

## Chapter 5: Joint Probability Distributions and Random Samples

Curtis Miller

2018-06-13

### Introduction

WE MAY NATURALLY INQUIRE about collections of random variables that are related to each other in some way. For instance, we may record an individual's height and weight, calling these random variables  $X$  and  $Y$ , and ask if these are correlated, uncorrelated, or even independent characteristics, and describe a probability model that accounts for the relationship in these two characteristics.

Additionally, we may have a large collection of random variables, say  $X_1, X_2, \dots, X_n$  which will be used to estimate some essential quantity of a distribution, such as the mean  $\mu$ . We compute some quantity based on this collection of random variables, such as  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , or any other  $T = T(X_1, \dots, X_n)$ . This quantity, dependent on random variables, is itself a random variable, and we call it a statistic. Being a statistic it has its own probability distribution, with its own mean and variance and cdf, and we can use the distribution of the statistic to make statements about the process that generated the original dataset  $X_1, \dots, X_n$ . It is here when probability theory begins to turn into statistical theory.

### Section 1: Jointly Distributed Random Variables

Suppose  $X$  and  $Y$  are two discrete random variables. Their **joint probability mass function** is described below:

This can be used to compute  $\mathbb{P}((X, Y) \in A)$  for an event  $A$ :

From this we can compute the **marginal probability mass functions**,  $p_X(x)$  and  $p_Y(y)$ , for  $X$  and  $Y$  respectively.

These represent the probability distribution of  $X$  and  $Y$  respectively regardless of what value the other rv takes.

We can also compute what is known as the **conditional probability mass function of  $Y$  given  $X = x$** , which represents the probability distribution of  $Y$  when we know that  $X = x$ . The **conditional probability mass function of  $X$  given  $Y = y$**  is defined in a similar manner.

---

*Example 1*

A fair six-sided die is rolled; let  $X$  represent the number of pips shown. At the same time, a fair coin is flipped, and  $Y(\omega) = 1$  if the coin lands heads-up, and  $Y(\omega) = 2$  if the coin lands tails-up. The joint pmf of  $X$  and  $Y$  is

1. Compute  $\mathbb{P}(X < Y)$ .

2. Compute  $\mathbb{P}(\text{Both } X \text{ and } Y \text{ are even})$

3. Find the marginal pmfs for both  $X$  and  $Y$ .

4. Find the conditional distributions  $p_{X|Y}(x|y)$  and  $p_{Y|X}(y|x)$ .

```

library(discreteRV)

##
## Attaching package: 'discreteRV'

## The following object is masked from 'package:base':
##
##      %in%

library(magrittr) # Adds the %>% operator

XY <- jointRV(list(1:6, 1:2), probs = rep(1/12, times = 12))
(X <- marginal(XY, 1)) # The relationship between X and Y is still preserved

## Random variable with 6 outcomes
##
## Outcomes  1  2  3  4  5  6
## Probs     1/6 1/6 1/6 1/6 1/6 1/6

(Y <- marginal(XY, 2))

## Random variable with 2 outcomes
##
## Outcomes  1  2
## Probs     1/2 1/2

joint(X, Y)

## Random variable with 12 outcomes
##
## Outcomes  1,1  1,2  2,1  2,2  3,1  3,2  4,1  4,2  5,1  5,2  6,1  6,2
## Probs     1/12 1/12 1/12 1/12 1/12 1/12 1/12 1/12 1/12 1/12 1/12 1/12

P(X < Y)

## [1] 0.08333333

P((X %in% c(2, 4, 6)) %AND% (Y %in% c(2))) # Both even

## [1] 0.25

X | Y == 2 # Gets a conditional random variable

## Random variable with 6 outcomes
##
## Outcomes  1  2  3  4  5  6
## Probs     1/6 1/6 1/6 1/6 1/6 1/6

```

```
YgivenX <- function(x) {Y | X == x}
XgivenY <- function(y) {X | Y == y}
```

### YgivenX(2)

```
## Random variable with 2 outcomes
##
## Outcomes  1  2
## Probs     1/2 1/2
```

### XgivenY(2)

```
## Random variable with 6 outcomes
##
## Outcomes  1  2  3  4  5  6
## Probs     1/6 1/6 1/6 1/6 1/6 1/6
```

---

Now suppose that  $X$  and  $Y$  are continuous random variables. Much is the same; we work with a **joint probability density function, marginal probability density functions, and conditional probability density functions.**

---

*Example 2*

A company sells bags of “deluxe” mixed nuts, containing almonds, cashews, and peanuts. One bag is five pounds, and the joint pdf for the amount of almonds  $X$  and cashews  $Y$  in the bag (in pounds) is given below:

(We don't need to worry about the amount of peanuts; this is simply  $5 - X - Y$  and thus is completely determined given  $X$  and  $Y$ .)  
The region on which the pdf is illustrated below:

1. Customers buying bags of “deluxe” mixed nuts complain when 60% of the nuts in the bag are peanuts. Compute the probability this occurs.

2. Find the marginal distributions of  $X$  and  $Y$ . Use this to compute  $\mathbb{E}[X]$ ,  $\mathbb{E}[Y]$ ,  $\text{Var}(X)$ , and  $\text{Var}(Y)$ .

3. Find the conditional pdfs  $f_{X|Y}(x|y)$  and  $f_{Y|X}(y|x)$ . Use this to compute  $\mathbb{P}(X > 2|Y = 2)$ .

---

We say that two random variables  $X$  and  $Y$  are **independent** if

---

*Example 3*

Are the random variables in the previous two example independent?  
Explain why or why not.



`independent(X, Y) # For Example 1`

`## [1] TRUE`

---

We can generalize our definitions from two to  $n$  rvs,  $X_1, \dots, X_n$ .

We say that  $X_1, \dots, X_n$  are independent if for *every* subset  $X_{i_1}, \dots, X_{i_k}$  of the collection, we have

If  $X_1, \dots, X_n$  are independent and each has the same pmf/pdf that  $X_1, \dots, X_n$  are **independent and identically distributed**, often abbreviated *i.i.d.*<sup>1</sup>

<sup>1</sup> This is a typical assumption about a dataset in statistics.

