The Expectation and Variance of Linear Functions of Random Variables

April 13, 2020

1 Linear functions of Random Variables

In this section we consider n random variables $Y_1, Y_2, \ldots, Y_n$, and compute the expectation of linear functions of them. In general, to compute the expectation of a function of $Y_1, Y_2, \ldots, Y_n$ we would consider the jpdf (or jpmf) $f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n)$ (or $p_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n)$) and then integrate (or sum) the function of the random variables times the jpdf (or jpmf). This can be quite complicated, fortunately when we consider a linear function it simplifies quite a bit.

Theorem 1.1. *Let $Y_1, Y_2, \ldots, Y_n$ be random variables and $a_1, a_2, \ldots, a_n$ be real number, then*

$$\mathbb{E}[a_1Y_1 + a_2Y_2 + \ldots + a_nY_n] = a_1\mathbb{E}[Y_1] + a_2\mathbb{E}[Y_2] + \ldots + a_n\mathbb{E}[Y_n]$$

or more compactly written:

$$\mathbb{E}[\sum_{i=1}^{n} a_iY_i] = \sum_{i=1}^{n} a_i\mathbb{E}[Y_i]$$

Each the quantities $\mathbb{E}[Y_i]$ can be computed from just the marginal distributions, instead of the entire joint distribution, which is generally much easier.

Proof. We'll consider the proof just in the case $n = 2$, but the general case is not harder, just more to write. We'll also do the continuous case, but the discrete case follows exactly the same, if integrals are replaced by sums and pdfs by pmfs.

$$\begin{aligned}
\mathbb{E}[a_1Y_1 + a_2Y_2] &= \iint (a_1y_1 + a_2y_2) f_{Y_1,Y_2}(y_1, y_2)\ dy_1dy_2 \\
&= \iint a_1y_1 f_{Y_1,Y_2}(y_1, y_2)\ dy_1dy_2 + \iint a_2y_2 f_{Y_1,Y_2}(y_1, y_2)\ dy_1dy_2 \\
&= a_1 \iint y_1 f_{Y_1,Y_2}(y_1, y_2)\ dy_1dy_2 + a_2 \iint y_2 f_{Y_1,Y_2}(y_1, y_2)\ dy_1dy_2 \\
&= a_1 \int y_1 \left(\int f_{Y_1,Y_2}(y_1, y_2)\ dy_2 \right) dy_1 + a_2 \int y_2 \left(\int f_{Y_1,Y_2}(y_1, y_2)\ dy_1 \right) dy_2 \\
&= a_1 \int y_1 f_{Y_1}(y_1) dy_1 + a_2 \int y_2 f_{Y_2}(y_2) dy_2 \\
&= a_1\mathbb{E}[Y_1] + a_2\mathbb{E}[Y_2]
\end{aligned}$$

The first line is the definition of the expectation. The next two lines are properties of the integral. In the next line we write the double integral as at iterated integral. Then we integrate out one of the variables to get the marginal distribution. Then we're left with the the expectation of the individual random variables.

□

This theorem holds true no matter how related the random variables are. We have a similar, but a bit more complicated, formula for the variance of the sum.

THEOREM 1.2. *Let $Y_1,Y_2,...,Y_n$ be random variables and $a_1,a_2,...,a_n$ be real number, then*

$$\text{Var}[\sum_{i=1}^{n} a_i Y_i] = \sum_{i=1}^{n} a_i^2 \text{Var}[Y_i] + \sum_{i \neq j} a_i a_j \text{Cov}(Y_i,Y_j) = \sum_{i=1}^{n} a_i^2 \text{Var}[Y_i] + 2\sum_{i<j} a_i a_j \text{Cov}(Y_i,Y_j)$$

The finally equality is using that $\text{Cov}(Y_i,Y_j) = \text{Cov}(Y_j,Y_i)$.

Let's write this out more explicitly when $n=2$,

$$\begin{aligned}\text{Var}(a_1Y_1+a_2Y_2) &= a_1^2\text{Var}(Y_1)+a_2^2\text{Var}(Y_2)+a_1a_2\text{Cov}(Y_1,Y_2)+a_2a_1\text{Cov}(Y_2,Y_1)\\ &= a_1^2\text{Var}(Y_1)+a_2^2\text{Var}(Y_2)+2a_1a_2\text{Cov}(Y_1,Y_2)\end{aligned}$$

So to compute the variance of a sum, we must also consider the covariance between the random variables.

