1 The Expectation of a Binomial R.V.

The ***random variable X*** follows a binomial probability distribution with ***parameters*** n and p. Prove that:

$$E[X] = np \tag{1}$$

There are different ways of proving equation (1), and we are going to introduce a few here.

1.1 Proof using the definition

Proof. This is the proof we had during the lecture. According to the definition of *expectation*, we have

$$\begin{aligned} E[\boldsymbol{X}] &= \sum_{k=0}^{n} k p_{\boldsymbol{X}}(k) \\ &= 0 \cdot \binom{n}{0} p^0 (1-p)^n + 1 \cdot \binom{n}{1} p^1 (1-p)^{n-1} + \cdots \\ &+ n \cdot \binom{n}{n} p^n (1-p)^0 \end{aligned} \tag{2}$$

Note the first term $0 \cdot \binom{n}{0} p^0 (1-p)^n = 0$. Therefore, equation (2) becomes:

$$\begin{aligned} E[\boldsymbol{X}] &= \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \cdot p^k (1-p)^{n-k} \end{aligned} \tag{3}$$

Now, let $a = k - 1$, since $k = 1, 2, 3, ..., n$. Then $a = 0, 1, 2, ..., n - 1$.

Therefore, equation (3) becomes:

$$
\begin{aligned}
E[\boldsymbol{X}] &= \sum_{a=0}^{n-1} \frac{n!}{a!(n-a-1)!} \cdot p^{a+1} \cdot (1-p)^{n-a-1} \\
&= \sum_{a=0}^{n-1} \frac{n \cdot (n-1)!}{a!(n-1-a)!} \cdot p \cdot p^{a} \cdot (1-p)^{n-1-a} \qquad (4)
\end{aligned}
$$

Both n and p are constant, so we can take them out:

$$
E[\boldsymbol{X}] = np \sum_{a=0}^{n-1} \frac{(n-1)!}{a!(n-1-a)!} \cdot p^{a} \cdot (1-p)^{n-1-a} \qquad (5)
$$

Now we let $b = n-1$, then equation (5) becomes:

$$
\begin{aligned}
E[\boldsymbol{X}] &= np \sum_{a=0}^{b} \frac{b!}{(b-a)!} p^{a} (1-p)^{b-a} \\
&= np
\end{aligned}
$$

□

1.2 Proof using linearity of expectation

1.2.1 Joint Probability Mass Function

Before we begin, we need to look at ***joint probability mass functions*** (joint PMF). A joint PMF is just like the PMF we learned during the lecture, but we are looking at more than one discrete random variable here. This is very useful, because in real life we are often interested in more than two random variables at the same time. For example, we often want to measure the height and the weight of the same person, the expressions of many genes from the same cell *etc.* Don't worry if you find it difficult at this stage, since we will talk about this again in the later section of the course.

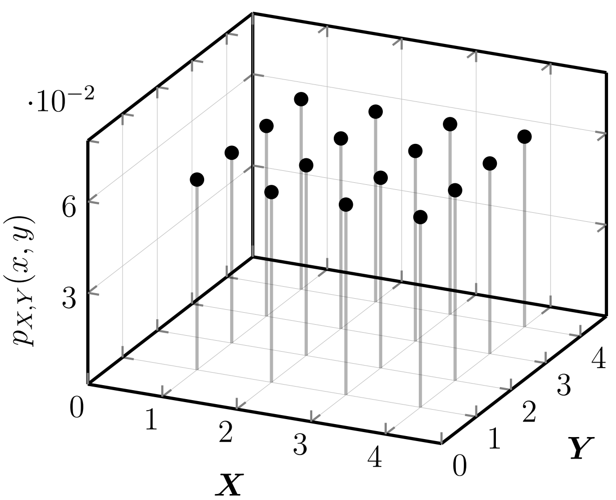
The simplest example of a joint PMF will be the one with only two discrete random variables. Let's use the example of rolling two fair tetrahedral dice, one blue and one red, for the demonstration. We could ask the question like this: what is the probability of having the outcome that the blue die is 1 *AND* the red die is 4?

If we let $\boldsymbol{X}$ represent the result of the blue die, and $\boldsymbol{Y}$ the red. The question becomes: what is the probabilty of $X = 1$ ***AND*** $Y = 4$? This can be simply denoted as $P(\boldsymbol{X} = 1, \boldsymbol{Y} = 4)$. Since both $\boldsymbol{X}$ and $\boldsymbol{Y}$ need to be considered, we actually need the joint PMF of $\boldsymbol{X}$ and $\boldsymbol{Y}$ to solve this.

