Probability Distributions and Moment Matching

ESS 575 Models for Ecological Data

N. Thompson Hobbs

February 7, 2019

Graduate Degree Program in Ecology
Department of Ecosystem Science and Sustainability
Colorado State University

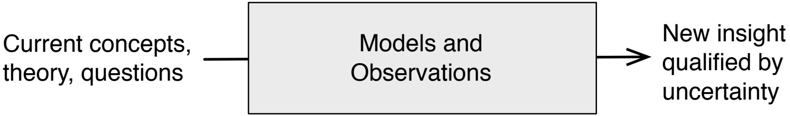
Housekeeping

Errors in text:

```
http://www.stat.colostate.edu/~hooten/papers/pdf/
Hobbs_Hooten_Bayesian_Models_2015_errata.pdf
```

Note in particular that the plot for a cumulative distribution function is wrong.

Motivation: A general approach to scientific research



Roadmap

- The rules of probability
  - conditional probability and independence
  - the law of total probability
  - the chain law of probability
- Directed acyclic graphs (Bayesian networks)
- Probability distributions for discrete and continuous random variables
- Marginal distributions
- Moment matching

What you must know and why

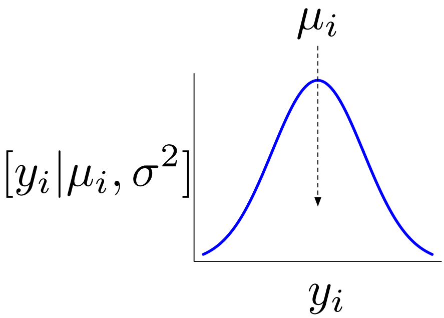
Concept to be taught	Why do you need to understand this concept?
Conditional probability	It is the foundation for Bayes' Theorem and all inferences we will make.
The law of total probability	Basis for the denominator of Bayes' Theorem $[y]$
Factoring joint distributions	This is the procedure we will use to build models.
Independence	Allows us to simplify fully factored joint distributions.
Probability distributions	Our toolbox for fitting models to data and representing uncertainty
Moments	The way we summarize distributions
Marginal distributions	Bayesian inference is based on marginal distributions of unobserved quantities.
Moment matching	Allows us to embed the predictions of models into any statistical distribution

Motivation: The essence of Bayes

Bayesian analysis is the *only* branch of statistics that treats all unobserved quantities as random variables. We seek to understand the characteristics of the probability distributions governing the behavior of these random variables.

Motivation: models of data

$$\mu_i = g(\boldsymbol{\theta}, x_i)$$



A model of the data describes our ideas about how the data arise.

Motivation: flexibility in analysis

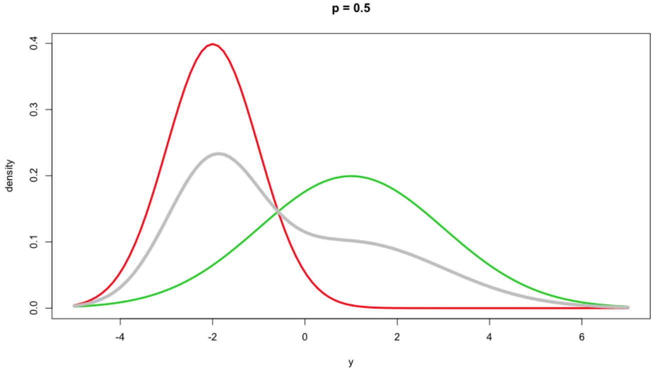
Deterministic models	Types of data
general linear	real numbers
nonlinear	non-negative real numbers
differential equations	counts
difference equations	0 to 1
auto-regressive	0 or 1
occupancy	counts in categories
state-transition	proportions in categories
integral-projection	ordinal categories

univariate and multivariate

Motivation: flexibility in analysis

Probability model	Support for random variable
normal	real numbers
multivariate normal	real numbers (vectors)
lognormal	non-negative real numbers
gamma	non-negative real numbers
beta	0 to 1 real numbers
Bernoulli	0 or 1
binomial	counts in 2 categories
Poisson	counts
multinomial	counts in $>$ 2 categories
negative binomial	counts
Dirichlet	proportions in $\geq$ 2 categories
Cauchy	real numbers