---

*Example 4*

Compute  $\mathbb{P}(\min\{X_1, \dots, X_n\} \geq x)$  if  $X_1, \dots, X_n$  are *i.i.d.*

---

We can generalize the binomial distribution we saw before the **multinomial distribution**. We have  $r$  categories, and a single observation belongs to category  $i$  with probability  $p_i$ . We count how many observations belong to category  $i$ ; this gives  $X_i$ . Then the vector  $(X_1, \dots, X_r)$  follows the multinomial distribution, or  $(X_1, \dots, X_r) \sim \text{MULTINOM}(p_1, \dots, p_r)^2$ , and we have the pmf

<sup>2</sup> The binomial distribution is a particular instance of the multinomial distribution, when  $r = 2$ . We omit the count of tails, which we may call  $X_2$ , as it's redundant information given  $X_1$ .

### *Section 2: Expected Values, Covariance, and Correlation*

Expectations involving two random variables are defined similarly to the univariate cases.

---

#### *Example 5*

Reconsider the random variables in Examples 1 and 2. Compute  $\mathbb{E}[XY]$  for both cases.



$E(X * Y)$  # For Example 1's random variables

## [1] 5.25

---

One measure of the relationship between two random variables is the **covariance**.

The covariance is positive if the two random variables tend to be large together, while the covariance is negative if one rv tends to be large when the other tends to be small. If  $\text{Cov}(X, Y) = 0$ , then  $X$  and  $Y$  are **uncorrelated**.<sup>3</sup>

We also have the following shortcut formula for the covariance:

<sup>3</sup> Due to the relationship between the covariance and the variance, we sometimes see the notation  $\text{Cov}(X, Y) = \sigma_{XY}$ .

It's obtained in a similar manner to shortcut formulas found for computing  $\text{Var}(X)$ .<sup>4</sup>

Notice that the covariance is *not* insensitive to the units of the random variable; in fact, we can compute the covariance  $\text{Cov}(aX + b, cY + d)$ :

<sup>4</sup> From this it's clear that  $\text{Cov}(X, X) = \text{Var}(X)$ .

Changing the units changes the covariance. A unit-free measure of the relationship between  $X$  and  $Y$  is the **correlation**.<sup>5</sup>

<sup>5</sup> The usual Greek letter representing correlation is  $\rho$ .

The correlation is 1 if there is a perfect positive *linear* relationship between  $X$  and  $Y$ , -1 if there is a perfect negative *linear* relationship, and 0 if there is no *linear* relationship between  $X$  and  $Y$ .<sup>6</sup> Thus  $|\rho|$  determines the *strength* of the relationship<sup>7</sup> between  $X$  and  $Y$  and  $\text{sign}(\rho)$  determines the *direction* of the relationship.<sup>8</sup>

<sup>6</sup> Notice the emphasis on the word "linear"; there can be a relationship between  $X$  and  $Y$  that would make their correlation small yet there could still be a strong nonlinear relationship linking the two variables.

<sup>7</sup> We can classify the strength of the relationship between rvs using completely arbitrary cutoffs; specifically, we could say that if  $|\rho| < 0.3$  there is no notable correlation, if  $|\rho| > 0.7$  there is a strong correlation, and otherwise the correlation is weak.

<sup>8</sup> There is a sample statistic for estimating  $\rho$  from paired data  $(x_1, y_i)$ :

$$r = \frac{1}{s_x s_y (n-1)} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

The interpretation is the same. We do not discuss the sample statistic in this course.

---

### Example 6

Compute the covariance and correlation for the random variables mentioned in Examples 1 and 2.

```
# For Example 1
(sigma_xy <- E(X * Y) - E(X) * E(Y))

## [1] 0
```

---

If  $X$  and  $Y$  are independent, then  $\text{Cov}(X, Y) = 0$ .<sup>9</sup> The converse is *not* true in general, as the following example shows.

<sup>9</sup> As a consequence, we can equivalently say that  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$  when  $X$  and  $Y$  are independent.

### Example 7

The point  $(U, V)$  is equally likely to be any of the points in the sample space  $\mathcal{S} = \{(1, 1), (1, -1), (-1, 2), (-1, -2)\}$ . Compute  $\text{Cov}(U, V)$ . Are  $U$  and  $V$  independent?

---

```

(UV <- jointRV(list(c(-1, 1), c(-2, -1, 1, 2)),
                probs = c(1/4, 0, 0, 1/4, 0, 1/4, 1/4, 0)))

## Random variable with 4 outcomes
##
## Outcomes -1,-2 -1,2 1,-1 1,1
## Probs     1/4  1/4  1/4  1/4

(U <- marginal(UV, 1))

## Random variable with 2 outcomes
##
## Outcomes -1 1
## Probs     1/2 1/2

(V <- marginal(UV, 2))

## Random variable with 4 outcomes
##
## Outcomes -2 -1 1 2
## Probs     1/4 1/4 1/4 1/4

E(U*V) - E(U) * E(V)

## [1] 0

independent(U, V)

## [1] FALSE

```

---

We say  $(X_1, X_2)$  follows the bivariate Normal distribution, or  $(X_1, X_2) \sim \text{BINORM}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ , if the joint pdf of  $X_1$  and  $X_2$  is

The pdf of the bivariate Normal distribution is illustrated below.

```

library(mvtnorm)
library(lattice)

my.settings <- list(superpose.polygon = list(border = "transparent"))

```

```

points <- data.frame("x" = rep(seq(-3, 3, length.out = 100), times = 100),
                    "y" = rep(seq(-3, 3, length.out = 100), each = 100))
points$z <- apply(points, 1, function(r) {dmvnorm(r)})
head(points)

