The expectation and variance of a random variable

Tan Do

Vietnamese-German University

Lecture 9

In this lecture

- Expectation of a random variable
- Variance of a random variable

Expectations of discrete random variables

The expected value or expectation of a discrete random variable with a probability mass function $P(X = x_i) = p_i$ is

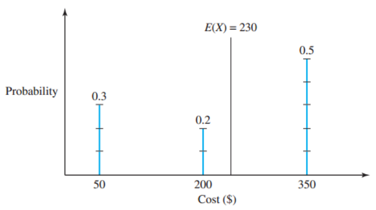
$$E(X) = \sum_i p_i x_i.$$

$E(X)$ measures the average value taken by the random variable X, which is also known as the mean of X.

Example (Machine breakdowns) The expected repair cost is

$$E(\text{cost}) = 50 \times 0.3 + 200 \times 0.2 + 350 \times 0.5 = 230.$$

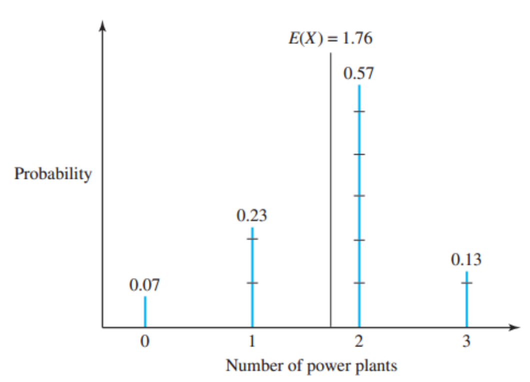
So in the long run, the repairs will cost an average of about \$230 each.



Example (Power plant operation) The expected number of power plants generating electricity is

$$E(X) = 0 \times 0.07 + 1 \times 0.23 + 2 \times 0.57 + 3 \times 0.13 = 1.76.$$

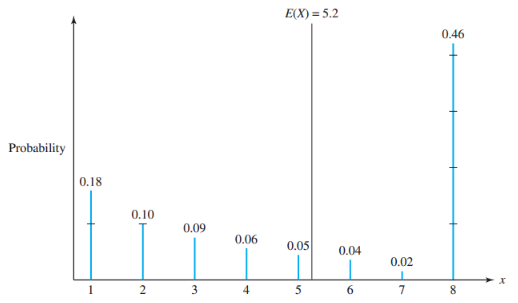
So in the long run, the average number of power plants generating electricity is 1.76 at a particular point in time.



Example (Personnel recruitment) The expected number of applicants interviewed is

$$E(X) = 1\times 0.18 + 2\times 0.1 + 3\times 0.09 + 4\times 0.06 + 5\times 0.05 + 6\times 0.04 + 7\times 0.02 + 8\times 0.46 = 5.$$

This expected value provides some indication of how many of the 8 applicants will actually have to be interviewed under the company's interviewing strategy.



Exercise Can you think of a new interviewing strategy to reduce the average number of applicants interviewed?

Games of chance

Die rolling If a fair die is rolled, the expected value of the outcome is

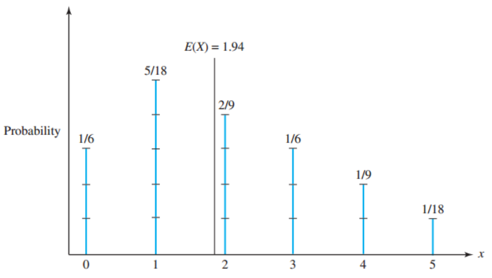
$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5.$$

The probability mass function of the positive difference between the scores is given by the table

x_i	0	1	2	3	4	5
p_i	1/6	5/18	2/9	1/6	1/9	1/18

The expected difference is

$$E(X) = 0 \times \frac{1}{6} + 1 \times \frac{5}{18} + 3 \times \frac{1}{6} + 4 \times \frac{1}{9} + 5 \times \frac{1}{18} = \frac{35}{18} = 1.94.$$



Question If you want to play a game in which you pay the player the dollar amount of the difference in the scores, how much should you charge a person to play?

Answer The long-run average difference in the scores is 1.94. To make a profit, you must charge a person more than $1.94.

Suppose you charge $2 per game. It is expected that sometimes you win, sometimes you lose or sometimes you break even. But in the long run, you expect the profit will be $2 - $1.94 per game, i.e., 6 cents per game. The more people who play the better.

Of course, this assumes that the dice are fair. If they are not, the result may be different.