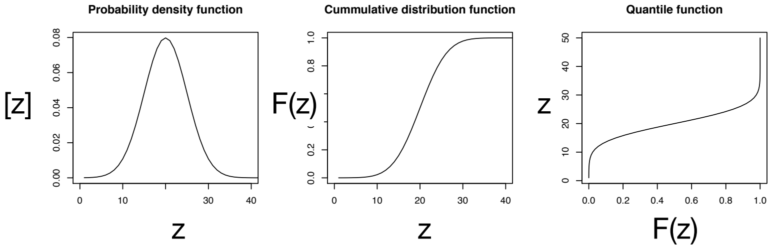
Motivation: flexibility in analysis



Work flow: probability distributions

- General properties and definitions (today)
  - discrete random variables
  - continuous random variables
- Specific distributions (cheat sheet and lab exercises)
- Marginal distributions (Thursday and lab exercise)
- Moment matching (Thursday and lab exercise)

How will we use probability distributions?



Used to fit models to data, to represent uncertainty in processes and parameters, and to portray prior information

Used to make inference

Key points for today

1. What makes a function a probability mass function or a probability density function?
2. How to compute moments of distributions?
3. How to approximate moments of distributions from random draws?
4. Relationships among:
    - 4.1 probability mass function
    - 4.2 probability density function
    - 4.3 cumulative distribution function
    - 4.4 quantile function

Board work on general concepts of probability distributions with previous slide on screens.

Key points for today

1. What makes a function a probability mass function or a probability density function?
2. How to compute moments of distributions?
3. How to approximate moments of distributions from random draws.
4. Relationships among:
   4.1 probability mass function
   4.2 probability density function
   4.3 cumulative distribution function
   4.4 quantile function

Break here for marginal distributions and moment matching

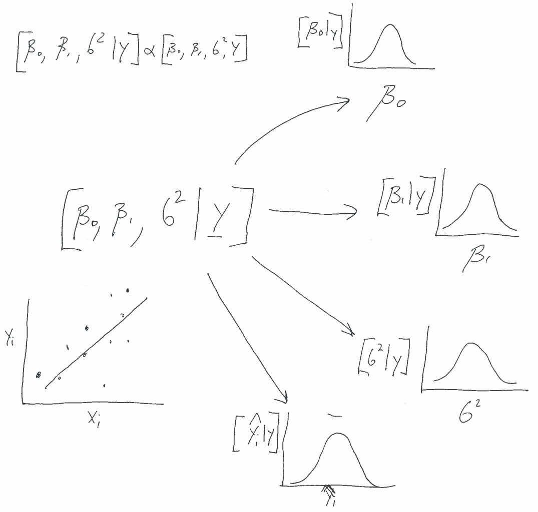
Work flow: probability distributions

- General properties and definitions
  - discrete random variables
  - continuous random variables
- Specific distributions (in lab)
- Marginal distributions (Thursday and lab exercise)
- Moment matching (Thursday and lab exercise)

Key points

1. Marginal distributions summarize multivariate, joint distributions as univariate distributions.
2. Moment matching allows us to compute moments of distributions in terms of parameters and parameters of distributions in terms of moments.

How will we use marginal distributions?



Marginal distributions of discrete random variables

$$
\begin{array}{cccccc}
 & a_1 & a_2 & a_3 & a_4 & \Pr(B)\downarrow \\
b_1 & \frac{2}{20} & 0 & \frac{1}{20} & \frac{3}{20} & \frac{6}{20} \\
b_2 & 0 & \frac{4}{20} & 0 & \frac{6}{20} & \frac{10}{20} \\
b_3 & \frac{1}{20} & \frac{1}{20} & \frac{1}{20} & 0 & \frac{3}{20} \\
b_4 & 0 & 0 & 0 & \frac{1}{20} & \frac{1}{20} \\
\Pr(A)\rightarrow & \frac{3}{20} & \frac{5}{20} & \frac{2}{20} & \frac{10}{20} & \frac{20}{20}
\end{array}
$$

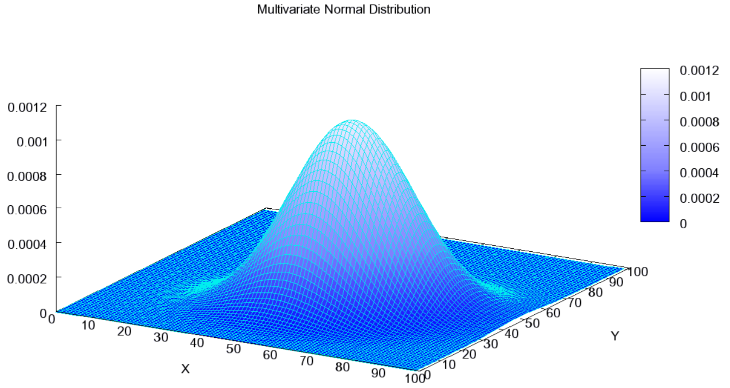
Marginal distributions of discrete random variables

