University of Nevada, Reno

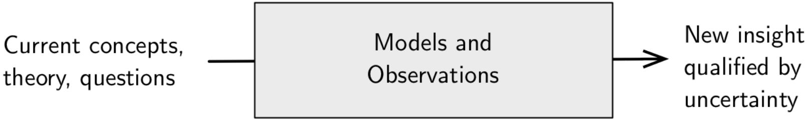
Lecture 5: Probability Distributions and Moment Matching

Perry Williams, PhD

NRES 779

Bayesian Hierarchical Modeling in Natural Resources

Motivation: A general approach to scientific research



Road-map

- The rules of probability
  - Conditional probability and independence
  - The law of total probability
  - The chain rule 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

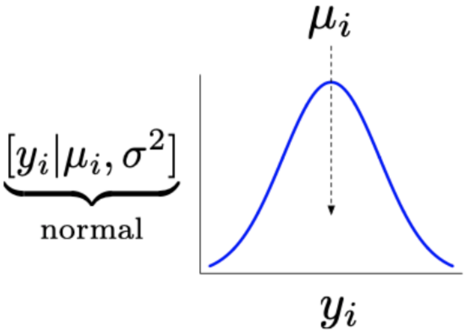
Concept to be taught	Why do you need to understand this concept?
Conditional probability	The foundation for Bayes' Theorem
The law of total probability	Basis for the denominator of Bayes' Theorem $[y]$
Independence	Allows us to simplify fully factored joint distributions.
Factoring	The procedure for building hierarchical models
Chain rule of probability	The algebra of hierarchical factoring joint distributions
Probability distributions	Our toolbox for applying the Bayesian approach
Moments	Allow us to summarize properties of probability distributions. Basis for inference from MCMC.
Marginal distributions	Bayesian inference is based on marginal distributions of unobserved quantities.
Moment matching	Allows us to embed deterministic models into any probability 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.

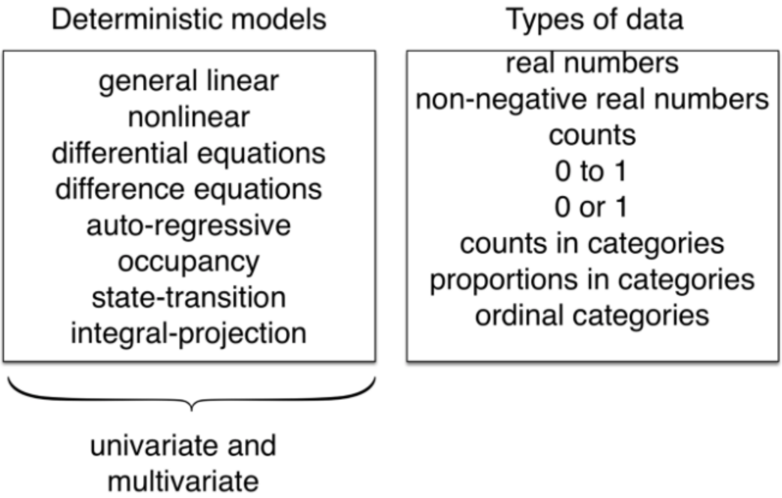
Motivation: Models of data

$\boldsymbol{\theta} = (\beta_0, \beta_1)'$
$\boldsymbol{\mu}_i = f(x_i, \boldsymbol{\theta}) = \beta_0 + \beta_1 x_1$



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

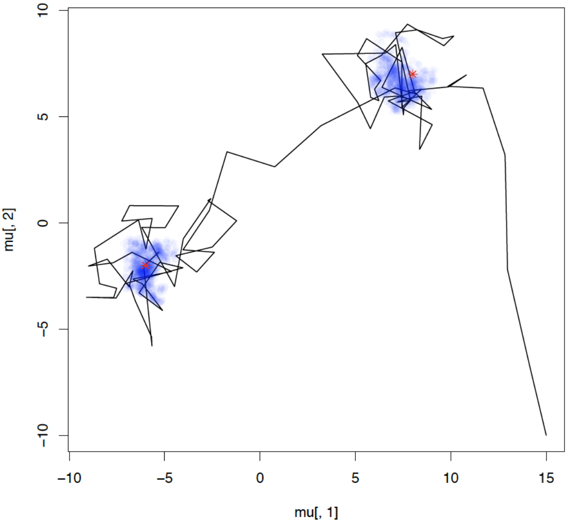
Motivation: Flexibility in analyses



Motivation: Flexibility in analyses

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 ≥ 2 categories
Cauchy	real numbers