Proof. Once again we'll write out the $n=2$ case, because it's less terms to write, but doesn't change much. This theorem actually follows from the previous theorem on expectations, once we write it out the write way, so we can consider the continuous and discrete case at the same time.

We begin with the definition of the variance, we don't use the expanded form because we want to keep the random variable and it's expectation together.

$$\begin{aligned}\text{Var}(a_1Y_1+a_2Y_2) &= \mathbb{E}[\Big(a_1Y_1+a_2Y_2-\mathbb{E}[a_1Y_1+a_2Y_2]\Big)^2]\\ &= \mathbb{E}[\Big(a_1(Y_1-\mathbb{E}[Y_1])+a_2(Y_2-\mathbb{E}[Y_2])\Big)^2]\end{aligned}$$

Then we expand the square

$$\mathbb{E}[\Big(a_1(Y_1-\mathbb{E}[Y_1])\Big)^2+\Big(a_2(Y_2-\mathbb{E}[Y_2])\Big)^2+2\Big(a_1(Y_1-\mathbb{E}[Y_1])\Big)\Big(a_2(Y_2-\mathbb{E}[Y_2])\Big)]$$

Then writing the expectation of the sum as the sum of expectations and pulling the coefficient a_i out we have

$$a_1^2\text{Var}(Y_1)+a_2^2\text{Var}(Y_2)+2a_1a_2\text{Cov}(Y_1,Y_j)$$

As desired.

□

Beyond the variance of a linear function of random variable we can also consider the covariance of two functions of two different sets of random variables

THEOREM 1.3. *Let $Y_1,Y_2,...,Y_n$ and $X_1,X_2,...,X_m$ be random variables and $a_1,a_2,...,a_n$ and $b_1,b_2,...,b_m$ be real numbers. Then*

$$\text{Cov}(\sum_{i=1}^{n} a_i Y_i, \sum_{j=1}^{m} b_j X_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i,X_j)$$

The proof is basically the same as the previous theorem, instead of considering $\mathbb{E}[\left(a_1Y_1+a_2Y_2-\mathbb{E}[a_1Y_1+a_2Y_2]\right)^2]$ we start with $\mathbb{E}[\left(\sum_i a_iY_i-\mathbb{E}[a_iY_i]\right)\left(\sum_j b_jX_j-\mathbb{E}[b_jX_j]\right)]$ and then expand and rearrange in a similar fashion.

In the book Examples 5.25 and 5.26, shows some straight forward applications of these theorems in the case of 2 random variables. So let's look at some applications of these theorems, that make some of the random variables we've considered much easier to understand. We'll first consider an important special case.

2 Applications of the theorem

It turns out that many times we can represent a complicated random variable as a sum of simpler random variables. We can than compute the expectation of the complicated random variable by computing the expectation of the simpler random variables and summing them up. We can also compute the variance of the complicated random variable by computing both the variance of the simpler random variables and their covariances. In the next example, we'll consider a general case of independent random variables. Then in the next examples we'll give explicit examples, where a random variable we've seen before can be represented as a sum of simpler independent random variables. After than we'll do some similar exams where the simpler random variables we're summing are no longer independent.

EXAMPLE 2.1. *Let $X_1,...,X_n$ be independent, identically distribution (iid) random variables. (This means the rv are independent and their pdf/pmf are all the same). Let $\mathbb{E}[X_1]=\mu$ and $\mathrm{Var}(X_i)=\sigma^2$.*

(a) Compute $\mathbb{E}[\sum_i^n X_i]$.

(b) Compute $\mathrm{Var}[\sum_i^n X_i]$.

Solution:

(a)

$$\mathbb{E}[\sum_{i=1}^n X_i]=\sum_{i=1}^n \mathbb{E}[X_i]=\sum_{i=1}^n \mu=n\mu$$

(b)

$$\mathrm{Var}[\sum_{i=1}^n X_i]=\sum_{i=1}^n \mathrm{Var}[X_i]+\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)=\sum_{i=1}^n \sigma^2+0=n\sigma^2$$

Where we used that the covariance of independent random variables is 0.

