

# 01 Foundations

Shravan Vasishth

25-29 March 2019

# Preview: Steps in Bayesian analysis

The way we will conduct data analysis is as follows.

- Given data, specify a *likelihood function*.
- Specify *prior distributions* for model parameters.
- Using software, derive *marginal posterior distributions* for parameters given likelihood function and prior density.
- Simulate parameters to get *samples from posterior distributions* of parameters using some *Markov Chain Monte Carlo (MCMC) sampling algorithm*.
- Evaluate whether model makes sense, using *model convergence* diagnostics, fake-data simulation, *prior predictive* and *posterior predictive* checks, and (if you want to claim a discovery) calibrating true and false discovery rates.
- Summarize *posterior distributions* of parameter samples and make your scientific decision.

# Bayes' rule

A and B are events. Conditional probability is defined as follows:

$$P(A|B) = \frac{P(A, B)}{P(B)} \text{ where } P(B) > 0 \quad (1)$$

This means that  $P(A, B) = P(A|B)P(B)$ .

Since  $P(B, A) = P(A, B)$ , we can write:

$$P(B, A) = P(B|A)P(A) = P(A|B)P(B) = P(A, B). \quad (2)$$

Rearranging terms:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (3)$$

This is Bayes' rule

# Random variable theory

A random variable  $X$  is a function  $X : S \rightarrow \mathbb{R}$  that associates to each outcome  $\omega \in S$  exactly one number  $X(\omega) = x$ .

$S_X$  is all the  $x$ 's (all the possible values of  $X$ , the support of  $X$ ). I.e.,  $x \in S_X$ . We can also sloppily write  $X \in S_X$ .

Good example: number of coin tosses till H

- $X : \omega \rightarrow x$
- $\omega$ : H, TH, TTH, ... (infinite)
- $x = 0, 1, 2, \dots; x \in S_X$

# Random variable theory

Every discrete (continuous) random variable  $X$  has associated with it a **probability mass (distribution) function (pmf, pdf)**. I.e., PMF is used for discrete distributions and PDF for continuous. (I will sometimes use lower case for pdf and sometimes upper case. Some books use pdf for both discrete and continuous distributions.)

$$p_X : S_X \rightarrow [0, 1] \quad (4)$$

defined by

$$p_X(x) = P(X(\omega) = x), x \in S_X \quad (5)$$

# Random variable theory

Probability density functions (continuous case) or probability mass functions (discrete case) are functions that assign probabilities or relative frequencies to all events in a sample space.

The expression

$$X \sim f(\cdot) \tag{6}$$

means that the random variable  $X$  has pdf/pmf  $g(\cdot)$ . For example, if we say that  $X \sim N(\mu, \sigma^2)$ , we are assuming that the pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \tag{7}$$

# Random variable theory

We also need a **cumulative distribution function** or cdf because, in the continuous case,  $P(X=\text{some point value})$  is zero and we therefore need a way to talk about  $P(X \text{ in a specific range})$ . cdfs serve that purpose.

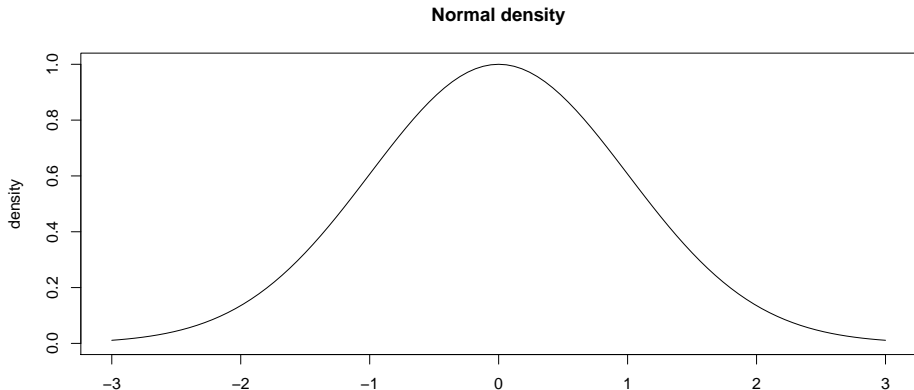
In the continuous case, the cdf or distribution function is defined as:

$$P(X < x) = F(X < x) = \int_{-\infty}^x f(x) dx \quad (8)$$

# Random variable theory

$$f(x) = \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (9)$$

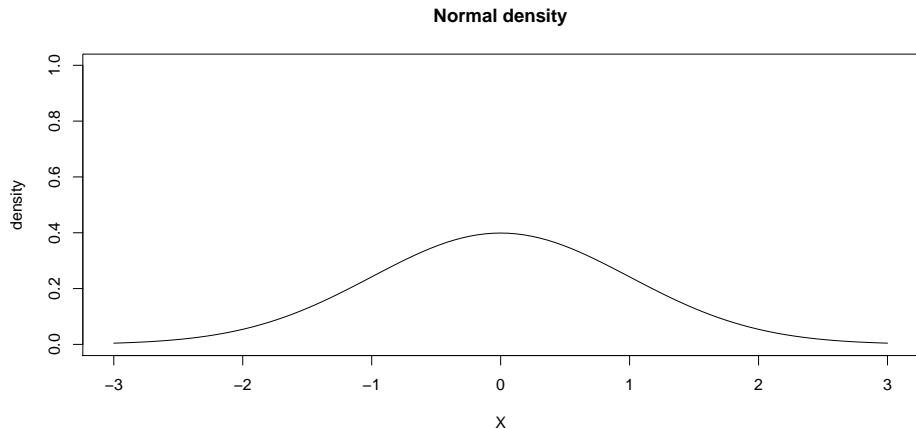
This is the “kernel” of the normal pdf, and it doesn’t sum to 1:





# Random variable theory

Adding a normalizing constant makes the above kernel density a pdf.



# Random variable theory

Recall that a random variable  $X$  is a function  $X : S \rightarrow \mathbb{R}$  that associates to each outcome  $\omega \in S$  exactly one number  $X(\omega) = x$ .  $S_X$  is all the  $x$ 's (all the possible values of  $X$ , the support of  $X$ ). I.e.,  $x \in S_X$ .

$X$  is a continuous random variable if there is a non-negative function  $f$  defined for all real  $x \in (-\infty, \infty)$  having the property that for any set  $B$  of real numbers,

$$P\{X \in B\} = \int_B f(x) dx \quad (10)$$

# Distributions

```
if ( !('devtools' %in%  
      installed.packages()) )  
  install.packages("devtools")  
  
devtools::install_github("bearloga/tinydensR")
```

Then, run

```
library(tinydensR)  
univariate_discrete_addin()
```

or

```
univariate_continuous_addin()
```

# Binomial distribution

If we have  $x$  successes in  $n$  trials, given a success probability  $p$  for each trial.  
If  $x \sim \text{Bin}(n, p)$ .

$$P(x \mid n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \quad (11)$$

The mean is  $np$  and the variance  $np(1 - p)$ .

```
dbinom(x, size, prob, log = FALSE)
### cdf:
pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
### quantiles:
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
### pseudo-random generation of samples:
rbinom(n, size, prob)
```

# The Poisson distribution

This is a distribution associated with “rare events”, for reasons which will become clear in a moment. The events might be:

- traffic accidents,
- typing errors, or
- customers arriving in a bank.

For psychology and linguistics, one application is in eye tracking: modeling number of fixations.

# The Poisson distribution

Let  $\lambda$  be the average number of events in the time interval  $[0, 1]$ . Let the random variable  $X$  count the number of events occurring in the interval. Then:

$$f_X(x) = \mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (12)$$

# Uniform distribution

A random variable ( $X$ ) with the continuous uniform distribution on the interval  $(\alpha, \beta)$  has PDF

$$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha < x < \beta, \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

The associated R function is `dunif(min = a, max = b)`. We write  $X \sim \text{unif}(\text{min} = a, \text{max} = b)$ . Due to the particularly simple form of this PDF we can also write down explicitly a formula for the CDF  $F_X$ :

# Uniform distribution

$$F_X(a) = \begin{cases} 0, & a < 0, \\ \frac{a-\alpha}{\beta-\alpha}, & \alpha \leq t < \beta, \\ 1, & a \geq \beta. \end{cases} \quad (14)$$

$$E[X] = \frac{\beta + \alpha}{2} \text{ and } \text{Var}(X) = \frac{(\beta - \alpha)^2}{12} \quad (15)$$