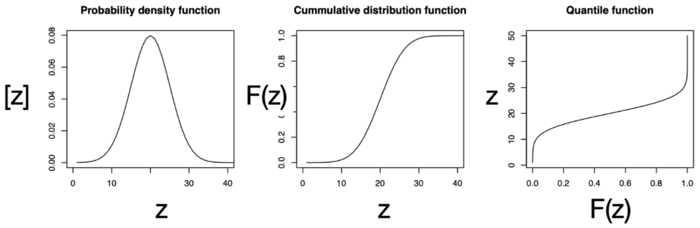
Motivation: Flexibility in analyses



Work Flow: Probability distributions

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

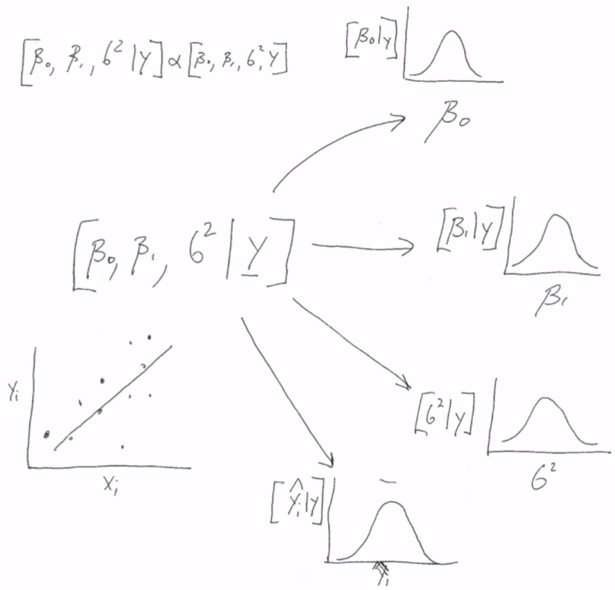
Learning Objectives

1. What makes a function a probability mass function (pmf) or a probability density function (pdf)?
2. How to compute moments of distributions?
3. How to approximate moments of distributions from random draws?
4. Relationships among:
    1. probability mass function
    2. probability density function
    3. cumulative distribution function
    4. quantile function

Key Points

1. Marginal distributions summarize multivariate, joint distributions as uni-variate 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

	a_1	a_2	a_3	a_4
b_1	$\frac{2}{20}$	0	$\frac{1}{20}$	$\frac{3}{20}$
b_2	0	$\frac{4}{20}$	0	$\frac{6}{20}$
b_3	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	0
b_4	0	0	0	$\frac{1}{20}$

Marginal Distributions of Discrete Random Variables

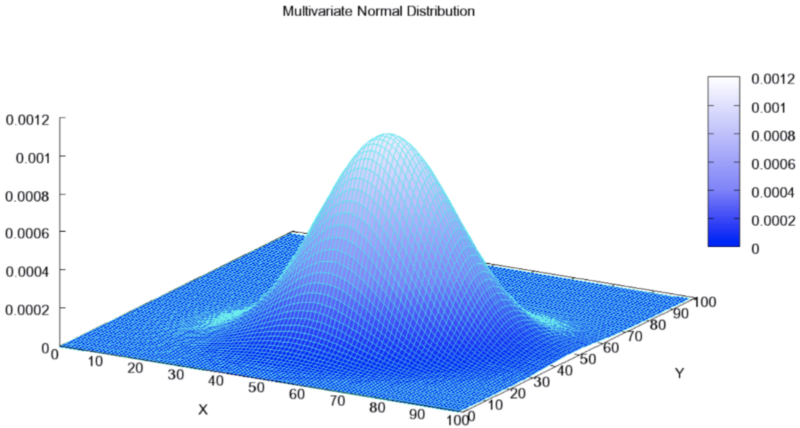
	a_1	a_2	a_3	a_4	$\Pr(B)\downarrow$
b_1	$\frac{2}{20}$	0	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{6}{20}$
b_2	0	$\frac{4}{20}$	0	$\frac{6}{20}$	$\frac{10}{20}$
b_3	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	0	$\frac{3}{20}$
b_4	0	0	0	$\frac{1}{20}$	$\frac{1}{20}$
$\Pr(A)\rightarrow$	$\frac{3}{20}$	$\frac{5}{20}$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

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
  - $\sum_B [A, B]$ 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

Joint Distributions of Continuous Random Variables



Marginal distributions of continuous random variables

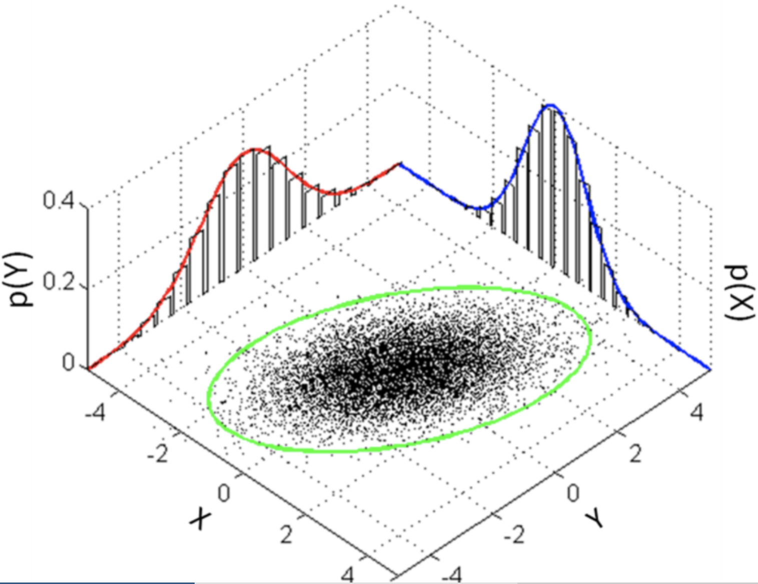
Exercise: if A and B are continuous random variables, and we have a function $[A, B]$ that gives their joint probability density, what is the marginal distribution of A? Of B?
Write two equivalent equations for these marginal distributions.

