Contents

Some Useful Properties of Normal PDFs

The content of this **Extra Reading Material** is a bit long, so I put a summary at the beginning of this document:

1 Summary

- The normal PDF satisfies the normalisation axiom:

$$\int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x)\mathrm{d}x = 1$$

- The mean of a normal random variable is μ: $\mathbb{E}[X] = \mu$
- The variance of a normal random variable is σ^2: $\mathrm{var}(X) = \sigma^2$
- A function of a normal random variable is still normal:

$$X \sim \mathcal{N}(\mu, \sigma^2) \text{ and } Y = aX + b\text{, then}$$

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)\text{, where } a, b \text{ are constants}$$

- The variance of the sum of independent random variables is the sum of their variances:

$$X_1, X_2, \cdots, X_{n-1}, X_n \text{ are independent, then}$$

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i)$$

- The sum of two independent normal random variables is still normal:

$$X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)\text{, and } X, Y \text{ are independent}$$

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

2 $\int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x)\mathrm{d}x = 1$

The title means that the normal PDF is a valid probabilistic model. Therefore, we need to prove that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = 1 \tag{1}$$

Proof. The exponent of e is a bit complicated and annoying. A common trick of reduce the complexity is by ***change of variables***. Let

$$t = \frac{x-\mu}{\sqrt{2}\sigma} \tag{2}$$

and then we have

$$\begin{aligned} \mathrm{d}t &= \frac{1}{\sqrt{2}\sigma}\mathrm{d}x \\ \Rightarrow \mathrm{d}x &= \sqrt{2}\sigma\mathrm{d}t \end{aligned} \tag{3}$$

Put equations (2) and (3) into the left-hand side of equation (1), we have

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2} \sqrt{2}\sigma \,\mathrm{d}t \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2}\mathrm{d}t = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1 \end{aligned} \tag{4}$$

Note that the blue part of equation (4) is the famous ***Gaussian integral***, and its value is $\sqrt{\pi}$. □

3 $E[\boldsymbol{X}] = \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x)\mathrm{d}x = \mu$

The mean of a normal random variable is μ. By definition, we want to calculate:

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \tag{5}$$

Again, use the same ***change-of-variable*** trick and put equations (2) and (3) into equation (5), we have:

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2} \sqrt{2}\sigma \mathrm{d}t$$

$$= \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) \frac{1}{\sqrt{\pi}} e^{-t^2} \mathrm{d}t = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) e^{-t^2} \mathrm{d}t$$

$$= \frac{1}{\sqrt{\pi}} \left(\int_{-\infty}^{\infty} \sqrt{2}\sigma t e^{-t^2} \mathrm{d}t + \int_{-\infty}^{\infty} \mu e^{-t^2} \mathrm{d}t \right)$$

$$= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \int_{-\infty}^{\infty} t e^{-t^2} \mathrm{d}t + \mu \int_{-\infty}^{\infty} e^{-t^2} \mathrm{d}t \right) \tag{6}$$

Now look at the blue part of equation (6). It consists of two terms. The first term is $\sqrt{2}\sigma \int_{-\infty}^{\infty} t e^{-t^2} \mathrm{d}t$. The function te^{-t^2} is apparently an ***odd function***. If we integrate from $-\infty$ to ∞, the area under the curve below and above the axis will cancel out[1], so the the first term is 0. The second term contains a Gaussian integral $\int_{-\infty}^{\infty} e^{-t^2} \mathrm{d}t = \sqrt{\pi}$. Therefore equation (6) becomes:

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \frac{1}{\sqrt{\pi}} (0 + \mu\sqrt{\pi}) = \mu$$

Therefore, we have finished the calculation:

$$E[\boldsymbol{X}] = \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x)\mathrm{d}x = \mu$$

[1] This is not very strict. Strictly speaking, this is an ***improper integral*** what we should do here is to separate the integral: $\int_{-\infty}^{\infty} te^{-t^2} = \int_{-\infty}^{0} te^{-t^2} + \int_{0}^{\infty} te^{-t^2}$. then we show that they both converge.