Let's start with the sample space. The sample space of the die rolls will be a set of (x, y) tuples, because we need to consider the values of both $\boldsymbol{X}$ and $\boldsymbol{Y}$ at the same time. The sample space is:

$$\begin{aligned}\Omega = \{&(1,1),(1,2),(1,3),(1,4)\\&(2,1),(2,2),(2,3),(2,4)\\&(3,1),(3,2),(3,3),(3,4)\\&(4,1),(4,2),(4,3),(4,4)\}\end{aligned}$$

If you want to plot the joint PMF, you can add a Z-axis to present the probability of each tuple, like shown on the right-hand side. Since all outcomes are equally likely in the sample space, we could use the ***discrete uniform law*** to assign the probability of each outcome as $\frac{1}{16}$. Therefore, the answer is $P(\boldsymbol{X} = 1, \boldsymbol{Y} = 4) = \frac{1}{16}$. I think you get the idea that the joint PMF is defined as:



$$p_{\boldsymbol{X},\boldsymbol{Y}}(x,y) = P(\boldsymbol{X} = x, \boldsymbol{Y} = y)$$

We could also ask: what is $P(X = 1)$? In this situation, we only care about $\boldsymbol{X}$. We can simply add up the probabilities of all outcomes that satisfy $X = 1$. The outcomes that satisfy $X = 1$ are $\{(1, 1), (1, 2), (1, 3), (1, 4)\}$. Therefore:

$$\begin{aligned} p_{\boldsymbol{X},\boldsymbol{Y}}(1, y) =& P(\boldsymbol{X} = 1, \boldsymbol{Y} = 1)+ \\ & P(\boldsymbol{X} = 1, \boldsymbol{Y} = 2)+ \\ & P(\boldsymbol{X} = 1, \boldsymbol{Y} = 3)+ \\ & P(\boldsymbol{X} = 1, \boldsymbol{Y} = 4) = \frac{4}{16} = \frac{1}{4} \end{aligned}$$

From the above example, we could see that, for any particular value $X = k$, we have:

$$\begin{aligned} P(\boldsymbol{X} = k) &= \sum_y p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{X} = k, \boldsymbol{Y} = y) \\ &= \sum_y P(\boldsymbol{X} = k, \boldsymbol{Y} = y) \end{aligned} \tag{6}$$

In the same way, we have:

$$\begin{aligned} P(\boldsymbol{Y} = k) &= \sum_x p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{X} = x, \boldsymbol{Y} = k) \\ &= \sum_x P(\boldsymbol{X} = x, \boldsymbol{Y} = k) \end{aligned} \tag{7}$$

1.2.2 Linearity of Expectation

Linearity of expectation is the property that the expectation of the sum of a sequence of random variables is equal to the sum of expectations of each random variable. As usual, always start with the simplest example when we are facing a new problem. In this case, the simplest case is two random variables. We basically want to prove $E[\boldsymbol{X} + \boldsymbol{Y}] = E[\boldsymbol{X}] + E[\boldsymbol{Y}]$.

Proof. According to the definition of the expectation:

$$
\begin{aligned}
E[\boldsymbol{X}+\boldsymbol{Y}] &= \sum_x \sum_y [(x+y)\cdot P(\boldsymbol{X}=x,\boldsymbol{Y}=y)] \\
&= \sum_x \sum_y [x\cdot P(\boldsymbol{X}=x,\boldsymbol{Y}=y)] + \sum_x \sum_y [y\cdot P(\boldsymbol{X}=x,\boldsymbol{Y}=y)] \\
&= \sum_x x \sum_y P(\boldsymbol{X}=x,\boldsymbol{Y}=y) + \sum_y y \sum_x P(\boldsymbol{X}=x,\boldsymbol{Y}=y) \qquad (8)
\end{aligned}
$$

According to equations (6) and (7), we have:

$$
\begin{aligned}
E[\boldsymbol{X}+\boldsymbol{Y}] &= \sum_x x \underbrace{\sum_y P(\boldsymbol{X}=x,\boldsymbol{Y}=y)}_{P(\boldsymbol{X}=x)} + \sum_y y \underbrace{\sum_x P(\boldsymbol{X}=x,\boldsymbol{Y}=y)}_{P(\boldsymbol{Y}=y)} \\
&= \sum_x x\cdot P(\boldsymbol{X}=x) + \sum_y y\cdot P(\boldsymbol{Y}=y) \\
&= E[\boldsymbol{X}] + E[\boldsymbol{Y}] \qquad (9)
\end{aligned}
$$

