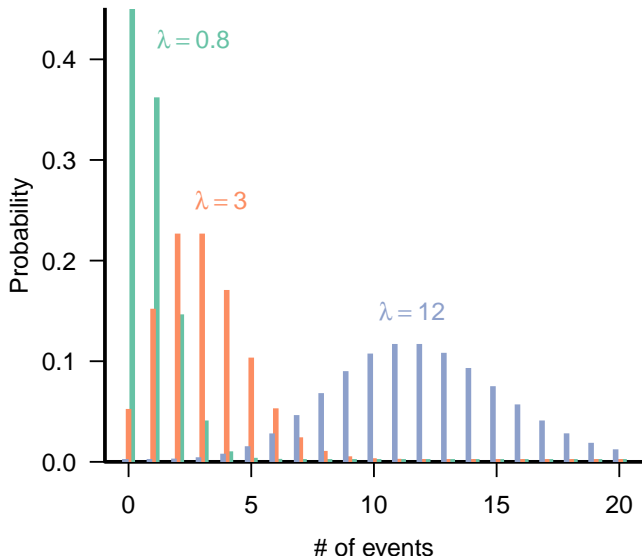
Mean: λ (or rt)

Variance: λ (or rt)

CV : $1/\sqrt{\lambda}$ (or $1/\sqrt{rt}$)

Conjugate prior: Gamma

Poisson



Negative Binomial

For binomial trials, the number of failures before n successes.

In ecology, most often used because it is discrete like the Poisson but the variance can be greater than the mean (*overdispersed*).

Range: discrete, $x \geq 0$

Distribution:

$$P(X = x) = \frac{(n + x - 1)!}{(n - 1!)x!} p^n (1 - p)^x$$

or

$$\frac{\Gamma(k + x)}{\Gamma(k)x!} (k/(k + \mu))^k (\mu/(k + \mu))^x$$

Parameters:

p ($0 < p < 1$) probability per trial [prob]

or μ (real, positive) expected number of counts [mu]

n (positive integer) number of successes awaited [size]

or k (real, positive), overdispersion parameter [size]

(= shape parameter of underlying heterogeneity)

Negative Binomial

R: `dnbinom`, `pnbinom`, `qnbinom`, `rnbinom`

Mean: $\mu = n(1 - p)/p$

Variance: $\mu + \mu^2/k = n(1 - p)/p^2$

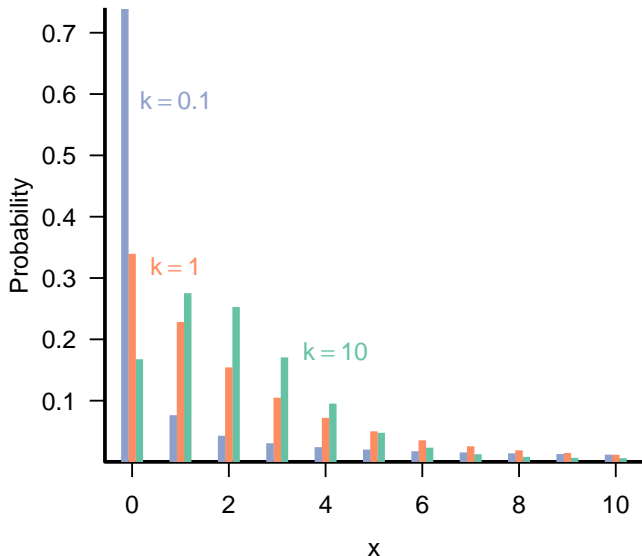
CV: $\sqrt{\frac{(1+\mu/k)}{\mu}} = 1/\sqrt{n(1 - p)}$

Conjugate prior: No simple conjugate prior (Bradlow et al. 2002)

To use the 'ecology' parameterization in R you *must* name `mu` explicitly.

The negative binomial is also the result of a Poisson sampling process where λ is Gamma-distributed.

Negative Binomial ($\mu = 2$ all cases)



Continuous Probability Distributions

Uniform distribution

Constant probability across a range with limits a and b

Standard uniform, $U(0,1)$, frequently used as building block.

Range: $a \leq x \leq b$

Distribution: $1/(b - a)$

R: `dunif`, `punif`, `qunif`, `runif`

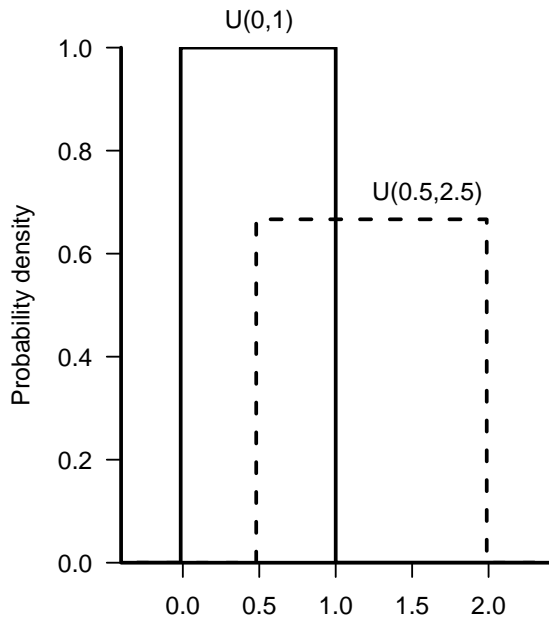
Parameters: minimum (a) and maximum (b) limits (real)
[`min`, `max`]

Mean: $(a + b)/2$

Variance: $(b - a)^2/12$

CV: $(b - a)/((a + b)\sqrt{3})$

Uniform distribution



Normal Distribution

Arises from adding things together.

