As we have seen in the previous two lectures, random variables can have complicated probability distributions, corresponding to a complex behaviour.

In the present lecture, we focus on the two most fundamental quantities to approximate the behavior of a random variable: the **mean** (also called expectation or expected value) and the **variance**. These two quantities are in general not sufficient to fully determine the probability distribution function of a random variable, but serve as a simple and valuable summary.

# 1 Expected value

## 1.1 Discrete random variable

### 1.1.1 Introductory example

Imagine you are in a casino, and planning on playing $N$ rounds of a game such that

- The probability of losing 20$ is 0.2.
- The probability of losing 10$ is 0.3.
- The probability of neither losing money nor winning money is 0.1.
- The probability of winning 10$ is 0.2.
- The probability of winning 20$ is 0.2.

You are wondering what the expected gain/loss per game is, in dollars.

An intuitive reasoning for calculating this quantity is as follows. Let $N_{-20}$ be the number of games in which you would lose 20$, $N_{-10}$ be the number of games in which you would lose 10$, $N_0$ be the number of games in which you neither win nor lose money, $N_{10}$ be the number of games in which you would win 10$, and $N_{20}$ be the number of games in which you would win 20$.

By construction, you have

$$N_{-20} + N_{-10} + N_0 + N_{10} + N_{20} = N$$

The amount of money won or lost after these $N$ games would be:

$$S_N = (-20) \cdot N_{-20} + (-10) \cdot N_{-10} + 0 \cdot N_0 + 10 \cdot N_{10} + 20 \cdot N_{20}$$

Hence, the *average* amount of money won or lost *per game* would be:

$$s_N = \frac{S_N}{N} = (-20) \cdot \frac{N_{-20}}{N} + (-10) \cdot \frac{N_{-10}}{N} + 0 \cdot \frac{N_0}{N} + 10 \cdot \frac{N_{10}}{N} + 20 \cdot \frac{N_{20}}{N}$$

Now, in the limit $N \to +\infty$, we expect

$$\frac{N_{-20}}{N} \xrightarrow[N\to+\infty]{} 0.2 \ , \quad \frac{N_{-10}}{N} \xrightarrow[N\to+\infty]{} 0.3 \ , \quad \frac{N_0}{N} \xrightarrow[N\to+\infty]{} 0.1 \ , \quad \frac{N_{10}}{N} \xrightarrow[N\to+\infty]{} 0.2 \ , \quad \frac{N_{20}}{N} \xrightarrow[N\to+\infty]{} 0.2$$

So the average amount of money won or lost per game in the limit $N \to \infty$ is:

$$s_\infty = \lim_{N\to+\infty} s_N = (-20) \cdot 0.2 + (-10) \cdot 0.3 + 0 \cdot 0.1 + 10 \cdot 0.2 + 20 \cdot 0.2 = -1\$$$

This intuitive example illustrates why we define the expected value as we will do in the following section.

### 1.1.2 Definition

The **expectation** of a discrete random variable $X$ taking values $a_1, a_2, \ldots$ with probability mass function $p_X$ is the number

$$\boxed{E[X] = \sum_i a_i P(X = a_i) = \sum_i a_i p_X(a_i)} \tag{1}$$

Note that $E[X]$ is also called the **mean** or **expected value**.

In our casino example, $X$ was the amount of money one can win or lose in a round, with possible values $\{-20, -10, 0, 10, 20\}$. We had

$$p_X(-20) = 0.2 \ , \ p_X(-10) = 0.3 \ , \ p_X(0) = 0.1 \ , \ p_X(10) = 0.2 \ , \ p_X(20) = 0.2$$

so

$$E[X] = (-20)p_X(-20) + (-10)p_X(-10) + 0p_X(0) + 10p_X(10) + 20p_X(20) = s_\infty$$

#### 1.1.3 Example

Consider the discrete random variable $X$ corresponding to the result of rolling a fair dice. $X$ takes on values in $\{1, 2, 3, 4, 5, 6\}$, and each result has probability $\frac{1}{6}$. Hence,

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

### 1.2 Continuous random variable

#### 1.2.1 Extending the definition from discrete random variables to continuous random variables

We have introduced the concept of expected value for discrete random variables, and would now like to extend the concept to continuous random variables. To do so, we take some continuous random variable $X$, and approximate it by a discrete random variable $Y$ for which we can compute $E[Y]$. This is done with the following steps:

- For simplicity, we take $X$ such that its probability density function $f$ is zero outside the interval $[0, 1]$.
- We take some integer $N$ large, and subdivide the interval $[0, 1]$ in $N$ equal subintervals of size $\frac{1}{N}$.
- We consider a discrete random variable $Y$ which takes on the values of the right end of each interval: $Y$ takes on values in
$$\left\{\frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\right\}$$
- We construct a probability mass function for $Y$ from the probability density function of $X$ as follows:
$$\forall k = 1, \ldots, N \ , \qquad p_Y\left(\frac{k}{N}\right) = P\left(Y = \frac{k}{N}\right) = P\left(\frac{k-1}{N} \leq X \leq \frac{k}{N}\right)$$
$Y$ is thus constructed in such a way that as $N$ increases, $Y$ approximates $X$ better and better. Now, by definition of the probability density function of $X$,
$$p_Y\left(\frac{k}{N}\right) = \int_{\frac{k-1}{N}}^{\frac{k}{N}} f(x)dx$$
In the limit of $N$ large, which is the limit we are interested in here, the intervals of integration become very small, and the approximation of the integrals by the area of a rectangle with width $\frac{1}{N}$ and height $f(\frac{k}{N})$ become excellent:
$$p_Y\left(\frac{k}{N}\right) = \int_{\frac{k-1}{N}}^{\frac{k}{N}} f(x)dx \approx \frac{1}{N} f\left(\frac{k}{N}\right) \qquad \text{(valid for N large)}$$
- Now that the probability mass function of $Y$ has been constructed, we can compute the expectation of $Y$:
$$E[Y] = \sum_{k=1}^{N} \frac{k}{N} p_Y\left(\frac{k}{N}\right) \approx \sum_{k=1}^{N} \frac{k}{N}\frac{1}{N} f\left(\frac{k}{N}\right) = \sum_{k=1}^{N} \frac{1}{N}\frac{k}{N} f\left(\frac{k}{N}\right)$$

In the right-hand side of this approximate equality, we recognize a Riemann sum approximation of the integral $\int_0^1 f(x)dx$, the approximation becoming better and better as $N \to \infty$, at the same time as $Y$ approximates $X$ better and better. This justifies the following definition for the expectation of a continuous random variable.

### 1.2.2 Definition

The **expectation** of a continuous random variable $X$ with probability density function $f$ is the number

$$\boxed{E[X] = \int_{-\infty}^{\infty} x f(x) dx} \tag{2}$$

As before, $E[X]$ is also often called the **mean** of $X$, or the **expected value** of $X$.

### 1.2.3 Examples

- The uniform distribution

  Consider a uniformly distributed continuous random variable $X$ with distribution $U(a, b)$.

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_{a}^{b} x f(x) dx \\ &= \int_{a}^{b} \frac{x}{b-a} dx = \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$

- The normal distribution

  Consider a normally distributed continuous random variable $X$ with distribution $N(\mu, \sigma^2)$.

$$\begin{aligned} E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx + \frac{\mu}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= -\frac{\sigma}{\sqrt{2\pi}} \underbrace{\left[ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right]_{-\infty}^{+\infty}}_{=0} + \mu \cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx}_{=\int_{-\infty}^{+\infty} f(x) dx = 1} \end{aligned}$$

  Thus,

$$E[X] = \mu$$

  The mean of a normally distributed random variable $X$ is the parameter $\mu$ in $N(\mu, \sigma^2)$, as already mentioned without proof in Lecture 4.

- The log-normal distribution

  Consider a log-normally distributed continuous random variable $X$ with parameters $\mu$ and $\sigma^2$.

$$\begin{aligned} E[X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{+\infty} x \cdot \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{+\infty} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} dx \\ &\underset{dz = \frac{dx}{\sigma x}}{\overset{z = \frac{\ln x - \mu}{\sigma}}{=}} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \sigma e^{\mu} e^{\sigma z} e^{-\frac{z^2}{2}} dz = \frac{e^{\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z^2 + 2\sigma z)} dz \\ &= \frac{e^{\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z+\sigma)^2} e^{\frac{\sigma^2}{2}} dz = e^{\mu + \frac{\sigma^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du}_{=\int_{-\infty}^{+\infty} \phi(u) du = 1} \end{aligned}$$

  So we see that for the lognormal distribution,

$$E[X] = e^{\mu + \frac{\sigma^2}{2}}$$

### 1.2.4 Important note on the existence of expected values

Thus far, we operated under the asumption that the integral $\int_{-\infty}^{+\infty} xf(x)dx$ is always well defined., i.e. is always convergence, as is its counterpart, the sum $\sum_i a_i p_X(a_i)$ for discrete random variables. This is not always true. There are random variables for which the expected value does not exist.