```

```

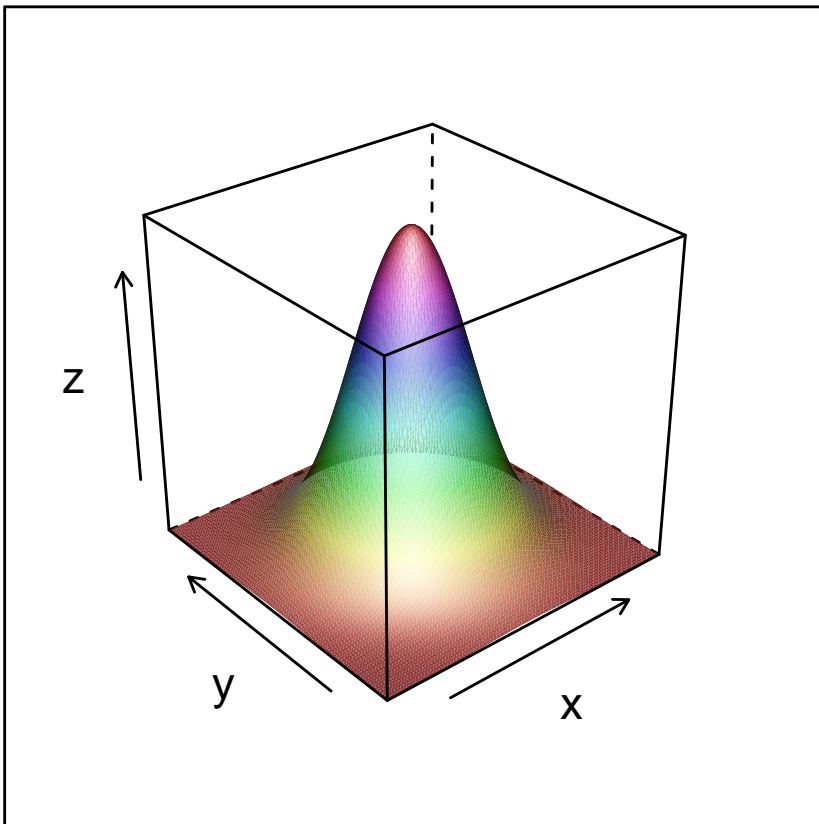
##           x y           z
## 1 -3.000000 -3 1.964128e-05
## 2 -2.939394 -3 2.351445e-05
## 3 -2.878788 -3 2.804818e-05
## 4 -2.818182 -3 3.333337e-05
## 5 -2.757576 -3 3.946923e-05
## 6 -2.696970 -3 4.656321e-05

```

```

wireframe(z ~ x * y, data = points, lines = FALSE,
           col = "transparent", shade = TRUE)

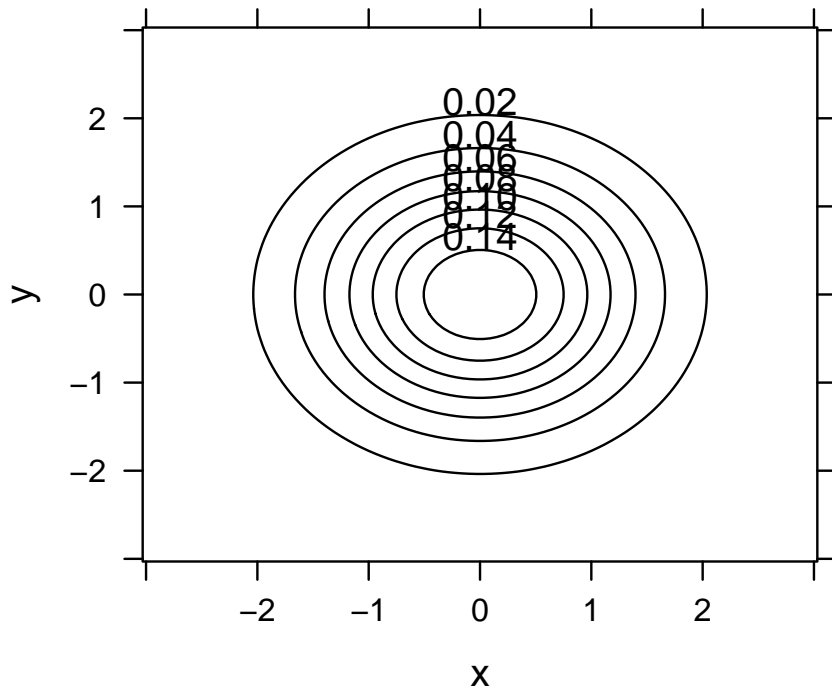
```



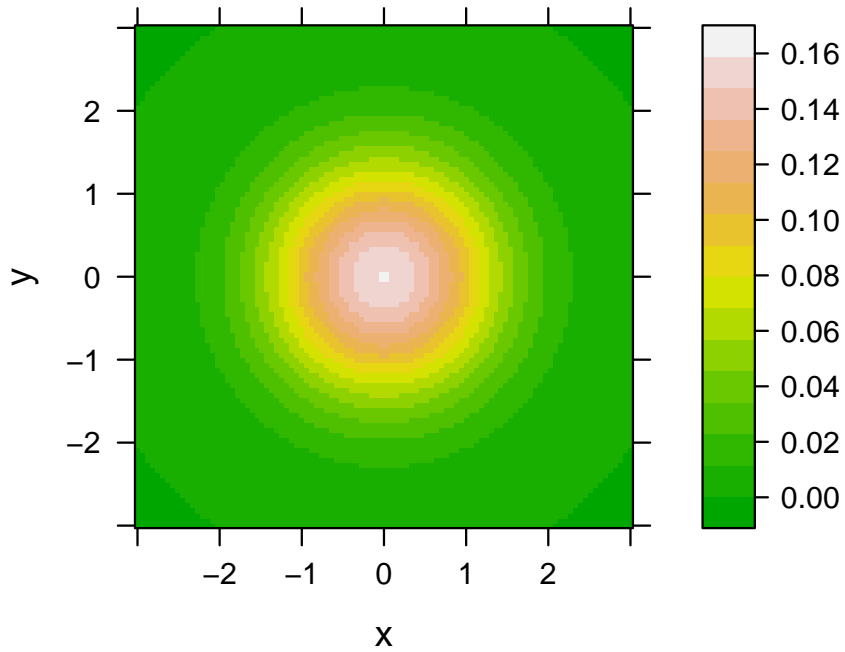
```

contourplot(z ~ x * y, data = points)

```



```
levelplot(z ~ x * y, data = points, drape = TRUE,  
          col.regions = terrain.colors(100))
```





If you were to slice the pdf in any direction, the resulting plot would be another Normal distribution. Specifically, the marginal distributions  $f_{X_1}(x_1)$  and  $f_{X_2}(x_2)$  are Normal distributions, and conditional distributions  $f_{X_1|X_2}(x_1|x_2)$  and  $f_{X_2|X_1}(x_2|x_1)$  are all all Normal distributions.

