5 Expectation and variance

In Section 3 and 4 we encountered a wide range of random variables. It is often useful to describe or summarise their behaviour and in this section we introduce several quantities to do this. We pay particular attention to two of key importance: the expectation and the variance of a random variable. Both quantities will play a crucial role in proving central results in Section 6.

5.1 Expectation

Definition 5.1 (Expectation – discrete case)**.** Let X be a discrete random variable. The *expectation* of X is given by

$$\mathbb{E}[X] := \sum_{x \in S_X} x \cdot \mathbb{P}(X = x),$$

provided the series $\sum_{x \in S_X} |x| \cdot \mathbb{P}(X = x)$ converges. If the sum $\sum_{x \in S_X} |x| \cdot \mathbb{P}(X = x)$ diverges then the expectation is undefined [1].

Remark 5.2. Let X be a discrete random variable.

- The expectation of X is often called the *mean* or the *average* of X.
- If S_X is a finite set then $\sum_{x \in S_X} |x| \cdot \mathbb{P}(X = x)$ converges, and so the expectation always exists. This might not be true if $|S_X|$ is infinite (see Example 5.10 below).

Example 5.3. Let X satisfy $\mathbb{P}(X = 2) = 0.4, \mathbb{P}(X = 4) = 0.1$ and $\mathbb{P}(X = 10) = 0.5$. Then

$$\mathbb{E}[X] = 2 \cdot (0.4) + 4 \cdot (0.1) + 10 \cdot (0.5) = 6.2.$$

When betting on a game of chance, an important random variable is the amount of money gained over the game. The expectation is then often used to decide if the game is worth playing.

Example 5.4. Take two kings and four aces, shuffle the six cards thoroughly and then draw two without replacement. We win £1.25 if we draw two aces; otherwise, we pay £1. Should we play?

Let X denote the amount gained in a game (negative if we lose). The question gives $\mathbb{P}(X = 1.25) = \binom{4}{2}/\binom{6}{2} = 0.4$ and $\mathbb{P}(X = -1) = 0.6$, so

$$\mathbb{E}[X] = (1.25) \cdot \mathbb{P}(X = 1.25) + (-1) \cdot \mathbb{P}(X = -1) = (1.25) \cdot (0.4) + (-1) \cdot (0.6) = -0.1.$$

As the expected gain is negative, on average we will loss money and so should not play [2].

[1] If a series is absolutely convergent then the order in which we sum its elements does not effect the answer, proven in 1SAS. This is why absolute convergence is important here.

[2] Hopefully this sounds like an good criteria to make a decision about whether to play but at the moment this relies only on our intuition. We will obtain a more rigorous justification in Section 6.

Let us calculate the expectations of some familiar random variables.

Example 5.5 (Constant). If X has $\mathbb{P}(X = c) = 1$ for some $c \in \mathbb{R}$ then $\mathbb{E}[X] = c \cdot \mathbb{P}(X = c) = c$.

Example 5.6 (Discrete uniform). Let X follow the uniform distribution on $\{1, \dots, n\}$, that is $\mathbb{P}(X = k) = 1/n$ for all $k = 1, \dots, n$. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} i \cdot \mathbb{P}(X = i) = \sum_{i=1}^{n} \frac{i}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

Example 5.7 (Binomial). On problem sheet 4 you will show that if $X \sim \mathrm{bin}_{n,p}$ then $\mathbb{E}[X] = np$.

Example 5.8 (Geometric). Let $X \sim \mathrm{geo}_p$ with $p \in (0,1)$. Here there is a small trick. Taking $q = 1-p$

$$p \cdot \mathbb{E}[X] = \mathbb{E}[X] - q \cdot \mathbb{E}[X] = \sum_{k=1}^{\infty} kpq^{k-1} - \sum_{k=1}^{\infty} kpq^{k} = \sum_{k=1}^{\infty} kpq^{k-1} - \sum_{k=2}^{\infty} (k-1)pq^{k-1} = \sum_{k=1}^{\infty} pq^{k-1} = 1.$$

The third equality here follows by a change of variable and the final equality follows by summing a geometric series. Rearranging we get $\mathbb{E}[X] = \frac{1}{p}$.

Example 5.9 (Poisson). Let $X \sim \mathrm{Poi}_\lambda$ with $\lambda > 0$. Then,

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \overset{\ell = k-1}{=} \lambda e^{-\lambda} \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Example 5.10 (No expectation). Let X be a discrete random variable with $\mathbb{P}(X = n) = \frac{1}{n(n+1)}$ for $n \in \mathbb{N}$. These probabilities do indeed sum to one since:

$$\sum_{n=1}^{m} \frac{1}{n(n+1)} = \sum_{n=1}^{m} \frac{1}{n} - \sum_{n=1}^{m} \frac{1}{n+1} = 1 - \frac{1}{m+1} \to 1, \qquad \text{as } m \to \infty.$$

However $\sum_{n=1}^{\infty} n \cdot \mathbb{P}(X = n) = \sum_{n=1}^{\infty} \frac{1}{n+1}$ which diverges [3]. Therefore $\mathbb{E}[X]$ is not defined in this case.

The expectation of continuous random variables is defined very similarly.

Definition 5.11 (Expectation – continuous case). Let X be a continuous random variable with density f_X. Then the *expectation* of X is given by

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x \cdot f_X(x) dx,$$

provided $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx$ exists. If $\int_{-\infty}^{\infty} |x| \cdot f_X(x) dx$ does not exist then $\mathbb{E}[X]$ is undefined.

Example 5.12 (Continuous uniform). Let $X \sim \mathrm{unif}[a,b]$ with density $f_X(x) = (b-a)^{-1}$ for $x \in [a,b]$ and $f_X(x) = 0$ for $x \notin [a,b]$. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = (b-a)^{-1} \int_{a}^{b} x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

[3] As the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

Example 5.13 (Exponential)**.** Let $X \sim \exp_\lambda$ where $\lambda > 0$. Then X has density function $f_X(x) = \lambda \cdot e^{-\lambda x}$ for $x > 0$ and $f_X(x) = 0$ otherwise. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x}dx = \big[x(-e^{-\lambda x})\big]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}dx = \frac{1}{\lambda},$$

where we used integration by parts in the second last equality.

