Expectation, Variance and Moments

Averages

Consider the rolling of a fair six-sided die. We are interested in the ***average value*** of the result. One approach is the following:

Since each result (numbers 1,2,3,4,5,6) occurs with probability 1/6, we take the weighted sum:

$$\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5.$$

The average result of a die roll is then 3.5, even though this result itself can never occur.

As we shall see later, there are also other ways of thinking of an average (such as the ***median*** or the ***modes*** of a distribution).

Expectation

4.1. Definition. Let (X, f_X) be a discrete random variable. Then the ***expected value*** or ***expectation*** of X is

$$\mathrm{E}[X] := \sum_{x \in \Omega} x \cdot f_X(x).$$

provided that the sum (possibly series, if Ω is infinite) on the right converges absolutely.

We often write μ_X or simply μ for the expectation.

4.2. Example. We will prove later that the expectation for a geometric distribution X is

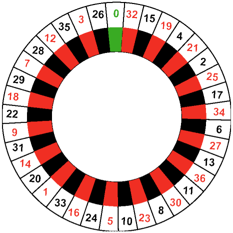
$$\mathrm{E}[X] = \frac{1}{p}$$

and for a binomial distribution Y the expectation is

$$\mathrm{E}[Y] = np.$$

Roulette

Wheel of fortune. Wikimedia Commons. Wikimedia Foundation. Web. 1 October 2012



Roulette call bets by Darsie. Wikimedia Commons. Wikimedia Foundation. Web. 1 October 2012

In (European) roulette, the player may bet an amount x on a number. If the ball lands on that number, he receives 36 times his initial bet, $36x$; if the ball lands on a different number, he loses his bet. There are 37 numbers on the wheel, so the expected winnings are

$$E[W] = \underbrace{\frac{1}{37}\cdot(-x) + \cdots + \frac{1}{37}\cdot(-x)}_{\text{36 times}} + \frac{1}{37}\cdot(36-1)x$$
$$= -\frac{1}{37}x.$$

A game is said to be ***fair*** if the expected winnings are zero. The addition of the green zero to the Roulette wheel makes the game into an unfair game for the player and ensures the casino's profit.

American Roulette wheels actually have two zeroes (green "0" and "00")!

St. Petersburg Paradox

Suppose someone offers you the following game: he will flip a fair coin, and if heads come up on the first toss, you receive 2 RMB. If the first toss comes up tails and the second toss comes up heads, he will give you 4 RMB; if only the third toss yields heads, you receive 8 RMB; and so on. Thus, if the first heads comes up on the nth toss, you will receive 2^n RMB.

Question. What is a fair price to pay in order to enter into the game? In other words, what are the expected winnings? How much would you pay to be allowed to play the game?

St. Petersburg Paradox

Calculating the expectation, we see that the probability of the first head coming up on the nth toss is $1/2^n$. Then

$$E[W] = \frac{1}{2}\cdot 2 + \frac{1}{4}\cdot 4 + \frac{1}{8}\cdot 8 + \cdots = \infty.$$

The expected value is infinite! (More precisely, the expectation doesn't exist in the sense of Definition 4.1.)

Hence, you should be willing to pay any finite amount (such as 1,000,000 RMB) to participate in the game.

The fact that most people would not pay nearly as much is known as the ***St. Petersburg paradox***.

Functions of Random Variables

Given a random variable X, we may consider functions of X. For example,

$$Y := X^2$$

represents the random variable obtained by squaring the values of X. In the case of a discrete random variable, it is not difficult to find the density function f_Y of $Y = \varphi(X)$ where $\varphi\colon \Omega \to \mathbb{R}$ represents a suitable function:

$$f_Y(y) = P[Y = y] = P[\varphi(X) = y]$$

In particular, if $y \notin \operatorname{ran}\varphi$, then $P[\varphi(X) = y] = 0$ and hence $f_Y(y) = 0$. Furthermore, since X is discrete,

$$P[\varphi(X) = y] = \sum_{\substack{x \in \Omega \\ \varphi(x) = y}} f_X(x).$$

(Since the outcomes $X = x$ for different values of x are mutually exclusive, their probabilities can be summed.)

