Expectation of a random variable

Definition. Suppose X is a discrete random variable with probability function $f(x)$. Then the **expected value** of X is given by

$$E(X) := \sum_{x \in range(X)} x \cdot f(x).$$

- $E(X)$ is sometimes referred to as the **mean**, or the **expectation**, or the **first moment** of X. It is often denoted by μ.
- Note that the mean is completely determined if we know p.f. Thus if

$$X \sim Y,$$

then X and Y must have the same means.

Interpretations

1. $E(X)$ is a weighted average of possible values of the r.v. X, where the outcomes that are more likely to occur are assigned higher weights.

   It is designed to summarize a particular aspect of a probability distribution: its "average" or "typical" outcome.

Example:

p.f.	1	2	3	4	5	6	μ
f_1	0.3	0.25	0.15	0.1	0.1	0.1	2.75
f_2	1/6	1/6	1/6	1/6	1/6	1/6	3.50
f_3	0.1	0.1	0.1	0.15	0.25	0.3	4.25

2. The often used description "expected value" can be misleading since the average value $E(X)$ may be a value which X never takes!

   As an example, consider X that takes values -1 and 1 with probabilities 1/2 each. Then,

$$E(X) = 0.$$

3. In this course we always assume that $E(X)$ exists and is finite, but there are some probability distributions for which this is not the case.

Clicker Question(s).

3. Important interpretation: $E(X)$ is what the average (called the **sample mean**) of many independent realizations of a random variable X would approach if the number of repetitions was increasing to infinity (Law of Large Numbers).

Let $X_1, X_2, X_3, \ldots$ be independent[1] and identically distributed random variables with the common mean $\mu = E(X_1) = E(X_2) = \cdots < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \quad \overset{\text{as } n \to \infty}{\longrightarrow} \quad \mu.$$

Thus, if $x_1, x_2, x_3, \ldots, x_n$ are observed values of $X_1, X_2, .., X_n$, then for large n the sample mean

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

should be close to μ.

[1] We will look at the formal definition of "independent" variables in Chapter 9

Example: For the following outcomes when rolling a six sided die 200 times:

$$421654441125 \ldots 143122$$

we get the sample mean $\frac{1}{200}\sum_{i=1}^{200} x_i = 3.370$. Compare this with the expected value of X_1, which is $\mu = 3.5$.

When we repeat the experiment, we get the outcomes

$$561254412225 \ldots 336136,$$

which give the sample mean 3.495.

Such experiments can be easily implemented in *R*:

```
r=sample(1:6, 200, replace=T)
mean(r)
```

Summary statistics

- To summarize a set of data values (often called a sample), we usually use the sample mean. Two other common summary statistics are the median and mode.

  Definition 14 The **median** of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.

  If there are an even number of observations, we go half way between the middle two values.

  Definition 15 The **mode** of a sample is the value which occurs most often.

  There is no guarantee there will be only a single mode.

- Median and mode can be also defined for model distributions.

Example. When we order the data

$$2, 3, 7, 3, 9, 10, 5$$

we get

$$2\ 3\ 3\ 5\ 7\ 9\ 10$$

which gives

$$\text{mean} = 5.57, \quad \text{median} = 5, \quad \text{mode} = 3.$$

For the data

$$2, 3, 7, 3, 9, 10, 5, 6$$

we get

$$\text{mean} = 5.625, \quad \text{median} = 5.5, \quad \text{mode} = 3.$$

Properties of expectation

1. Expectation is a linear operation: for any constants a and b we have

$$E(aX + b) = aE(X) + b.$$

   More generally, for two r.v.'s X and Y:

$$E(aX + bY) = aE(X) + bE(Y).$$

2. If a r.v. X is such that $a \leq X \leq b$ for two constants a and b, then

$$a \leq E(X) \leq b.$$

3. If $X \geq 0$, then
$$E(X) \geq 0.$$

4. If two random variables X and Y are defined on the same sample space S and satisfy
$$X(s) \leq Y(s) \text{ for each } \quad s \in S,$$
then
$$E(X) \leq E(Y).$$

Clicker Question(s).

- Often we are interested in a random variable Y that is a function of another random variable X:

$$Y := g(X),$$

where g is given function $g : \mathbb{R} \to \mathbb{R}$.

For example, $Y = X^2$, where X is the result of a fair six sided die roll.

  - Note that the general properties (1)–(4) of expectation apply directly to $g(X)$. For example

$$E[ag(X) + b] = aE[g(X)] + b$$

    for given constants a and b.