4 $\mathrm{var}(\boldsymbol{X}) = \sigma^2$

Since we know that

$$E[\boldsymbol{X}] = E[\boldsymbol{X}^2] - (E[\boldsymbol{X}])^2 = E[\boldsymbol{X}^2] - \mu^2$$

What is left for us is to compute $E[\boldsymbol{X}^2]$. By definition, we have:

$$E[\boldsymbol{X}^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \tag{7}$$

Let's use the same trick again by letting $t = \frac{x-\mu}{\sqrt{2}\sigma}$. Then we have:

$$x^2 = (\sqrt{2}\sigma t + \mu)^2 = 2\sigma^2 t^2 + 2\sqrt{2}\sigma\mu t + \mu^2 \tag{8}$$

Now we put equations (3) and (8) into equation (7), we have:

$$\begin{aligned} E[\boldsymbol{X}^2] &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (2\sigma^2 t^2 + 2\sqrt{2}\sigma\mu t + \mu^2) e^{-t^2} \sqrt{2}\sigma \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}\sigma} \left(\int_{-\infty}^{\infty} 2\sqrt{2}\sigma^3 t^2 e^{-t^2} \mathrm{d}t + \int_{-\infty}^{\infty} 4\sigma^2 \mu t e^{-t^2} \mathrm{d}t + \int_{-\infty}^{\infty} \sqrt{2}\sigma\mu^2 e^{-t^2} \mathrm{d}t \right) \end{aligned} \tag{9}$$

The blue part of equation (9) consists of three terms. Let's look at them one by one in reverse order, because the last two terms are easier to compute. First, let's look at the third term, there is a Gaussian integral there:

$$\int_{-\infty}^{\infty} \sqrt{2}\sigma\mu^2 e^{-t^2} \mathrm{d}t = \sqrt{2}\sigma\mu^2 \int_{-\infty}^{\infty} e^{-t^2} \mathrm{d}t = \sqrt{2}\sigma\mu^2 \cdot \sqrt{\pi} = \sqrt{2\pi}\sigma\mu^2 \tag{10}$$

Then let's look at the second term:

$$\int_{-\infty}^{\infty} 4\sigma^2 \mu t e^{-t^2} \mathrm{d}t = 4\sigma^2\mu \int_{-\infty}^{\infty} t e^{-t^2} \mathrm{d}t = 0 \tag{11}$$

We have already seen this when we were computing the mean. The integration from $-\infty$ to ∞ is 0.

Finally, let's look at the first term:

$$\int_{-\infty}^{\infty} 2\sqrt{2}\sigma^3 t^2 e^{-t^2}\mathrm{d}t = 2\sqrt{2}\sigma^3 \int_{-\infty}^{\infty} t^2 e^{-t^2}\mathrm{d}t \tag{12}$$

The integral in equation (12) is a slightly more difficult to do. First, we notice that $t^2e^{-t^2}$ is an ***even function***, so $\int_{-\infty}^{\infty} t^2e^{-t^2}\mathrm{d}t = 2\cdot\int_0^{\infty} t^2e^{-t^2}\mathrm{d}t$, roughly[2]. Then equation (12) becomes:

$$\int_{-\infty}^{\infty} 2\sqrt{2}\sigma^3 t^2 e^{-t^2}\mathrm{d}t = 4\sqrt{2}\sigma^3 \int_{0}^{\infty} t^2 e^{-t^2}\mathrm{d}t \tag{13}$$

Now we need to use ***integration by parts***, which tells us:

$$\int \boldsymbol{u}(x)\boldsymbol{v}'(x) = \boldsymbol{u}(x)\boldsymbol{v}(x) - \int \boldsymbol{u}'(x)\boldsymbol{v}(x)\mathrm{d}x$$

Note that $\left(e^{-t^2}\right)' = -2te^{-t^2}$. We cloud re-write the integral part of equation (13) and we have:

$$\int_0^{\infty} t^2 e^{-t^2}\mathrm{d}t = \int_0^{\infty}\left(-\frac{1}{2}t\right)\cdot\left(-2te^{-t^2}\right)\mathrm{d}t \tag{14}$$