Expectation of a Function of a Random Variable

If $X\colon S \to \Omega$ is a random variable with density f_X, $\varphi\colon \Omega \to \mathbb{R}$ a function and $Y = \varphi(X)$, then

$$\begin{aligned}
\mathrm{E}[Y] &= \sum_{y\in\mathbb{R}} y \cdot f_Y(y) = \sum_{y\in\mathbb{R}} \sum_{\substack{x\in\Omega \\ \varphi(x)=y}} y \cdot f_X(x) \\
&= \sum_{\substack{(x,y)\in\Omega\times\mathbb{R} \\ y=\varphi(x)}} y \cdot f_X(x) = \sum_{x\in\Omega} \varphi(x) f_X(x).
\end{aligned}$$

We have hence proved the following result:

4.3. Lemma. Let (X, f_X) be a discrete random variable and $\varphi\colon \Omega \to \mathbb{R}$ some function. Then the expected value of $\varphi \circ X$ is

$$\mathrm{E}[\varphi \circ X] = \sum_{x\in\Omega} \varphi(x) \cdot f_X(x).$$

provided that the sum (series) on the right converges absolutely.

Some Properties of the Expectation

By taking $\varphi(x) = c \in \mathbb{R}$ (constant) and $\varphi(x) = c \cdot x$ for some $c \in \mathbb{R}$, we immediately obtain

$$\mathrm{E}[c] = c, \qquad\qquad \mathrm{E}[cX] = c\,\mathrm{E}[X]$$

for any discrete random variable X.

Given two random variables X and Y their values can be added to yield a new random variable $X + Y$. (This sort of function of multiple random variables will be discussed in more detail in a later section.) For now we give, without proof, the following result:

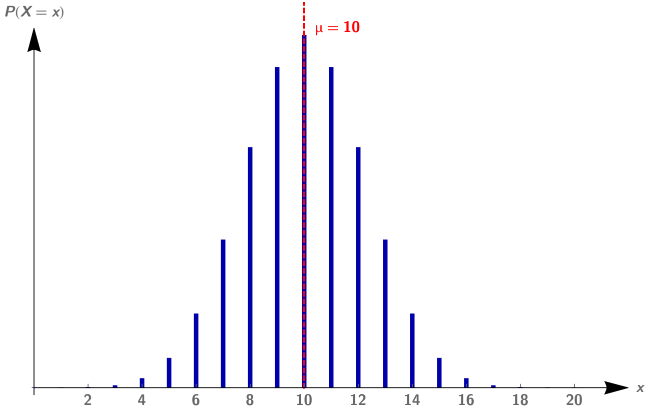
4.4. Theorem. Let X and Y be random variables. Then

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y].$$

Location

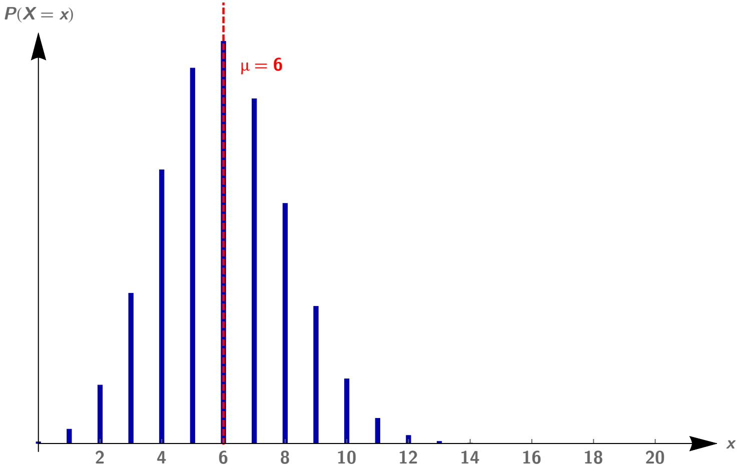
The expectation can be seen as a measure of ***location*** of a distribution: it indicates where the values of a random variable are concentrated.