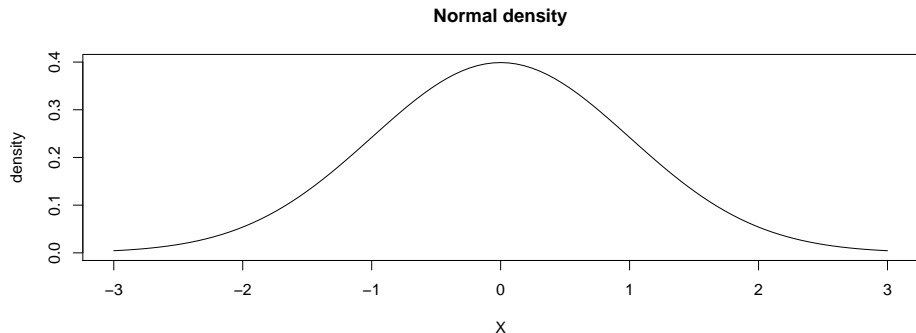
```
dunif(x, min = 0, max = 1, log = FALSE)
punif(q, min = 0, max = 1, lower.tail = TRUE,
      log.p = FALSE)
qunif(p, min = 0, max = 1, lower.tail = TRUE,
      log.p = FALSE)
runif(n, min = 0, max = 1)
```



# Normal distribution

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \quad (16)$$

We write  $X \sim \text{norm}(\text{mean} = \mu, \text{sd} = \sigma)$ , and the associated R function is `dnorm(x, mean = 0, sd = 1)`.



# Normal distribution

If  $X$  is normally distributed with parameters  $\mu$  and  $\sigma^2$ , then  $Y = aX + b$  is normally distributed with parameters  $a\mu + b$  and  $a^2\sigma^2$ .

## Standard or unit normal random variable:

If  $X$  is normally distributed with parameters  $\mu$  and  $\sigma^2$ , then  $Z = (X - \mu)/\sigma$  is normally distributed with parameters 0, 1.

We conventionally write  $\Phi(x)$  for the CDF:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{\frac{-y^2}{2}} dy \quad \text{where } y = (x - \mu)/\sigma \quad (17)$$

# Normal distribution

If  $Z$  is a standard normal random variable (SNRV) then

$$P\{Z \leq -x\} = P\{Z > x\}, \quad -\infty < x < \infty \quad (18)$$

Since  $Z = ((X - \mu)/\sigma)$  is an SNRV whenever  $X$  is normally distributed with parameters  $\mu$  and  $\sigma^2$ , then the CDF of  $X$  can be expressed as:

$$F_X(a) = P\{X \leq a\} = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right) \quad (19)$$

