Moment Matching

Models for Socio-Environmental Data

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

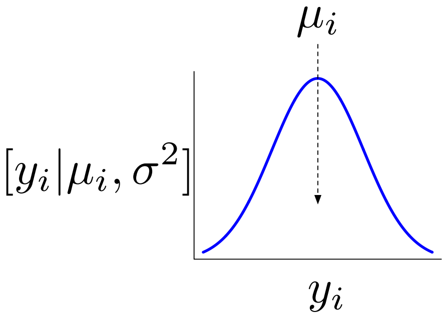
December 9, 2020

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)

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

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

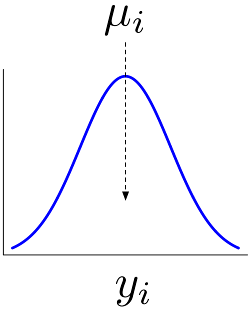
$$\mu_i = g(\boldsymbol{\theta}, x_i)$$

A familiar approach

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

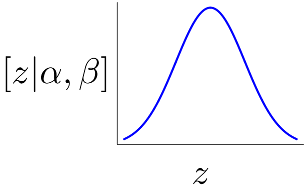
$$\mu_i = g(\boldsymbol{\theta}, x_i) = \beta_0 + \beta_1 x_i$$

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



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.

$$
\begin{aligned}
\alpha &= f_1(\mu, \sigma^2) \\
\beta &= f_2(\mu, \sigma^2)
\end{aligned}
$$

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

Moment matching

$$
\begin{aligned}
\boldsymbol{\mu}_i &= g(\boldsymbol{\theta}, x_i) \\
\boldsymbol{\alpha} &= f_1(\boldsymbol{\mu}_i, \boldsymbol{\sigma}^2) \\
\boldsymbol{\beta} &= f_2(\boldsymbol{\mu}_i, \boldsymbol{\sigma}^2) \\
& [y_i|\boldsymbol{\alpha}, \boldsymbol{\beta}]
\end{aligned}
$$

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

Moment matching the beta distribution

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

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

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

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

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

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

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,

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

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

which allows us to use

$$\boldsymbol{\mu}_i = g(\boldsymbol{\theta}, x_i) \quad (3)$$

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

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 Moment Matching Lab