We have  $\mathbb{E}[X_1] = \mu_1$ ,  $\mathbb{E}[X_2] = \mu_2$ ,  $\text{SD}(X_1) = \sigma_1$ ,  $\text{SD}(X_2) = \sigma_2$ , and  $\text{Corr}(X_1, X_2) = \rho$ . Crucially, when  $(X_1, X_2)$  follows a bivariate Normal distribution,  $\text{Cov}(X_1, X_2)$  *does* imply independence!<sup>10</sup>

---

### Example 8

Let  $H_C$  represent the height of a son and  $H_F$  the height of the son's father (in inches). Suppose

$$(H_C, H_F) \sim \text{BINORM}(69.2, 69.2, 2.6, 2.6, 0.4)$$

1. What are the marginal distributions of  $H_C$  and  $H_F$ ?

<sup>10</sup> This is not the same as saying that if two random variables are Normally distributed and uncorrelated they are independent. Joint normality does not follow from the normality of the marginal distributions; for example, if we choose  $Z_1 \sim N(0, 1)$  and  $Z_2 = SZ_1$  with  $\mathbb{P}(S = 1) = \mathbb{P}(S = -1) = \frac{1}{2}$ , then  $Z_2 \sim N(0, 1)$ , and the marginal distributions of  $(Z_1, Z_2)$  are thus standard Normal distributions, and  $\text{Cov}(Z_1, Z_2) = 0$ . However,  $Z_1$  and  $Z_2$  are obviously *not* independent since if we know  $Z_1$  then we know  $Z_2$  differs from  $Z_1$  by at most a sign.

2. Suppose a person's father is 78 inches tall. Find an equal-tailed<sup>11</sup> interval such that the probability the child's height is in this interval is 0.95.

<sup>11</sup> In general, when we say an interval is equal-tailed, we mean that the probability that the random variable is too small to be in the region is equal to the probability that the random variable is too large. We need this restriction in order to have a unique solution; otherwise, there could be an infinite number of solutions.

```

# Marginal distributions are trivial; let's worry about the conditional
mu1 <- 69.2; mu2 <- 69.2; sigma1 <- 2.6; sigma2 <- 2.6; rho <- 0.4
h <- 78
(mu_2g1 <- (mu1 - rho * mu2 * sigma1 / sigma2) + rho * sigma1 * h / sigma2)

## [1] 72.72

(sigma_2g1 <- sqrt((1 - rho^2) * sigma1^2))

## [1] 2.382939

qnorm(0.025, mean = mu_2g1, sd = sigma_2g1) # Lower bound

## [1] 68.04952

qnorm(0.975, mean = mu_2g1, sd = sigma_2g1) # Upper bound

## [1] 77.39048

```

---

### Section 5: The Distribution of a Linear Combination<sup>12</sup>

Consider a collection of  $n$  random variables  $X_1, \dots, X_n$  and numerical constants  $a_1, \dots, a_n$ . The rv  $Y$  is a **linear combination** of the random variables  $X_1, \dots, X_n$  if  $Y$  is of the form

**Proposition 1.** Suppose  $\mathbb{E}[X_i] = \mu_i$  and  $\text{Var}(X_i) = \sigma_i^2$ . The following facts are true:

<sup>12</sup> In my opinion, Section 5 of this chapter is a more logical successor of Section 2; we will come back to Section 3 later.

Thus we have the important property about expectations; they are linear operators, to use linear algebra language. The variance is not a linear operator (although it is a sublinear operator), but the covariance is a bilinear operator (linear in both its arguments).

**Corollary 1.** *Suppose  $X_1$  and  $X_2$  are two independent random variables. Then:*

We can weaken the independence assumption to simply being uncorrelated and the variance computation will still be true.

**Proposition 2.** *Suppose  $X_1$  and  $X_2$  are two Normal random variables. Then*

**Corollary 2.** *A linear combination of Normal random variables also follows a Normal distribution.*

---

*Example 9*

Suppose  $X_1, \dots, X_n$  are i.i.d. random variables. Compute the expected value, variance, and standard deviation of  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

*Example 10*

Suppose  $X_1, \dots, X_n$  are i.i.d. random variables. Compute the expected value, variance, and standard deviation of  $T = \sum_{i=1}^n X_i$

There are several random variables where we know the distribution of sums of those random variables. Below is a summary:

---

*Section 3: Statistics and Their Distributions*

We will call the collection  $X_1, \dots, X_n$  a **random sample** if it consists of i.i.d. random variables. We will call any quantity we can compute from a random sample a **statistic**. Before the dataset is observed, a statistic is a random quantity, with its own distribution, referred to as the **sampling distribution**; statistics in this random state are usually referred to using upper-case letters, while the observed statistic (after we have a dataset) is usually referred to using lower-case letters.

Examples of statistics include:

*Example 11*

Let  $X_1, \dots, X_n$  be i.i.d.r.v. with  $X_1 \sim \text{Ber}(p)$ . What is the sampling distribution of  $\bar{X}$ ?

*Example 12*

Let  $X_1, \dots, X_n$  be i.i.d.r.v. with  $X_1 \sim N(\mu, \sigma)$ . What is the sampling distribution of  $\bar{X}$ ? Use the sampling distribution to find an interval such that  $\mathbb{P}(l(\bar{X}) \leq \mu \leq u(\bar{X})) = 1 - \alpha$ .

---

One approach to finding information about the sampling distribution of statistics is to use **simulations**. We generate  $K$  random samples of size  $n$ ,  $X_{1,k}, \dots, X_{n,k}$ ,  $k \in [K]$ . For each sample we compute the statistic of interest  $T(X_{1,k}, \dots, X_{n,k}) = T_k$ , and then study the random sample  $T_1, \dots, T_K$ .

