

# Probability Concepts and Distributions

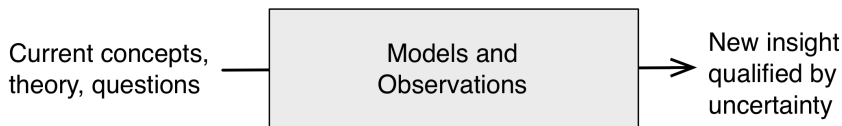
## Models for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, N. Thompson Hobbs

June 2, 2019



## 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

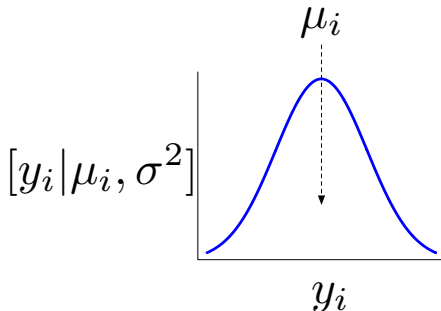
Concept to be taught	Why do you need to understand this concept?
Conditional probability	It is the foundation for Bayes' Theorem and all inferences we will make.
The law of total probability	Basis for the denominator of Bayes' Theorem $[y]$
Factoring joint distributions	This is the procedure we will 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.

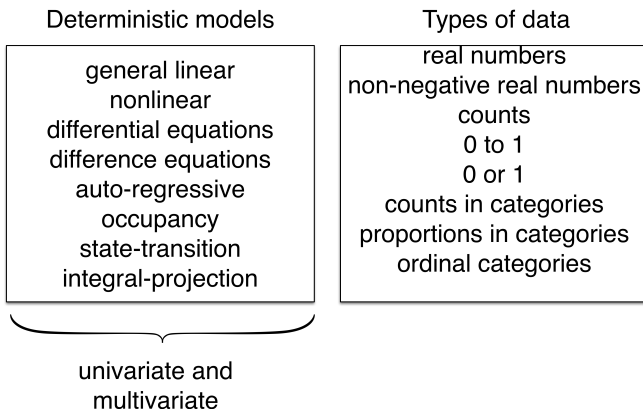
## Motivation: models of data

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



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

## Motivation: flexibility in analysis



# Motivation: flexibility in analysis

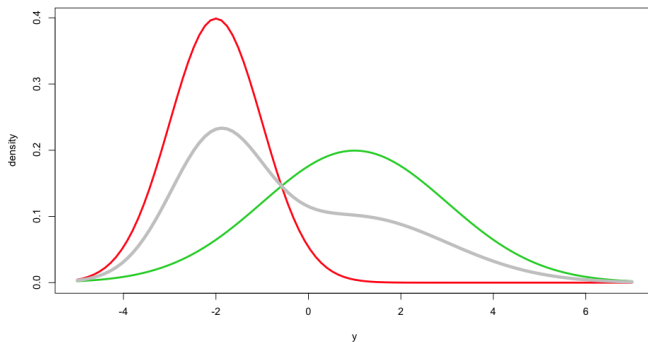
Probability model	Support for random variable
normal	real numbers
multivariate normal	real numbers (vectors)
lognormal	non-negative real numbers
gamma	non-negative real numbers
beta	0 to 1 real numbers
Bernoulli	0 or 1
binomial	counts in 2 categories
Poisson	counts
multinomial	counts in $> 2$ categories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



## Motivation: flexibility in analysis



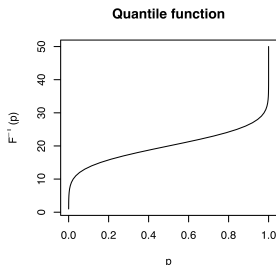
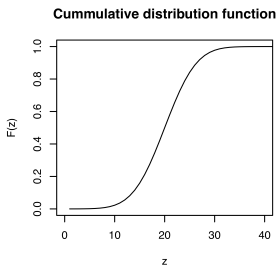
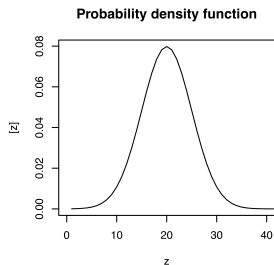
$p = 0.5$



