Probability Distributions and Moment Matching

Models for Socio-Environmental Data

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

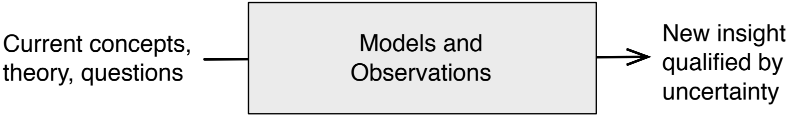
May 29, 2018

SESYNC

Roadmap

- The rules of probability
  - conditional probability
  - independence
  - the law of total probability
- Factoring joint probabilities
- Probability distributions for discrete and continuous random variables
- Marginal distributions
- Moment matching

Motivation: A general approach to scientific research



Motivation: Why do we need to know this stuff?

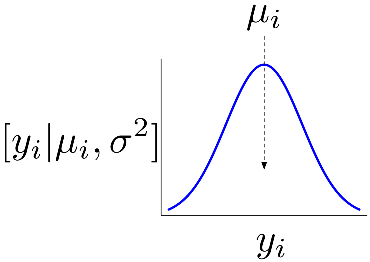
Concept to be taught	**Why do you need to understand this concept?**
Conditional probability	It is the foundation for Bayes' Theorem and all inferences we will make.
The law of total probability	Basis for the denominator of Bayes' Theorem $[y]$
Factoring joint distributions	This is the procedure we will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

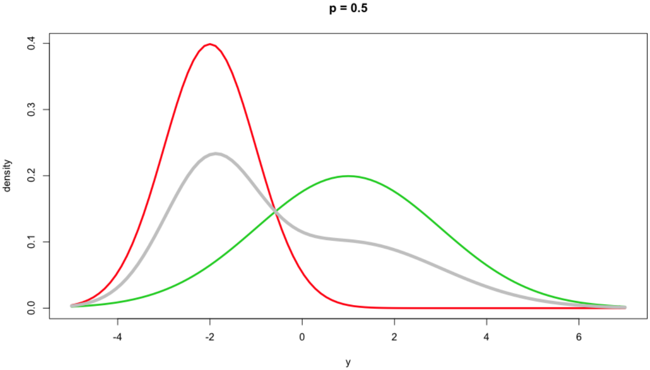
Deterministic models	Types of data
general linear	real numbers
nonlinear	non-negative real numbers
differential equations	counts
difference equations	0 to 1
auto-regressive	0 or 1
occupancy	counts in categories
state-transition	proportions in categories
integral-projection	

univariate and multivariate

Motivation: flexibility in analysis

Probability model	Support for random variable
normal	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

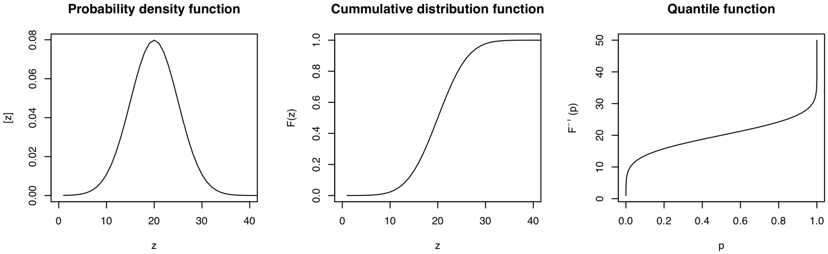
Motivation: flexibility in analysis



Work flow: probability distributions

- General properties and definitions
  - discrete random variables
  - continuous random variables
- Specific distributions
- Marginal distributions
- Moment matching

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 inferences about models.

A bit about notation

Ours	Others	Meaning
$[z]$	$f(Z=z)$	The probability or probability density of z.
$[z\|\alpha,\beta]$	$p(z\|\alpha,\beta), P(z\|\alpha,\beta)$	The probability or probability density of z conditional on α and β.
$[z]$	$p(z), P(z)$	The probability or probability density of z where its parameters are *numeric*. Used for priors.
$z \sim \text{Poisson}(\lambda)$	$z \sim \text{Poisson}(\lambda)$	z is distributed Poisson with mean λ
$z \sim \text{beta}(\alpha,\beta)$	$z \sim \text{beta}(\alpha,\beta)$	z is distributed beta with parameters α and β.

Board work on probability distributions for discrete random variables

- Probability mass function, z, PMF (also called probability function, discrete destiny function)
  - notation $[z], f(z)$
  - $[z]$ is a function that returns the probability of a specific value of the random variable $= z$
  - Support of random variable z is defined as all values of z for which $[z] > 0$ and defined.
  - requirements to be a PMF
    - $[z] \geq 0$
    - $\sum_{z \in s} z = 1$, where s is the support of the random variable
  - moments of PMF
    - first moment, the expected value (or mean) $= \mathsf{E}(z) = \mu = \sum_{z \in s} z[z]$, approximated from many (n) random draws from $[z]$ using $\mathsf{E}(z) \simeq \frac{1}{n} \sum_{i=1}^{n} z_i$
    - second central moment, the variance $=$ $\mathsf{E}\left((z-\mu)^2\right) = \sigma^2 = \sum_{z \in s}(z-\mu)^2[z]$, approximated from many (n) random draws from $[z]$ using $\mathsf{E}\left((z-\mu)^2\right) \simeq \frac{1}{n} \sum_{i=1}^{n}(z_i-\mu)^2$
- cumulative distribution function for z
- quantile function for z