Example 5.14. Let $f(x) = \frac{1}{4}|\sin(x)|$ for $x \in [0, 2\pi]$ and $f(x) = 0$ elsewhere. Then f gives a density function (check this!) and if X is a random variable with this density

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx = \frac{1}{4} \cdot \int_0^{2\pi} x \cdot |\sin(x)|dx = \frac{1}{4} \cdot \int_0^{\pi} x \cdot \sin(x)dx - \frac{1}{4} \cdot \int_{\pi}^{2\pi} x \cdot \sin(x)dx.$$

Integration by parts (with $u = x$ and $v = -\cos(x)$) gives $\int_a^b x\sin(x)dx = [-x \cdot \cos(x) + \sin(x)]_a^b$. Thus

$$\mathbb{E}[X] = \frac{1}{4} \cdot [-x \cdot \cos(x) + \sin(x)]_0^{\pi} - \frac{1}{4} \cdot [-x \cdot \cos(x) + \sin(x)]_{\pi}^{2\pi} = \frac{(\pi + 0)}{4} - \frac{(-2\pi - \pi)}{4} = \pi.$$

5.2 Key properties of expectation

The next theorem, together with Remark 5.16), is absolutely central to *why* expectation is so useful.

Theorem 5.15. *Let X, Y be discrete or continuous random variables with well-defined expectations.*

(i) Given any $a, b \in \mathbb{R}$, we have $\mathbb{E}[aX + bY] = a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y]$. (linearity of expectation)

(ii) If X, Y are independent then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

We postpone the proof of Theorem 5.15 to end of the section so that we can highlight some aspects of the theorem and see some applications.

Remark 5.16. The following points are of key importance.

- Linearity of expectation (Theorem 5.15 (i)) extends by induction on n to give that for any $n \geq 1$, $a_1, \ldots, a_n \in \mathbb{R}$ and random variables $X_1, \ldots, X_n$ with well-defined expectations, we have

$$\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n].$$

- From Section 3 we know that if X, Y are independent random variables then $f(X), g(Y)$ are also independent random variables for any functions $f, g : \mathbb{R} \to \mathbb{R}$. It follows from Theorem 5.15 (ii) that $\mathbb{E}[X^2Y^3] = \mathbb{E}[X^2] \cdot \mathbb{E}[Y^3]$, that $\mathbb{E}[\sin(X) \cdot e^Y] = \mathbb{E}[\sin(X)] \cdot \mathbb{E}[e^Y]$, etc. [4].

- Independence is key in (ii); in general for random variables X, Y we have $\mathbb{E}[X \cdot Y] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

[4] Provided all these expectations are well defined.

Example 5.17 (Dice)**.** Let $X_1, \dots, X_n$ be random variables with the uniform distribution on $\{1, \dots, 6\}$ (representing fair dice rolls). Then the random variable $Y = X_1 + \cdots + X_n$ has $S_Y = \{n, n+1, \dots, 6n-1, 6n\}$ but $\mathbb{P}(Y = k)$ is quite tricky to calculate. However $\mathbb{E}[Y] = \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X_1]$ by Theorem 5.15(i). But $\mathbb{E}[X_1] = 3.5$ by Example 5.6, so $\mathbb{E}[Y] = 3.5n$.

Example 5.18 (Sum of digits)**.** Let X be uniformly distributed on $\{0, 1, \dots, 999\}$. Let Y be its sum of digits. What is $\mathbb{E}[Y]$? We can write $X = X_1 + 10X_2 + 100X_3$, where $X_i \in \{0, 1, \dots, 9\}$ follows the uniform distribution on this set. In particular, $\mathbb{E}[X_i] = (0 + 1 + 2 + \cdots + 9)/10 = 4.5$. As $Y = X_1 + X_2 + X_3$, we find $\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] = 13.5$ using linearity of expectation.

Example 5.19 (Expectation of the normal distribution)**.** We first focus on the standard normal, $\mathcal{N} \sim N(0, 1)$. Using that $xe^{-x^2/2}$ is an odd function, by symmetry

$$\mathbb{E}[\mathcal{N}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} xe^{-x^2/2} dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} xe^{-x^2/2} dx = 0. \tag{1}$$

Now to consider $X \sim N(\mu, \sigma^2)$. By Proposition 4.19 we have $\mu + \sigma\mathcal{N} \sim N(\mu, \sigma^2)$ and it follows from Theorem 5.15 *(i)* that $\mathbb{E}[X] = \mathbb{E}[\mu + \sigma\mathcal{N}] = \mathbb{E}[\mu] + \sigma \cdot \mathbb{E}[\mathcal{N}] = \mu$, by (1).

To further exploit linearity of expectation we need the notion of a Bernoulli random variable.

Definition 5.20 (Bernoulli distribution)**.** The *Bernoulli distribution with parameter* p, with $p \in [0, 1]$, is the probability distribution on $\{0, 1\}$ given by

$$\mathrm{Ber}_p(1) = p, \qquad \mathrm{Ber}_p(0) = 1 - p.$$

A random variable X follows the Bernoulli distribution with parameter p if $S_X = \{0, 1\}$ and $\mathbb{P}(X = k) = \mathrm{Ber}_p(k)$ for all $k \in S_X = \{0, 1\}$. In this case we write $X \sim \mathrm{Ber}_p$. (Note that $\mathrm{bin}_{1,p} = \mathrm{Ber}_p$.)

Example 5.21 (Bernoulli)**.** If $X \sim \mathrm{Ber}_p$ then $\mathbb{E}[X] = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p$.

Example 5.22. Let $X \sim \mathrm{Ber}_{0.5}$ and set $Y = 1 - X$. Then $Y \sim \mathrm{Ber}_{0.5}$ and $X \cdot Y = 0$. This gives $\mathbb{E}[X \cdot Y] = 0 \neq 1/4 = \mathbb{E}[X] \cdot \mathbb{E}[Y]$; see the third bullet point in Remark 5.16.

In computing expectations, it is often possible to 'break up' a random variable X into a sum of simpler ones $X = \sum_{i=1}^{n} X_i$ and then use linearity of expectation to find $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$. The following three examples illustrate this method, which is very applicable.

Example 5.23 (Binomial)**.** Let $X \sim \mathrm{bin}_{n,p}$. We will show $\mathbb{E}[X] = np$. We know that X counts the number of occurrences of independent events $A_1, \dots, A_n$, where each event happens with probability p. Thus, $X = X_1 + X_2 + \cdots + X_n$ with Bernoulli random variables $X_1, \dots, X_n$, where $X_i = 1$ if and only if A_i occurs. Linearity of expectation and Example 5.21 give $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np$.

