Probability Review – 2.1 Fundamental Stuff

2.1.1 Randomness

- Unpredictability
- Outcomes we can’t predict are random
- Represents an inability to predict
- Example: rolling two dice

Sample Space

- Set of all outcomes of interest
- Dice example

Event

- Subset of outcomes
- Example: rolling higher than a 10

2.1.2 Probability

- Between 0 and 1 (or a percentage)
- "The probability of an event is the proportion of times it occurs in the long run"
- Probability of rolling 7, 12, or higher than 10?

2.2 Random Variables

- Translates random outcomes into numerical values
- Die roll has numerical meaning
- RVs are human-made
- Example: temperature in Celsius, Fahrenheit, Kelvin
- RVs can be discrete or continuous
- A continuous RV always has an infinite number of possibilities
- Probability of temp. being -20 tomorrow?
- Random variable vs. the *realization* of a random variable

2.3 Probability function

Probability function = probability distribution = probability distribution function (PDF) = probability mass function (PMF) = **probability function**

- Usually an equation
- Probability function: (i) lists all possible numerical values the RV can take; (ii) assigns a probability to each value.
- Prob. function contains all possible knowledge we can have about an RV
- 2.3.1 Example: die roll

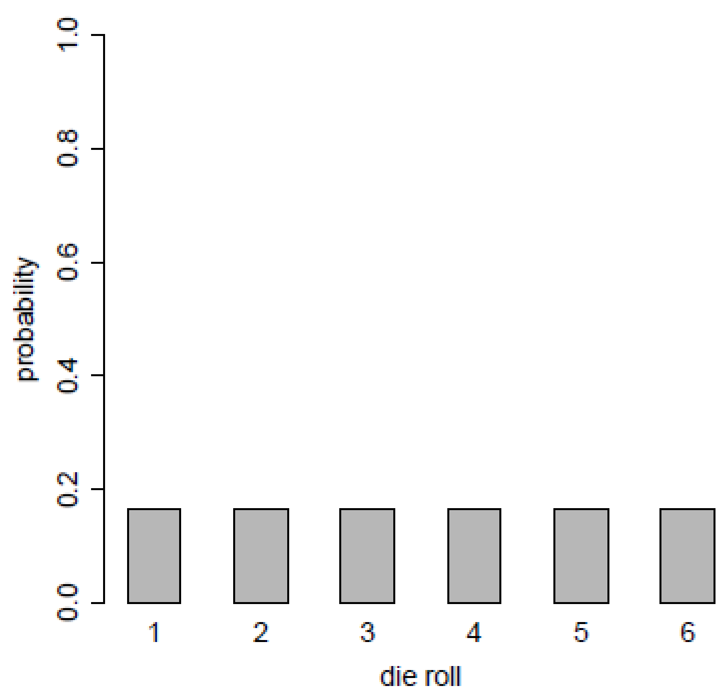
$$Pr(Y = y) = \frac{1}{6}; y = 1, \ldots, 6 \tag{2.2}$$

- 2.3.2 Example: a normal RV

$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y-\mu)^2}{2\sigma^2} \tag{2.3}$$

- Probability function for die roll in a picture:

Figure 2.1: Probability function for the result of a die roll



2.3.3 Probabilities of events

Probability function can be used to calculate the probability of events occurring.