As an example, consider the discrete random variable $X$ which takes on the values $2^n$ with probability $P(X = 2^n) = \frac{1}{2^n}$, where $n$ is a natural number.

We have

$$E[X] = \sum_{n=1}^{+\infty} 2^n \cdot \frac{1}{2^n} = \sum_{n=1}^{+\infty} 1$$

We see that this sum is actually not well defined, as it diverges, so the expected value of this random variable does not exist.

For continuous random variables, a standard example is the *Cauchy distribution*, with probability density function

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad , \quad x \in \mathbb{R}$$

For a continuous random variable $X$ with this distribution

$$E[X] = \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx = \frac{1}{2\pi} \left[\ln(1+x^2)\right]_{-\infty}^{+\infty}$$

The integral is not convergent, so the expectation of this random variable does not exist.

# 2 Change of variables

We have seen in the previous section that the computation of the expected values of continuous random variables can involve challenging integrals, as did the computation of some cumulative distribution functions. Changes of variables can sometimes make these computations easier. This is one reason for bringing up the topic in this section. Another reason is that it will help us define the other key concept of this lecture, namely the *variance*.

## 2.1 Change of variables for probability mass functions

To understand how change of variables determine distributions, we start with the simpler case of discrete random variables.

Let $X$ be a discrete random variable which takes on the values $a_1, a_2, \ldots$, and $g$ be a one-to-one function on $\mathbb{R}$. We define the random variable $Y = g(X)$. $Y$ is a discrete random variable, which takes on the values $g(a_1)$, $g(a_2)$, ....

If the probability mass function $p_X$ of $X$ is given by

$$\begin{cases} p_X(a_1) = p_1 \\ p_X(a_2) = p_2 \\ \vdots \end{cases}$$

The probability mass functio $p_Y$ of $Y$ is then given by

$$\begin{cases} p_Y(g(a_1)) = p_1 = p_X(a_1) \\ p_Y(g(a_2)) = p_2 = p_X(a_2) \\ \vdots \end{cases}$$

which may be rewritten in the following more useful form:

$$\boxed{\begin{cases} p_Y(b_1) = p_1 = p_X(g^{-1}(b_1)) \\ p_Y(b_2) = p_2 = p_X(g^{-1}(b_2)) \\ \vdots \end{cases}}$$

**Example**

Let $X$ be the discrete random variable corresponding to the possible results of a fair dice. We can write:

$$\begin{cases} p_X(1) = \frac{1}{6} \\ p_X(2) = \frac{1}{6} \\ p_X(3) = \frac{1}{6} \\ p_X(4) = \frac{1}{6} \\ p_X(5) = \frac{1}{6} \\ p_X(6) = \frac{1}{6} \\ p_X(a) = 0 \quad \text{for} \quad a \notin \{1,2,3,4,5,6\} \end{cases}$$

The discrete random variable $Y = X^2$ takes on values in $\{1, 4, 9, 16, 25, 36\}$, and has the following probability mass function:

$$\begin{cases} p_Y(1) = p_X(\sqrt{1}) = \frac{1}{6} \\ p_Y(4) = p_X(\sqrt{4}) = \frac{1}{6} \\ p_Y(9) = p_X(\sqrt{9}) = \frac{1}{6} \\ p_Y(16) = p_X(\sqrt{16}) = \frac{1}{6} \\ p_Y(25) = p_X(\sqrt{25}) = \frac{1}{6} \\ p_Y(6) = p_X(\sqrt{36}) = \frac{1}{6} \\ p_Y(a) = 0 \quad \text{for} \quad a \notin \{1,4,9,16,25,36\} \end{cases}$$

## 2.2 Change of variables for probability density functions

Let us consider a one-to-one function $g$ on $\mathbb{R}$, a continuous random variable $X$ which takes values in $\mathbb{R}$, and a second continuous random variable $Y$ defined by $Y = g(X)$.