# Normal distribution

The standardized version of a normal random variable  $X$  is used to compute specific probabilities relating to  $X$  (it is also easier to compute probabilities from different CDFs so that the two computations are comparable).

```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE,
      log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE,
      log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

# Beta distribution

This is a generalization of the continuous uniform distribution.

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

# Beta distribution

We write  $X \sim \text{beta}(\text{shape1} = \alpha, \text{shape2} = \beta)$ . The associated R function is `=dbeta(x, shape1, shape2)=`.

The mean and variance are

$$E[X] = \frac{a}{a+b} \text{ and } \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}. \quad (20)$$

## *t* distribution

A random variable  $X$  with PDF

$$f_X(x) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi} \Gamma(r/2)} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}, \quad -\infty < x < \infty \quad (21)$$

is said to have Student's  $t$  distribution with  $r$  degrees of freedom, and we write  $X \sim t(\text{df} = r)$ . The associated R functions are `dt`, `pt`, `qt`, and `rt`, which give the PDF, CDF, quantile function, and simulate random variates, respectively.

We will just write:

$X \sim t(\mu, \sigma^2, r)$ , where  $r$  is the degrees of freedom ( $n - 1$ ), where  $n$  is sample size.

# Jointly distributed random variables

## Discrete case

Consider two discrete random variables  $X$  and  $Y$  with PMFs  $f_X$  and  $f_Y$  that are supported on the sample spaces  $S_X$  and  $S_Y$ , respectively. Let  $S_{X,Y}$  denote the set of all possible observed **pairs**  $(x, y)$ , called the **joint support set** of  $X$  and  $Y$ . Then the **joint probability mass function** of  $X$  and  $Y$  is the function  $f_{X,Y}$  defined by

$$f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y), \quad \text{for } (x, y) \in S_{X,Y}. \quad (22)$$

Every joint PMF satisfies

$$f_{X,Y}(x, y) > 0 \text{ for all } (x, y) \in S_{X,Y}, \quad (23)$$

and

$$\sum_{(x,y) \in S_{X,Y}} f_{X,Y}(x, y) = 1. \quad (24)$$



# Jointly distributed random variables

## Discrete case

It is customary to extend the function  $f_{X,Y}$  to be defined on all of  $\mathbb{R}^2$  by setting  $f_{X,Y}(x,y) = 0$  for  $(x,y) \notin S_{X,Y}$ .

The PMFs  $f_X$  and  $f_Y$  are called the **marginal PMFs** of  $X$  and  $Y$ , respectively. If we are given only the joint PMF then we may recover each of the marginal PMFs by using the Theorem of Total Probability: observe

$$f_X(x) = \mathbb{P}(X = x), \quad (25)$$

$$= \sum_{y \in S_Y} \mathbb{P}(X = x, Y = y), \quad (26)$$

$$= \sum_{y \in S_Y} f_{X,Y}(x, y). \quad (27)$$

# Jointly distributed random variables

## Discrete case

By interchanging the roles of  $X$  and  $Y$  it is clear that

$$f_Y(y) = \sum_{x \in S_X} f_{X,Y}(x, y). \quad (28)$$

Given the joint PMF we may recover the marginal PMFs, but the converse is not true. Even if we have **both** marginal distributions they are not sufficient to determine the joint PMF; more information is needed.

Associated with the joint PMF is the **joint cumulative distribution function**  $F_{X,Y}$  defined by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad \text{for } (x, y) \in \mathbb{R}^2.$$

# Jointly distributed random variables

## Discrete case

### Example:

Roll a fair die twice. Let  $X$  be the face shown on the first roll, and let  $Y$  be the face shown on the second roll. For this example, it suffices to define

$$f_{X,Y}(x,y) = \frac{1}{36}, \quad x = 1, \dots, 6, \quad y = 1, \dots, 6.$$

The marginal PMFs are given by  $f_X(x) = 1/6$ ,  $x = 1, 2, \dots, 6$ , and  $f_Y(y) = 1/6$ ,  $y = 1, 2, \dots, 6$ , since

$$f_X(x) = \sum_{y=1}^6 \frac{1}{36} = \frac{1}{6}, \quad x = 1, \dots, 6,$$

and the same computation with the letters switched works for  $Y$ .

# Jointly distributed random variables

## Continuous case

For random variables  $X$  and  $Y$ , the **joint cumulative pdf** is

$$F(a, b) = P(X \leq a, Y \leq b) \quad -\infty < a, b < \infty \quad (29)$$

The **marginal distributions** of  $F_X$  and  $F_Y$  are the CDFs of each of the associated RVs:

- 1 The CDF of  $X$ :

$$F_X(a) = P(X \leq a) = F_X(a, \infty) \quad (30)$$

- 2 The CDF of  $Y$ :

$$F_Y(b) = P(Y \leq b) = F_Y(b, \infty) \quad (31)$$

# Jointly distributed random variables

## Continuous case

### Definition

**Jointly continuous:** Two RVs  $X$  and  $Y$  are jointly continuous if there exists a function  $f(x, y)$  defined for all real  $x$  and  $y$ , such that for every set  $C$ :

$$P((X, Y) \in C) = \iint_{(x,y) \in C} f(x, y) dx dy \quad (32)$$

$f(x, y)$  is the **joint PDF** of  $X$  and  $Y$ .

# Jointly distributed random variables

## Continuous case

Every joint PDF satisfies

$$f(x, y) \geq 0 \text{ for all } (x, y) \in S_{X,Y}, \quad (33)$$

and

$$\iint_{S_{X,Y}} f(x, y) \, dx \, dy = 1. \quad (34)$$

# Jointly distributed random variables

## Continuous case

For any sets of real numbers  $A$  and  $B$ , and if  $C = \{(x, y) : x \in A, y \in B\}$ , it follows from equation~32 that