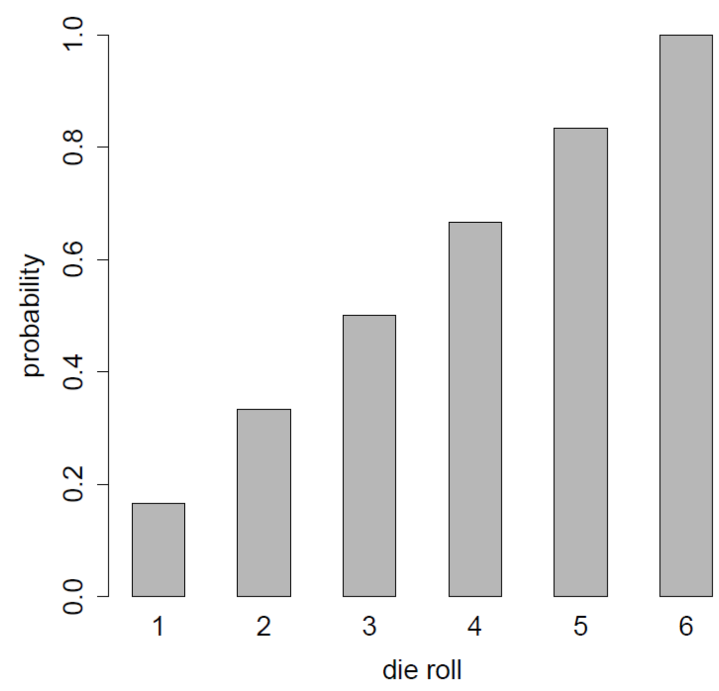
Example. Let Y be the result of a die roll. What is the probability of rolling higher than 3?

$$Pr(Y > 3) = Pr(Y = 4) + Pr(Y = 5) + Pr(Y = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

2.3.4 Cumulative distribution function (CDF)

- CDF is related to the probability function
- It's the prob. that the RV is *less than or equal to* a particular value
- In a picture:

Figure 2.2: Cumulative density function for the result of a die roll



2.4 Moments of a random variable

- “Moment” refers to a concept in physics
- 1^{st} moment is the mean
- 2^{nd} (central) moment is the variance
- 3^{rd} is skewness
- 4^{th} is kurtosis
- Covariance and correlation is a mixed moment

Moments summarize information about the RV. Moments are obtained from the ________________.

2.4.1 Mean (expected value)

- Value that is expected
- Average through repeated realizations of the RV
- Determined from the probability function (do some math to it)
- Mean is summarized info that is already contained in the prob. function

- Let Y be the RV
- Mean of Y = expected value of Y = $\mu_Y = E[Y]$
- If Y is discrete:

The mean is the weighted average of all possible outcomes, where the weights are the probabilities of each outcome.

The equation for the mean of *Y* (*Y* is discrete):

$$\mathrm{E}[Y] = \sum_{i=1}^{K} p_i Y_i \qquad (2.5)$$

where p_i is the probability of the i^{th} event, Y_i is the value of the i^{th} outcome, and K is the total number of outcomes (K can be infinite). Study this equation. It is a good way of understanding what the mean is.

Exercise: calculate the mean die roll.

What are the *properties* of the mean?

The equation for the mean of *y* (*y* is continuous):

Let y be a random variable. The mean of y is

$$\mathrm{E}[y] = \int y f(y)\, \mathrm{d}y$$

If y is normally distributed, then $f(y)$ is equation (2.3), and the mean of y turns out to by μ. You do not need to integrate for this course, but you should have some idea about how the mean of a continuous random variable is determined from its probability function.

The *mean* is different from the *median* and the *mode*, although all are measures of central tendency.

The mean is different from the sample mean or sample average. The mean comes from the probability function. The sample mean/average comes from a sample of data.

2.4.3 Variance

- Measure of the *spread* or *dispersion* of a RV
- Denoted by σ^2. The variance of y would be σ_y^2 and the variance of X would be σ_X^2
- Variance is the expected squared difference of a variable from its mean
- Equation:

2.4.3 Variance

- Measure of the *spread* or *dispersion* of a RV
- Denoted by σ^2. The variance of y would be σ_y^2 and the variance of X would be σ_X^2
- Variance is the expected squared difference of a variable from its mean
- Equation:

$$\mathrm{Var}(Y) = \mathrm{E}[(Y - \mathrm{E}[Y])^2] \tag{2.6}$$

When Y is a discrete random variable, then equation (2.6) becomes

$$\mathrm{Var}(Y) = \sum_{i=1}^{K} p_i \times (Y_i - \mathrm{E}[Y_i])^2 \tag{2.7}$$

- For variance (the 2nd moment), we are taking the expectation of a squared term
- For skewness (the 3rd moment), we would take the expectation of a cubed term, etc.