Marginal distributions of continuous random variables

- If we have a function $[A, B]$ specifying the joint probability of the continuous random variables A and B, then
  - $\int_A [A, B] dA = \int_A [B|A][A] dA$ is the marginal probability of B
  - $\int_B [A, B] dB = \int_B [A|B][B] dB$ 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 random variable.

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

Marginal distributions of continuous random variables



Moment Matching

Motivation: Flexibility in analyses

Deterministic models

general linear
nonlinear
differential equations
difference equations
auto-regressive
occupancy
state-transition
integral-projection

univariate and
multivariate

Types of data

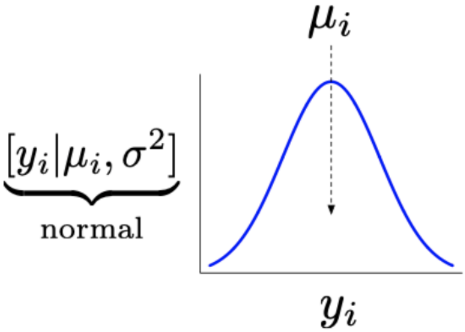
real numbers
non-negative real numbers
counts
0 to 1
0 or 1
counts in categories
proportions in categories
ordinal categories

Motivation: Flexibility in analyses

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 ≥ 2 categories
Cauchy	real numbers

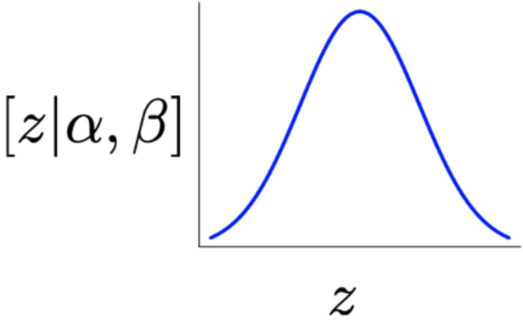
A familiar approach

$\boldsymbol{\theta} = (\beta_0, \beta_1)'$
$\boldsymbol{\mu}_i = f(x_i, \boldsymbol{\theta}) = \beta_0 + \beta_1 x_1$



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.

$$\alpha = m_1(\mu, \sigma^2)$$
$$\beta = m_2(\mu, \sigma^2)$$

We can use the functions to "match" the moments to the parameters.

Moment Matching

$$\mu_i = f(\theta, x_i)$$
$$\alpha = m_1(\mu_i, \sigma^2)$$
$$\beta = m_2(\mu, \sigma^2)$$
$$[y_i|\alpha, \beta]$$

Moment Matching the Gamma Distribution

- The gamma distributions is:

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

- The mean of the gamma distributions is:

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

- The variance of the gamma distributions is:

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

Exercise

- The mean of the gamma distributions is:

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

- The variance of the gamma distributions is:

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

If you were to choose a mean and variance for a prior that was gamma distributed, how would you calculate α and β?

Answer

1. $\mu = \frac{\alpha}{\beta}$
2. $\sigma^2 = \frac{\alpha}{\beta^2}$
3. Solve (1) for β, substitute for β in (2), solve for α
4. $\alpha = \frac{\mu^2}{\sigma^2}$
5. Substitute rhs of (4) for α in (2), solve for β
6. $\beta = \frac{\mu}{\sigma^2}$

Moment Matching the Beta Distributions

The Beta distribution gives the probability density of random variables with support on 0–1.

$$[z|\alpha,\beta] = \frac{z^{\alpha-1}(1-z)^{\beta_1}}{B(\alpha,\beta)}$$
$$B = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

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

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

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

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

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,

$$\begin{aligned}\mu &= \frac{\alpha}{\alpha+\beta}\\ \alpha &= \frac{\mu\beta}{1-\mu},\end{aligned}$$

which allows us to use

$$\begin{aligned}\mu_i &= g(\boldsymbol{\theta}, x_i)\\ y_i &\sim \text{beta}\left(\frac{\mu_i\beta}{1-\mu_i}, \beta\right)\end{aligned}$$

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 $= \sigma^2_{\text{log}}$, the variance of the random variable on the log scale.
We often moment match the median of the lognormal distribution:

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

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

Problems Continued

Work on Lab section on Moment Matching.