EXAMPLE 2.2. *Let Y be a Binomial(n,p) random variable. Write Y as a sum of independent random variables, and then compute the expectation and variance of Y.*

Solution:

Consider n independent Bernoulli trials, with probability of success p. Recall this means each trial can be a success or failure, and the probability of a success is p. On each of the trials we define a new random variable as follows: if the i^{th} trial is a success then let $X_i=1$, otherwise it is 0 (Recall these are called Bernoulli random variables).

The main trick of this section is that we use the simple random variables X_i (they are simple because they only take two values, 0 or 1) to represent the somewhat complicated Binomial random variable (recall computing it's expectation was tricky, and it's variance was so long to compute that I just skipped it in class).

Indeed, letting $Y=\sum_{i=1}^{n} X_i$ be the total number of successes, this is true because we're just adding up a 1 for each success and a 0 for each failure, we see Y is $Binomial(n,p)$ as desired. In fact, every Binomial random variable can be thought as just adding up n independent random variables that are 0 or 1.

Then we can use the previous example to compute the expectation and variance of Y, once we know the expectation and variance of X_1. This is an easy computation because X_1 is such a simple random variable.

$$\mathbb{E}[X_1]=1*p+0*(1-p)=p$$

$$\mathrm{Var}(X_1)=\mathbb{E}[X_1^2]-\mathbb{E}[X_1]^2=\left(1^2*p+0^2*(1-p)\right)-p^2=p(1-p)$$

So we see

$$\mathbb{E}[Y]=\mathbb{E}[\sum_{i=1}^{n} X_i]=\sum_{i=1}^{n}\mathbb{E}[X_i]=\sum_{i=1}^{n} p=np$$

$$\mathrm{Var}[Y]=\sum_{i=1}^{n}\mathrm{Var}[X_i]+\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)=\sum_{i=1}^{n}\sigma^2+0=n\sigma^2=np(1-p).$$

We already knew this, but this is a much simpler computation. As a reminder the complicated way to do these computations would be:

$$\mathbb{E}[Y]=\sum_{k=0}^{n} k\frac{k!}{n!(n-k)!}p^k(1-p)^{n-k}$$

$$\mathrm{Var}[Y]=\sum_{k=0}^{n} k^2\frac{k!}{n!(n-k)!}p^k(1-p)^{n-k}-\left(\sum_{k=0}^{n} k\frac{k!}{n!(n-k)!}p^k(1-p)^{n-k}\right)^2$$

Which are not so easy to simplify.

EXAMPLE 2.3. *Consider a sequence of independent Bernoulli random trails, each with probability of success p. Let $k>1$ be an integer and let Y be trial on which the k^{th} success occurs. (We briefly saw that Y is called a negative binomial random variable, this is a generalization of the geometric random variable).*

Compute the expectation and variance of Y.

Solution:

We already know that the distribution of the first success is a geometric random variable. The time of the second success can be represented as the sum of two independent geometric random variables, the first geometric r.v. representing the time until the first success and then the time between first success to the second success is just the second geometric r.v. We can continue this by noting the time between any two successes is a geometric random variable.

So let $X_1, X_2, ..., X_k$ be independent geometric(p) random variables, then $Y = \sum_{i=1}^{k} X_i$, gives the distribution of the time of the k^{th} success.

Once again, because the trials are independent the time between the first success and the second success is a new geometric random variable, independent of X_1, the location of the first success. The same follows for the time between the j^{th} success and the $(j+1)^{th}$ success for any $1 \leq j \leq (k-1)$.

With this observation, and recalling $\mathbb{E}[X_1] = 1/p$ and $\text{Var}[X_1] = (1-p)/p^2$ we have

$$\mathbb{E}[Y] = \frac{k}{p}, \ \text{Var}[Y] = \frac{k(1-p)}{p^2}$$

We now consider an example where the random variables that we add are not independent.

EXAMPLE 2.4. *Consider an urn with N balls, r of which are red. Pick n balls without replacement and let Y be the number of red balls picked.*

Write Y as a sum of random variables, and then compute the expectation and variance of Y. (Recall Y is a Hypergeometric random variable).