We can let $\boldsymbol{u}(t) = -\frac{1}{2}t$ and $\boldsymbol{v}'(t) = -2te^{-t^2}$, so $\boldsymbol{v}(x) = e^{-t^2}$. Therefore, by using integration by parts, equation (14) becomes (note there is a Gaussian integral):

$$\begin{aligned}
\int_0^{\infty} t^2e^{-t^2}\mathrm{d}t &= \left[-\frac{1}{2}t\cdot e^{-t^2}\right]_0^{\infty} - \int_0^{\infty} -\frac{1}{2}e^{-t^2}\mathrm{d}t = \left[-\frac{t}{2e^{t^2}}\right]_0^{\infty} + \frac{1}{2}\int_0^{\infty} e^{-t^2}\mathrm{d}t \\
&= \lim_{m\to\infty}\left[-\frac{t}{2e^{t^2}}\right]_0^{m} + \frac{1}{2}\cdot\frac{\sqrt{\pi}}{2} \\
&= \lim_{m\to\infty}\left(-\frac{m}{2e^{m^2}}\right) - \left(-\frac{0}{2e^{0^2}}\right) + \frac{1}{4}\sqrt{\pi} \qquad (15)
\end{aligned}$$

[2]Once again, we see an ***improper integral***. Strictly speaking, what we should do here is $\int_{-\infty}^{\infty} t^2e^{-t^2}\mathrm{d}t = \int_{-\infty}^{0} t^2e^{-t^2}\mathrm{d}t + \int_0^{\infty} t^2e^{-t^2}\mathrm{d}t$, and show they both converge and their values are the same.

The first limit is in the $\frac{\infty}{\infty}$ form, so we could use the ***L'Hopital's rule***. Then equation (15) becomes:

$$\int_0^{\infty} t^2 e^{-t^2} \mathrm{d}t = \lim_{m \to \infty} \left(-\frac{m'}{(2e^{m^2})'} \right) - 0 + \frac{1}{4}\sqrt{\pi}$$

$$= \lim_{m \to \infty} \left(-\frac{1}{4me^{m^2}} \right) + \frac{1}{4}\sqrt{\pi}$$

$$= 0 + \frac{1}{4}\sqrt{\pi} = \frac{1}{4}\sqrt{\pi} \tag{16}$$

Put equation (16) into equation (13), we have solved the first term from the blue part of equation (9):

$$\int_{-\infty}^{\infty} 2\sqrt{2}\sigma^3 t^2 e^{-t^2} \mathrm{d}t = 4\sqrt{2}\sigma^3 \int_0^{\infty} t^2 e^{-t^2} \mathrm{d}t$$

$$= 4\sqrt{2}\sigma^3 \cdot \frac{1}{4}\sqrt{\pi} = \sqrt{2\pi}\sigma^3 \tag{17}$$

Finally, put equations (17), (11) and (10) into equation (9), we have:

$$E[\boldsymbol{X}^2] = \frac{1}{\sqrt{2\pi}\sigma}(\sqrt{2\pi}\sigma^3 + 0 + \sqrt{2\pi}\sigma\mu^2) = \sigma^2 + \mu^2 \tag{18}$$

Now, we can easily get

$$\mathrm{var}(\boldsymbol{X}) = E[\boldsymbol{X}^2] - (E[\boldsymbol{X}])^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

5 A Linear Function of A Continuous Random Variable

5.1 The General Case

Let's first have a look what happens to a ***random variable*** in general when we apply a linear function to it. Let $\boldsymbol{X}$ be a continuous random variable with a PDF $f_X(x)$. Let the random variable $\boldsymbol{Y}$ be:

$$\boldsymbol{Y} = a\boldsymbol{X} + b$$

where $a \neq 0$. What is the PDF of $\boldsymbol{Y}$?

