Various Proofs About Random Variables

BIO210 Biostatistics

Extra Reading Material for Lecture 10

Xi Chen
School of Life Sciences
Southern University of Science and Technology

Fall, 2025

1 About Calculating $\mathbb{E}\left[g(\boldsymbol{X})\right]$

Recall that a ***random variable*** is a ***function*** that maps an outcome from the ***sample space*** to a real-valued number. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be discrete random variables, and

$$\boldsymbol{Y} = g(\boldsymbol{X})$$

Then

$$\mathbb{E}\left[\boldsymbol{Y}\right] = \sum_{x} g(x)\mathbb{P}_{\boldsymbol{X}}(x)$$

Where $\mathbb{P}_{\boldsymbol{X}}(x)$ is the ***probability mass function*** (PMF) of the random variable $\boldsymbol{X}$.

Proof. According to the definition of ***expectation***, we have

$$\mathbb{E}\left[\boldsymbol{Y}\right] = \sum_{y} y\mathbb{P}_{\boldsymbol{Y}}(y) \tag{1}$$

Based on the relationship between the random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, we have

$$\mathbb{P}_{\boldsymbol{Y}}(y) = \sum_{x \,|\, g(x)=y} \mathbb{P}_{\boldsymbol{X}}(x) \tag{2}$$

Put the equation (2) into (1), we have

$$\begin{aligned} \mathbb{P}_{\boldsymbol{Y}}(y) &= \sum_{y} y \sum_{x \,|\, g(x)=y} \mathbb{P}_{\boldsymbol{X}}(x) \\ &= \sum_{y} \sum_{x \,|\, g(x)=y} y\mathbb{P}_{\boldsymbol{X}}(x) \\ &= \sum_{y} \sum_{x \,|\, g(x)=y} g(x)\mathbb{P}_{\boldsymbol{X}}(x) \\ &= \sum_{x} g(x)\mathbb{P}_{\boldsymbol{X}}(x) \end{aligned}$$

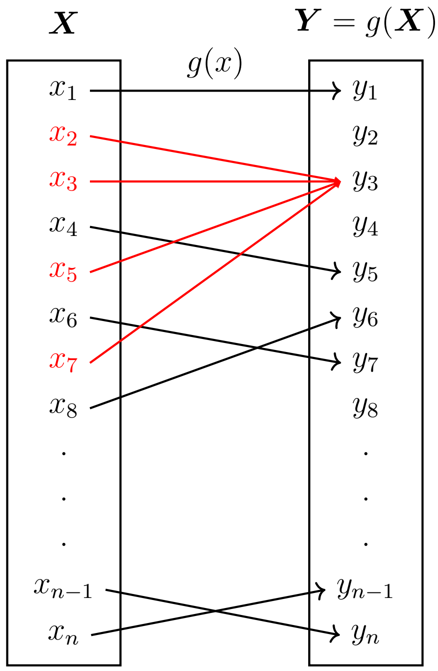
□

If you are not comfortable with the $\sum$ notation, you can look at the diagram at the right. Since

$$\mathbb{E}[\boldsymbol{Y}] = \sum_y y\mathbb{P}_{\boldsymbol{Y}}(y)$$

We need to calculate every single $y_i\mathbb{P}_{\boldsymbol{Y}}(y_i)$ in order to get $\mathbb{E}[\boldsymbol{Y}]$. Here, we use the coloured (red) example as a demonstration.

In this case, we are calculating $y_3 \cdot \mathbb{P}_{\boldsymbol{Y}}(\boldsymbol{Y} = y_3)$. Note $\mathbb{P}_{\boldsymbol{Y}}(\boldsymbol{Y} = y_3)$ is the sum of all x_i such that $g(x_i) = y_3$. In this specific example, they are $\{x_2, x_3, x_5, x_7\}$. Therefore, we have:



$$\begin{aligned} y_3 \cdot \mathbb{P}_{\boldsymbol{Y}}(\boldsymbol{Y} = y_3) &= y_3 \cdot [\mathbb{P}_{\boldsymbol{X}}(x_2) + \mathbb{P}_{\boldsymbol{X}}(x_3) + \mathbb{P}_{\boldsymbol{X}}(x_5) + \mathbb{P}_{\boldsymbol{X}}(x_7)] \\ &= y_3 \cdot \mathbb{P}_{\boldsymbol{X}}(x_2) + y_3 \cdot \mathbb{P}_{\boldsymbol{X}}(x_3) + y_3 \cdot \mathbb{P}_{\boldsymbol{X}}(x_5) + y_3 \cdot \mathbb{P}_{\boldsymbol{X}}(x_7) \end{aligned}$$

Note $y_3 = g(x_2) = g(x_3) = g(x_5) = g(x_7)$, then the above equation becomes:

$$\begin{aligned} y_3 \cdot \mathbb{P}_{\boldsymbol{Y}}(\boldsymbol{Y} = y_3) &= g(x_2)\mathbb{P}_{\boldsymbol{X}}(x_2) + g(x_3)\mathbb{P}_{\boldsymbol{X}}(x_3) + g(x_5)\mathbb{P}_{\boldsymbol{X}}(x_5) + g(x_7)\mathbb{P}_{\boldsymbol{X}}(x_7) \\ &= \sum_{i \in \{2,3,5,7\}} g(x_i)\mathbb{P}_{\boldsymbol{X}}(x_i) \end{aligned}$$

2 About $\mathbb{E}[\alpha\boldsymbol{X} + \beta]$

The linear function $\alpha\boldsymbol{X} + \beta$ scale the random variable $\boldsymbol{X}$ by a constant factor α and shift everything by a constant factor β. Therefore, we intuitively should expect that:

$$\mathbb{E}[\alpha\boldsymbol{X} + \beta] = \alpha\mathbb{E}[\boldsymbol{X}] + \beta$$

Proof. Using the property that $\mathbb{E}[g(\boldsymbol{X})] = \sum_x g(x)\mathbb{P}_{\boldsymbol{X}}(x)$, we have:

$$\begin{aligned}\mathbb{E}[\alpha\boldsymbol{X} + \beta] &= \sum_x (\alpha x + \beta)\mathbb{P}_{\boldsymbol{X}}(x) = \sum_x [\alpha x\mathbb{P}_{\boldsymbol{X}}(x) + \beta\mathbb{P}_{\boldsymbol{X}}(x)] \\ &= \sum_x \alpha x\mathbb{P}_{\boldsymbol{X}}(x) + \sum_x \beta\mathbb{P}_{\boldsymbol{X}}(x)\end{aligned}$$

Since α and β are constants, we could take them out from the summation:

$$\mathbb{E}[\alpha\boldsymbol{X} + \beta] = \alpha\sum_x x\mathbb{P}_{\boldsymbol{X}}(x) + \beta\sum_x \mathbb{P}_{\boldsymbol{X}}(x)$$

By definition, $\sum_x x\mathbb{P}_{\boldsymbol{X}}(x) = \mathbb{E}[\boldsymbol{X}]$ and $\sum_x \mathbb{P}_{\boldsymbol{X}}(x) = 1$, we have:

$$\mathbb{E}[\alpha\boldsymbol{X} + \beta] = \alpha\mathbb{E}[\boldsymbol{X}] + \beta$$

□

3 About $\mathbb{V}\mathrm{ar}(\boldsymbol{X})$

The definition is:

$$\mathbb{V}\mathrm{ar}(\boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2\right]$$

Let's look at this bit by bit. First, for any given random variable $\boldsymbol{X}$, $\mathbb{E}[\boldsymbol{X}]$ is a constant value. Then $(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2$ is a function of the random variable $\boldsymbol{X}$, so $(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2$ is also an random variable. Therefore, it is reasonable to ask: what is the expectation of $(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2$? This is basically $\mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2]$, and it represents how much $\boldsymbol{X}$ deviates from its mean. Very often, it is actually easier to calculate the variance using the following formula:

$$\mathbb{V}\mathrm{ar}(\boldsymbol{X}) = \mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}[\boldsymbol{X}])^2$$