Example 5.24 (Hypergeometric)**.** We show that if $X \sim \mathrm{hyp}_{n,r,t}$ where $n, r \leq t$ then $\mathbb{E}[X] = n\left(\frac{r}{t}\right)$. One could compute this by hand, exploiting identities for binomial coefficients (exercise!). Instead we present a computation-free proof.

Recall that X counts the number of red balls in a sample of size n from an urn containing r red balls and t balls in total. Label the red balls from 1 to r and the remaining balls from $r+1$ to t. Let $X_i = 1$

if the i-th ball is in the sample and $X_i = 0$ otherwise. Then the number of red balls $X = X_1 + \cdots + X_r$. By linearity of expectation,

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_r]. \tag{2}$$

But $\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \binom{t-1}{n-1}/\binom{t}{n} = n/t$ as we draw n balls without replacement and are interested in the event we draw a specific ball. Putting this value into (2) gives $\mathbb{E}[X] = n \cdot \left(\frac{r}{t}\right)$.

Example 5.25 (Birthday paradox)**.** Place 50 people in a room and assume that their birthdays are uniformly distributed over the whole year, independently of each other. Let X denote the number of days in the year on which someone in the group has their birthday. What is $\mathbb{E}[X]$?

The distribution of X is intricate, and we will not even think of studying it. Instead, for $i = 1, \ldots, 365$, let A_i be the event that somebody celebrates on day i. Then, $\mathbb{P}\left(A_i^c\right) = (364/365)^{50}$. Hence, $\mathbb{P}\left(A_i\right) = 1 - (364/365)^{50}$. As we can write $X = X_1 + X_2 + \cdots + X_{365}$ with Bernoulli random variables $X_1, \ldots, X_{365}$, where $X_i = 1$ if and only if A_i occurs, linearity of expectation gives

$$\mathbb{E}[X] = 365 \cdot \mathbb{E}[X_1] = 365 \cdot \mathbb{P}(A_1) = 365 \cdot \left(1 - (364/365)^{50}\right) = 46.786 \quad [3\text{dp}].$$

Remark 5.26. The fact that linearity of expectation works *without* independence is a huge strength. Note that in the previous two examples, the random variables $\{X_i\}$ were not independent.

Example 5.27 (Dice)**.** We roll a fair die repeatedly and let X be the number of rolls necessary for all faces to show up at least once. What is $\mathbb{E}[X]$?

To calculate this, for each $i \in \{1, \ldots, 6\}$ let T_i denote the roll when the i-th <u>new</u> face first appears [5]. In particular, we have $T_1 = 1$ and $T_6 = X$. Then we can write

$$X = T_1 + (T_2 - T_1) + (T_3 - T_2) + (T_4 - T_3) + (T_5 - T_4) + (T_6 - T_5).$$

By linearity of expectation, our calculation now reduces to calculating $\mathbb{E}[T_{i+1} - T_i]$ for $i \in \{1, \ldots, 5\}$.

What is the distribution of $T_{i+1} - T_i$ if $i \in \{1, \ldots, 5\}$? At time T_i, we have seen i faces. The number of rolls until the next new face appears follows the geometric distribution with parameter $p_i = (6-i)/6$. Hence, $T_{i+1} - T_i \sim \text{geo}_{p_i}$ and $\mathbb{E}\left[T_{i+1} - T_i\right] = 1/p_i = 6/(6-i)$. By linearity of expectation, we find

$$\mathbb{E}[X] = 1 + \sum_{i=1}^{5} \mathbb{E}[T_{i+1} - T_i] = 1 + \sum_{i=1}^{5} \frac{6}{6-i} = 1 + 6 \sum_{i=1}^{5} \frac{1}{i} = 14.7$$

This calculation is a variant of the *coupon collector* problem, which tends to appear quite often [6].

The following result is also very useful in calculating expectations.

Lemma 5.28. Let $g : \mathbb{R} \to \mathbb{R}$ be a function.

[5] Note: T_i is <u>not</u> the first time that face i appears.

[6] You might like to see the Wikipedia article for more information here.

(i) If X is a discrete random variable, we have

$$\mathbb{E}[g(X)] = \sum_{x \in S_X} g(x) \cdot \mathbb{P}(X = x),$$

provided $\sum_{x \in S_X} |g(x)| \cdot \mathbb{P}(X = x)$ converges.

(ii) If X is a continuous random variable with density f_X then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx,$$

provided $\int_{-\infty}^{\infty} |g(x)| \cdot f_X(x) dx$ converges.

Again we postpone the proof, to first see how it applies.

Example 5.29. Lemma 5.28 might look a little odd, but it is very convenient. To see why, note that if X is a discrete random variable then by Definition 5.1 we have

$$\mathbb{E}[X^2] = \sum_{x \in S_{X^2}} x \cdot \mathbb{P}(X^2 = x).$$

This might require determining S_{X^2} and probabilities $\mathbb{P}(X^2 = x)$. Lemma 5.28 shows this is not necessary; taking $g(x) = x^2$ we also have $\mathbb{E}[X^2] = \sum_{x \in S_X} x^2 \cdot \mathbb{P}(X = x)$.

Example 5.30 (Karate)**.** We chop a stick of length 1 into two pieces at a uniformly chosen point. What is the expected length of the longer of the two pieces?

Letting $X \sim \text{unif}[0, 1]$ be the breaking point, we want $\mathbb{E}[g(X)]$ where $g(x) = \max(x, 1 - x)$. Thus,

$$\mathbb{E}[Y] = \int_0^1 \max(x, 1 - x) \cdot 1 dx = \int_0^{0.5} (1 - x) dx + \int_{0.5}^1 x dx = \left[\frac{-(1 - x)^2}{2}\right]_0^{0.5} + \left[\frac{x^2}{2}\right]_{0.5}^1 = \frac{3}{4}.$$

5.2.1 The proofs of Theorem 5.15 and Lemma 5.28.

Proof of Theorem 5.15. We only consider the case of discrete random variables in the proof. To simplify the notation, we write $S_X = \{x_1, x_2, \ldots\}$ and $S_Y = \{y_1, y_2, \ldots\}$, noting that these sets could be finite or infinite.