Exercise: calculate the variance of a die roll

What are the *properties* of the variance?

Exercise: I change the sides of the die to equal 2,4,6,8,10,12. What is the mean and variance of the die roll?

Exercise: What is the mean and variance of the sum of two dice?

2.4.5 Covariance

- Measures the relationship between two random variables
- Random variables Y and X have a *joint* probability function
- Joint prob. func.: (i) lists all possible combos of Y and X; (ii) assign a probability to each combination
- A useful summary of a joint probability function is the *covariance*
- The covariance between Y and X is the expected difference of Y from its mean, multiplied by the expected difference of X from its mean
- Covariance tells us something about how two variables are *related*, or how they *move together*
- Tells us about the direction and strength of the relationship between two variables

$$\text{Cov}(Y, X) = \text{E}[(Y - \mu_Y)(X - \mu_X)] \tag{2.8}$$

The covariance between Y and X is often denoted as σ_{YX}. Note the following properties of σ_{YX}:

- σ_{YX} is a measure of the *linear* relationship between Y and X. Non-linear relationships will be discussed later.
- $\sigma_{YX} = 0$ means that Y and X are linearly independent.
- If Y and X are independent (neither variable causes the other), then $\sigma_{YX} = 0$. The converse is not necessarily true (because of non-linear relationships).
- The $\text{Cov}(Y, Y)$ is the $\text{Var}(Y)$.
- A positive covariance means that the two variables tend to differ from their mean in the *same* direction.
- A negative covariance means that the two variables tend to differ from their mean in the *opposite* direction.

2.4.6 Correlation

- Correlation usually denoted by ρ
- Similar to covariance, but is easier to interpret

$$\rho_{YX} = \frac{\text{Cov}(Y, X)}{\sqrt{\text{Var}(Y)\text{Var}(X)}} = \frac{\sigma_{YX}}{\sigma_Y \sigma_X} \tag{2.9}$$

The difficulty in interpreting the value of covariance is because $-\infty < \sigma_{YX} < \infty$. Correlation transforms covariance so that it is bound between -1 and 1. That is, $-1 \leq \rho_{YX} \leq 1$.

- $\rho_{YX} = 1$ means perfect positive linear association between Y and X.
- $\rho_{YX} = -1$ means perfect negative linear association between Y and X.
- $\rho_{YX} = 0$ means no linear association between Y and X (linear independence).

2.4.7 Conditional distribution

- Joint distribution – 2 RVs
- Conditional distribution – fix (condition on) one of those RVs
- Condition expectation – the mean of one RV after the other RV has been "fixed"

Let Y be a discrete random variable. Then, the conditional mean of Y given some value for X is

$$\mathrm{E}(Y|X=x) = \sum_{i=1}^{K}(p_i|X=x)Y_i \tag{2.10}$$

- If the two RVs are independent, the conditional distribution is the same as the *marginal* distribution

Example: Blizzard and cancelled midterm

Suppose that you have a midterm tomorrow, but there is a possibility of a blizzard. You are wondering if the midterm might be cancelled.