Given the probability density function $f_X$ of $X$, let us construct the probability density function $f_Y$ of $Y$. To do so, we consider the cumulative distribution function $F_Y$ and $F_X$ of $Y$ and $X$:

$$F_Y(x) = P(Y \leq x) = P(g(X) \leq x)$$

Now, since $g$ is one-to-one, it is strictly monotone. Let us first assume it is strictly increasing. Then

$$P(g(X) \leq x) = P(X \leq g^{-1}(x)) = F_X(g^{-1}(x))$$

Now, we know that the cumulative distribution functions are the antiderivatives of the probability density functions. Thus,

$$f_Y(x) = \frac{d}{dx}\left[F_Y(x)\right] = \frac{d}{dx}\left[F_X\left(g^{-1}(x)\right)\right]$$

Applying the chain rule, we find in the case when $g$ is strictly increasing,

$$f_Y(x) = \frac{d}{dx}\left(g^{-1}(x)\right) \cdot f_X(g^{-1}(x))$$

Let us now instead consider the case when $g$ is strictly decreasing. We can then write

$$P(g(X) \leq x) = P(X \geq g^{-1}(x)) = 1 - F_X\left(g^{-1}(x)\right)$$

Thus, for the case when $g$ is strictly decreasing,

$$f_Y(x) = \frac{d}{dx}\left[F_Y(x)\right] = -\frac{d}{dx}\left(g^{-1}(x)\right) f_X(g^{-1}(x))$$

We will now show that the two different formulae can be synthetized into a single formula, by observing that

- When $g$ is strictly increasing, $\frac{d}{dx}(g^{-1}(x)) > 0$, so that

$$\frac{d}{dx}(g^{-1}(x)) > 0 = \left|\frac{d}{dx}(g^{-1}(x)) > 0\right|$$

- When $g$ is strictly decreasing, $-\frac{d}{dx}(g^{-1}(x)) > 0$, so that

$$-\frac{d}{dx}(g^{-1}(x)) > 0 = \left|\frac{d}{dx}(g^{-1}(x)) > 0\right|$$

The two situations can therefore indeed be combined into a single expression

$$\boxed{\text{For } Y = g(X) \quad , \quad f_Y(x) = f_X\left(g^{-1}(x)\right)\left|\frac{d}{dx}\left(g^{-1}(x)\right)\right|}$$

Observe that this formula guarantees that $f_Y(x) \geq 0$ for all $x$, as desired.

Example: This technique of changing variable is in fact the method we used (without saying it explicitly) to derive the lognormal density function from the normal density function. Let us now verify the general formula we just derived does indeed work in that case.

The starting point is the probability density function for the normal distribution,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

and the lognormal distribution $Y$ is such that $Y = e^X$. In the notation of our derivation above, we thus have

$$g(x) = e^x \quad \Leftrightarrow \quad g^{-1}(x) = \ln x$$

According to our general formula, we can write

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}\left|\frac{d}{dx}(\ln x)\right| = \frac{1}{\sqrt{2\pi\sigma^2}\,x} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}$$

where we could drop the symbol for the absolute value on the right-hand side because the log-normal distribution is only non-zero and defined for $x > 0$.

We see that the general formula indeed yields the correct formula, as expected.

## 2.3 Expected value and change of variable

- Discrete random variables

  Let $X$ and $Y$ be two discrete random variables such that $Y = g(X)$. The expected value of $X$ is

$$E[X] = \sum_i a_i p_X(a_i)$$

  The expected value of $Y$ is

$$E[Y] = \sum_i g(a_i) p_Y(g(a_i)) = \sum_i g(a_i) p_X(a_i)$$

  In other words, for discrete random variables, we derived the general formula

$$\boxed{E[g(X)] = \sum_i g(a_i) p_X(a_i)} \tag{3}$$

- Continuous random variables

  Let $X$ and $Y$ be two continuous random variables such that $Y = g(X)$. We can write

$$
\begin{aligned}
E[Y] &= \int_{-\infty}^{+\infty} x f_Y(x)dx = \int_{-\infty}^{+\infty} x f_X(g^{-1}(x)) \left| \frac{d}{dx}\left(g^{-1}(x)\right) \right| dx \\
&\overset{u=g^{-1}(x)}{\underset{dx=g'(u)du}{=}} \int_{-\infty}^{+\infty} g(u) f_X(u) \left| \frac{1}{g'(u)} \frac{du}{du} \right| |g'(u)| du \\
&= \int_{-\infty}^{+\infty} g(u) f_X(u) du
\end{aligned}
$$

  We thus proved that for a continuous random variable $X$, we have

$$
\boxed{E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx} \tag{4}
$$

  Note that in order to derive these latest expressions, we used results we had proved only for functions $g : \mathbb{R} \to \mathbb{R}$ which are one-to-one. However, these formulae can actually be shown to hold even if the function $g$ is not one-to-one (we will not prove this in this class).