The graph below shows the values of the probability density function for a binomial random variable with $n = 20$ and $p = 0.5$:



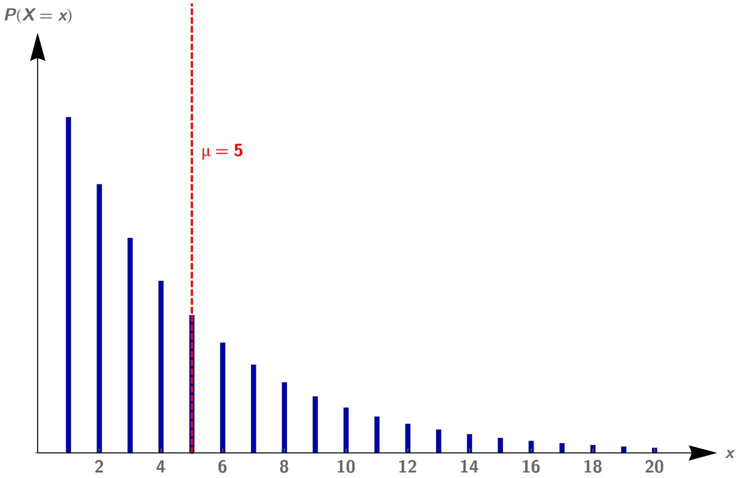
Location

For comparison, here is the graph of the values of the probability density function for a binomial random variable with $n = 20$ and $p = 0.3$:



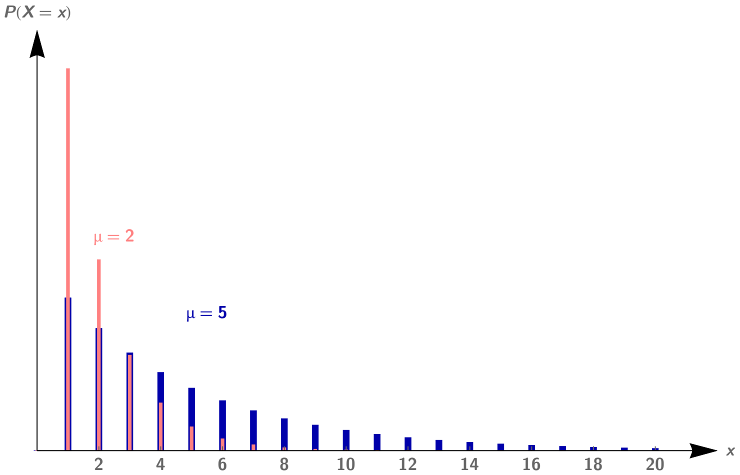
Location

Finally, this is the graph of the values of the probability density function for a geometric random variable with $p = 0.2$:



Dispersion

The ***dispersion*** measures how much the values of a random variables deviate from their mean. The example below shows two geometric random variables with $p = 0.5$ and $p = 0.2$, respectively.



Variance and Standard Deviation

One possible way to measure the dispersion of a random variable is the ***variance***, which is the

mean square deviation from the mean.

Given X, the deviation from the mean is $X - \mathrm{E}[X]$. The mean square deviation is hence

$$\mathrm{Var}[X] := \mathrm{E}\big[(X - \mathrm{E}[X])^2\big],$$

which is defined as long as the right-hand side exists.

The variance is often denoted by σ_X^2 or just σ^2.

The ***standard deviation*** is defined as

$$\sigma_X = \sqrt{\mathrm{Var}[X]}.$$

Some Properties of the Variance

Using the properties of the mean, we can derive the useful formula

$$\begin{aligned}
\operatorname{Var}[X] &= \mathrm{E}\big[(X - \mathrm{E}[X])^2\big] \\
&= \mathrm{E}\big[X^2 - 2\,\mathrm{E}[X]\cdot X + \mathrm{E}[X]^2\big] \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2.
\end{aligned}$$

It is then easy to check that for any constant $c \in \mathbb{R}$,

$$\operatorname{Var}[c] = 0, \qquad\qquad \operatorname{Var}[cX] = c^2 \operatorname{Var}[X]$$

where c by itself is interpreted as a random variable whose values are constant and cX is interpreted in the obvious way.