To prove (i) let $Z = aX + bY$ with $S_Z = \big\{ax_i + by_k : i \geq 1, k \geq 1\big\}$. We have

$$\begin{aligned} \mathbb{E}[Z] = \sum_{z \in S_Z} z \cdot \mathbb{P}\big(Z = z\big) &= \sum_{z \in S_Z} z \cdot \mathbb{P}\Bigg(\bigcup_{i \geq 1, k \geq 1 : ax_i + by_k = z} \big\{X = x_i, Y = y_k\big\} \Bigg) \\ &= \sum_{z \in S_Z} z \cdot \Bigg(\sum_{i \geq 1, k \geq 1 : ax_i + by_k = z} \mathbb{P}\big(X = x_i, Y = y_k\big) \Bigg) \\ &= \sum_{i \geq 1} \sum_{k \geq 1} \big(ax_i + by_k\big) \cdot \mathbb{P}\big(X = x_i, Y = y_k\big). \end{aligned}$$

The series for $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are absolutely convergent, so we can rearrange this sum by 1SAS results:

$$\begin{aligned}\mathbb{E}[Z] &= a \cdot \sum_{i\geq 1} x_i \cdot \Big(\sum_{k\geq 1} \mathbb{P}\big(X = x_i, Y = y_k\big)\Big) + b \cdot \sum_{k\geq 1} y_k \cdot \Big(\sum_{i\geq 1} \mathbb{P}\big(X = x_i, Y = y_k\big)\Big) \\ &= a \cdot \sum_{i\geq 1} x_i \cdot \mathbb{P}(X = x_i) + b \cdot \sum_{k\geq 1} y_k \cdot \mathbb{P}(Y = y_k) \\ &= a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y].\end{aligned}$$

To prove (ii), note that by replacing $aX + bY$ with $X \cdot Y$ in (i), the analogue of the calculation for $\mathbb{E}[Z]$ gives

$$\begin{aligned}\mathbb{E}[X \cdot Y] = \sum_{i\geq 1}\sum_{k\geq 1} x_i y_k \cdot \mathbb{P}(X = x_i, Y = y_k) &= \sum_{i\geq 1}\sum_{k\geq 1} x_i y_k \cdot \Big(\mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_k)\Big) \\ &= \Big(\sum_{i\geq 1} x_i \cdot \mathbb{P}(X = x_i)\Big) \cdot \Big(\sum_{k\geq 1} y_k \cdot \mathbb{P}(Y = y_k)\Big) \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y].\end{aligned}$$

The second equality above is the key point at which independence is used (recalling Definition 3.33). The third equality is clear if the series are finite, but holds for infinite series by results from 1SAS, as both series are absolutely convergent. The final equality holds by definition of $\mathbb{E}[X]$ and $\mathbb{E}[Y]$. □

Proof of Lemma 5.28. We only prove *(i)*. By definition of $\mathbb{E}[g(X)]$, we have

$$\begin{aligned}\mathbb{E}[g(X)] = \sum_{y\in S_{g(X)}} y \cdot \mathbb{P}\big(g(X) = y\big) &= \sum_{y\in S_{g(X)}} y \cdot \mathbb{P}\big(X \in \{x \in S_X : g(x) = y\}\big) \\ &= \sum_{y\in S_{g(X)}} y \cdot \Bigg(\sum_{x\in S_X : g(x)=y} \mathbb{P}(X = x)\Bigg) \\ &= \sum_{x\in S_X} g(x) \cdot \Bigg(\sum_{y\in S_{g(X)} : g(x)=y} \mathbb{P}(X = x)\Bigg) \\ &= \sum_{x\in S_X} g(x) \cdot \mathbb{P}(X = x).\end{aligned}$$

The fourth equality holds by switching the order of summations (i.e. the order of x and y) and the final equality holds by noting that $\big|\{y \in S_{g(X)} : g(x) = y\}\big| = |\{g(x)\}| = 1$. □

5.3 Variance

Definition 5.31 (Variance - discrete case)**.** Let X be a discrete random variable with well-defined expectation. The *variance* of X is given by

$$\mathbb{V}\mathrm{ar}(X) := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \sum_{x\in S_X} (x - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = x), \tag{3}$$

provided the series in (3) converges. The *standard deviation* of X is then given by

$$\sigma_X := \sqrt{\mathbb{V}\text{ar}(X)}.$$

If the series in (3) diverges then both $\mathbb{V}\text{ar}(X)$ and σ_X are undefined.

Remark 5.32. Let X be a discrete random variable with well-defined expectation and variance.

- The variance of X is *always non-negative*, by definition in (3).
- Both the variance and the standard deviation give a measure of the typical distance of the random variable X from $\mathbb{E}[X]$. These quantities have different benefits:
  (i) typically $|X - \mathbb{E}[X]|$ is not much larger than σ_X (proven in Section 6), while
  (ii) variance behaves much better *algebraically* (see Proposition 5.39 and Theorem 5.50 below).

Example 5.33. We return to the game from Example 5.4. As X satisfies $\mathbb{P}(X = 1.25) = 0.4$, $\mathbb{P}(X = -1) = 0.6$ and $\mathbb{E}[X] = -0.1$, we deduce

$$\mathbb{V}\text{ar}(X) = (1.25 - (-0.1))^2 \cdot 0.4 + (-1 - (-0.1))^2 \cdot 0.6 = 1.215.$$

Example 5.34 (Constant random variable)**.** We return to Example 5.5 with a random variable X satisfying $\mathbb{P}(X = c) = 1$ for some $c \in \mathbb{R}$. Then, $\mathbb{E}[X] = c$ and $\mathbb{V}\text{ar}(X) = (c - c)^2 \cdot \mathbb{P}(X = x) = 0$. This makes sense as there is no variation in the values which X can attain.

Example 5.35. Let X follow the uniform distribution on $\{49, 50, 51\}$ and Y follow the uniform distribution on $\{1, \ldots, 99\}$. Then $\mathbb{E}[X] = \mathbb{E}[Y] = 50$. However the random variable X attains values very close to $\mathbb{E}[X]$ and has a small standard deviation; $\sigma_X = \sqrt{2/3}$. The random variable Y is evenly spread on $\{1, \ldots, 99\}$, and σ_Y is much larger; $\sigma_Y = 28.57\ldots$ as shown in Example 5.40.