**Important illustrations**

- Consider a continuous random variable $X$, and the linear change of variable $g(x) = ax + b$, with $a$ and $b$ real numbers.

$$
E[g(X)] = \int_{-\infty}^{+\infty} (ax+b) f_X(x) dx = a \underbrace{\int_{-\infty}^{+\infty} x f(x) dx}_{=E[X]} + b \underbrace{\int_{-\infty}^{+\infty} f(x) dx}_{=1}
$$

  We thus showed that

$$
\boxed{E[aX + b] = aE[X] + b} \tag{5}
$$

  It is a good exercise for you to verify that this formula also holds for discrete random variables.

- **Linearity of expectation**

  Consider two functions $g$ and $h$ from $\mathbb{R}$ to $\mathbb{R}$, a continuous random variable $X$, and suppose we are interested in the random variable $Y = (ag + bh)(X) = ag(X) + bh(X)$, with $a$ and $b$ real numbers. We have

$$
E[ag(X) + bh(X)] = \int_{-\infty}^{+\infty} (ag(x) + bh(x)) f_X(x) dx = a \underbrace{\int_{-\infty}^{+\infty} g(x) f_X(x) dx}_{E[g(X)]} + b \underbrace{\int_{-\infty}^{+\infty} h(x) f_X(x) dx}_{E[h(X)]}
$$

  Hence,

$$
\boxed{E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]}
$$

  This formula generalizes equation (5), and will soon be useful to derive an alternative expression for the variance.

## 3 Variance of a random variable

As we said, the mean of a random variable is its most basic property, but it is not a complete characterization of the random variable. It is easy to see this. Consider for example two discrete random variables $X$ and $Y$ with the following probability mass functions:

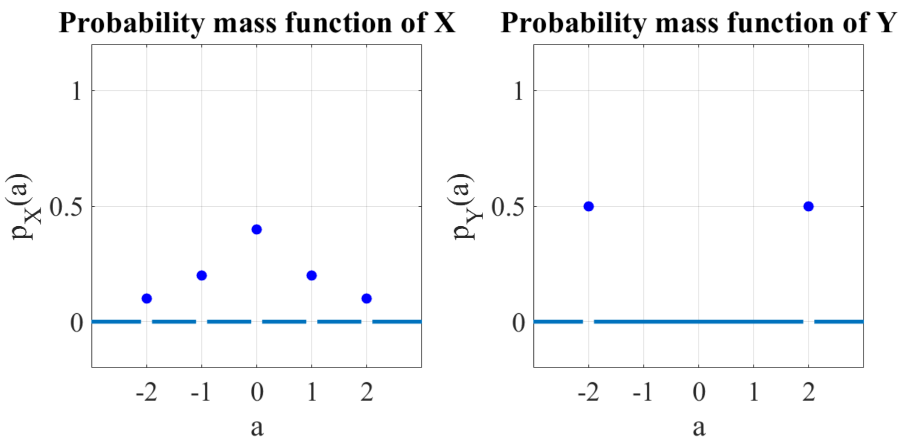


Figure 1: Two discrete random variables $X$ and $Y$ with the same expectation, $E[X] = E[Y]$, but different variance, $\text{Var}(X) \neq \text{Var}(Y)$.

$$\begin{cases} p_X(-2) = \frac{1}{10} \\ p_X(-1) = \frac{2}{10} \\ p_X(0) = \frac{4}{10} \\ p_X(1) = \frac{2}{10} \\ p_X(2) = \frac{1}{10} \\ p_X(a) = 0 \quad \text{for} \;\; a \notin \{-2,-1,0,1,2\} \end{cases} \qquad \text{and} \qquad \begin{cases} p_Y(-2) = \frac{1}{2} \\ p_Y(2) = \frac{1}{2} \\ p_Y(a) = 0 \quad \text{for} \;\; a \notin \{-2,2\} \end{cases}$$

It is easy to verify that $E[X] = E[Y] = 0$. Yet their spreads about the mean is quite different, as can be seen from their probability mass functions, plotted in Figure 1.