Proof. We start from the definition, use the property of $\mathbb{E}[g(\boldsymbol{X})]$, and expand the thing inside the parentheses:

$$\begin{aligned}\mathbb{V}\text{ar}\left(\boldsymbol{X}\right) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\left[\boldsymbol{X}\right])^2\right] &= \sum_x (x - \mathbb{E}\left[\boldsymbol{X}\right])^2 \mathbb{P}_{\boldsymbol{X}}(x) \\ &= \sum_x \left[x^2 \mathbb{P}_{\boldsymbol{X}}(x) - 2x\mathbb{E}\left[\boldsymbol{X}\right] \mathbb{P}_{\boldsymbol{X}}(x) + (\mathbb{E}\left[\boldsymbol{X}\right])^2 \mathbb{P}_{\boldsymbol{X}}(x)\right] \\ &= \sum_x x^2 \mathbb{P}_{\boldsymbol{X}}(x) - \sum_x 2x\mathbb{E}\left[\boldsymbol{X}\right] \mathbb{P}_{\boldsymbol{X}}(x) + \sum_x (\mathbb{E}\left[\boldsymbol{X}\right])^2 \mathbb{P}_{\boldsymbol{X}}(x)\end{aligned}$$

Since $\mathbb{E}\left[\boldsymbol{X}\right]$ is a constant value, we can take it out of the summation:

$$\mathbb{V}\text{ar}\left(\boldsymbol{X}\right) = \sum_x x^2 \mathbb{P}_{\boldsymbol{X}}(x) - 2\mathbb{E}\left[\boldsymbol{X}\right] \sum_x x\mathbb{P}_{\boldsymbol{X}}(x) + (\mathbb{E}\left[\boldsymbol{X}\right])^2 \sum_x \mathbb{P}_{\boldsymbol{X}}(x)$$

Now, note that $\sum_x x^2 \mathbb{P}_{\boldsymbol{X}}(x) = \mathbb{E}\left[\boldsymbol{X}^2\right]$ and $\sum_x x\mathbb{P}_{\boldsymbol{X}}(x) = \mathbb{E}\left[\boldsymbol{X}\right]$ by definition, and $\sum_x \mathbb{P}_{\boldsymbol{X}}(x) = 1$. Therefore, we have:

$$\begin{aligned}\mathbb{V}\text{ar}\left(\boldsymbol{X}\right) &= \mathbb{E}\left[\boldsymbol{X}^2\right] - 2\mathbb{E}\left[\boldsymbol{X}\right] \cdot \mathbb{E}\left[\boldsymbol{X}\right] + (\mathbb{E}\left[\boldsymbol{X}\right])^2 \\ &= \mathbb{E}\left[\boldsymbol{X}^2\right] - 2(\mathbb{E}\left[\boldsymbol{X}\right])^2 + (\mathbb{E}\left[\boldsymbol{X}\right])^2 \\ &= \mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}\left[\boldsymbol{X}\right])^2\end{aligned}$$

□

4 About $\mathbb{V}\text{ar}\left(\alpha\boldsymbol{X} + \beta\right)$

Again, $\alpha\boldsymbol{X} + \beta$ means scaling the random variable $\boldsymbol{X}$ by a constant factor α and shifting everything by a constant factor β. When scaling the values, the scaling factor will be exaggerated by the square operation in the variance formula. When shifting the values, the shape of the distribution does not change, so the variance will not be affected by the shifting factor. We should not be surprised that:

$$\mathbb{V}\text{ar}\left(\alpha\boldsymbol{X} + \beta\right) = \alpha^2 \mathbb{V}\text{ar}\left(\boldsymbol{X}\right)$$

Proof. We can start with the definition:

$$\mathbb{V}\text{ar}\left(\alpha\boldsymbol{X} + \beta\right) = \mathbb{E}\left[\left(\ (\alpha\boldsymbol{X} + \beta) - \mathbb{E}\left[\alpha\boldsymbol{X} + \beta\right]\ \right)^2\right]$$