Definition 5.36 (Variance - continuous case)**.** Let X be a random variable with density f_X, and well-defined expectation. Then the *variance* of X is given by

$$\mathbb{V}\text{ar}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 \cdot f_X(x) dx,$$

provided the integral exist and the corresponding *standard deviation* by $\sigma_X = \sqrt{\mathbb{V}\text{ar}(X)}$.

Remark 5.37. More generally, for $g : \mathbb{R} \to \mathbb{R}$, if $g(X)$ has a well-defined expectation, then

$$\mathbb{V}\text{ar}(g(X)) = \int_{-\infty}^{\infty} (g(x) - \mathbb{E}[g(X)])^2 \cdot f_X(x) dx, \quad \sigma_{g(X)} = \sqrt{\mathbb{V}\text{ar}(g(X))}.$$

Example 5.38. Coming back to Example 5.30, recall that we had $X \sim \text{unif}[0, 1]$ and were interested in $g(X)$ where $g(x) = \max(x, 1 - x)$. Our calculation gave $\mathbb{E}[g(X)] = 3/4$.

$$\begin{aligned}\mathbb{V}\text{ar}(g(X)) = \int_0^1 \big(\max(x, 1-x) - 3/4\big)^2 dx &= \int_0^{1/2} \big(1 - x - 3/4\big)^2 dx + \int_{1/2}^1 \big(x - 3/4\big)^2 dx \\ &= \Big[-\frac{(1/4 - x)^3}{3}\Big]_0^{1/2} + \Big[\frac{(x - 3/4)^3}{3}\Big]_{1/2}^1 = \frac{4\big(\frac{1}{4}\big)^3}{3} = \frac{1}{48}.\end{aligned}$$

5.4 Key properties of variance

Proposition 5.39. Let X be a discrete or continuous random variable with well-defined expectation and variance. Then the following hold:

(i) $\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$,

(ii) given $a, b \in \mathbb{R}$ we have $\mathbb{V}\text{ar}(aX + b) = a^2 \cdot \mathbb{V}\text{ar}(X)$,

(iii) $\mathbb{V}\text{ar}(X) = 0$ if and only if $\mathbb{P}(X = x_0) = 1$ for some $x_0 \in \mathbb{R}$.

Proof. We will prove (i)–(iii) assuming X is a discrete random variable. For (i) note that

$$\begin{aligned}\mathbb{V}\text{ar}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] &= \mathbb{E}\big[X^2 - 2 \cdot X \cdot \mathbb{E}[X] + \mathbb{E}[X]^2\big] \\ &= \mathbb{E}\big[X^2\big] - 2\mathbb{E}\big[X\big] \cdot \mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2.\end{aligned}$$

The second equality holds by expanding $(X - \mathbb{E}[X])^2$ and third holds by Theorem 5.15 *(i)*.

To see (ii) note that by Theorem 5.15 (i) we have $\mathbb{E}[aX + b] = a \cdot \mathbb{E}[X] + b$. It follows that

$$\mathbb{V}\text{ar}(aX + b) = \mathbb{E}\Big[\big(aX + b - \mathbb{E}[aX + b]\big)^2\Big] = \mathbb{E}\Big[\big(aX - a \cdot \mathbb{E}[X]\big)^2\Big] = a^2 \mathbb{V}\text{ar}(X).$$

To prove (iii), note that if $\mathbb{P}(X = x_0) = 1$ then given $y \in \mathbb{R} \setminus \{x_0\}$ we have $\mathbb{P}(X = y) \le \mathbb{P}(X \neq x_0) = 0$. It follows that $\mathbb{E}[X] = x_0 \cdot 1 + 0 = x_0$ and $\mathbb{V}\text{ar}(X) = (x_0 - \mathbb{E}[X])^2 \cdot 1 + 0 = 0$.

On the other hand, suppose $\mathbb{V}\text{ar}(X) = 0$. Then given any $y \in S_X$ we have

$$(y - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = y) \le \sum_{x \in S_X} \big(x - \mathbb{E}[X]\big)^2 \cdot \mathbb{P}(X = x) = \mathbb{V}\text{ar}(X) = 0.$$

Thus if $y \neq \mathbb{E}[X]$ then $\mathbb{P}(X = y) = 0$. It follows that $\mathbb{P}(X = \mathbb{E}[X]) + 0 = \sum_{x \in S_X} \mathbb{P}(X = x) = \mathbb{P}(\Omega) = 1$, and so $\mathbb{P}(X = \mathbb{E}[X]) = 1$. $\square$

Example 5.40 (Discrete uniform)**.** Let X follow the uniform distribution on $\{1, \ldots, n\}$. Then by Example 5.6 we have $\mathbb{E}[X] = (n + 1)/2$. As

$$\mathbb{E}[X^2] = \sum_{i=1}^{n} \frac{i^2}{n} = \frac{n(n + 1)(2n + 1)}{6n} = \frac{(n + 1)(2n + 1)}{6},$$

by Proposition 5.39 (i) we obtain that $\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{12}(n^2 - 1)$.

Example 5.41 (Bernoulli)**.** If $X \sim \text{Ber}_p$ then $\mathbb{E}[X] = p$ by Example 5.21 and we find that $\mathbb{V}\text{ar}(X) = (1 - p)^2 \cdot p + p^2 \cdot (1 - p) = p(1 - p)$.

Example 5.42 (Binomial)**.** $\mathbb{V}\text{ar}(X) = np(1 - p)$ for $X \sim \text{bin}_{n,p}$ (see problem sheet 4).

Example 5.43 (Hypergeometric)**.** If $X \sim \text{hyp}_{n,r,t}$ then $\mathbb{V}\text{ar}(X) = n\big(\frac{r}{t}\big)\big(\frac{t-r}{t}\big)\big(\frac{t-n}{t-1}\big)$. We will not prove this, but if you are interested, see Example 8.30 in *Introduction to Probability* by Anderson, Seppäläinen and Valkó.