Solution: For $i = 1, 2, ..., n$, let X_i be 1 if the i^{th} ball is red and 0 otherwise. Then $Y = \sum_{i=1}^{n} X_i$. This is similar to the Binomial case, if X_i is 1 it means a red ball was picked on this trial, so we're adding up a 1 for each time we pulled a red ball and 0 for each time we didn't. When we do this we've added up all the red balls picked.

So to compute the expectation and variance of Y we need to compute the expectation and variance of X_1, as well as the covariance of X_1 and X_2.

Note the 1 and 2, could be any two indices between 1 and n, the X_i's all have the same distribution. This fact is not entirely obvious, but follows from the symmetry of the problem, if I had chosen a different labeling, for example X_1 is based on the fifth ball, X_2 is based on the first ball, and so on, then the distribution of the X_i's wouldn't change. Since they all have the same distribution (pmf) they also have the same expectation, variance and covariances.

$$E[X_i] = \frac{r}{N}$$

because there are r red balls and N balls in total. So that $E[X_1] = \frac{r}{N}$, should be clear, but all the X_i's have the same expectation.

$$Var[X_i] = \frac{r}{N}(1 - \frac{r}{N})$$

This computation follows exactly as in the Binomial case, using that $\mathbb{P}(X_i = 1) = r/N$ and $\mathbb{P}(X_i = 0) = 1 - r/N$.

The new part comes when we compute the covariance, as the random variables are no longer independent.

Since $\text{Cov}(X_1,X_2)=\mathbb{E}[X_1X_2]-\mathbb{E}[X_1]\mathbb{E}[X_2]$ and we just did $\mathbb{E}[X_1],\mathbb{E}[X_2]$ we need to compute the first term.

The new random variable X_1X_2 will be 1 if both X_1 and X_2 are 1 and 0 otherwise. So to compute its expectation we need the probability it is 1.

$$\mathbb{P}(X_1=1,X_2=1)=\mathbb{P}(X_2=1|X_1=1)\mathbb{P}(X_1=1)=\frac{r-1}{N-1}\frac{r}{N}$$

So

$$\text{Cov}(X_1,X_2)=\frac{r-1}{N-1}\frac{r}{N}-\frac{r}{N}\frac{r}{N}=\frac{r}{N}\left(\frac{r-1}{N-1}-\frac{r}{N}\right)=\frac{r}{N}\left(\frac{-N+r}{N(N-1)}\right)=-\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{1}{N-1}$$

Now we put this all together:

$$\mathbb{E}[Y]=\sum_{i=1}^{n}\mathbb{E}[X_i]=\frac{nr}{N}$$

$$\begin{aligned}\text{Var}(Y)&=\sum_{i=1}^{n}\text{Var}(X_i)+2\sum_{i<j}^{n}\text{Cov}(X_i,X_j)\\&=n\frac{r}{N}(1-\frac{r}{N})-(n)(n-1)\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{1}{N-1}\\&=n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)\end{aligned}$$

The last line follows from having a simplifying the expression in the second line. The second line uses that the sum $\sum_{i<j}^{n}$ consists of $\frac{n(n-1)}{2}=\binom{n}{2}$ terms that are all the same. There are $n(n-1)$ ways to pick two distinct numbers from 1 to n, then you divide by two to make sure the first number is bigger.

Obviously, this was a fairly involved calculation, but to try to compute the variance directly form the pmf, was basically not tractable.

The idea to compute the expectation and variance of the hypergeometric random variable can be used in many different problems.

EXAMPLE 2.5. *There are N people, labeled $1,2,\ldots,N$, each with a snow ball, each person chooses a different person uniformly at random and throws a snowball at them, independently. Let Y be the number of people who don't get hit by at least one snowball. Let Y_i be 1 on the event that the i^{th} person is not hit by a snowball and 0 otherwise. Note that the Y_i's are not independent, but they do have the same distribution (pmf).*

Note that $Y=\sum_{i=N}Y_i$.

(a) Compute $\mathbb{E}[Y_i]$.

(b) *Compute* $\mathrm{Var}[Y_i]$.

(c) *Compute* $\mathbb{E}[Y_iY_j]$.

(d) *Compute* $\mathbb{E}[Y]$.

(e) *Compute* $\mathrm{Var}[Y]$.

