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

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

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

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

- 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)

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

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

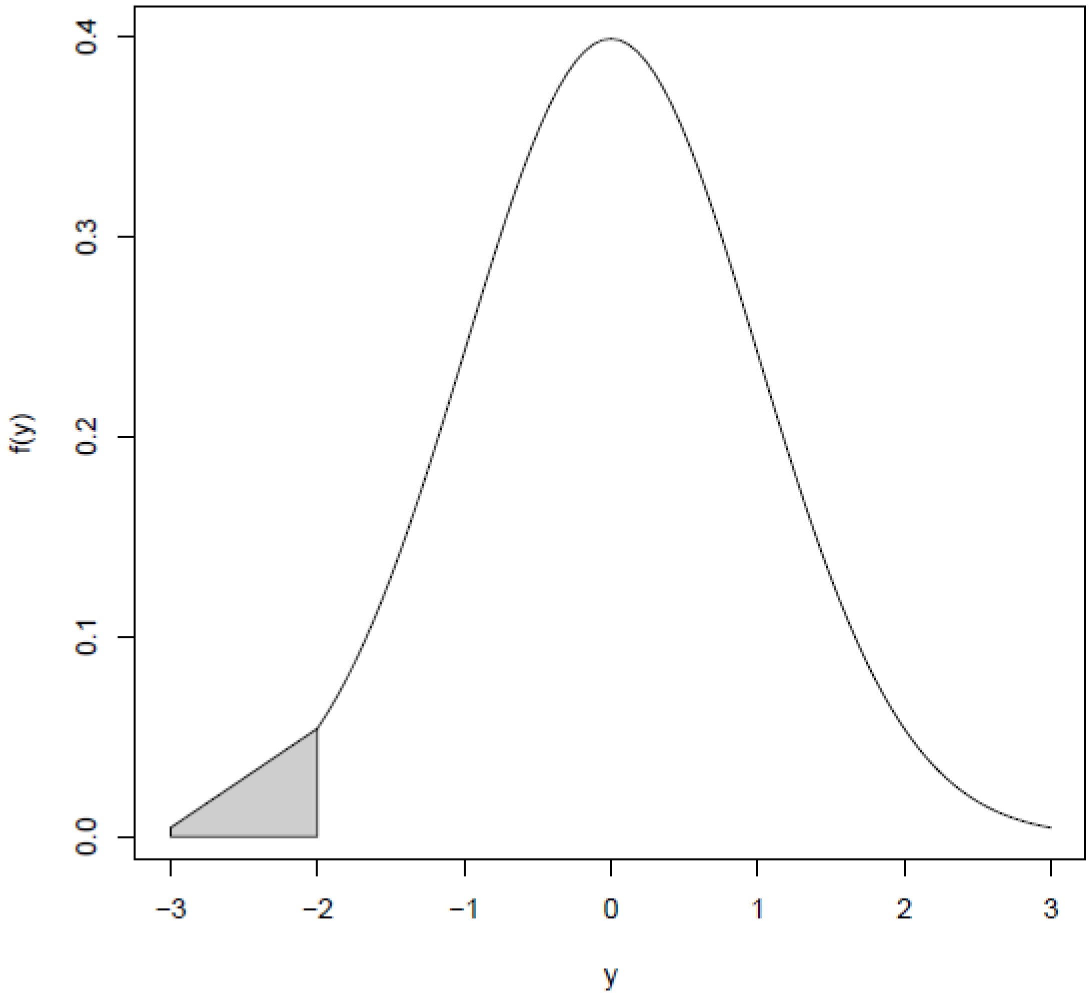
2.5.2 The standard normal distribution