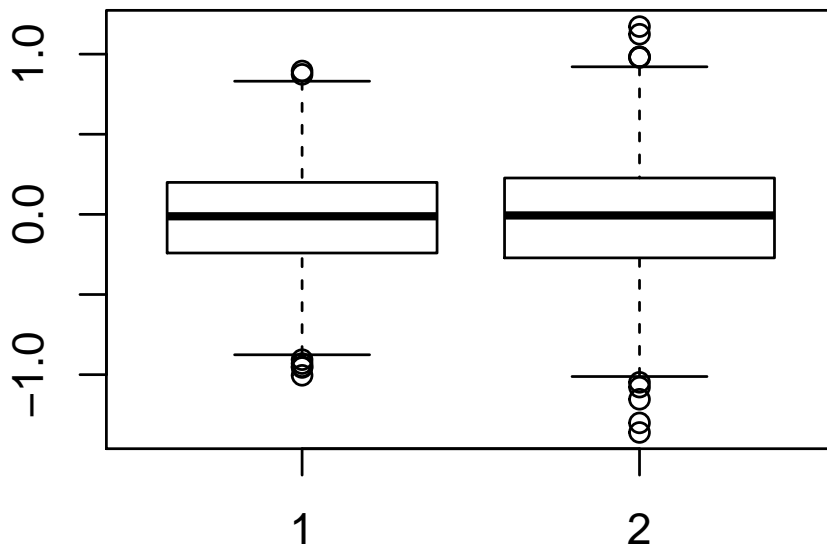
---

### Example 13

For the Normal distribution, we could estimate the parameter  $\mu$  using either the sample mean  $\bar{X}$  or the sample median  $\tilde{X}$ . What are the properties of these two statistics' sampling distributions? What are their respective shapes? Which has a smaller variance?

Let's suppose  $X_1 \sim N(0,1)$ , then conduct a simulation study to compare these statistics. We'll look at  $n = 10$  and use  $K = 1000$  samples.

```
# Generate 1000 random samples of size ten, storing them in a 10x1000 matrix
datamat <- replicate(1000, rnorm(10))
sim_mean <- apply(datamat, 2, mean)
sim_med <- apply(datamat, 2, median)
boxplot(sim_mean, sim_med)
```



```
summary(sim_mean)

##      Min. 1st Qu.  Median    Mean 3rd Qu.
## -1.00415 -0.24089 -0.01189 -0.01956  0.19948
##      Max.
##      0.89499
```

```
summary(sim_med)
```

```
##      Min.   1st Qu.   Median     Mean  
## -1.360863 -0.271161 -0.006523 -0.018751  
##   3rd Qu.     Max.  
##  0.227032  1.171895
```

```
var(sim_mean)
```

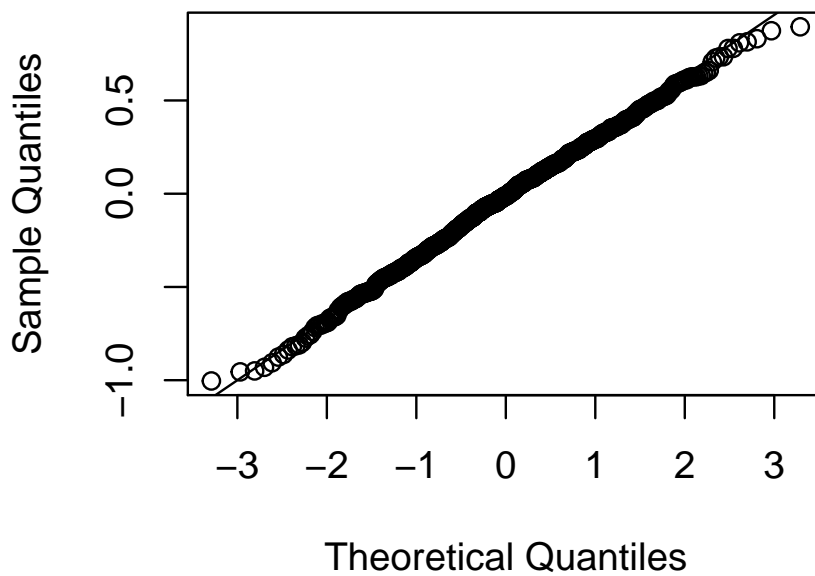
```
## [1] 0.1017097
```

```
var(sim_med)
```

```
## [1] 0.1414632
```

```
qqnorm(sim_mean); qqline(sim_mean)
```

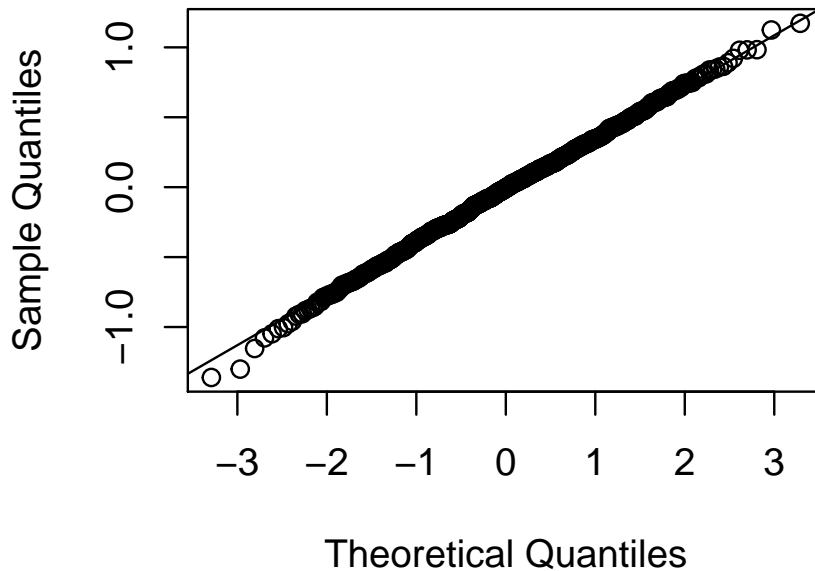
### Normal Q–Q Plot



```
qqnorm(sim_med); qqline(sim_med)
```



## Normal Q–Q Plot



### Example 14

Let's now consider the sample mean of random samples  $U_1, \dots, U_n$  with  $U_1 \sim \text{UNIF}(0,1)$ . What can we say about the distribution of the sample mean  $\bar{U}$  as the sample size  $n$  gets large?