- **Important:** typically, for a general non-linear function g, we have

$$E[g(X)] \neq g(E[X]).$$

For example, in general

$$E(X^2) \neq (E(X))^2.$$

- How do we find $E[g(X)]$?
  - one way of finding $E[g(X)]$ is to apply the definition of expectation to the r.v. $Y := g(X)$. This, however, requires finding the p.f. f_Y of Y.
  - a faster way is described by the following theorem.

Theorem 17 Let X be a discrete random variable with p.f. $f_X(x)$. Then the expected value of some function $g(X)$ of X is given by

$$E[g(X)] = \sum_{x \in range(X)} g(x) f_X(x).$$

Proof:

$$\begin{aligned} E[g(X)] &= \sum_{y \in range(Y)} y f_Y(y) \\ &= \sum_{y \in range(Y)} y \sum_{x : g(x) = y} f_X(x) \\ &= \sum_{y \in range(Y)} \sum_{x : g(x) = y} y f_X(x) \\ &= \sum_{y \in range(Y)} \sum_{x : g(x) = y} g(x) f_X(x) \\ &= \sum_{x \in range(X)} g(x) f_X(x). \end{aligned}$$

Clicker Question(s).

Applications of Expectations

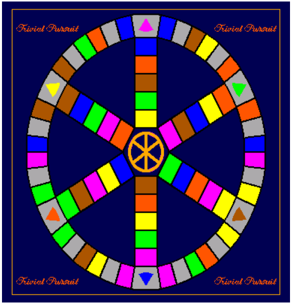
Example. A lottery is conducted in which 7 numbers are drawn without replacement between the numbers 1 and 50. A player wins the lottery if the numbers selected on their ticket match all 7 of the drawn numbers. A ticket to play the lottery costs \$1, and the jackpot is valued at \$5,000,000.

(i) Compute the expected payout for this bet.

(ii) Compute the expected value of the player's net winnings from a single play[2].

[2] See also Problems 9, 10 and 11 in Section 7.5

Example. When playing trivial pursuit, to finish the game one must roll a specified number on a single fair six sided die in order to land on the middle tile. Once in range of the middle tile, the probability of doing this on any one roll is $1/6$. Let X denote the number of rolls required in order to finish the game. Compute[3] $E[X]$.



[3] See also Problem 5 in 7.5

Example - Double down strategy. Consider the following strategy to win \$1. In the game of roulette, bet \$1 on black. If you win, your net winning is \$1:

$$\$2 - \$1 = \$1.$$

If you lose, bet \$2. If you win on your second bet, your net winning is \$1:

$$\$4 - \$2 - \$1 = \$1.$$

Otherwise, bet \$ 2^2 on next bet. Proceed in this way. If you win on the k^{th} bet you will have a return of

$$2^k - \sum_{i=0}^{k-1} 2^i = 1$$

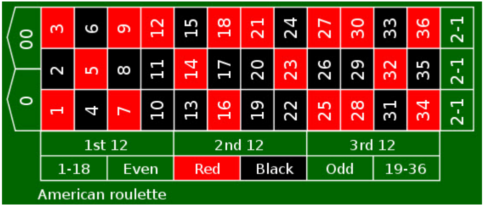
dollars.

Theorem. The probability that this betting system will result in you winning \$1 is one... so long as you have infinite money.

American Roulette Wheel:

1. Numbers 1-36. Half are black, other half red
2. Two green slots: 0 00

Bet	Payout
Black/Red	1 to 1
Odd/Even	1 to 1
1st-3rd Dozen	2 to 1
Columns	2 to 1
Any number (including 0,00)	35 to 1



Means of Selected Probability Models

We would like to compute $E[W]$, $E[X]$, $E[Y]$, and $E[Z]$ when

(i) $W \sim Binomial(n, p)$
(ii) $X \sim Hyp(N, r, n)$
(iii) $Y \sim NB(k, p)$
(iv) $Z \sim Poi(\mu)$

- If $W \sim Binomial(n,p)$, then $E[W] = np$.

$$
\begin{aligned}
E(W) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n(n-1)!}{x(x-1)!(n-x)!} p p^{x-1} (1-p)^{(n-1)-(x-1)} \\
&= np(1-p)^{n-1} \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{-(x-1)} \\
&= np(1-p)^{n-1} \sum_{x=1}^{n} \binom{n-1}{x-1} (\frac{p}{1-p})^{x-1} \\
&= np(1-p)^{n-1} \sum_{y=0}^{n-1} \binom{n-1}{y} (\frac{p}{1-p})^{y} \\
&= np(1-p)^{n-1} (1 + \frac{p}{1-p})^{n-1} = np(1-p)^{n-1} \frac{1}{(1-p)^{n-1}} \\
&= np.
\end{aligned}
$$