Sum of a large number of independent samples from the same distribution is approximately normal.

Limit of many distributions (binomial, Poisson, negative binomial, Gamma).

Range: all real values

Distribution: $\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

R: `dnorm`, `pnorm`, `qnorm`, `rnorm`

Parameters:

- μ (real), mean [`mean`]

- σ (real, positive), standard deviation [`sd`]

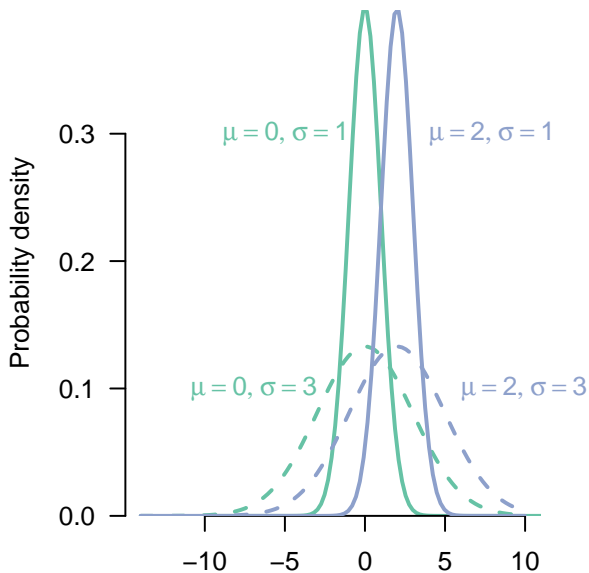
Mean: μ

Variance: σ^2

CV: σ/μ

Conjugate prior: Normal (μ); Gamma ($1/\sigma^2$)

Normal distribution



Gamma

Distribution of waiting times until a certain number of events occurs.

Continuous counterpart to the negative binomial.

Gamma is very useful. Continuous positive variable with large variance and (possible) skew.

Range: positive real values

R: `dgamma`, `pgamma`, `qgamma`, `rgamma`

Distribution: $\frac{1}{s^a \Gamma(a)} x^{a-1} e^{-x/s}$

Parameters:

s (real, positive), scale: length per event [scale]

or r (real, positive), rate = $1/s$; rate at which events occur [rate]

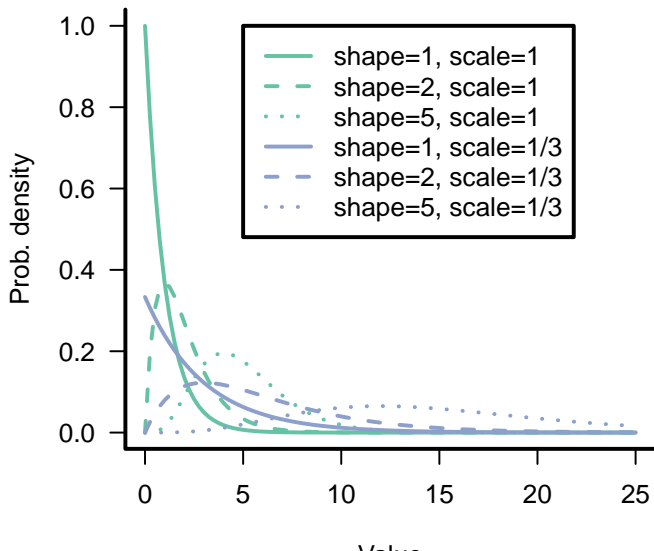
a (real, positive), shape: number of events [shape]

Mean: as or a/r

Variance: as^2 or a/r^2

CV: $1/\sqrt{a}$

Gamma



Beta

Continuous distribution related to the binomial.

Distribution of *probability* of success in a binomial trial with $a - 1$ successes and $b - 1$ failures.

Very useful in modeling probabilities or proportions.