Expectations of continuous random variables

The expected value or expectation of a continuous random variable with a probability density function $f(x)$ is

$$E(X) = \int_{\text{state space}} x f(x)\, dx.$$

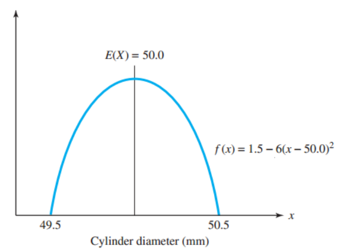
$E(X)$ measures the average value taken by the random variable X, which is also known as the mean of X.

Example (Metal cylinder production) The expected diameter of a metal cylinder is

$$E(X) = \int_{49.5}^{50.5} x(1.5 - 6(x-50)^2)\,dx$$

Let $y = x - 50$. Then

$$\begin{aligned} E(X) &= \int_{-0.5}^{0.5} (y+50)(1.5-6y^2)\,dy \\ &= \int_{-0.5}^{0.5} (-6y^3 - 300y^2 + 1.5y + 75)\,dy \\ &= \left[-\frac{3}{2}y^4 - 100y^3 + 0.75y^2 + 75y\right]_{-0.5}^{0.5} = 50. \end{aligned}$$

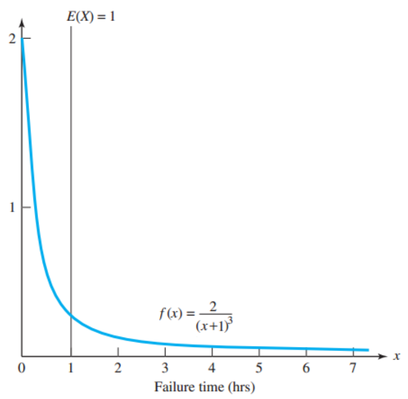


Example (Battery failure times) The expected battery failure time is

$$E(X) = \int_0^\infty \frac{2x}{(x+1)^3}\,dx = \int_0^\infty \frac{2}{(x+1)^2} - \frac{2}{(x+1)^3}\,dx$$

$$= \left[\frac{-2}{x+1} + \frac{1}{(x+1)^2}\right]_0^\infty = 0 - (-1) = 1.$$

This expected value indicates that the batteries fail on average after an hour of operation.

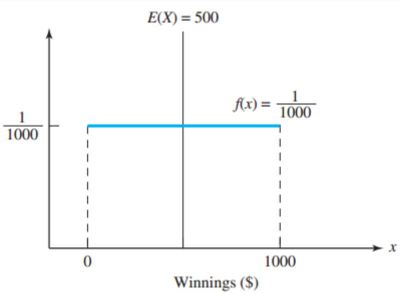


Games of chance

Dial spinning game The expected winnings are

$$E(X) = \int_0^{1000} \frac{x}{1000}\,dx = \frac{x^2}{2000}\Big|_0^{1000} = 500.$$

A fair price for playing the game is thus \$500.



Medians of random variables

The median of a continuous random variable X with a cumulative distribution function $F(x)$ is the values x in the state space for which

$$F(x) = 0.5.$$

Intuitively, the median gives information about the "middle" value of the random variable.

Example (Metal cylinder production) The median value of the metal cylinder diameters is the solution to

$$F(x) = 1.5x - 2(x-50)^3 - 74.5 = 0.5$$

which is $x = 50$.

In this case, the median is the same as the expectation, since the probability density function is symmetric about $x = 50$.

In general, if the p.d.f $f(x)$ is symmetric about $x = \mu$, then both the median and the expectation are equal to μ.

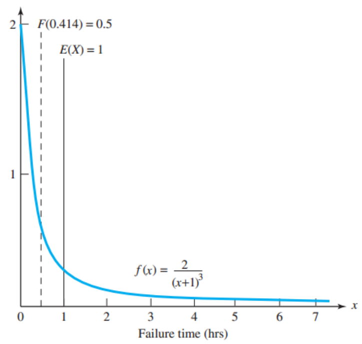
Example (Battery failure times) The median value of the battery failure times is the solution to

$$F(x) = 1 - \frac{1}{(x+1)^2} = 0.5$$

which is $x = \sqrt{2} - 1 = 0.414\,(\text{hours}) = 25\,(\text{mins})$.

In this case, the expected lifetime (1 hour) is much longer than the median lifetime (25 mins). The reason is that if a battery last for more than 25 mins, it is likely that it will last much longer.

Both the expectation and the median provide useful information about a random variable. For example, if an engineer has one application and wants to know how many batteries are required to keep it running over a long period of time, say a day, then the expectation must be used. So the engineer will need about 24 batteries on average.



On the other hand, if the engineer has many appliances and wants to know how many of them will last for more than 30 mins on newly charge batteries, then the median must be used. So the number will be about less than half of the appliances.