## Work flow: probability distributions

- ▶ General properties and definitions (today)
  - ▶ discrete random variables
  - ▶ continuous random variables
- ▶ Specific distributions (cheat sheet and Probability Lab 2)
- ▶ Marginal distributions (Probability Lab 3)
- ▶ Moment matching (Probability Lab 4)

# 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

# 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.

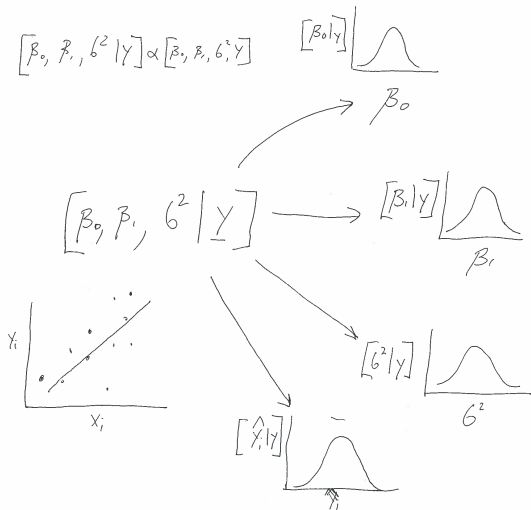
Break here for marginal distributions and moment matching

# Work flow: probability distributions

- ▶ General properties and definitions
  - ▶ discrete random variables
  - ▶ continuous random variables
- ▶ Specific distributions (cheat sheet and Probability Lab 2)
- ▶ Marginal distributions (Probability Lab 3)
- ▶ Moment matching (Probability Lab 4)



## How will we use marginal distributions?



# Marginal distributions of discrete random variables

	$a_1$	$a_2$	$a_3$	$a_4$	$\Pr(B) \downarrow$
$b_1$	$\frac{2}{20}$	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	$\frac{1}{20}$	$\frac{1}{20}$
$\Pr(A) \rightarrow$	$\frac{3}{20}$	$\frac{5}{20}$	$\frac{2}{20}$	$\frac{10}{20}$	$\frac{20}{20}$

## Marginal distributions of discrete random variables

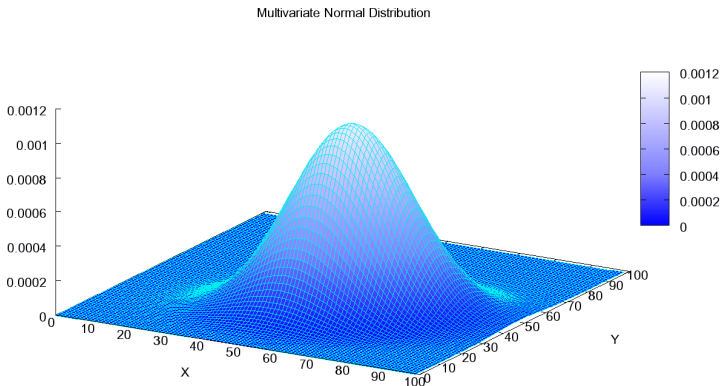
If we have a function  $[A, B]$  specifying the joint probability of the discrete random variables A and B, then