Based on the same principle, equation (9) can be extended to a more general case:

$$
E\left[\sum_{i=1}^{n} c_i\boldsymbol{X}_i\right] = \sum_{i=1}^{n} c_i E[\boldsymbol{X}_i]
$$

or

$$
E[c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2 + \cdots + c_n\boldsymbol{X}_n] = c_1E[\boldsymbol{X}_1] + c_2E[\boldsymbol{X}_2] + \cdots + c_nE[\boldsymbol{X}_n]
$$

Finally, the binomial distribution can be treated as the sum of n Bernoulli trials. Therefore, let the binomial random variable $\boldsymbol{X} = \boldsymbol{Y}_1 + \boldsymbol{Y}_2 + \cdots + \boldsymbol{Y}_n$, where $\boldsymbol{Y}_i$ are n independent Bernoulli random variables with the parameter p. We know that the expectation of a Bernoulli random variable is p. We have:

$$E[\boldsymbol{X}] = E\left[\sum_{i=1}^{n} \boldsymbol{Y}_i\right] = \sum_{i=1}^{n} E[\boldsymbol{Y}_i]$$
$$= \sum_{i=1}^{n} p = np$$

□

1.3 Proof using the expansion of the binomial coefficient

Proof. We know that:

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

Take the derivative of the above equation with respect to p:

$$n(p+q)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} p^{k-1} q^{n-k}$$

Multiply p on both sides of the above equation, we have:

$$np(p+q)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

The above equation is true for any given p and q. Let $q = 1-p$. The left-hand side become np, and the right-hand side becomes $\sum_{k=0}^{n} k\binom{n}{k} p^k (1-q)^{n-k}$, which is the definition of the expectation of a binomial random variable. Therefore, we have:

$$np = E[\boldsymbol{X}]$$

□

2 The Variance of a Binomial R.V.

For simply notation, we will let $q = 1 - p$. We can start with the definition of the ***variance***:

$$\begin{aligned} \text{var}(\boldsymbol{X}) &= E\left[(\boldsymbol{X} - E(\boldsymbol{X}))^2\right] = E\left[(\boldsymbol{X} - np)^2\right] \\ &= E[\boldsymbol{X}^2 - 2np \cdot \boldsymbol{X} + n^2p^2] \end{aligned} \tag{10}$$

According to the *linearity of exptection*, equation (10) becomes:

$$\begin{aligned} \text{var}(\boldsymbol{X}) &= E[\boldsymbol{X}^2] - E[2np \cdot \boldsymbol{X}] + E[n^2p^2] \\ &= E[\boldsymbol{X}^2] - 2np \cdot E[\boldsymbol{X}] + E[n^2p^2] \\ &= E[\boldsymbol{X}^2] - 2n^2p^2 + n^2p^2 \\ &= E[\boldsymbol{X}^2] - n^2p^2 \end{aligned} \tag{11}$$

I hope you can see that we have reached the stage where have talked about during the lecture: $E[\boldsymbol{X}] = E[\boldsymbol{X}^2] - (E[\boldsymbol{X}])^2$. We actually can start from here in the first place. Anyway, now we need to calculate $E[\boldsymbol{X}^2]$:

$$\begin{aligned} E[\boldsymbol{X}^2] &= \sum_x \boldsymbol{X}^2 \cdot p_{\boldsymbol{X}}(x) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k} = \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \end{aligned} \tag{12}$$

We can see that the extra term k is kind of annoying to have here. If we could remove it, that would be great. Now look at equation (3) when we derived the expectation:

$$E[\boldsymbol{X}] = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \cdot p^k q^{n-k} \tag{3}$$

We can do equation (12) − (3), and use a similar thought, we get:

$$
\begin{aligned}
&E[\boldsymbol{X}^2] - E[\boldsymbol{X}] \\
&\qquad = \sum_{k=1}^{n}(k-1)\,\frac{n!}{(k-1)!(n-k)!}p^k q^{n-k} \\
&\qquad = \sum_{k=2}^{n}(k-1)\,\frac{n!}{(k-1)!(n-k)!}p^k q^{n-k} \\
&\qquad = \sum_{k=2}^{n}\frac{n!}{(k-2)!(n-k)!}p^k q^{n-k} \\
&\qquad = n(n-1)p^2\sum_{k=2}^{n}\frac{(n-2)!}{(k-2)!\,[(n-2)-(k-2)]!}p^{k-2}q^{(n-2)-(k-2)} \qquad (13)
\end{aligned}
$$