Example 5.44 (Geometric)**.** Let $X \sim \text{geo}_p$ where $p \in (0,1)$. Letting $q := 1-p$ we have

$$\mathbb{E}[X^2] = \sum_{k=1}^{\infty} k^2(1-q)q^{k-1} = \sum_{k=0}^{\infty}(k+1)^2 q^k - \sum_{k=1}^{\infty} k^2 q^k = 2\sum_{k=1}^{\infty} kq^k + \sum_{k=0}^{\infty} q^k = 2 \cdot \Big(\frac{q}{p}\Big) \cdot \mathbb{E}[X] + \frac{1}{p}.$$

Since $\mathbb{E}[X] = 1/p$ by Example 5.8, we find $\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}$.

Example 5.45 (Poisson)**.** Let $X \sim \text{Poi}_\lambda$ with $\lambda > 0$. In Example 5.9 we saw that $\mathbb{E}[X] = \lambda$. A direct computation of $\mathbb{E}[X^2]$ is cumbersome, and it is better to consider $\mathbb{E}[X(X-1)]$:

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} \overset{\ell=k-2}{=} \lambda^2 e^{-\lambda}\sum_{\ell=0}^{\infty}\frac{\lambda^\ell}{\ell!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2.$$

We obtain $\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Example 5.46. (Exercise!) Taking $Y = g(X)$ as in Example 5.30, you can check that $\mathbb{V}\text{ar}(Y) = 1/48$. Similarly for the random variable X from Example 5.14, we have $\mathbb{V}\text{ar}(X) = \pi^2/2 - 2$.

Example 5.47 (Continuous uniform distribution)**.** Let $X \sim \text{unif}[a,b]$. In Example 5.12 we had already shown that $\mathbb{E}[X] = (a+b)/2$. To find the variance, compute

$$\mathbb{E}[X^2] = (b-a)^{-1}\int_a^b x^2 dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}.$$

Thus, $\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{(b-a)^2}{12}$.

Example 5.48 (Exponential)**.** Let $X \sim \exp_\lambda$ where $\lambda > 0$. Then

$$\mathbb{E}[X^2] = \int_0^{\infty} \lambda x^2 e^{-\lambda x}dx = \Big[-2xe^{-\lambda x}\Big]_0^{\infty} + 2\int_0^{\infty} xe^{-\lambda x}dx = 0 + \frac{2\mathbb{E}[X]}{\lambda} = \frac{2}{\lambda^2},$$

using Example 5.13. This gives $\mathbb{V}\text{ar}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{\lambda^2}$.

Example 5.49 (Normal)**.** Again let $\mathcal{N} \sim N(0,1)$ with density $f(x) = (2\pi)^{-1/2}e^{-x^2/2}, x \in \mathbb{R}$, recalling that $\mathbb{E}[\mathcal{N}] = 0$ from 5.19. To compute $\mathbb{V}\text{ar}(\mathcal{N}) = \mathbb{E}[\mathcal{N}^2] - \big(\mathbb{E}[\mathcal{N}]\big)^2 = \mathbb{E}\big[\mathcal{N}^2\big]$, we perform integration by parts with $u = x, v' = xe^{-x^2/2}$:

$$\mathbb{V}\text{ar}(\mathcal{N}) = \mathbb{E}[\mathcal{N}^2] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x \cdot xe^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\left(\Big[-xe^{-x^2/2}\Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx\right) = \int_{-\infty}^{\infty} f(x)dx = 1.$$

As $\mu + \sigma\mathcal{N} \sim N(\mu,\sigma^2)$, it follows from Proposition 5.39 *(ii)* that $\mathbb{V}\text{ar}(X) = \sigma^2$ for $X \sim N(\mu,\sigma^2)$.

The following theorem shows that variance is additive for independent random variables.

Theorem 5.50. *Let $X_1, \dots, X_n$ be discrete or continuous random variables with well-defined expectations and variances. If $X_1, \dots, X_n$ are independent, then*

$$\mathbb{V}ar(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathbb{V}ar(X_i).$$

Proof. We again prove this only for discrete random variables. It is convenient to set $\mu_i = \mathbb{E}[X_i]$ for each $i \in [n]$. By linearity of expectation we have $\mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mu_i$. Secondly, given $1 \le i \ne j \le n$ the random variables X_i and X_j are independent, so the random variables $X_i - \mu_i$ and $X_j - \mu_j$ are also independent by Proposition 3.38. Thus

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\big(\sum_{i=1}^n X_i\big) = \mathbb{E}\Big[\Big(\sum_{i=1}^n X_i - \mathbb{E}[\sum_{i=1}^n X_i]\Big)^2\Big] &= \mathbb{E}\Big[\Big(\sum_{i=1}^n (X_i - \mu_i)\Big)^2\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^n (X_i - \mu_i) \cdot \sum_{j=1}^n (X_j - \mu_j)\Big] \\
&= \sum_{i=1}^n \mathbb{E}\Big[\Big(X_i - \mu_i\Big)^2\Big] + \sum_{i=1}^n \sum_{j=1:j\ne i}^n \mathbb{E}\big[(X_i - \mu_i) \cdot (X_j - \mu_j)\big] \\
&= \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i) + \sum_{i=1}^n \sum_{j=1:j\ne i}^n \mathbb{E}\big[(X_i - \mu_i)\big] \cdot \mathbb{E}\big[(X_i - \mu_i)\big] \\
&= \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X).
\end{aligned}$$

The second last equality here uses that $\mathbb{V}\mathrm{ar}(X_i) = \mathbb{E}\big[\big(X_i - \mu_i\big)^2\big]$ for all $i = 1, \dots, n$ and that by Theorem 5.15 *(ii)*, as $X_i - \mu_i$ and $X_j - \mu_i$ are independent for $i \ne j$. The final inequality holds as $\mathbb{E}\big[X_i - \mu_i\big] = 0$ for all $i = 1, \dots, n$. ☐

Example 5.51. Let X, Y be independent random variables with $\mathbb{V}\mathrm{ar}(X) = 2$ and $\mathbb{V}\mathrm{ar}(Y) = 5$. What is $\mathbb{V}\mathrm{ar}(3X - 5Y)$? By Proposition 3.38 the random variables $3X, -5Y$ are independent and so

$$\mathbb{V}\mathrm{ar}(3X - 5Y) = \mathbb{V}\mathrm{ar}(3X) + \mathbb{V}\mathrm{ar}(-5Y) = 3^2\mathbb{V}\mathrm{ar}(X) + (-5)^2\mathbb{V}\mathrm{ar}(Y) = 18 + 125 = 143.$$

Remark 5.52. Let X, Y be random variables with well-defined variances. In general, we do **not** have $\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$. For example, let X be an arbitrary random variable with well-defined variance $\mathbb{V}\mathrm{ar}(X) > 0$ and set $Y = -X$. Then $0 = \mathbb{V}\mathrm{ar}(X + Y) \ne \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) = 2\mathbb{V}\mathrm{ar}(X)$ where we used Proposition 5.39(ii) with $a = -1, b = 0$ in the last step.