We will create 1000 samples for  $n \in \{5, 20, 80\}$ , then compare the distributions.

```
sizes <- c(2, 5, 20, 80)
k <- 1000

datasets <- lapply(sizes, function(n) {
  replicate(k, runif(n))
})
names(datasets) <- sizes
str(datasets)

## List of 4
## $ 2 : num [1:2, 1:1000] 0.3575 0.0616 0.6279 0.0724 0.015 ...
## $ 5 : num [1:5, 1:1000] 0.7801 0.4316 0.0868 0.2669 0.5255 ...
## $ 20: num [1:20, 1:1000] 0.851 0.697 0.347 0.513 0.173 ...
## $ 80: num [1:80, 1:1000] 0.436 0.665 0.468 0.503 0.514 ...

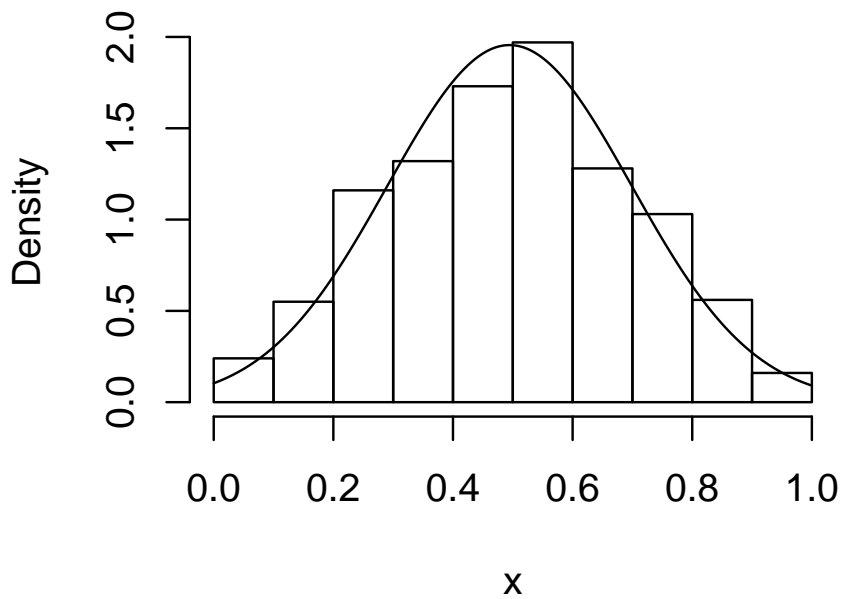
sim_mean_unif <- lapply(datasets, function(d) {apply(d, 2, mean)})
str(sim_mean_unif)
```

```
## List of 4
## $ 2 : num [1:1000] 0.21 0.35 0.479 0.228 0.804 ...
## $ 5 : num [1:1000] 0.418 0.716 0.779 0.407 0.491 ...
## $ 20: num [1:1000] 0.48 0.377 0.504 0.568 0.483 ...
## $ 80: num [1:1000] 0.513 0.495 0.492 0.5 0.503 ...

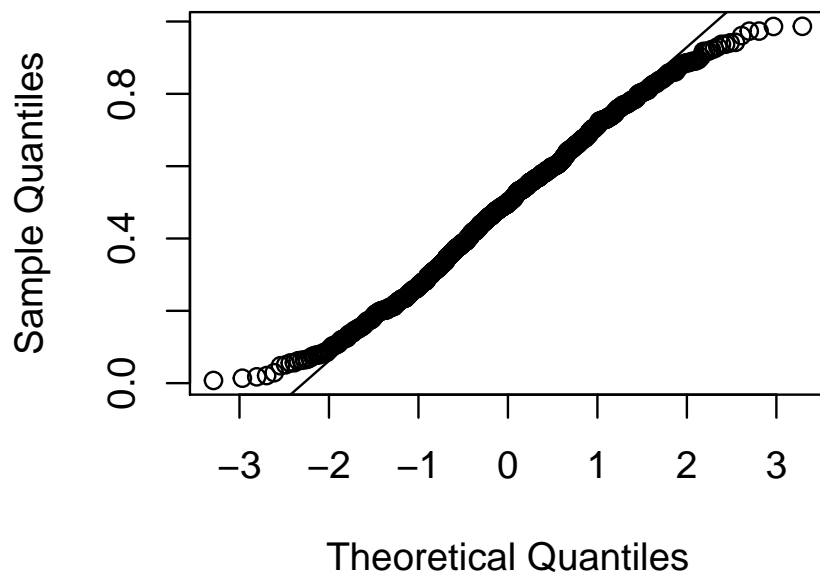
for (x in sim_mean_unif) {
  print(summary(x))
  hist(x, freq = FALSE)
  lines(seq(0, 1, length.out = 1000),
        dnorm(seq(0, 1, length.out = 1000), mean = mean(x), sd = sd(x)))
  qqnorm(x); qqline(x)
}

##      Min.   1st Qu.   Median     Mean   3rd Qu.
## 0.007424 0.349114 0.499931 0.494561 0.641655
##      Max.
## 0.986755
```