Let $a = k-2$ and $b = n-2$. Since $k = 2,3,4,...,n$, then $a = 0,1,2,...,n-2$ and equation (13) becomes:

$$
\begin{aligned}
E[\boldsymbol{X}^2] - E[\boldsymbol{X}] &= n(n-1)p^2\sum_{a=0}^{b}\frac{b!}{a!(b-a)!}p^a q^{b-a} \\
&= n(n-1)p^2\binom{b}{a}\sum_{a=0}^{b}p^a q^{b-a} = n(n-1)p^2 \qquad (14)
\end{aligned}
$$

Therefore, we have:

$$
E[\boldsymbol{X}^2] = n(n-1)p^2 + E[\boldsymbol{X}] = n(n-1)p^2 + np \qquad (15)
$$

Now that we have the value of $E[\boldsymbol{X}^2]$, put equation (15) into equation (11), and finally, we have:

$$
\begin{aligned}
\mathrm{var}(\boldsymbol{X}) &= E[\boldsymbol{X}^2] - n^2p^2 = n(n-1)p^2 + np - n^2p^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 = np - np^2 \\
&= np(1-p) = npq \qquad (16)
\end{aligned}
$$

3 The Expectation of a Poisson R.V.

Before we begin, you need to know the *Taylor/Maclaurin series* of e^x:

$$e^x = \frac{x^0}{1!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{17}$$

Now we start with the definition of the expectation:

$$\begin{aligned} E[\boldsymbol{X}] &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \sum_{k=1}^{\infty} \cdot \frac{\lambda^k}{(k-1)!} \cdot e^{-\lambda} \\ &= \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \cdot \frac{\lambda^{k-1}}{(k-1)!} \end{aligned} \tag{18}$$

Let $a = k - 1$. Since $k = 1, 2, 3, ...$, then $a = 0, 1, 2, ...$ Therefore, equation (18) becomes:

$$E[\boldsymbol{X}] = \lambda \cdot e^{-\lambda} \cdot \sum_{a=0}^{\infty} \cdot \frac{\lambda^a}{a!} \tag{19}$$

Put equation (17) into equation (19), we have:

$$E[\boldsymbol{X}] = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda \tag{20}$$

4 The Variance of a Poisson R.V.

We will used a similar strategy when we derived the variance for the binomial random variable. We start with:

$$\begin{aligned} E[\boldsymbol{X}(\boldsymbol{X}-1)] &= \sum_{k=0}^{\infty} k(k-1) \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \sum_{k=2}^{\infty} k(k-1) \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \\ &= \sum_{k=2}^{\infty} \cdot \frac{\lambda^k}{(k-2)!} \cdot e^{-\lambda} = \lambda^2 \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \cdot e^{-\lambda} \end{aligned} \tag{21}$$

Let $a = k - 2$. Since $k = 2, 3, 4, ...$, then $a = 0, 1, 2, ...$ Equation (21) becomes:

$$E[\boldsymbol{X}(\boldsymbol{X}-1)] = \lambda^2 \cdot \sum_{a=0}^{\infty} \frac{\lambda^a}{a!} \cdot e^{-\lambda} \tag{22}$$

Note the 2nd term $\sum_{a=0}^{\infty} \frac{\lambda^a}{a!} \cdot e^{-\lambda}$ is the sum of the entire sample space of a Poisson distribution, so it sums up to 1. Then equation (22) becomes:

$$\begin{aligned} E[\boldsymbol{X}(\boldsymbol{X}-1)] &= \lambda^2 \\ E[\boldsymbol{X}^2 - \boldsymbol{X}] &= \lambda^2 \end{aligned} \tag{23}$$

Using linearity of expectation, equantion (23) becomes:

$$\begin{aligned} E[\boldsymbol{X}^2] - E[\boldsymbol{X}] &= \lambda^2 \\ E[\boldsymbol{X}^2] &= \lambda^2 + E[\boldsymbol{X}] = \lambda^2 + \lambda \end{aligned} \tag{24}$$

Finally, we have:

$$\begin{aligned} var(\boldsymbol{X}) &= E[\boldsymbol{X}^2] - (E[\boldsymbol{X}])^2 = \lambda^2 + \lambda - \lambda^2 \\ &= \lambda \end{aligned} \tag{25}$$