Sum of Independent Random Variables

BIO210 Biostatistics

Extra reading material for Lecture 27

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

March 2022

There are many situations that different random variables get added together. Therefore, it is important to know how to compute the probability of the sum of different random variables. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two independent random variables. Now let the random variable $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$. We want to know the probability distribution of $\boldsymbol{Z}$ in terms of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

1 The Discrete Case

As usual, we start with something simple. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are discrete random variables, the situation is straightforward. We have:

$$\begin{aligned} p_{\boldsymbol{X}}(x) &= P(\boldsymbol{X} = x) \\ p_{\boldsymbol{Y}}(y) &= P(\boldsymbol{Y} = y) \end{aligned}$$

Now we could derive the PMF of $\boldsymbol{Z}$ as follows, which involves in finding the probability for all possible values of $\boldsymbol{Z}$. Say, we want to calculate $P(\boldsymbol{Z} = 3)$, how do we do this? We need to find all possible pairs of $(\boldsymbol{X} = x, \boldsymbol{Y} = y)$ that satisfy $x + y = 3$, *e.g.* (1,2) (2,1) (-1,4) *etc.*. That is:

$$P(\boldsymbol{Z} = 3) = \sum_{\{(x,y)\,|\,x+y=3\}} P(\boldsymbol{X} = x, \boldsymbol{Y} = y)$$

Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, then $P(\boldsymbol{X} = x, \boldsymbol{Y} = y) = p_{\boldsymbol{X}}(x) \cdot p_{\boldsymbol{Y}}(y)$. Now, in a more general term, we can find the PMF of $\boldsymbol{Z}$ as follows:

$$\begin{aligned} p_{\boldsymbol{Z}}(z) &= \sum_{\{(x,y)\,|\,x+y=z\}} P(\boldsymbol{X} = x, \boldsymbol{Y} = y) = \sum_x P(\boldsymbol{X} = x, \boldsymbol{Y} = z - x) \\ &= \sum_x p_{\boldsymbol{X}}(x) p_{\boldsymbol{Y}}(z - x) \end{aligned}$$

The formula $p_{\boldsymbol{Z}}(z) = \sum_x p_{\boldsymbol{X}}(x) p_{\boldsymbol{Y}}(z - x)$ is called the **convolution** formula.

2 The Continuous Case

Now, let's look at the continuous case. In this situation, we have $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two independent continuous random variables with known PDFs. Now we want to derive the PDF of the random varible $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$. Since we already know the discrete case, we can actually guess the formula in the continuous case, which is:

$$f_{\boldsymbol{Z}}(z) = \int_{-\infty}^{+\infty} f_{\boldsymbol{X}}(x) f_{\boldsymbol{Y}}(z-x)dx$$

Now let's justify the above formula. Let's first look at $\boldsymbol{Z}$ when $\boldsymbol{X}$ takes some specific value, say 3, meaning that we are looking at $\boldsymbol{Z}$ conditioned on $\boldsymbol{X} = 3$. We have $x = 3$ and $z = y + 3$. Then we want to figure out:

$$f_{\boldsymbol{Z}\,|\,\boldsymbol{X}}(z\,|\,3) = f_{\boldsymbol{Y}+3\,|\,\boldsymbol{X}}(z\,|\,3)$$

Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, we can remove the condition:

$$f_{\boldsymbol{Z}\,|\,\boldsymbol{X}}(z\,|\,3) = f_{\boldsymbol{Y}+3\,|\,\boldsymbol{X}}(z\,|\,3) = f_{\boldsymbol{Y}+3}(z)$$

Now, $\boldsymbol{Y}+3$ is just $\boldsymbol{Y}$ with a constant added to it. Remember this is equivalent of shifting the PDF to the right by the constant. In this specific case, $\boldsymbol{Y}+3$ is just $\boldsymbol{Y}$ shifted to the right by 3. Therefore, we have:

$$f_{\boldsymbol{Z}\,|\,\boldsymbol{X}}(z\,|\,3) = f_{\boldsymbol{Y}+3\,|\,\boldsymbol{X}}(z\,|\,3) = f_{\boldsymbol{Y}+3}(z) = f_{\boldsymbol{Y}}(z-3)$$

To make this in a more general case, we have the conditional PDF:

$$f_{\boldsymbol{Z}\,|\,\boldsymbol{X}}(z\,|\,x) = f_{\boldsymbol{Y}}(z-x)$$

Therefore, the joint PDF of $\boldsymbol{Z}$ and $\boldsymbol{X}$ are:

$$f_{\boldsymbol{X},\boldsymbol{Z}}(x,z) = f_{\boldsymbol{X}}(x) \cdot f_{\boldsymbol{Z}\,|\,\boldsymbol{X}}(z\,|\,x) = f_{\boldsymbol{X}}(x) f_{\boldsymbol{Y}}(z-x)$$

Now we have the joint PDF of $\boldsymbol{X}$ and $\boldsymbol{Z}$, but remember what we really want is the PDF of $\boldsymbol{Z}$. We can easily get this by integrating all possible x from the joint PDF to get the marginal PDF of $\boldsymbol{Z}$, which is what we want originally:

$$f_{\boldsymbol{Z}}(z) = \int_{-\infty}^{+\infty} f_{\boldsymbol{X},\boldsymbol{Z}}(x,z)dx = \int_{-\infty}^{+\infty} f_{\boldsymbol{X}}(x) f_{\boldsymbol{Y}}(z-x)dx$$

Hence, we have justified our guess.

3 The Sum of Independent Normal Random Variables

Let's just start with simplest case: the sum of two normal random variables. Let $\boldsymbol{X} \sim \mathcal{N}(\mu_x, \sigma_x)$ and $\boldsymbol{Y} \sim \mathcal{N}(\mu_y, \sigma_y)$ be two independent normal random variables. We want to derive the PDF of $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$.

We know:

$$f_{\boldsymbol{X}}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \text{ and } f_{\boldsymbol{Y}}(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}$$

Now start with the PDF of $\boldsymbol{Z}$:

$$\begin{aligned}
f_{\boldsymbol{Z}(z)} &= \int_{-\infty}^{+\infty} f_{\boldsymbol{X}}(x) f_{\boldsymbol{Y}}(z-x)dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(z-x-\mu_y)^2}{2\sigma_y^2}} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{2\pi}\sigma_y} e^{-\dfrac{(x-\mu_x)^2}{2\sigma_x^2} - \dfrac{(z-x-\mu_y)^2}{2\sigma_y^2}} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_x\sigma_y} e^{-\dfrac{\sigma_y^2(x-\mu_x)^2 + \sigma_x^2(z-x-\mu_y)^2}{2\sigma_x^2\sigma_y^2}} dx
\end{aligned}$$

Now we just need to be patient and manipulate the formula. With some algebra, we can get:

$$f_{\boldsymbol{Z}}(z) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2 + \sigma_y^2}} e^{-\dfrac{[z-(\mu_x+\mu_y)]^2}{2(\sigma_x^2+\sigma_y^2)}}$$

Apparently, $\boldsymbol{Z} \sim \mathcal{N}(\mu = \mu_x + \mu_y, \sigma = \sqrt{\sigma_x^2 + \sigma_y^2})$. Check this Wikipedia page if you are interested in the algebra manipulation.