Note that $\mathbb{E}[\alpha \boldsymbol{X} + \beta] = \alpha\mathbb{E}[\boldsymbol{X}] + \beta$, so we have:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(\alpha \boldsymbol{X} + \beta) &= \mathbb{E}\left[\Big((\alpha \boldsymbol{X} + \beta) - (\alpha\mathbb{E}[\boldsymbol{X}] + \beta) \Big)^2\right] \\
&= \mathbb{E}\left[(\alpha \boldsymbol{X} + \beta - \alpha\mathbb{E}[\boldsymbol{X}] - \beta)^2\right] \\
&= \mathbb{E}\left[(\alpha \boldsymbol{X} - \alpha\mathbb{E}[\boldsymbol{X}])^2\right] \\
&= \mathbb{E}\left[\alpha^2 \boldsymbol{X}^2 - 2\alpha^2 \boldsymbol{X}\mathbb{E}[\boldsymbol{X}] + \alpha^2(\mathbb{E}[\boldsymbol{X}])^2\right] \\
&= \sum_x [\alpha^2 x^2 - 2\alpha^2 x\mathbb{E}[\boldsymbol{X}] + \alpha^2(\mathbb{E}[\boldsymbol{X}])^2]\mathbb{P}_{\boldsymbol{X}}(x) \\
&= \sum_x \alpha^2 x^2 \mathbb{P}_{\boldsymbol{X}}(x) - \sum_x 2\alpha^2 x\mathbb{E}[\boldsymbol{X}]\mathbb{P}_{\boldsymbol{X}}(x) + \sum_x \alpha^2(\mathbb{E}[\boldsymbol{X}])^2\mathbb{P}_{\boldsymbol{X}}(x) \\
&= \alpha^2 \sum_x x^2 \mathbb{P}_{\boldsymbol{X}}(x) - 2\alpha^2\mathbb{E}[\boldsymbol{X}] \sum_x x\mathbb{P}_{\boldsymbol{X}}(x) + \alpha^2(\mathbb{E}[\boldsymbol{X}])^2 \sum_x \mathbb{P}_{\boldsymbol{X}}(x)
\end{aligned}
$$

Again, we already know that $\sum_x x^2\mathbb{P}_{\boldsymbol{X}}(x) = \mathbb{E}\left[\boldsymbol{X}^2\right]$, $\sum_x x\mathbb{P}_{\boldsymbol{X}}(x) = \mathbb{E}[\boldsymbol{X}]$ and $\sum_x \mathbb{P}_{\boldsymbol{X}}(x) = 1$. Therefore:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(\alpha \boldsymbol{X} + \beta) &= \alpha^2\mathbb{E}\left[\boldsymbol{X}^2\right] - 2\alpha^2\mathbb{E}[\boldsymbol{X}] \cdot \mathbb{E}[\boldsymbol{X}] + \alpha^2(\mathbb{E}[\boldsymbol{X}])^2 \\
&= \alpha^2\mathbb{E}\left[\boldsymbol{X}^2\right] - 2\alpha^2(\mathbb{E}[\boldsymbol{X}])^2 + \alpha^2(\mathbb{E}[\boldsymbol{X}])^2 \\
&= \alpha^2\mathbb{E}\left[\boldsymbol{X}^2\right] - \alpha^2(\mathbb{E}[\boldsymbol{X}])^2 \\
&= \alpha^2(\mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}[\boldsymbol{X}])^2)
\end{aligned}
$$

Remember we just proved that $\mathbb{V}\mathrm{ar}(\boldsymbol{X}) = \mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}[\boldsymbol{X}])^2$, so we have:

$$
\mathbb{V}\mathrm{ar}(\alpha \boldsymbol{X} + \beta) = \alpha^2\mathbb{V}\mathrm{ar}(\boldsymbol{X})
$$

□

To be honest, all those proofs will be much easier if we know ***Linearity of Expectation***, which will be covered in the next lecture.