Example 5.53 (Binomial)**.** Let $X \sim \mathrm{bin}_{n,p}$. As in Example 5.23 we have $X = \sum_{i=1}^n X_i$ where with $X_i \sim \mathrm{Ber}_p$ for all $i = 1, \dots, n$. By additivity of variances for independent random variables, since the random variables $X_1, \dots, X_n$ are independent (the events $A_1, \dots, A_n$ are independent), we obtain

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{V}\mathrm{ar}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i) = np(1-p),$$

where we used the expression for the variance of a Bernoulli random variable from Example 5.41.

For fixed $p \in [0, 1]$, the value $\mathbb{E}[X]$ in Example 5.53 is proportional to n as $n \to \infty$, while the standard deviation σ_X grows like $\sqrt{n}$. In particular, we find that the fluctuations of X are much smaller than $\mathbb{E}[X]$ for large n.

5.5 Median

Another quantity which plays an important role in statistics is the median.

Definition 5.54 (Median)**.** Let X be a discrete or continuous random variable. A value $m \in \mathbb{R}$ is called a *median* of X if $\mathbb{P}(X \geq m) \geq 1/2$ and $\mathbb{P}(X \leq m) \geq 1/2$.

Remark 5.55.

- In general, the median of a random variable is not necessarily unique.
- Like the expectation, the median gives an estimate of the 'typical value' of X. Statisticians tend to have a soft spot for the median and often choose it over the expectation to represent data. One reason is that it is more stable and less sensitive to extreme values than the expectation.
- On the downside, the median doesn't behave very well *algebraically* and the median analogues of the useful properties in Theorem 5.15 all tend to fail. In particular, the median is rarely linear.

Example 5.56 (Continuous uniform distribution)**.** Let $X \sim \mathrm{unif}[a, b]$ with distribution function

$$F_X(t) = \begin{cases} 0 & \text{if } t < a, \\ \frac{t-a}{b-a} & \text{if } a \leq t \leq b, \\ 1 & \text{if } t > b. \end{cases}$$

Solving $F_X(t) = 1/2$ shows that $m = (a + b)/2 = \mathbb{E}[X]$ is the unique median of X.

Example 5.57 (Discrete uniform distribution)**.** Let X have the uniform distribution on $\{1, \ldots, n\}$. Given $x \in \mathbb{R}$ with $1 \leq x \leq n$, we have $\mathbb{P}(X \leq x) = \lfloor x \rfloor / n$ and $\mathbb{P}(X \geq x) = (\lfloor n - x \rfloor + 1)/n$. If n is odd, the $\mathbb{E}[X] = (n+1)/2$ is the unique median of X. If n is even, however, then *every number* in the interval $[n/2, n/2 + 1]$ is a median of X.

Example 5.58. We roll a fair dice and as usual let $\Omega = \{1, \ldots, 6\}$ and $\mathbb{P}$ denote the uniform distribution on Ω. We also let $X : \Omega \to \mathbb{R}$ denote the outcome of the roll, so that $X(i) = i$. As seen above, any number in $[3, 4]$ is a median of X.

Now instead assign a new value K to the face 6, where K is some very large number (maybe $K = 1000$). Let Y be the outcome of rolling the modified fair dice. Then, for $K \geq 4$, any median of Y still lies in the interval $[3, 4]$. However, we find $\mathbb{E}[Y] = (K + 15)/6$ and so the mean changes drastically [7].

[7] One can show that, for any median m, we have $|\mathbb{E}[X] - m| \leq \sigma_X$, so median and expectation are close to each other if σ_X is small. (A proof of this inequality is difficult.)

5.6 Statistical applications

Hypothesis tests. Randomised double-blind clinical trials work as follows: a group of patients is randomly subdivided into two subgroups T (for treatment) and C (for control). Patients in group T are given a certain treatment while group C patients receive placebos. To minimise any kind of bias neither participants nor doctors know the partition into groups T and C.

In 1948 the first randomised clinical trial was held in the UK leading to a breakthrough in tuberculosis treatment using the antibiotic streptomycin. Six months of treatment saw the following results:

	improvement	no improvement	total
treatment (T)	39	16	55
no treatment (C)	17	35	52
total	56	51	107

The numbers look convincing, but one might still ask if such an extreme result could have occurred by coincidence. Assuming that the conditions of 56 of 107 patients were determined to improve (independently of any treatment), the number X of those patients in group T follows the hypergeometric distribution with $n = 55, r = 56$ and $t = 107$. Thus, $\mathbb{E}[X] = 55 \cdot 56/107 = 28.78\ldots$ The probability for a deviation of at least $39 - 28.78\ldots$ from the mean is

$$\mathbb{P}(X \leq 18) + \mathbb{P}(X \geq 39) = \sum_{k=4}^{18} \frac{\binom{56}{k} \cdot \binom{51}{55-k}}{\binom{107}{55}} + \sum_{k=39}^{56} \frac{\binom{56}{k} \cdot \binom{51}{55-k}}{\binom{107}{55}} \approx 0.001.$$

This probability is called the *p-value* associated with the data. In hypothesis testing, one fixes a *significance level* α (often $0.01, 0.05$ or 0.10) and rejects the *null hypothesis* (here: treatment has no effect) if the p-value is smaller than α. [8]

Mean estimation. A central problem in statistics is the following: given independent realisations $\hat{X}_1, \ldots, \hat{X}_n$ (that is, data) from an unknown distribution which has mean μ and variance $\sigma^2 \geq 0$, we would like to estimate μ. The obvious choice is the *sample mean*

$$\hat{\mu}_n = n^{-1}(\hat{X}_1 + \cdots + \hat{X}_n).$$

By linearity of expectation, we find $\mathbb{E}[\hat{\mu}_n] = \mu$. Further, by Proposition 5.39 (ii) and Theorem 5.50 we have $\mathbb{V}\mathrm{ar}(\hat{\mu}_n) = \sigma^2/n$. As $n \to \infty$, the fluctuations of $\hat{\mu}_n$ around its expectation μ become smaller giving that $\hat{\mu}_n$ is extremely likely [9] to be a good estimator for the true mean μ.