To characterize the spread of a probability mass function (for discrete random variables) or a probability density function (for continuous random variables) about the mean, we use the concept of *variance*.

### 3.1 Definition

The **variance** $\text{Var}(X)$ of a random variable $X$ is the number

$$\boxed{\text{Var}(X) = E[(X - E[X])^2]} \tag{6}$$

Note 1: Since $(X - E[X])^2 \geq 0$, the variance is **always a positive number**.

Note 2: Observe that $E[X]$ has the same dimension, i.e. "units", as $X$. Now,

$$(X - E[X])^2 = X^2 - 2XE[X] + E[X]^2$$

so $\text{Var}(X)$ has the same units as $X^2$. As a result, mathematicians and statisticians often like to use the **standard deviation** defined by

$$\boxed{s = \sqrt{\text{Var}(X)} \qquad \text{Standard deviation}} \tag{7}$$

which has the same units as $X$.

Note 3: It is easy to derive an alternative expression for the variance, which is often more convenient to use:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

Thus,

$$\boxed{\text{Var}(X) = E[X^2] - E[X]^2} \tag{8}$$

## 3.2 Examples

- Consider the discrete random variable $X$ which takes on the values $\{1,2,3,4,5,6\}$ with probability $\frac{1}{6}$ each.

  We already calculated that $E[X] = 3.5 = \frac{7}{2}$.

  Furthermore, using the change of variable formula, we have

$$E[X^2] = 1\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 9\cdot\frac{1}{6} + 16\cdot\frac{1}{6} + 25\cdot\frac{1}{6} + 36\cdot\frac{1}{6} = \frac{91}{6}$$

  Hence,

$$\mathrm{Var}(X) = \frac{91}{6} - \frac{49}{4} = \frac{70}{24} = \frac{35}{12} \approx 2.92$$

- Let $X$ be a Bernoulli random variable with parameter $p$.

$$\begin{aligned} E[X] &= 0\cdot(1-p) + 1\cdot p = p \\ E[X^2] &= 0^2\cdot(1-p) + 1^2\cdot p = p \\ \Rightarrow \quad \mathrm{Var}(X) &= p - p^2 = p(1-p) \end{aligned}$$

- Let $X$ be a normally distributed continuous random variable $X$ with distribution $N(\mu,\sigma^2)$.

  We already showed that $E[X] = \mu$. To compute the variance of $X$, it will be more convenient to evaluate $E[(X-\mu)^2]$. Using the change of variable formula, we can write

$$\begin{aligned} E[(X-\mu)^2] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &\underset{dz=dx}{\overset{z=x-\mu}{=}} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{1}{2}\frac{z^2}{\sigma^2}} dz \\ &\underset{\text{by parts}}{\overset{\text{Integration}}{=}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \left[-\sigma^2 z e^{-\frac{1}{2}\frac{z^2}{\sigma^2}}\right]_{-\infty}^{+\infty}}_{=0} + \sigma^2 \cdot \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\frac{z^2}{\sigma^2}} dz}_{=F_{N(0,\sigma)}(+\infty)=1} \\ &= \sigma^2 \end{aligned}$$

  The variance of a normally distributed random variable $X$ is the parameter $\sigma^2$ in $N(\mu,\sigma^2)$, as we already mentioned without justification in Lecture 4.

- Consider the continuous random variable $X$ with probability density function given by

$$f(x) = \begin{cases} \frac{3}{10}(3x - x^2) & \text{for } 0 \le x \le 2 \\ 0 & \text{for } x \notin [0,2] \end{cases}$$

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^2 x f(x) dx = \frac{3}{10}\int_0^2 (3x^2 - x^3) dx = \frac{3}{10}\left[x^3 - \frac{x^4}{4}\right]_0^2 = \frac{6}{5}$$

  Using the formula for a change of variable

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^2 x^2 f(x) dx = \frac{3}{10}\int_0^2 (3x^3 - x^4) dx = \frac{3}{10}\left[\frac{3}{4}x^4 - \frac{x^5}{5}\right]_0^2 = \frac{3}{10}(12 - \frac{32}{5}) = \frac{42}{25}$$

  Hence

$$\mathrm{Var}(X) = \frac{42}{25} - \frac{36}{25} = \frac{6}{25}$$