## Histogram of x

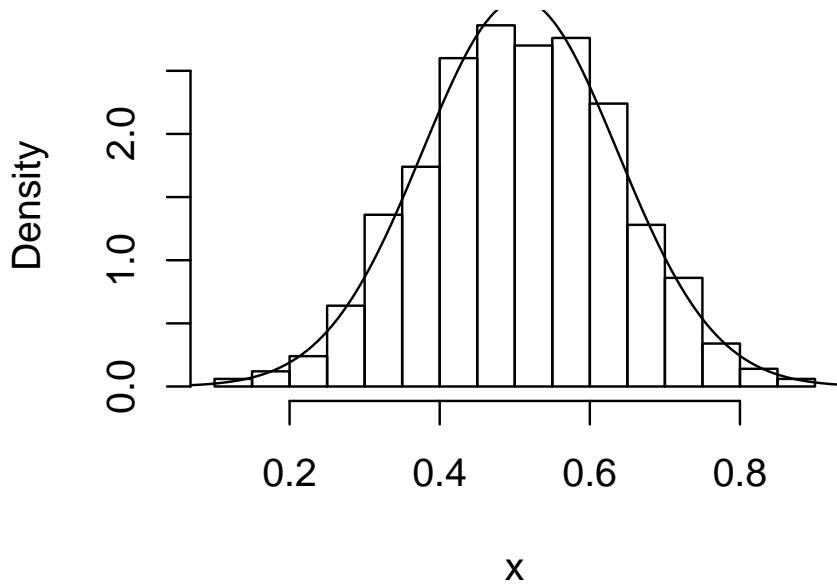


### Normal Q-Q Plot

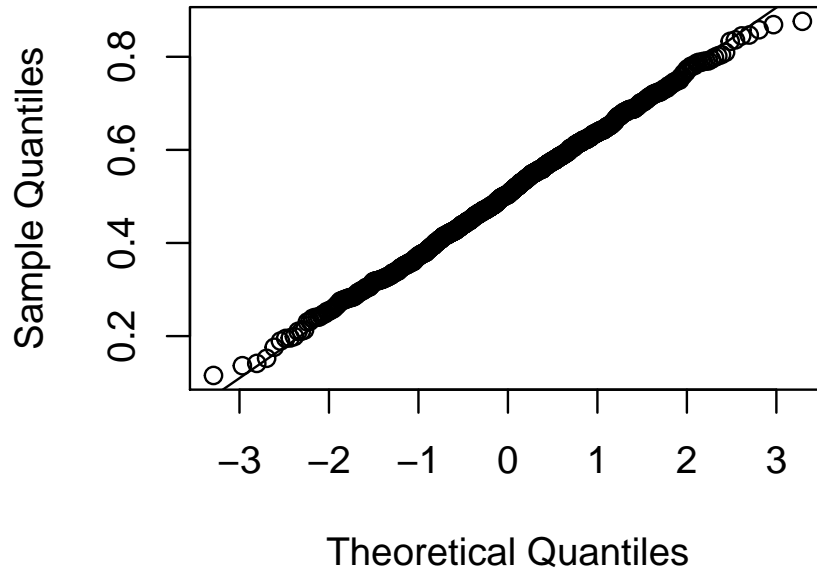


```
##   Min. 1st Qu.  Median    Mean 3rd Qu.
## 0.1154 0.4182  0.5044  0.5069  0.5971
##   Max.
## 0.8762
```

### Histogram of x

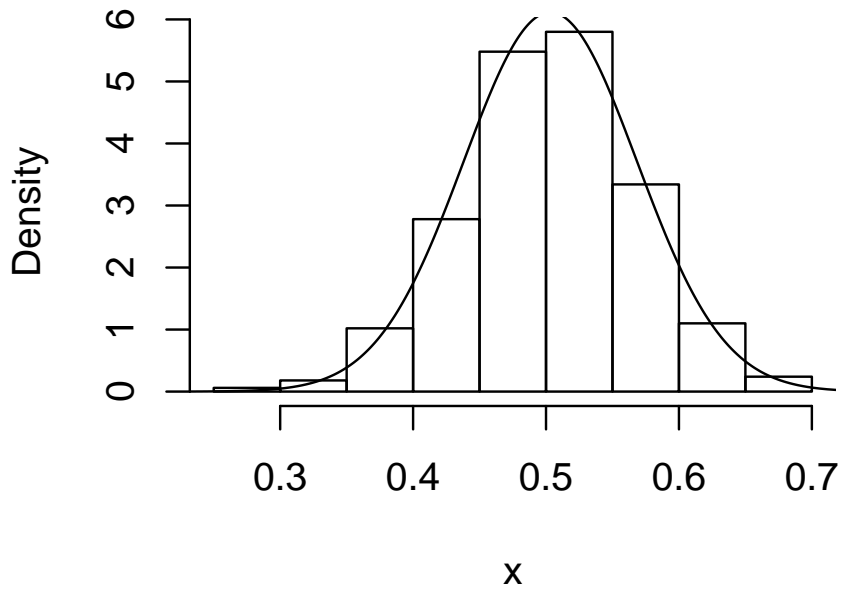


### Normal Q-Q Plot

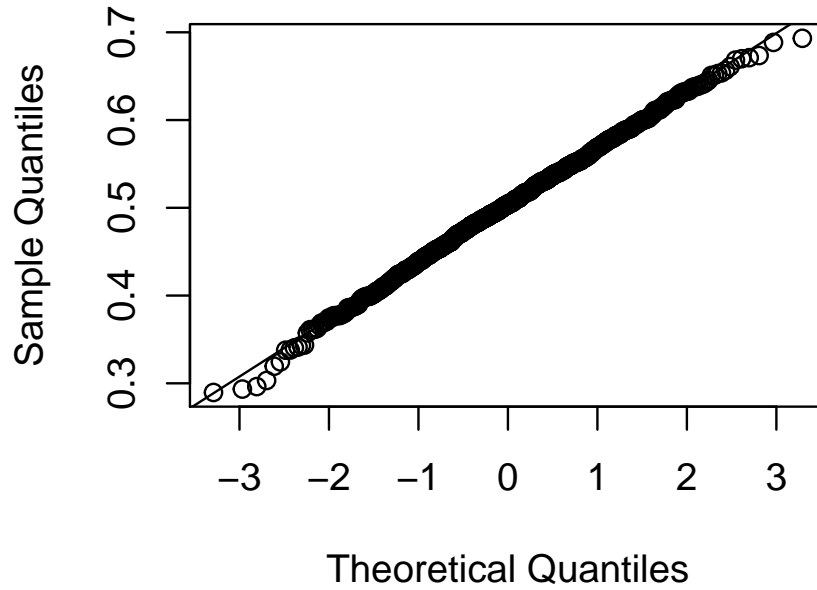


```
##   Min. 1st Qu.  Median    Mean 3rd Qu.
## 0.2896 0.4593  0.5041  0.5032  0.5473
##   Max.
## 0.6932
```