- If $Z \sim Poi(\mu)$, then $E[Z] = \mu$.

$$
\begin{aligned}
E(Z) &= \sum_{x=1}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} \\
&= \sum_{x=1}^{\infty} x \frac{\mu \mu^{x-1}}{x(x-1)!} e^{-\mu} \\
&= \mu e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} \\
&= \mu e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} \\
&= \mu e^{-\mu} e^{\mu} = \mu.
\end{aligned}
$$

Note: the technique (summation method) we use to find the mean is the same as the one we use to prove that $\sum_x f(x) = 1$.

- If $X \sim Hyp(N, r, n)$, then $E[X] = n\frac{r}{N}$.
- If $Y \sim NB(k, p)$, then $E[Y] = \frac{k(1-p)}{p}$.

Clicker Question(s).

Variance

The expectation $E(X)$ can be used to predict the value of a random variable X.
But one may wonder how much the random variable tends to deviate from its mean.
To measure the amount of variability of a r.v. X we may consider using:

(i) Average deviation

$$E[(X-\mu)] = E[X] - \mu = 0$$

(ii) Average absolute deviation:

$$E|X-\mu|$$

(iii) Average squared deviation:

$$E(X-\mu)^2$$

Definition. The **variance** of a random variable X, denoted by $Var(X)$, is defined as[4]

$$Var(X) = E[(X - E(X))^2].$$

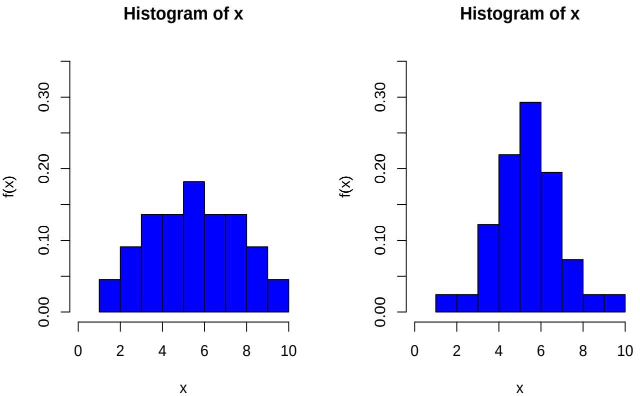
Thus, $Var(X)$ is the average square of the distance between a possible value of X and the mean $E(X)$. It is a widely used measure of variability.

- Larger values of $Var(X)$ indicate that the distribution is more "spread out" around the mean.
- Note that variance is completely determined if we know p.f. Thus if

$$X \sim Y,$$

then X and Y must have the same variances.

[4] we assume that it exists



Left panel: $\mu = 5.13,\ \sigma^2 = 4.65,\ \sigma = 2.16$

Right panel: $\mu = 5.14,\ \sigma^2 = 2.33,\ \sigma = 1.53$

Properties of variance:

- For any r.v. X
$$Var(X) \geq 0.$$

- For any r.v. X and constants a and b
$$Var(aX + b) = a^2 Var(X).$$

- Equivalent representations of variance
$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2 \end{aligned}$$

- $Var(X) = 0$ if and only if $P(X = E(X)) = 1$.

Example. Let X denote the outcome of a fair six sided die roll. Compute $Var(X)$.

We have,

x	1	2	3	4	5	6	$E(X) = 3.50$
x^2	1	4	9	16	25	36	$E(X^2) = 15.167$
p.f.	1/6	1/6	1/6	1/6	1/6	1/6	

Therefore,

$$Var(X) = 15.167 - (3.5)^2 = 2.92.$$

Clicker Question(s).

Exercise: Suppose X has CDF

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{4} & 0 \leq x < 1/2 \\ \frac{3}{4} & 1/2 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

Compute $E(X)$ and $Var(X)$.

Standard Deviation

Note that we should not compare directly the mean of a random variable with its variance, as their values are on different scales.
To compare the variability of a r.v. X with its mean, we use the standard deviation of X.

Definition. The **standard deviation** of a random variable X is defined as

$$SD(X) := \sqrt{Var(X)}.$$

Standard deviation is also denoted by σ.

Important: standard deviation does not have properties similar to those we have derived for variance.
Therefore, we typically work with variance first and only at the end we take the square-root.