- 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

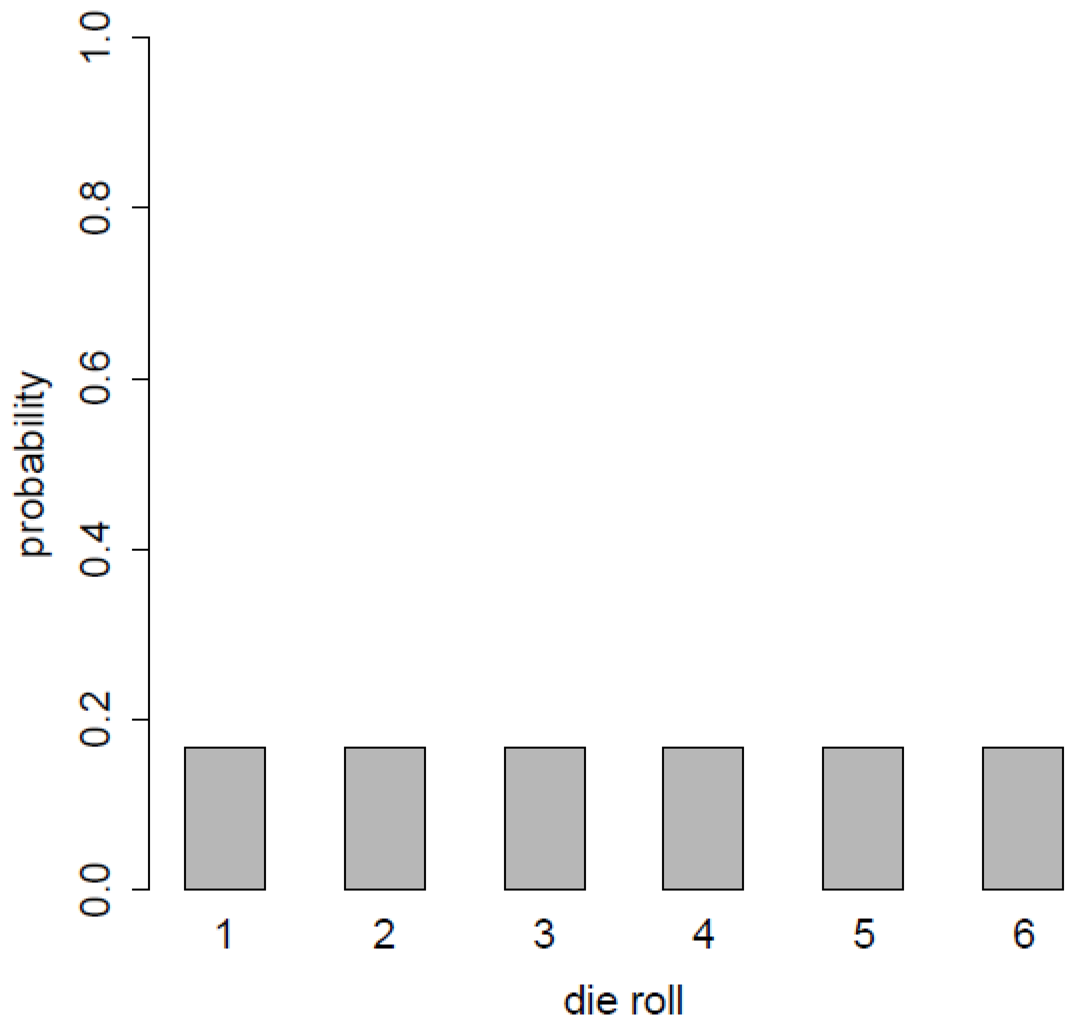


Figure 2.4: Probability function for the sum of two dice

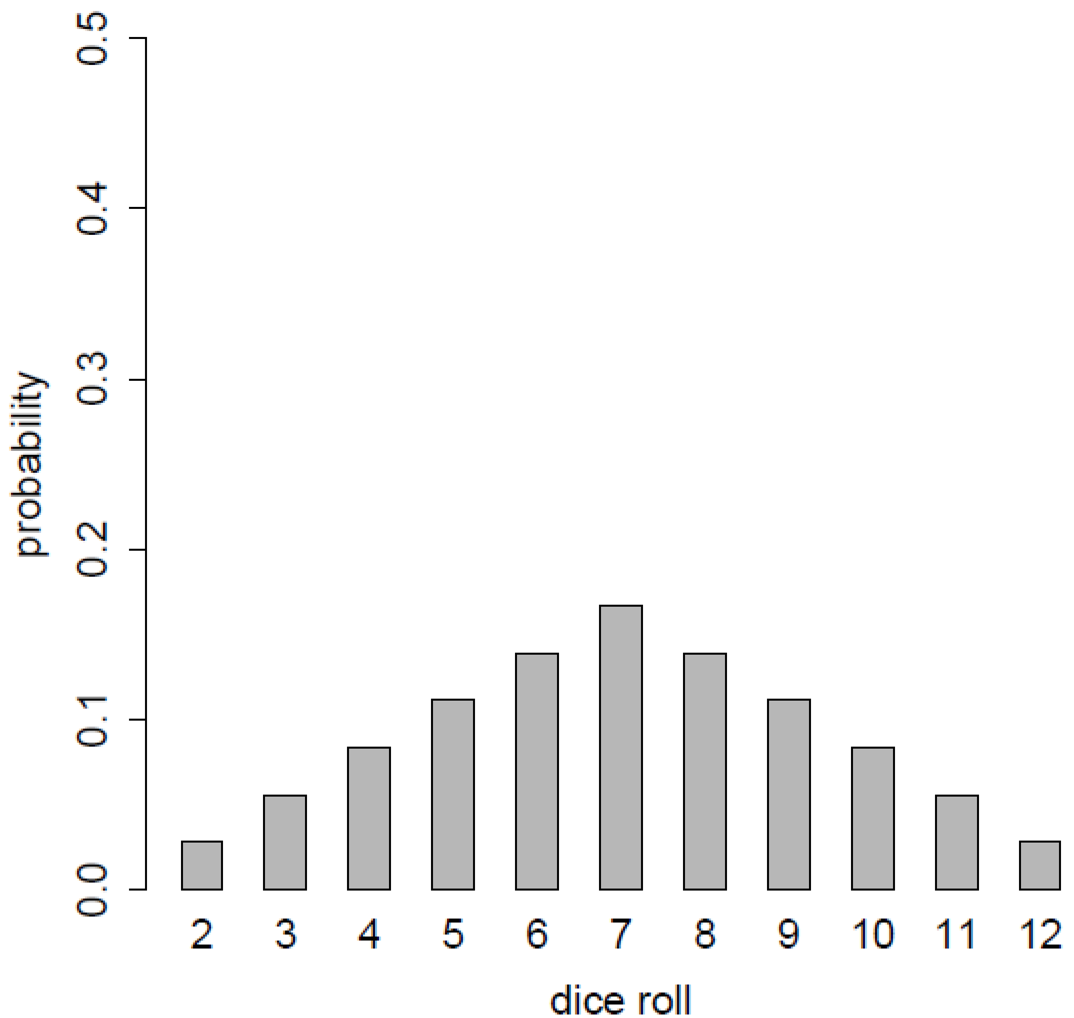


Figure 2.5: Probability function for three dice, and normal distribution