Board work on probability distributions for continuous random variables

- Probability density function, PDF, $[z]$
  - notation $[z], f(z), z \sim$ normal()
  - $[z]$ gives the *probability density* of a specific value of the random variable $= z$.
  - Support of random variable z is defined as all values of z for which $[z] > 0$ and defined.
  - requirements
    - $[z] \geq 0$
    - $\int_{-\infty}^{\infty} [z]dz = 1$
    - $\Pr(a < z < b) = \int_a^b [z]dz$
  - What is probability density?
  - moments
    - first moment, the expected value (or mean) $= \mathsf{E}(z) = \mu = \int_{-\infty}^{\infty} z[z]dz$, approximated from many (n) random draws from $[z]$ using $\mathsf{E}(z) \simeq \frac{1}{n}\sum_{i=1}^{n} z_i$
    - second central moment, the variance $= \mathsf{E}\left((z-\mu)^2\right) = \sigma^2 = \int_{-\infty}^{\infty}(z-\mu)^2[z]dz$, approximated from many (n) random draws from $[z]$ using $\mathsf{E}\left((z-\mu)^2\right) \simeq \frac{1}{n}\sum_{i=1}^{n}(z_i-\mu)^2$
- cumulative distribution function, CDF
- quantile function

Problems

1. Introduce cheat sheet
2. Do Probability distributions problems in lab (section VI)

Work flow: probability distributions

- General properties and definitions
  - discrete random variables
  - continuous random variables
- Specific distributions
- Marginal distributions
- Moment matching

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 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?

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

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

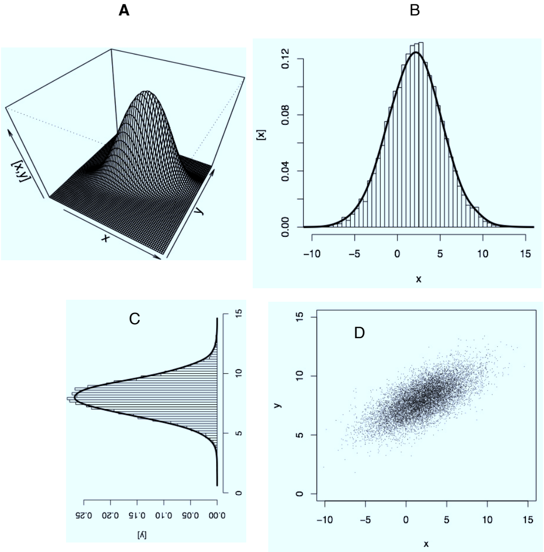
and

$\int_B [A, B]\, dB$ 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



Do marginal distribution problems

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	

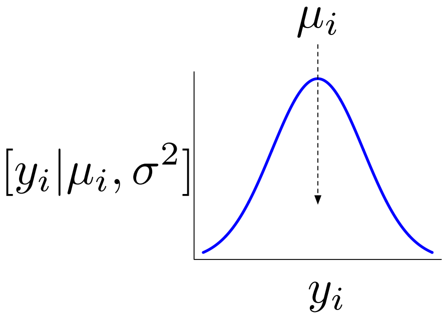
univariate and multivariate

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

Motivation: linking models to data

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



A familiar approach

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \\
\varepsilon_i &\sim \text{normal}(0, \sigma^2)
\end{aligned}
$$

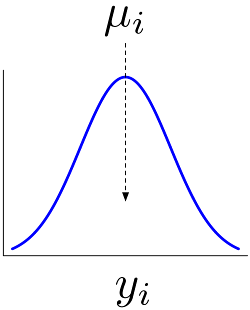
which is identical to

$$
\begin{aligned}
\mu_i &= \beta_0 + \beta_1 x_i \\
y_i &\sim \text{normal}(\mu_i, \sigma^2)
\end{aligned}
$$

A familiar approach

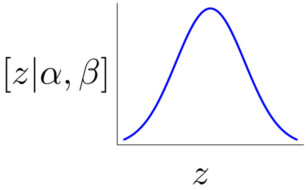
$$\boldsymbol{\theta} = (\beta_0, \beta_1)'$$
$$\mu_i = g(\boldsymbol{\theta}, x_i) = \beta_0 + \beta_1 x_i$$

$$\underbrace{[y_i|\mu_i, \sigma^2]}_{\text{normal}}$$



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 $0, ..., 1$.

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

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

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

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

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

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

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,

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{1}$$

$$\alpha = \frac{\mu\beta}{1-\mu}, \tag{2}$$

which allows us to use

$$\mu_i = g(\theta, x_i) \tag{3}$$

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

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:

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

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

Problems continued

Do section on Moment Matching