Standardized Random Variables

It is often useful to "standardize" a random variable by subtracting its mean and dividing by the standard deviation. If X is a given random variable, the standardized variable is hence

$$Y = \frac{X-\mu}{\sigma}.$$

We find that

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}\left[\frac{1}{\sigma}(X-\mu)\right] = \frac{1}{\sigma}\,\mathrm{E}\big[(X-\mu)\big] \\ &= \frac{1}{\sigma}\big(\mathrm{E}[X]-\mu\big) \\ &= 0 \end{aligned}$$

A similar calculation shows that $\mathrm{Var}[Y] = 1$.

Standardized Random Variables

It is often useful to "standardize" a random variable by subtracting its mean and dividing by the standard deviation. If X is a given random variable, the standardized variable is hence

$$Y = \frac{X - \mu}{\sigma}.$$

We find that

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}\left[\frac{1}{\sigma}(X-\mu)\right] = \frac{1}{\sigma}\,\mathrm{E}\big[(X-\mu)\big] \\ &= \frac{1}{\sigma}\big(\mathrm{E}[X] - \mu\big) \\ &= 0 \end{aligned}$$

A similar calculation shows that $\mathrm{Var}[Y] = 1$.

Standardized Bernoulli Variables

4.5. Example. Consider a Bernoulli random variable X which takes on values 0 and 1 with probability $p = 1/2$. Then

$$\mathrm{E}[X] = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$$

and

$$\begin{aligned}\mathrm{Var}[X] = \mathrm{E}[(X - 1/2)^2] &= \frac{1}{2}(0 - 1/2)^2 + \frac{1}{2}(1 - 1/2)^2 \\ &= \frac{1}{4}.\end{aligned}$$

Then the standardized random variable is

$$Y = \frac{X - 1/2}{1/2} = 2X - 1.$$

In other words, Y takes on the values 1 and -1, each with probability $1/2$.

Standardized Random Variables

Question. Given that $Y = (X - \mu)/\sigma$ has expectation zero, what is $\mathrm{E}[Y^2]$?

(a) 0
(b) 1
(c) $\mathrm{E}[Y^2]$ is not defined.
(d) $\mathrm{E}[Y^2]$ depends on X.

A quick calculation shows that

$$\begin{aligned}\mathrm{E}[Y] &= \mathrm{E}\left[\frac{1}{\sigma^2}(X-\mu)^2\right] = \frac{1}{\sigma^2}\,\mathrm{E}\big[(X-\mu)^2\big] \\ &= \frac{1}{\sigma^2}\,\mathrm{Var}[X] \\ &= 1.\end{aligned}$$

Ordinary and Central Moments

So far we have encountered the expectation, $\mathrm{E}[X]$, and the variance, $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$. The information contained in these two quantities is basically that of $\mathrm{E}[X]$ and $\mathrm{E}[X^2]$.

More generally, given a random variable X, the quantities

$$\mathrm{E}[X^n], \qquad\qquad n \in \mathbb{N},$$

are known as the ***n^{th} (ordinary) moments*** of X.

The quantities

$$\mathrm{E}\Big[\Big(\frac{X-\mu}{\sigma}\Big)^n\Big], \qquad\qquad n = 3, 4, 5, \dots,$$

are called the ***n^{th} central moments*** of X.

(Of course, not all moments may exist for a given random variable!)

The Moment-Generating Function

4.6. Definition. Let (X, f_X) be a random variable and such that the sequence of moments $\mathrm{E}[X^n]$, $n \in \mathbb{N}$, exists.

If the power series

$$m_X(t) := \sum_{k=0}^{\infty} \frac{\mathrm{E}[X^k]}{k!} t^k$$

has radius of convergence $\varepsilon > 0$, the thereby defined function

$$m_X(t) \colon (-\varepsilon, \varepsilon) \to \mathbb{R}$$

is called the ***moment-generating function*** for X.

The Moment-Generating Function

4.7. Theorem. Let $\varepsilon > 0$ be given such that $\mathrm{E}[e^{tX}]$ exists and has a power series expansion in t that converges for $|t| < \varepsilon$. Then the moment-generating function exists and

$$m_X(t) = \mathrm{E}[e^{tX}] \qquad\qquad \text{for } |t| < \varepsilon.$$

Furthermore,

$$\mathrm{E}[X^k] = \frac{d^k m_X(t)}{dt^k}\bigg|_{t=0}.$$

We can hence calculate the moments of X by differentiating the moment-generating function.