Table 2.1: Joint distribution for snow and a canceled midterm

	Midterm ($Y = 1$)	No Midterm ($Y = 0$)
Blizzard ($X = 1$)	0.05	0.20
No Blizzard ($X = 0$)	0.72	0.03

- What are the *marginal* probability distributions?
- What is E[Y]? What is E[$Y \mid X = 1$]?
- What is the covariance and correlation between X and Y?
- More exercises in the "Review Questions"

2.5 Some special probability functions

2.5.1 The normal distribution

- Common because of the “central limit theorem” (in a few slides)

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y-\mu)^2}{2\sigma^2} \quad (2.3)$$

- Mean of y is μ
- Variance of y is σ^2

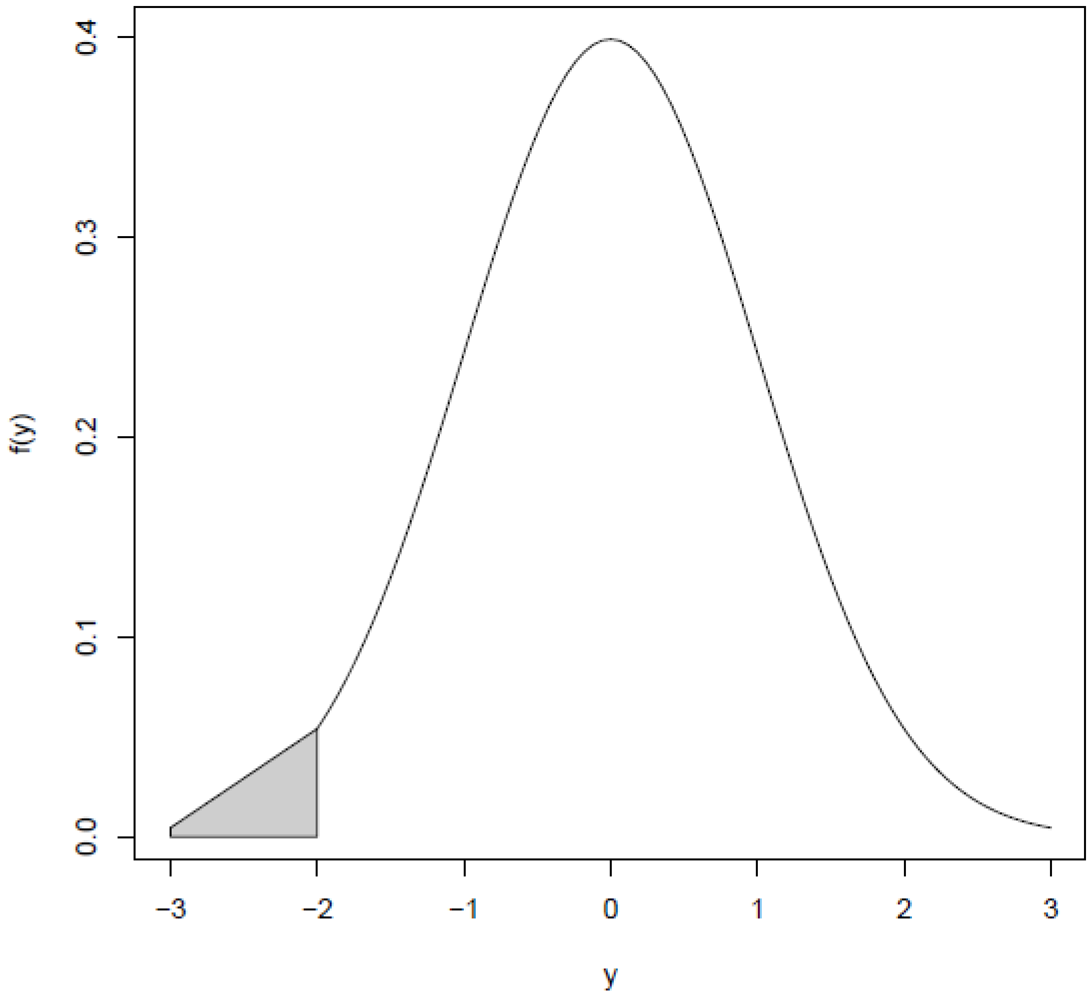
2.5.2 The standard normal distribution

- Special case of a normal distribution, where μ = 0 and σ^2 =1
- Equation 2.3 becomes:

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp \frac{-y^2}{2} \qquad (2.11)$$

- Any normal random variable can be "standardized"
- How to standardize?
- Standardizing has long been used in hypothesis testing (as we shall see)

Figure 2.3: Probability function for a standard normal variable, $p_{y<-2}$ in gray



2.5.3 The central limit theorem

- There are hundreds of different probability functions
- Examples: Poisson, Binomial, Generalized Pareto, Nakagami, Uniform
- So why is the normal distribution so important? Why are so many RVs normal?
- Answer: CLT
- CLT (loosely speaking) – if we add up enough RVs, the resulting sum tends to be normal

Exercise: draw the probability function for one die roll, then for the sum of two dice.

Figure 2.1: Probability function for the result of a die roll

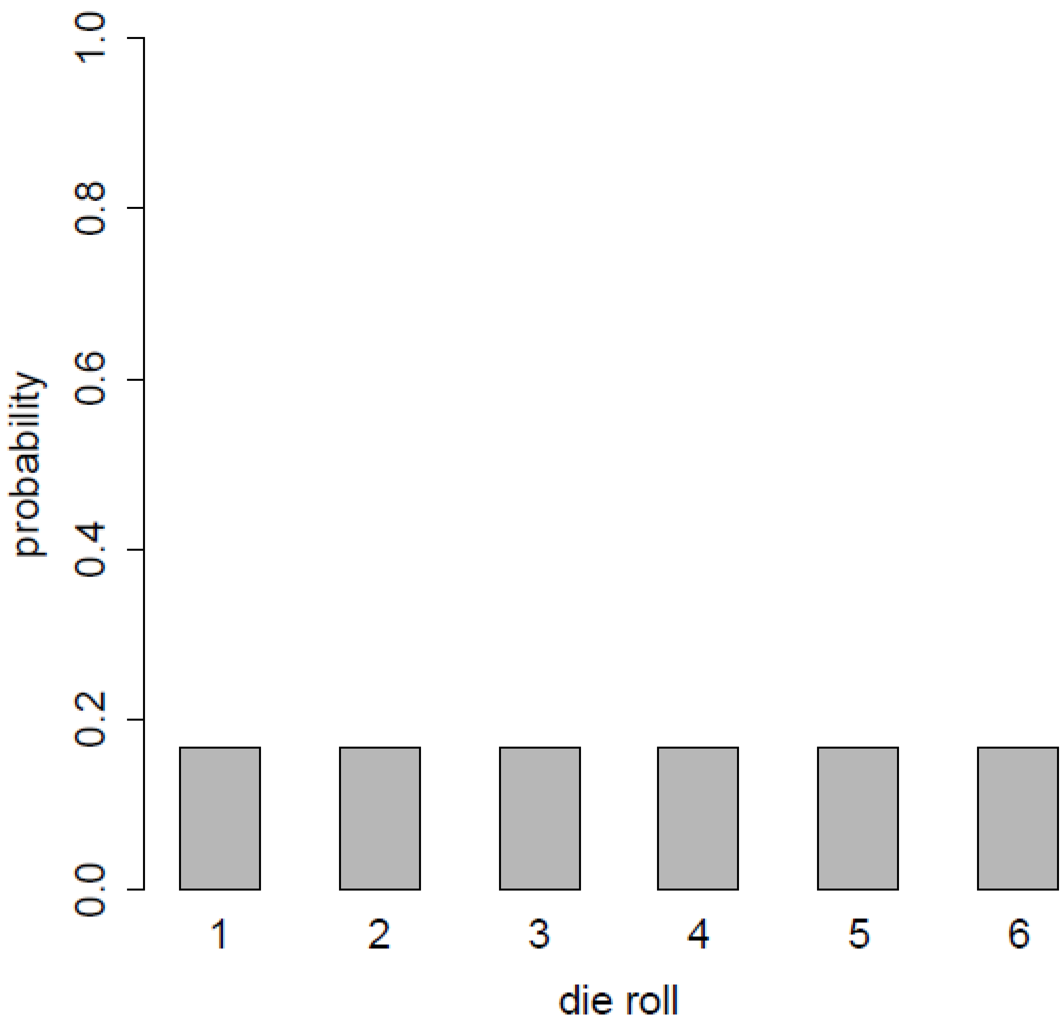


Figure 2.4: Probability function for the sum of two dice

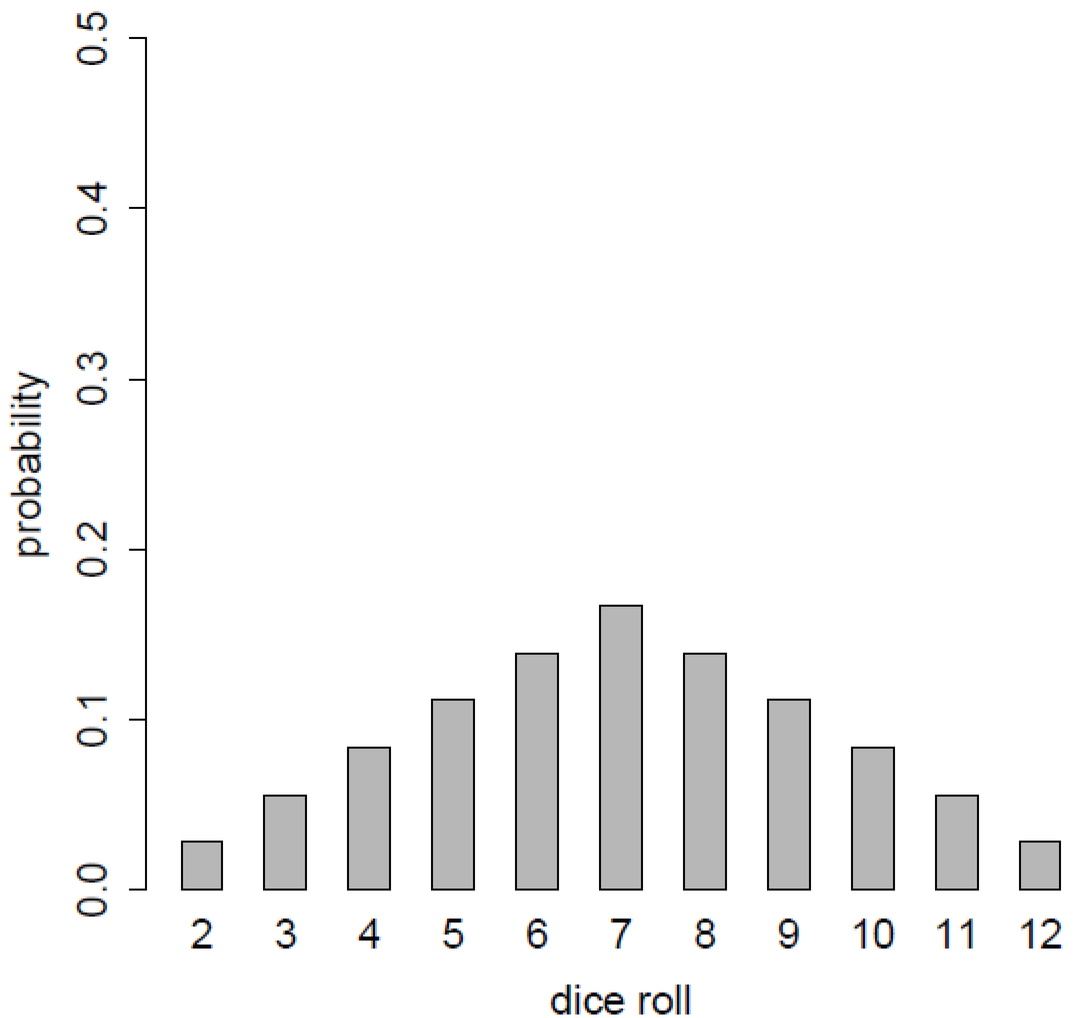


Figure 2.5: Probability function for three dice, and normal distribution

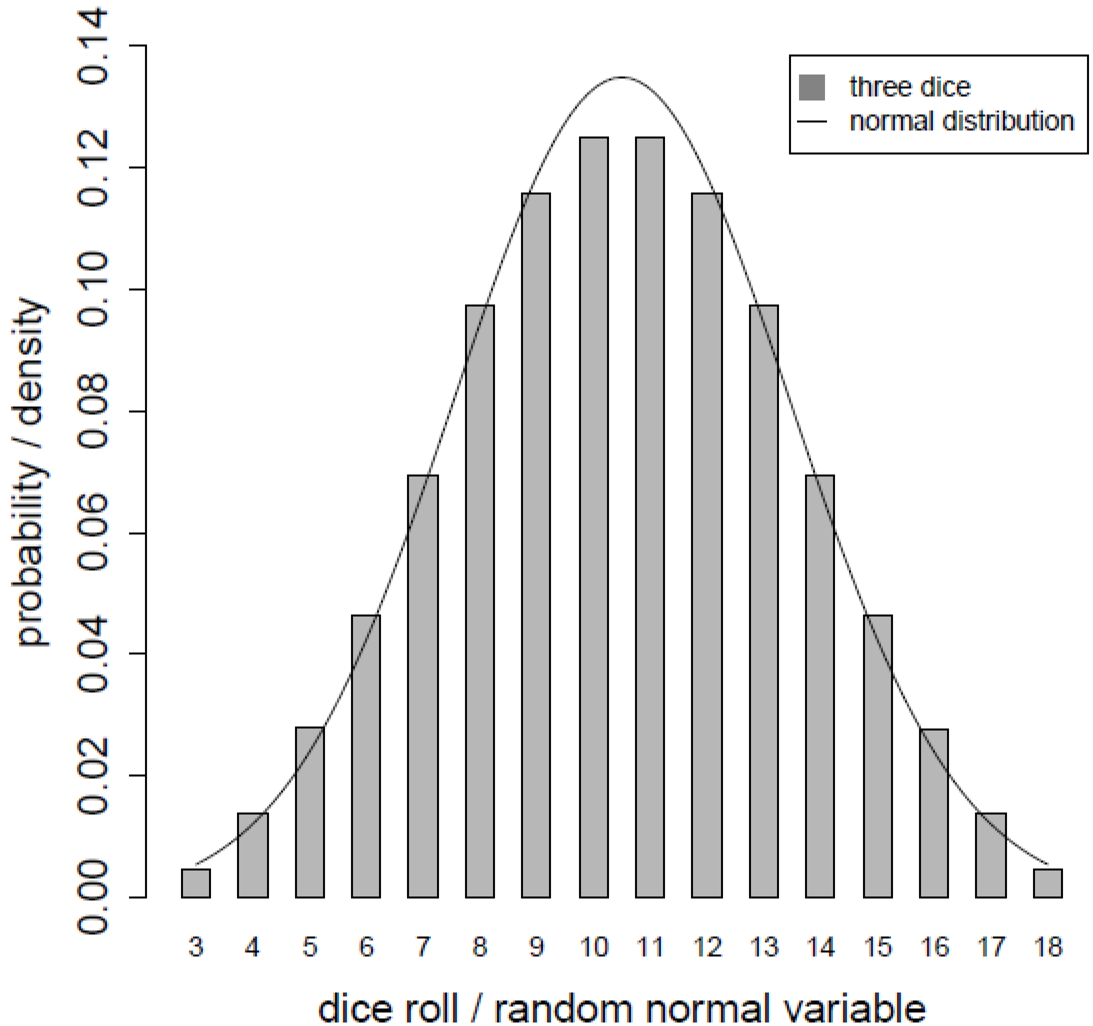