$\sum_A [A, B]$  is the marginal probability of B  
and

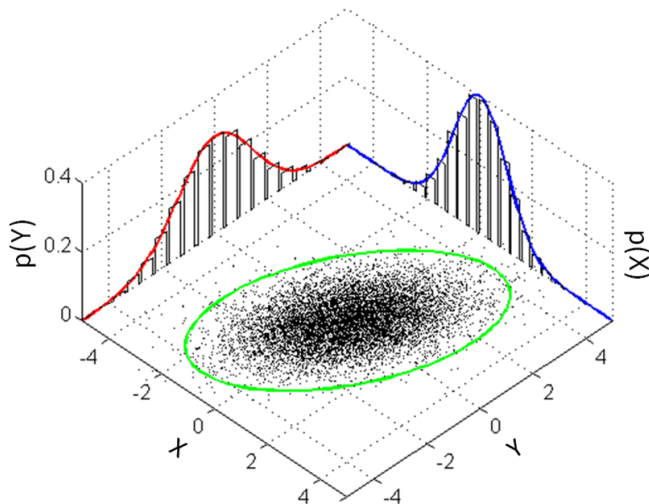
$\sum_B [A, B]$  is the marginal probability of A.

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

# Joint distribution of continuous random variables

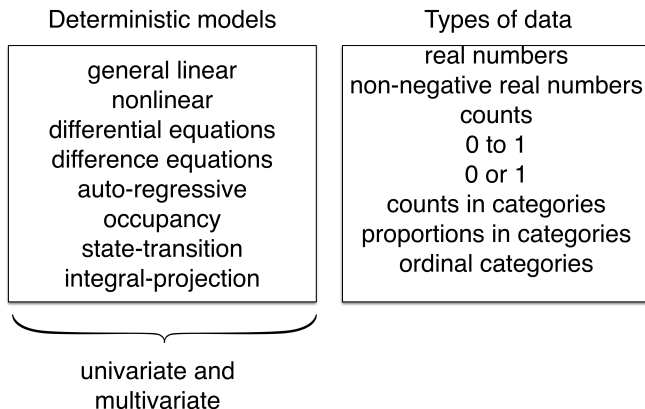


# Marginal distributions of continuous random variables



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

# Motivation: flexibility in analysis



## Motivation: flexibility in analysis

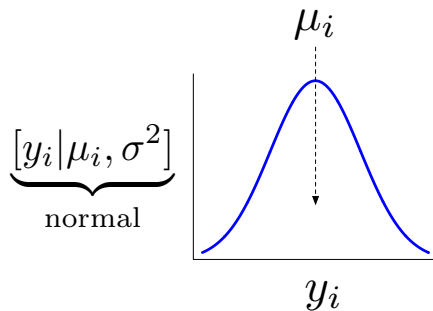
Probability model	Support for random variable
normal	all real numbers
lognormal	non-negative real numbers
gamma	non-negative real numbers
beta	0 to 1 real numbers
Bernoulli	0 or 1
binomial	counts in 2 categories
Poisson	counts
multinomial	counts in $> 2$ categories
negative binomial	counts
Dirichlet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

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

## A familiar approach

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

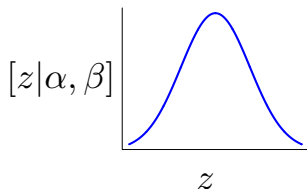
$$\mu_i = g(\boldsymbol{\theta}, x_i) = \beta_0 + \beta_1 x_i$$





# The problem

All distributions have parameters:



$\alpha$  and  $\beta$  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 <sup>-1</sup>

# 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.

$$\alpha = m_1(\mu, \sigma^2)$$

$$\beta = m_2(\mu, \sigma^2)$$

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

# Moment matching

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

# 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  $\alpha$  and  $\beta$  in terms of  $\mu$  and  $\sigma^2$ .

Note:  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \frac{dt}{t}$

# Moment matching the beta distribution

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

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

$$B = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

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

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

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

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

## 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  $\alpha$  in terms of  $\mu$  and  $\beta$ ,

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

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

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 =  $\sigma_{\log}^2$ , the variance of the random variable on the log scale

We often moment match the median the lognormal distribution:

$$\text{median} = \mu_i = g(\theta, x_i) \quad (5)$$

$$\mu = e^{\alpha} \quad (6)$$

$$\alpha = \log(\mu_i) \quad (7)$$

$$y_i \sim \text{lognormal}(\log(\mu_i), \sigma_{\log}^2) \quad (8)$$

In this case,  $\sigma^2$  remains on log scale.

## Problems continued

Do Moment Matching Lab