$$P((X \in A, Y \in B) \in C) = \int_B \int_A f(x, y) dx dy \quad (35)$$

# Jointly distributed random variables

## Continuous case

Note that

$$F(a, b) = P(X \in (-\infty, a], Y \in (-\infty, b]) = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy \quad (36)$$

Differentiating, we get the joint pdf:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) \quad (37)$$



# Jointly distributed random variables

## Marginal probability distribution functions

If  $X$  and  $Y$  are jointly continuous, they are individually continuous, and their PDFs are:

$$\begin{aligned} P(X \in A) &= P(X \in A, Y \in (-\infty, \infty)) \\ &= \int_A \int_{-\infty}^{\infty} f(x, y) dy dx \\ &= \int_A f_X(x) dx \end{aligned} \tag{38}$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \tag{39}$$

# Jointly distributed random variables

## Independent random variables

Random variables  $X$  and  $Y$  are independent iff, for any two sets of real numbers  $A$  and  $B$ :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \quad (40)$$

In the jointly continuous case:

$$f(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y \quad (41)$$

# Conditional distributions

## Discrete case

Recall that the conditional probability of  $B$  given  $A$ , denoted  $\mathbb{P}(B \mid A)$ , is defined by

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \quad \text{if } \mathbb{P}(A) > 0. \quad (42)$$

# Conditional distributions

## Discrete case

If  $X$  and  $Y$  are discrete random variables, then we can define the conditional PMF of  $X$  given that  $Y = y$  as follows:

$$\begin{aligned} p_{X|Y}(x | y) &= P(X = x | Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned} \tag{43}$$

for all values of  $y$  where  $p_Y(y) = P(Y = y) > 0$ .

## Discrete case

The **conditional cumulative distribution function** of  $X$  given  $Y = y$  is defined, for all  $y$  such that  $p_Y(y) > 0$ , as follows:

$$\begin{aligned} F_{X|Y} &= P(X \leq x \mid Y = y) \\ &= \sum_{a \leq x} p_{X|Y}(a \mid y) \end{aligned} \tag{44}$$

If  $X$  and  $Y$  are independent then

$$p_{X|Y}(x \mid y) = P(X = x) = p_X(x) \tag{45}$$

## Continuous case

If  $X$  and  $Y$  have a joint probability density function  $f(x, y)$ , then the conditional probability density function of  $X$  given that  $Y = y$  is defined, for all values of  $y$  such that  $f_Y(y) > 0$ , by

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)} \quad (46)$$

# Covariance and correlation

There are two very special cases of joint expectation: the **covariance** and the **correlation**. These are measures which help us quantify the dependence between  $X$  and  $Y$ .

## Definition

The **covariance** of  $X$  and  $Y$  is

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y). \quad (47)$$

Shortcut formula for covariance:

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y). \quad (48)$$

# Covariance and correlation

**The Pearson product moment correlation** between  $X$  and  $Y$  is the covariance between  $X$  and  $Y$  rescaled to fall in the interval  $[-1, 1]$ . It is formally defined by

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (49)$$



# Multivariate normal distributions

This is a very important distribution that we will need in linear mixed models.

Recall that in the univariate normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}\right\} \quad -\infty < x < \infty \quad (50)$$

# Multivariate normal distributions

We can write the power of the exponential as:

$$\left(\frac{x - \mu}{\sigma}\right)^2 = (x - \mu)(x - \mu)(\sigma^2)^{-1} = (x - \mu)(\sigma^2)^{-1}(x - \mu) = Q \quad (51)$$

Generalizing this to the multivariate case:

$$Q = (x - \mu)' \Sigma^{-1} (x - \mu) \quad (52)$$

# Multivariate normal distributions

So, for the multivariate case:

$$f(x) = \frac{1}{\sqrt{2\pi \det \Sigma}} e\{-Q/2\} \quad -\infty < x_i < \infty, i = 1, \dots, n \quad (53)$$

Properties of the multivariate normal (MVN) X:

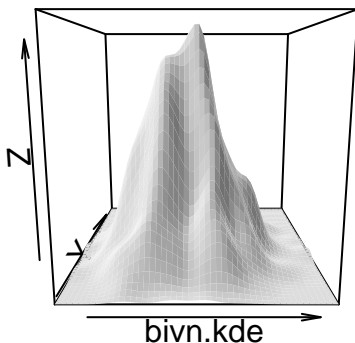
- Linear combinations of X are normal distributions.
- All subset's of X's components have a normal distribution.
- Zero covariance implies independent distributions.
- Conditional distributions are normal.

# Multivariate normal distributions

## Visualizing bivariate distributions

First, a visual of two uncorrelated RVs:

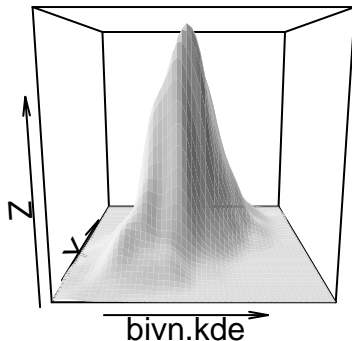
### Simulated bivariate normal density



# Multivariate normal distributions

And here is an example of a positively correlated case:

## Simulated bivariate normal density

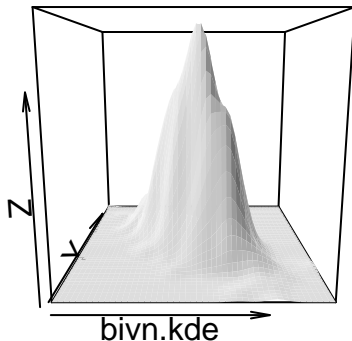


**Figure 3:** Visualization of two correlated random variables.

# Multivariate normal distributions

And here is an example with a negative correlation:

## Simulated bivariate normal density



**Figure 4:** Visualization of two negatively correlated random variables.

# Multivariate normal distributions

## Visualizing conditional distributions

You can run the following code to get a visualization of what a conditional distribution looks like when we take “slices” from the conditioning random variable:

```
for(i in 1:50){  
  plot(bivn.kde$z[i,1:50],type="l",ylim=c(0,0.1))  
  Sys.sleep(.5)  
}
```

# Maximum likelihood estimation

## Discrete case

Suppose the observed sample values are  $x_1, x_2, \dots, x_n$ . The probability of getting them is

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta) \quad (54)$$

i.e., the function  $f$  is the value of the joint probability **distribution** of the random variables  $X_1, \dots, X_n$  at  $X_1 = x_1, \dots, X_n = x_n$ . Since the sample values have been observed and are fixed,  $f(x_1, \dots, x_n; \theta)$  is a function of  $\theta$ . The function  $f$  is called a **likelihood function**.



# Maximum likelihood estimation

## Continuous case

Here,  $f$  is the joint probability **density**, the rest is the same as above.

## Definition

If  $x_1, x_2, \dots, x_n$  are the values of a random sample from a population with parameter  $\theta$ , the **likelihood function** of the sample is given by

$$L(\theta) = f(x_1, x_2, \dots, x_n; \theta) \quad (55)$$

for values of  $\theta$  within a given domain. Here,

$f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$  is the joint probability distribution or density of the random variables  $X_1, \dots, X_n$  at  $X_1 = x_1, \dots, X_n = x_n$ .

So, the method of maximum likelihood consists of maximizing the likelihood function with respect to  $\theta$ . The value of  $\theta$  that maximizes the likelihood function is the **MLE** (maximum likelihood estimate) of  $\theta$ .

# Finding maximum likelihood estimates for different distributions

## Example 1: Binomial

$$L(\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad (56)$$

Log likelihood:

$$\ell(\theta) = \log \binom{n}{x} + x \log \theta + (n - x) \log(1 - \theta) \quad (57)$$

# Finding maximum likelihood estimates for different distributions

Differentiating:

$$\ell'(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0 \quad (58)$$

Thus:

$$\hat{\theta} = \frac{x}{n} \quad (59)$$

# Finding maximum likelihood estimates for different distributions

## Example 2: Normal

Let  $X_1, \dots, X_n$  constitute a random variable of size  $n$  from a normal population with mean  $\mu$  and variance  $\sigma^2$ , find joint maximum likelihood estimates of these two parameters.

# Finding maximum likelihood estimates for different distributions

$$L(\mu; \sigma^2) = \prod N(x_i; \mu, \sigma) \quad (60)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right) \quad (61)$$

$$(62)$$

Taking logs and differentiating with respect to  $\mu$  and  $\sigma^2$ , we get:

$$\hat{\mu} = \frac{1}{n} \sum x_i = \bar{x} \quad (63)$$

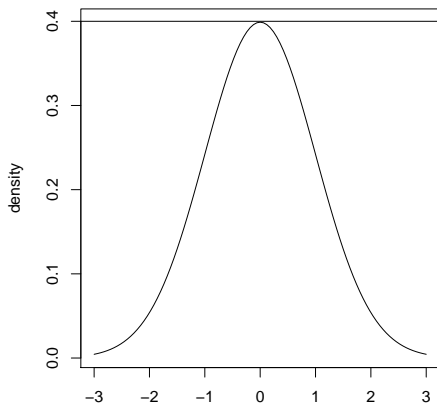
and

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad (64)$$

# Visualizing likelihood and maximum log likelihood for normal

For simplicity consider the case where  $N(\mu = 0, \sigma^2 = 1)$ .

Normal density



Normal density (log)