Variance estimation. In the setting of the previous example, we sometimes also want to estimate σ^2. The obvious choice would be $n^{-1}\big((\hat{X}_1 - \mu)^2 + \cdots + (\hat{X}_n - \mu)^2\big)$. Indeed, this quantity is a good

[8] This line of argument is known as *Fisher's exact test* after Sir Ronald Fisher (1890 - 1962), the founder of modern statistical science. You will study hypothesis tests in much greater detail in the Year 2 module 2S.

[9] The law of large numbers, discussed in Section 6, will make this statement more precise.

approximation for σ^2; however, in applications, it is useless since we do not know the true value of μ. Instead, we work with the *sample variance*

$$\hat{\sigma}_n^2 := n^{-1} \sum_{i=1}^{n} (\hat{X}_i - \hat{\mu}_n)^2.$$

By rearranging this expression (a one line calculation) we obtain that $\hat{\sigma}_n^2 = n^{-1} \big(\sum_{i=1}^n \hat{X}_i^2 \big) - \hat{\mu}_n^2$. Since $\mathbb{E}\big[\hat{\mu}_n^2\big] = \mathbb{V}\mathrm{ar}(\hat{\mu}_n) + (\mathbb{E}[\hat{\mu}_n])^2 = \sigma^2/n + \mu^2$ and $\mathbb{E}\big[\hat{X}_i^2\big] = \sigma^2 + \mu^2$, it follows that

$$\mathbb{E}[\hat{\sigma}_n^2] = \Big(\frac{n-1}{n} \Big) \cdot \sigma^2. \tag{4}$$

Similar to estimation of the mean, to show that $\hat{\sigma}_n^2$ is a good estimator for σ^2, we would like to prove that its variance tends to zero as $n \to \infty$. A lengthy calculation (a page or two in length) gives that

$$\mathbb{V}\mathrm{ar}\big(\hat{\sigma}_n^2\big) = \frac{\mathbb{E}\big[\hat{X}_i^4\big] - \sigma^4}{n} + \frac{\mathbb{E}\big[\hat{X}_i^4\big] + \sigma^4}{n^2} + \frac{\mathbb{E}\big[\hat{X}_i^4\big] - \sigma^4}{n^3}.$$

Provided $\mathbb{E}\big[\hat{X}_i^4\big] < \infty$ we obtain $\mathbb{V}\mathrm{ar}\big(\hat{\sigma}_n^2\big) \to 0$ as $n \to \infty$, and $\hat{\sigma}_n^2$ is very likely to be close to $\mathbb{E}\big[\hat{\sigma}_n^2\big]$.

Remark 5.59. In statistics, we often seek [10] an estimator $\hat{\theta}$ for a parameter θ. The estimator is said to be *unbiased* if θ is the average value (or expectation) of $\hat{\theta}$. From above $\mathbb{E}\big(\hat{\mu}_n\big) = \mu$ and so $\hat{\mu}_n$ is an unbiased estimator for μ. On the other hand, from (4) we see that $\hat{\sigma}_n^2$ is a *biased* estimator. This can be corrected, by instead taking the unbiased estimator

$$(n-1)^{-1} \sum_{i=1}^{n} (\hat{X}_i - \hat{\mu}_n)^2.$$

Most important takeaways in this chapter. You should

- know definitions of expectation, variance, standard deviation and median for discrete and continuous random variables,
- be familiar with the main properties of expectation and variance,
- be able to compute expectation, variance and median in simple examples,
- know expectation for binomial, hypergeometric, geometric, Poisson, discrete and continuous uniform, exponential and normal distribution,
- know variance of binomial, Poisson, geometric and normal distribution,
- be familiar with the Bernoulli distribution and know its expectation and variance,
- be able to exploit linearity of expectation to compute the expectation of complicated random variables,
- appreciate the concept of hypothesis tests.

[10] E.g. selecting a group of voters to estimate the proportion of an electorate in favour of a referendum result.

Tables of formula and comparisons

The following table extends the previous table from Section 4 (page 3), to include expressions for the expectation and variance.

	discrete r.v.	continuous r.v.
mass/density function	$p_X(k), k \in S_X$	$f_X(x), x \in \mathbb{R}$
distribution function F_X	$\sum_{k \in S_X, k \leq t} p_X(k)$ F_X is a step function	$\int_{-\infty}^{t} f_X(x)dx$ F_X is continuous
connection	$\mathbb{P}(X = k) = F_X(k) - \mathbb{P}(X < k)$	$f_X(x) = F'_X(x)$
expectation $\mathbb{E}[X]$	$\sum_{k \in S_X} k \cdot p_X(k)$	$\int_{-\infty}^{\infty} x \cdot f_X(x)dx$
expectation $\mathbb{E}[g(X)]$	$\sum_{k \in S_X} g(k) \cdot p_X(k)$	$\int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$
variance $\mathbb{V}\mathrm{ar}(X)$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2$

The second table below gives formulae for the expectation and variance of common random variables, and a reference to the example above in which the formula was derived [11].

distribution	parameters	expectation	Ex. no.	variance	Ex. no.
(discrete) uniform	n	$\frac{n+1}{2}$	5.6	$\frac{n^2-1}{12}$	5.40
Bernoulli	p	p	5.21	$p(1-p)$	5.41
binomial	n, p	np	5.23	$np(1-p)$	5.53
hypergeometric	n, r, t	$n\left(\frac{r}{t}\right)$	5.24	$n\left(\frac{r}{t}\right)\left(\frac{t-r}{t}\right)\left(\frac{t-n}{t-1}\right)$	5.43
geometric	p	$\frac{1}{p}$	5.8	$\frac{1-p}{p^2}$	5.44
Poisson	λ	λ	5.9	λ	5.45
(continuous) uniform	$a < b$	$\frac{a+b}{2}$	5.12	$\frac{(b-a)^2}{12}$	5.47
exponential	λ	$\frac{1}{\lambda}$	5.13	$\frac{1}{\lambda^2}$	5.48
Normal	μ, σ^2	μ	5.19	σ^2	5.49

[11] In the electronic version of these notes, the example number references all have hyperlinks, which might be convenient.