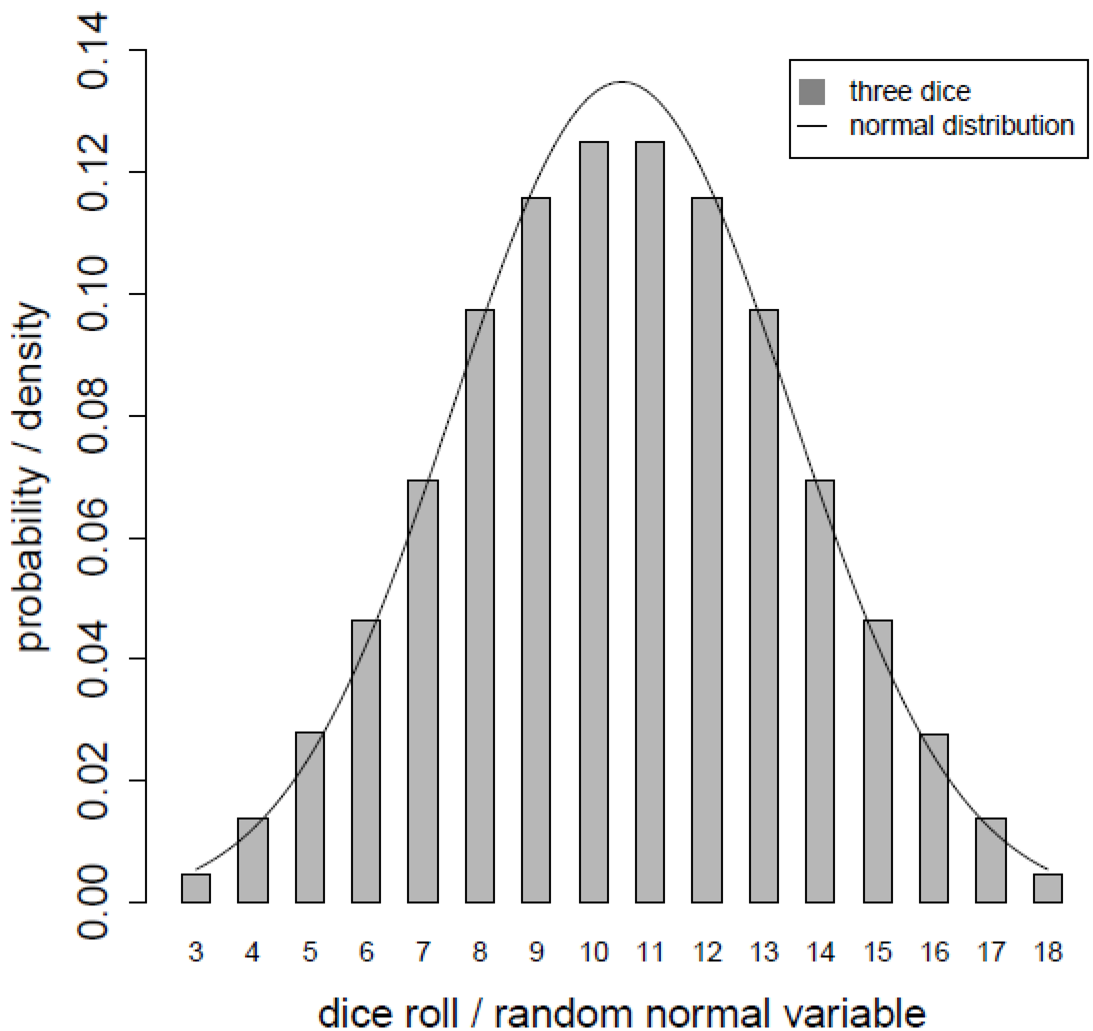
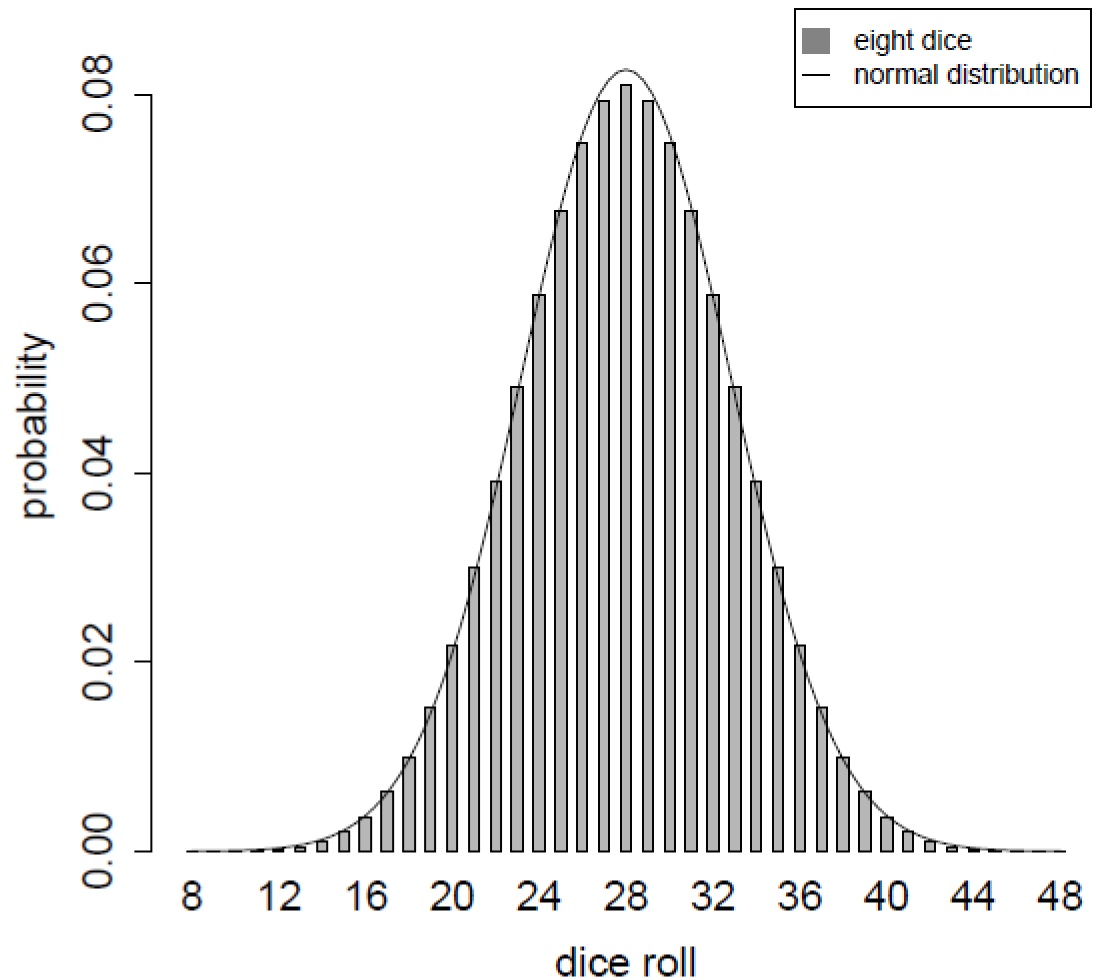


Figure 2.6: Probability function for eight dice, and normal distribution



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