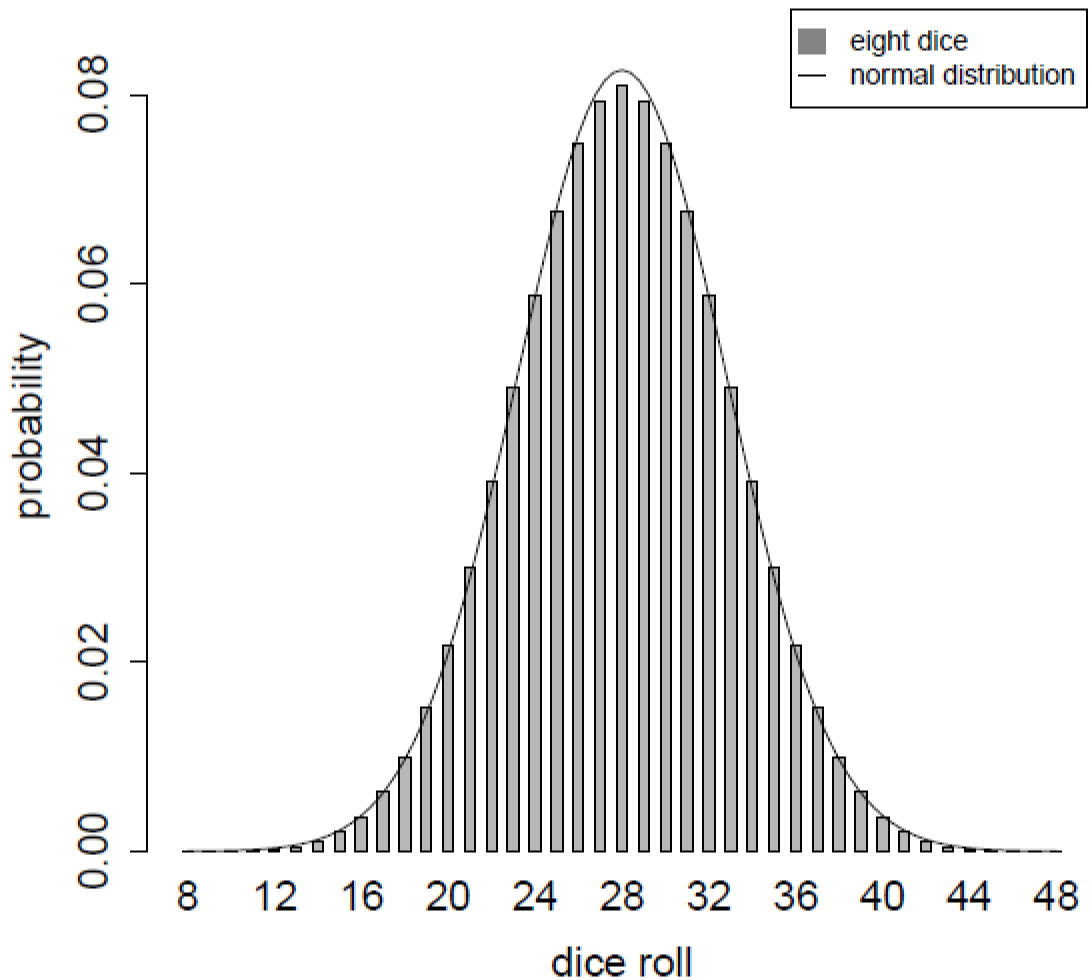


Figure 2.6: Probability function for eight dice, and normal distribution



2.5.4 The chi-square distribution

- Add to a normal RV – still normal
- Multiply a normal RV – still normal
- Square a normal RV – now it is chi-square distributed
- We will use the chi-square distribution for the F-test in a later chapter