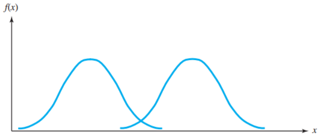
Definition and interpretation of variance

The variance of a random variable X is defined to be

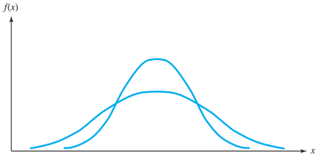
$$\mathrm{Var(X)} = \mathrm{E}((\mathrm{X} - \mathrm{E(X)})^2) = \mathrm{E}(\mathrm{X}^2) - (\mathrm{E(X)})^2.$$

The variance is a positive quantity that measures the spread of the distribution of the random variable about its mean value (expectation).

Larger values of the variance indicate that the distribution is more spread out.



Distributions with different mean but identical variances



Distributions with identical mean but different variances

The standard deviation σ of a random variable X is defined to be the square root of the variance:

$$\sigma = \sqrt{\mathrm{Var(X)}} \quad \text{or} \quad \mathrm{Var(X)} = \sigma^2.$$

Examples

Example (Machine breakdowns) We calculated before that $E(X) = \$230$.
The variance is

$$\mathrm{Var}(X) = E((X - E(X))^2) = \sum_i p_i \, (x_i - E(X))^2 = 17100.$$

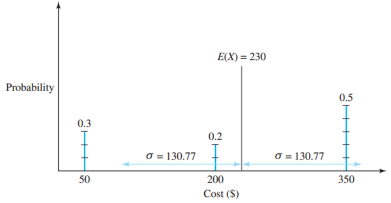
x_i	50	200	350
p_i	0.3	0.2	0.5

The standard deviation is

$$\sigma = \sqrt{\mathrm{Var}(X)} = \$130.77.$$

Alternatively, $E(X^2) = \sum_i p_i x_i^2 = 70000$, which gives

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 70000 - 230^2 = 17100.$$

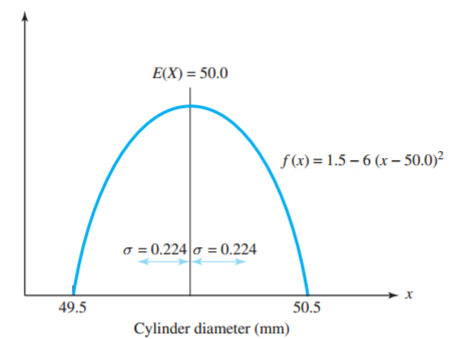


Example (Metal cylinder production) Recall that the mean cylinder diameter is $E(X) = 50$. So the variance is

$$\mathrm{Var}(X) = E((X-E(X))^2) = \int_{49.5}^{50.5} (x-50)^2 f(x)\, dx = \int_{49.5}^{50.5} (x-50)^2 \left(1.5-6(x-50)^2\right) dx = 0.05.$$

The standard deviation is

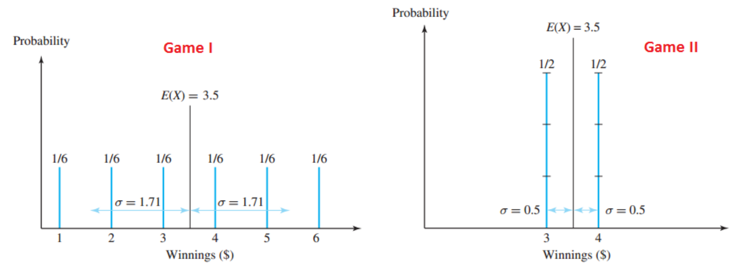
$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{0.05} = 0.224.$$



Games of chance

Die rolling Consider 2 games:

- Game I: A fair die is rolled and a player wins the dollar amount of the score obtained.
- Game II: The same die is rolled, but the player wins \$3 if a score of 1, 2 or 3 is obtained and wins \$4 if a score of 4, 5 or 6 is obtained.



In both games, the expected winnings are the same. So both games produce the same average revenue.

However, the variability in the winnings of Game I is clearly larger than the variability in the winnings of Game II. Consequently, the winnings of Game I should have larger variance than that of Game II.

Indeed, for Game I we have

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) = \sum_i p_i (x_i - 3.5)^2 \\ &= \frac{1}{6}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \ldots + \frac{1}{6}(6-3.5)^2 = \frac{35}{12},\end{aligned}$$

which gives $\sigma = \sqrt{35/12} = 1.71$.

For Game II we have

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) = \sum_i p_i (x_i - 3.5)^2 \\ &= \frac{1}{2}(3-3.5)^2 + \frac{1}{2}(4-3.5)^2 = \frac{1}{4},\end{aligned}$$

which gives $\sigma = \sqrt{1/4} = 0.5$.