### Histogram of x

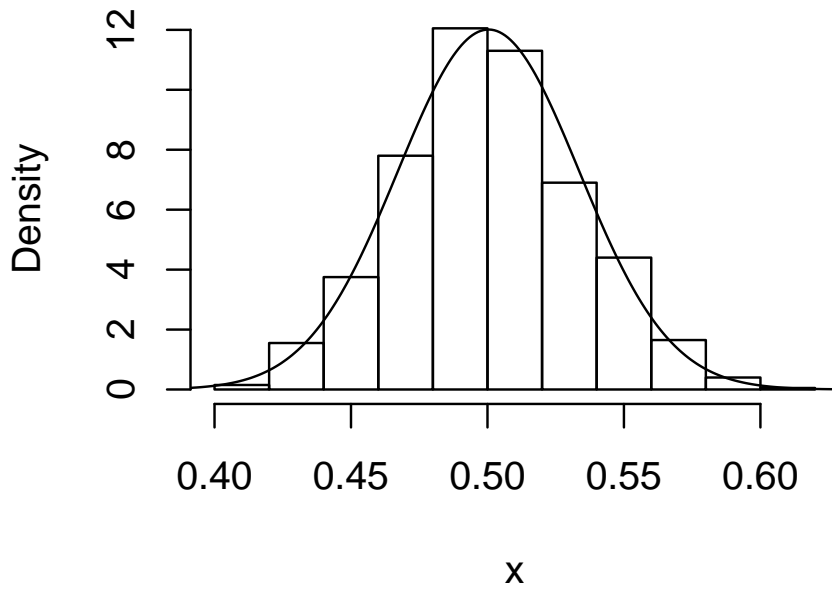


### Normal Q-Q Plot

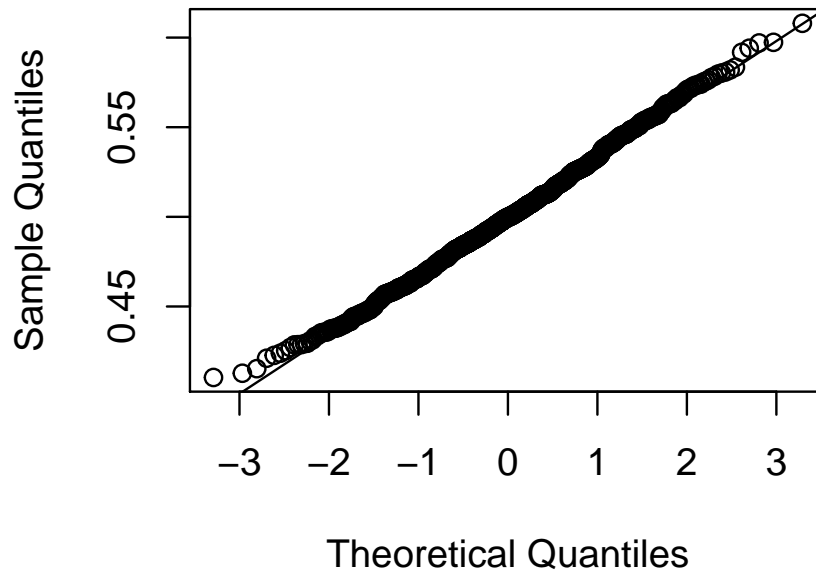


```
##      Min. 1st Qu.  Median      Mean 3rd Qu.
## 0.4104 0.4776 0.4998 0.5004 0.5218
##      Max.
## 0.6080
```

### Histogram of x



## Normal Q–Q Plot



### Section 4: The Distribution of the Sample Mean

There are two theorems that form the cornerstone of probability and statistics: the law of large numbers and the central limit theorem. The Law of Large Numbers (LLN) guarantees us that the sample mean will approximately equal the population mean, while the Central Limit Theorem (CLT) describes the distribution of the sample mean for large  $n$  when we have a mean and a variance.

**Theorem 1** (Law of Large Numbers). *Let  $X_1, X_2, \dots$  be a sequence of i.i.d.r.v. with  $\mathbb{E}[X_1] = \mu$ , and let  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Then the probability the (random) sequence  $\bar{X}_n$  converges to  $\mu$  is 1.*

**Theorem 2** (Central Limit Theorem). *Under the same assumptions as the Law of Large Numbers but with the additional assumption that  $\text{Var}(X_1) = \sigma^2$ ,  $\mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) \rightarrow \Phi(z)$  for all  $z$ . In other words, the distribution of  $\bar{X}_n$  is approximately  $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ , with the approximation improving as  $n \rightarrow \infty$ .*

More directly, the CLT describes the behavior of sums of i.i.d.r.v. as more random variables are summed. The CLT explains why some distributions—like the binomial distribution, the Poisson distribution, the gamma distribution, and the  $\chi^2$ -distribution—can be approximated with the Normal distribution as one of their parameters grows

large; these distributions can be interpreted as the distributions of sums of i.i.d.r.v.<sup>13</sup> and thus the CLT applies.<sup>14</sup>

Thanks to the CLT, we can describe the distribution of the sample mean without worrying about the exact distribution of the underlying data if the sample size  $n$  is large enough<sup>15</sup>, since the CLT says that the initial distribution is eventually “forgotten” by the sample mean.

---

### Example 15

The average customer visiting a grocery store spends  $X$  dollars, where  $\mathbb{E}[X] = 50$  and  $\text{SD}(X) = 55$ .<sup>16</sup> Every month about 30,000 purchases are made at the grocery store.

1. What will be the (approximate) distribution of the average purchase,  $\bar{X}$ ?

2. What is the (approximate) probability that the revenue of the grocery store in a month is less than \$1,485,000

<sup>13</sup> Specifically:  $\text{BIN}(n, p)$  is the sum of  $n$   $\text{Ber}(p)$  r.v.s;  $\text{POI}(n)$  is the sum of  $n$   $\text{POI}(1)$  r.v.s;  $\text{GAMMA}(\alpha, n)$  is the sum of  $n$   $\text{EXP}(\alpha)$  r.v.s; and  $\chi^2(n)$  is the sum of  $n$   $\chi^2(1)$  r.v.s.

<sup>14</sup> The CLT requires that  $\text{Var}(X_1) < \infty$ ; if this does not hold, the CLT no longer applies and its conclusion may not even be true.

<sup>15</sup> In general it's safe to use the CLT if  $n > 30$ .

<sup>16</sup> Notice  $X$  is non-negative but  $\text{SD}(X) > \mathbb{E}[X]$ . This can happen with skewed distributions.

```
55/sqrt(30000) # sd of xbar
```

```
## [1] 0.3175426
```

```
pnorm(1485000, mean = 50 * 30000, sd = 55 * sqrt(30000))
```

```
## [1] 0.05767537
```

---