Range: real, 0 to 1

R: dbeta, pbeta, qbeta, rbeta

Density: $s \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$

Parameters:

- a (real, positive), shape 1: number of successes +1 [shape1]

- b (real, positive), shape 2: number of failures +1 [shape2]

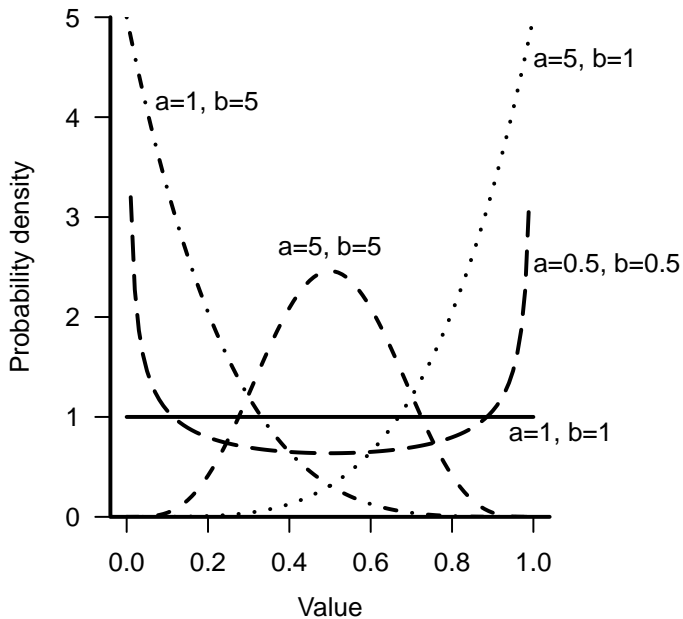
Mean: $a/(a+b)$

Mode: $(a-1)/(a+b-2)$

Variance: $ab/((a+b)^2)(a+b+1)$

CV: $\sqrt{(b/a)/(a+b+1)}$

Beta



Lognormal

Not a continuous analogue or limit of some discrete distribution.

Justification: as for normal, but for *product* of many iid variables.

Used in many situations where Gamma also fits, continuous, positive distribution with long tail or variance $>$ mean.

Range: positive real values

R: `dlnorm`, `plnorm`, `qlnorm`, `rlnorm`

Density: $\frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2 / (2\sigma^2)}$

Parameters:

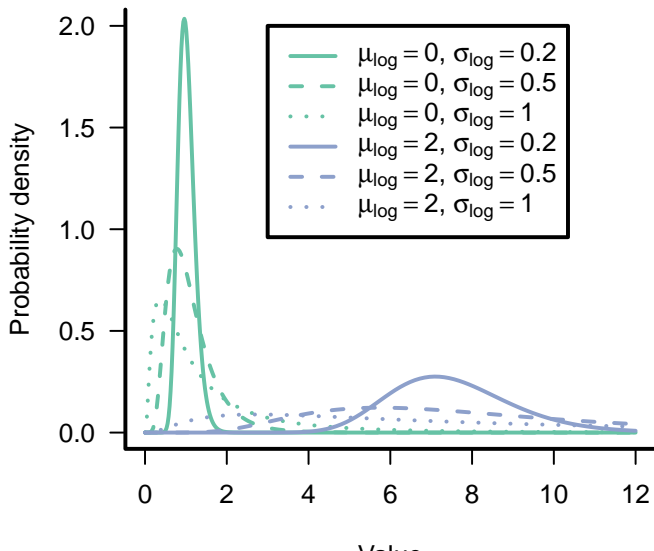
- μ (real): mean of the logarithm [`meanlog`]
- σ (real): standard deviation of the logarithm [`sdlog`]

Mean: $\exp(\mu + \sigma^2/2)$

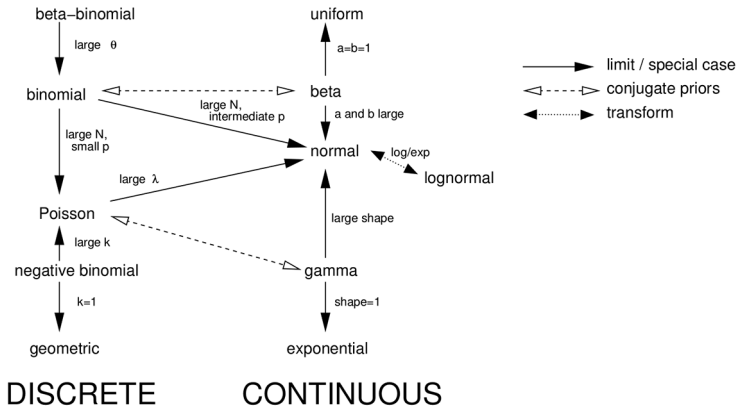
Variance: $\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$

CV: $\sqrt{\exp(\sigma^2) - 1}$ ($\approx \sigma$ when $\sigma < 1/2$)

Lognormal



Relationships among distributions



Other common distributions

Discrete

- Geometric (negative binomial with $k = 1$)
- Beta-binomial (binomial but with p being beta distributed)
- Hypergeometric (useful for sampling without replacement, finite population)
- Multivariate hypergeometric (similar to the multinomial)

Continuous

- Exponential (distribution of waiting times for a single event)
- Pareto (quantity whose log is exponentially distributed, power laws!)
- Chi square (distribution of a sum of squared standard normals)
- Student's t (ratio of a standard normal and the square root of a scaled chi square)
- F (ratio of two scaled chi-squares)
- Dirichlet (generalization of beta, for a vector that must sum to 1)
- Wishart (generalization of gamma, for a symmetric non-negative definite matrix)