Solution:

Since the Y_i's are random variables that only equal 0 or 1, $\mathbb{E}[Y_i]=\mathbb{P}(\{Y_i=1\})$. Since the Y_i's all have the same distribution we can just compute $\mathbb{E}[Y_1]$.

(a) Let E_i for $i=2,...,N$ be the event that the i^{th} person doesn't throw their snowball at person 1. Note that in order for Y_1 to equal 1, no one can throw their snowball at person 1. So $\{Y_i=1\}=\cap_{i=2}^N E_i$

$$\mathbb{E}[Y_i]=\mathbb{P}(\{Y_i=1\})=\mathbb{P}(\cap_{i=2}^N E_i)=\mathbb{P}(E_2)\mathbb{P}(E_3)\cdot\ldots\cdot\mathbb{P}(E_N)=\left(\frac{N-2}{N-1}\right)^{N-1}$$

Where the last equality uses the independence of the person each person chooses to throw at.

(b) Since Y_i can only equal 0 or 1, $\mathbb{E}[Y_i]=\mathbb{E}[Y_i^2]$, so

$$\mathrm{Var}[Y_i]=\mathbb{E}[Y_i^2]-\mathbb{E}[Y_i]^2=\left(\frac{N-2}{N-1}\right)^{N-1}-\left(\frac{N-2}{N-1}\right)^{2(N-1)}$$

(c) The random variable Y_iY_j is one if both the i^{th} and j^{th} person are not hit. Once again because the distribution are the same it suffices to compute Y_1Y_2.

$$\mathbb{E}[Y_iY_j]=\mathbb{P}(\{Y_1Y_2=1\})=\mathbb{P}(\{Y_1=1\}\cap\{Y_2=1\})=\mathbb{P}(\{Y_1=1\}|\{Y_2=1\})\mathbb{P}(\{Y_2=1\})$$

From part a) $\mathbb{P}(\{Y_2=1\})=\left(\frac{N-2}{N-1}\right)^{N-1}$. In a similar manner we have

$$\mathbb{P}(\{Y_1=1\}|\{Y_2=1\})=\mathbb{P}(\cap_{i=2}^N E_i|\{Y_2=1\})=\mathbb{P}(E_2|\{Y_2=1\})\mathbb{P}(E_3|\{Y_2=1\})\cdot\ldots\cdot\mathbb{P}(E_N|\{Y_2=1\})$$
$$=\left(\frac{N-2}{N-1}\right)\left(\frac{N-3}{N-2}\right)^{N-2}$$

where once again we use independence. The conditioning doesn't effect E_2 because they weren't going to throw it at themselves anyway this gives the $\left(\frac{N-2}{N-1}\right)$ term, but it does effect the remaining $N-2$ people, the conditioning means they only have $N-2$ to throw at.

So we have

$$\mathbb{E}[Y_iY_j]=\left(\frac{N-2}{N-1}\right)\left(\frac{N-3}{N-2}\right)^{N-2}\left(\frac{N-2}{N-1}\right)^{N-1}$$

(d) Now we have

$$\mathbb{E}[Y]=\sum_i^N\mathbb{E}[Y_i]=N\left(\frac{N-2}{N-1}\right)^{N-1}$$

(e)

$$\text{Var}[Y] = \sum_{i=1}^{N} \text{Var}[Y_i] + 2\sum_{i<j} \text{Cov}(Y_i, Y_j) = N\text{Var}[Y_1] + N(N-1)\text{Cov}(Y_1, Y_2)$$

$$= N\left(\left(\frac{N-2}{N-1}\right)^{N-1} - \left(\frac{N-2}{N-1}\right)^{2(N-1)}\right) + N(N-1)\left(\left(\frac{N-2}{N-1}\right)\left(\frac{N-3}{N-2}\right)^{N-2}\left(\frac{N-2}{N-1}\right)^{N-1} - \left(\frac{N-2}{N-1}\right)^{2(N-1)}\right)$$

where the last line uses that $\text{Cov}(Y_1, Y_2) = \mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2]$. The final equality on the first line uses that the sum $\sum_{i<j}^{n}$ consists of $\frac{n(n-1)}{2} = \binom{n}{2}$ terms that are all the same. Just like in the last problem.