If we have a function $[A, B]$ specifying the joint probability of the discrete random variables A and B, then
$\sum_A [A, B]$ is the marginal probability of B
and
$\sum_B [A, B]$ is the marginal probability of A.
This same idea applies to any number of jointly distributed random variables. We simply sum over all but one.

Marginal distributions of discrete random variables

Willow establishment example on board

Joint distribution of continuous random variables



Marginal distributions of continuous random variables

Exercise: If A and B are continuous random variables and we have a function [A,B] that gives their joint probability density, what is the marginal distribution of A? Of B?
Write two equivalent equations for these marginal distributions.

Marginal distributions of continuous random variables

If we have a function $[A, B]$ specifying the joint probability of the discrete random variables A and B, then

$\int_A [A, B] dA = \int_A [B|A][A] dA$ is the marginal probability of B

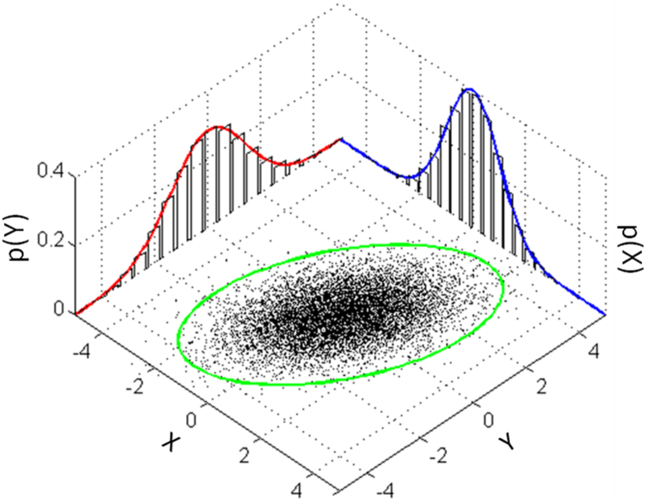
and

$\int_B [A, B] dB = \int_B [A|B][B] dB$ is the marginal probability of A.

This same idea applies to any number of jointly distributed random variables. We simply integrate over all but one.

Integrating over all but one random variable is often referred to as "integrating out."

Marginal distributions of continuous random variables



By Bscan - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=25235145

Motivation: flexibility in analysis

Deterministic models	Types of data
general linear	real numbers
nonlinear	non-negative real numbers
differential equations	counts
difference equations	0 to 1
auto-regressive	0 or 1
occupancy	counts in categories
state-transition	proportions in categories
integral-projection	ordinal categories

univariate and
multivariate

Motivation: flexibility in analysis

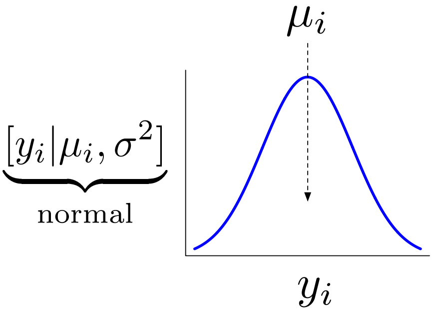
Probability model	Support for random variable
normal	all real numbers
lognormal	non-negative real numbers
gamma	non-negative real numbers
beta	0 to 1 real numbers
Bernoulli	0 or 1
binomial	counts in 2 categories
Poisson	counts
multinomial	counts in >2 categories
negative binomial	counts
Dirichlet	proportions in ≥ 2 categories
Cauchy	real numbers

$$\mu_i = g(\boldsymbol{\theta}, x_i)$$

A familiar approach

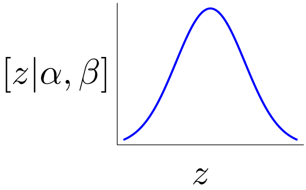
$$\boldsymbol{\theta} = (\beta_0, \beta_1)'$$

$$\mu_i = g(\boldsymbol{\theta}, x_i) = \beta_0 + \beta_1 x_i$$



The problem

All distributions have parameters:



α and β are parameters of the distribution of the random variable z.