For example, to find

$$SD(aX + b),$$

where a and b are given constants, we first use

$$Var(aX + b) = a^2 Var(X),$$

and then

$$SD(aX + b) = \sqrt{Var(aX + b)} = \sqrt{a^2 Var(X)} = |a| SD(X).$$

Variance of Binomial: For $X \sim Binomial(n, p)$, we have

$$Var(X) = np(1-p).$$

Since

$$Var(X) = E[X(X-1)] + E(X) - [E(X)]^2,$$

and $E(X) = np$, it suffices to find $E[X(X-1)]$.

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1)\frac{n(n-1)(n-2)!}{x(x-1)(x-2)!(n-x)!} p^2 p^{x-2}(1-p)^{(n-2)-(x-2)} \\
&= n(n-1)p^2(1-p)^{n-2} \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2}(1-p)^{-(x-2)} \\
&= n(n-1)p^2(1-p)^{n-2} \sum_{x=2}^{n} \binom{n-2}{x-2} (\frac{p}{1-p})^{x-2} \\
&= n(n-1)p^2(1-p)^{n-2} \sum_{y=0}^{n-2} \binom{n-2}{y} (\frac{p}{1-p})^{y}
\end{aligned}$$

By applying the Binomial theorem, we get

$$\begin{aligned} E[X(X-1)] &= n(n-1)p^2(1-p)^{n-2}\sum_{y=0}^{n-2}\binom{n-2}{y}\left(\frac{p}{1-p}\right)^y \\ &= n(n-1)p^2(1-p)^{n-2}\left(1+\frac{p}{1-p}\right)^{n-2} \\ &= n(n-1)p^2(1-p)^{n-2}\frac{1}{(1-p)^{n-2}} = n(n-1)p^2. \end{aligned}$$

Hence,

$$Var(X) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Variance of Poisson: For $X \sim Poi(\mu)$, we have

$$Var(X) = \mu.$$

Since

$$Var(X) = E[X(X-1)] + E(X) - [E(X)]^2,$$

and $E(X) = \mu$, it suffices to find $E[X(X-1)]$.

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)\frac{\mu^x}{x!}e^{-\mu} \\
&= \sum_{x=2}^{\infty} x(x-1)\frac{\mu^2\mu^{x-2}}{x(x-1)(x-2)!}e^{-\mu} \\
&= \mu^2 e^{-\mu}\sum_{x=2}^{\infty}\frac{\mu^{x-2}}{(x-2)!} \\
&= \mu^2 e^{-\mu}\sum_{y=0}^{\infty}\frac{\mu^y}{y!} \\
&= \mu^2 e^{-\mu}e^{\mu} = \mu^2.
\end{aligned}$$

Hence,

$$Var(X) = \mu^2 + \mu - \mu^2 = \mu.$$

Exercise. Suppose that X_n is binomial with the parameters n and p_n such that

$$np_n \to \mu \quad \text{as} \quad n \to \infty.$$

If $Y \sim Poi(\mu)$ show that

$$\lim_{n\to\infty} Var(X_n) = Var(Y).$$

- If $X \sim Hyp(N, r, n)$, then

$$Var(X) = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

- If $Z \sim NB(k, p)$, then $Var(Z) = \frac{k(1-p)}{p^2}$

Example. If your chances of winning a lottery prize in a single game is $p = 10^{-10}$, then the mean and the standard deviation of the number of games you have to play in order to win are

$$\frac{1-p}{p} + 1 = \frac{1}{p} = 10^{10} \quad \text{and} \quad \sqrt{\frac{(1-p)}{p^2}} \approx 10^{10}.$$

Clicker Question(s).

Skewness and Kurtosis

- Expectation and variance give a simple summary of the distribution.
- To capture other aspects of a p.f. we also consider

  Skewness:
  $$E\left[\frac{(X - E(X))}{\sqrt{Var(X)}}\right]^3$$
  It measures asymmetry of the distribution. For symmetric distributions it is zero.

  Kurtosis:
  $$E\left[\frac{(X - E(X))}{\sqrt{Var(X)}}\right]^4$$
  It measures how slowly the tails of the distribution decay to zero, that is, how likely it is that we get extreme observations.

Distribution without expectation

If X is a random variable with the probability function

$$f_X(x) = \frac{6}{\pi^2}\frac{1}{x^2}, \quad x = 1, 2, ...$$

then $E(X) = +\infty$, and $Var(X)$ is not defined.