The Moment-Generating Function

The basic idea behind the theorem is to write

$$m_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \operatorname{E}[X^n] = \operatorname{E}\Big[\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\Big] = \operatorname{E}[e^{tX}].$$

The interchange of the infinite series and the expectation is based on property (iii) of Theorem 4.4; however the fact that the series is infinite makes a rigorous justification a little difficult and we omit it here.

Differentiating term-by-term,

$$\frac{d^k m_X(t)}{dt^k} = \sum_{n=0}^{\infty} \frac{d^k}{dt^k} \frac{t^n}{n!} \operatorname{E}[X^n] = \sum_{n=k}^{\infty} \frac{t^{n-k}}{(n-k)!} \operatorname{E}[X^n].$$

At $t = 0$, only the first term survives, so $\frac{d^k m_X(t)}{dt^k}\Big|_{t=0} = \operatorname{E}[X^k]$.

M.G.F. for the Geometric Distribution

It turns out that the moment-generating function is uniquely associated to a given distribution: two random variables will have the same m.g.f. if and only if they have the same probability density function.
We now apply the previous discussion to the geometric distribution:

4.8. Proposition. Let (X, f_X) be a geometrically distributed random variable with parameter p. Then the moment-generating function for X is given by

$$m_X \colon (-\infty, -\ln q) \to \mathbb{R}, \qquad\qquad m_X(t) = \frac{pe^t}{1 - qe^t}$$

where $q = 1 - p$.

M.G.F. for the Geometric Distribution

Proof.

We have $f_X(x) = q^{x-1}p$ for $x \in \mathbb{N} \setminus \{0\}$. Then

$$m_X(t) = \mathsf{E}[e^{tX}] = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p = \frac{p}{q} \sum_{x=1}^{\infty} (qe^t)^x$$

This is a geometric series which converges for $|qe^t| = qe^t < 1$, i.e., for $t < -\ln q$. For such t, the limit is given by

$$\begin{aligned} m_X(t) &= \frac{p}{q} \sum_{x=1}^{\infty} (qe^t)^x = \frac{p}{q} \Big(\sum_{x=0}^{\infty} (qe^t)^x - 1 \Big) = \frac{p}{q} \Big(\frac{1}{1-qe^t} - 1 \Big) \\ &= \frac{p}{q} \frac{qe^t}{1-qe^t} = \frac{pe^t}{1-qe^t}. \end{aligned}$$

$\square$

Expectation and Variance for the Geometric Distribution

4.9. Lemma. Let (X, f_X) be a geometrically distributed random variable with parameter p. Then the expectation value and variance are given by

$$\mathrm{E}[X] = \frac{1}{p} \qquad \text{and} \qquad \mathrm{Var}[X] = \frac{q}{p^2}$$

where $q = 1 - p$.

Proof.

We use the moment-generating function to calculate the expectation value:

$$\begin{aligned} E[X] &= \frac{d}{dt}\bigg|_{t=0} m_X(t) = \frac{d}{dt}\bigg|_{t=0} \frac{p}{e^{-t} - q} \\ &= \frac{pe^t(1 - qe^t) + pq}{(e^{-t} - q)^2}\bigg|_{t=0} = \frac{p}{(1-q)^2} = \frac{1}{p}. \end{aligned}$$

The proof for the variance is similar and is left to the reader. $\square$

Expectation and Variance for the Binomial Distribution

4.10. Theorem. Let (X, f_X) be a binomial random variable with parameters n and p.

(i) The moment generating function of X is given by

$$m_X : \mathbb{R} \to \mathbb{R}, \qquad m_X(t) = (q + pe^t)^n, \qquad q = 1 - p.$$

(ii) $\mathrm{E}[X] = np$.

(iii) $\mathrm{Var}[X] = npq$.

The proof of this theorem is left as an exception.

The Mathematica commands for the expectation and variance are:

```
Mean[BinomialDistribution[n, p]]

n p

Variance[BinomialDistribution[n, p]]

n (1 - p) p
```