Types of parameters

Parameter name	Function
intensity, centrality, location	sets position on x axis
shape	controls dispersion and skew
scale, dispersion parameter	shrinks or expands width
rate	scale^{-1}

The problem

The normal and the Poisson are the only distributions for which the parameters of the distribution are the *same* as the moments. For all other distributions, the parameters are *functions* of the moments.

$$\begin{aligned} \alpha &= m_1(\mu, \sigma^2) \\ \beta &= m_2(\mu, \sigma^2) \end{aligned}$$

We can use these functions to “match”' the moments to the parameters.

Moment matching

$$
\begin{aligned}
\boldsymbol{\mu}_i &= g(\boldsymbol{\theta}, x_i) \\
\boldsymbol{\alpha} &= m_1(\boldsymbol{\mu}_i, \boldsymbol{\sigma}^2) \\
\boldsymbol{\beta} &= m_2(\boldsymbol{\mu}_i, \boldsymbol{\sigma}^2) \\
& [y_i | \boldsymbol{\alpha}, \boldsymbol{\beta}]
\end{aligned}
$$

Moment matching the gamma distribution

The gamma distribution: $[z|\alpha, \beta] = \frac{\beta^{\alpha} z^{\alpha-1} e^{-\beta z}}{\Gamma(\alpha)}$
The mean of the gamma distribution is

$$\mu = \frac{\alpha}{\beta}$$

and the variance is

$$\sigma^2 = \frac{\alpha}{\beta^2}.$$

Discover functions for α and β in terms of μ and σ^2.

Note: $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha} e^{-t} \frac{\mathrm{d}t}{t}$

Answer

$1)\mu = \frac{\alpha}{\beta}$
$2)\sigma^2 = \frac{\alpha}{\beta^2}$
Solve 1 for β, substitue for β in 2), solve for α :
$3)\ \alpha = \frac{\mu^2}{\sigma^2}$
Substitute rhs 3) for α in 2), solve for β :
$4)\ \beta = \frac{\mu}{\sigma^2}$

Moment matching the beta distribution

The beta distribution gives the probability density of random variables with support on $0, \ldots, 1$.

$$[z|\alpha,\beta] = \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha,\beta)}$$
$$B = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$\mu = \frac{\alpha}{\alpha+\beta}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\alpha = \frac{\mu^2 - \mu^3 - \mu\sigma^2}{\sigma^2}$$

$$\beta = \frac{\mu - 2\mu^2 + \mu^3 - \sigma^2 + \mu\sigma^2}{\sigma^2}$$

You need some functions...

```
#BetaMomentMatch.R
# Function for parameters from moments
shape_from_stats <- function(mu, sigma){
 a <-(mu^2-mu^3-mu*sigma^2)/sigma^2
 b <- (mu-2*mu^2+mu^3-sigma^2+mu*sigma^2)/sigma^2
shape_ps <- c(a,b)
return(shape_ps)
}
# Functions for moments from parameters
beta.mean=function(a,b)a/(a+b)
beta.var = function(a,b)a*b/((a+b)^2*(a+b+1))
```

Moment matching for a single parameter

We can solve for α in terms of μ and β,

$$\mu = \frac{\alpha}{\alpha+\beta} \quad (1)$$

$$\alpha = \frac{\mu\beta}{1-\mu}, \quad (2)$$

which allows us to use

$$\mu_i = g(\theta, x_i) \quad (3)$$

$$y_i \sim \text{beta}\left(\frac{\mu_i\beta}{1-\mu_i}, \beta\right) \quad (4)$$

to moment match the mean alone.

Moment matching for a single parameter

The first parameter of the lognormal $= \alpha$, the mean of the random variable on the log scale. The second parameter $= \sigma^2_{\log}$, the variance of the random variable on the log scale
We often moment match the median the lognormal distribution:

$$
\begin{aligned}
\text{median} = \mu_i &= g(\theta, x_i) && (5)\\
\mu &= e^{\alpha} && (6)\\
\alpha &= \log(\mu_i) && (7)\\
y_i &\sim \text{lognormal}(\log(\mu_i), \sigma^2_{\text{log}}) && (8)
\end{aligned}
$$

In this case, σ^2 remains on log scale.

Problems continued

Do section on Moment Matching