Again, we should start with something simple. Consider this: if $\boldsymbol{X}$ and $\boldsymbol{Y}$ were discrete random variables, the situation becomes straightforward. We would have:

$$p_{\boldsymbol{Y}}(y) = P(\boldsymbol{Y} = y) = P(a\boldsymbol{X} + b = y) = P\left(\boldsymbol{X} = \frac{y-b}{a}\right)$$

However, for continuous random variables, the probability of getting a specific value is 0. Therefore, it is not very helpful to use the strategy above. We need to work on intervals for continuous random variables. The trick here is to use the CDF to solve the problem

5.1.1 When $a > 0$

Consider the case where $a > 0$, we have:

$$\begin{aligned} F_{\boldsymbol{Y}}(y) &= P(\boldsymbol{Y} \leqslant y) = P(a\boldsymbol{X} + b \leqslant y) \\ &= P\left(\boldsymbol{X} \leqslant \frac{y-b}{a}\right) = F_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) \end{aligned}$$

Now, that tells us the CDF of $\boldsymbol{Y}$ in terms of the CDF of $\boldsymbol{X}$:

$$F_{\boldsymbol{Y}}(y) = F_{\boldsymbol{X}}\left(\frac{y-b}{a}\right)$$

Since the derivate of the CDF is the PDF, now we can simply find out the PDF by differentiating both sides of the above equation like this:

$$f_{\boldsymbol{Y}}(y) = F'_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) = \frac{1}{a} \cdot f_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) \tag{19}$$

5.1.2 When $a < 0$

Now consider the case where $a < 0$. Using the similar technique, we have:

$$\begin{aligned} F_{\boldsymbol{Y}}(y) &= P(\boldsymbol{Y} \leqslant y) = P(a\boldsymbol{X} + b \leqslant y) \\ &= P\left(\boldsymbol{X} \geqslant \frac{y-b}{a}\right) = 1 - P\left(\boldsymbol{X} \leqslant \frac{y-b}{a}\right) \\ &= 1 - F_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) \end{aligned}$$

Taking the derivate at the both sides of the above equation, we have:

$$f_{\boldsymbol{Y}}(y) = -F'_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) = -\frac{1}{a} \cdot f_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) \tag{20}$$

Combine the cases where $a > 0$ (19) and $a < 0$ (20), we have:

$$f_{\boldsymbol{Y}}(y) = \frac{1}{|a|} \cdot f_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) \tag{21}$$

5.2 A Linear Function of A Normal Random Variable

Now, consider the normal random variable $\boldsymbol{X} \sim \mathcal{N}(\mu, \sigma^2)$, what is the PDF of the random variable $\boldsymbol{Y} = a\boldsymbol{X} + b$? We are given that:

$$f_{\boldsymbol{X}}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Using equation (21) from the previous section, we have:

$$\begin{aligned} f_{\boldsymbol{Y}}(y) &= \frac{1}{|a|} \cdot f_{\boldsymbol{X}}\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma|a|} e^{-\dfrac{(y-b-a\mu)^2}{2a^2\sigma^2}} \end{aligned} \tag{22}$$

Re-write equation (22) a bit, we have:

$$f_{\boldsymbol{Y}}(y) = \frac{1}{\sqrt{2\pi} \cdot |a|\sigma} e^{-\dfrac{[y-(a\mu+b)]^2}{2(|a|\sigma)^2}} \tag{23}$$

From equation (23), we can easily see that $\boldsymbol{Y} \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

6 The Variance of The Sum of Independent Random Variables

In the **Extra Reading Material** from **Lesson 6**, we demonstrated the ***linearity of expectation***. I shows that the expectation of the sum of random variables is the sum of their expectations:

$$\mathbb{E}[X_1 + X_2 + \cdots + X_{n-1} + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots \mathbb{E}[X_{n-1}] + \mathbb{E}[X_n]$$

A natural question next is: what about the variance of the sum of different random variables? Again, like we have been discussing repeatedly: **whenever we start to do something new, *always, always* start with something simple to get an intuition**.

The simplest case is just the sum of two random variables. Let's have a look at the variance of the sum of two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$. By definition we have:

$$\text{var}(X + Y) = \mathbb{E}[(X + Y)^2] - [E(X + Y)]^2$$

Expand $(X + Y)^2$ and use linearity of expectation, we have:

$$\begin{aligned}
\text{var}(X + Y) &= \mathbb{E}[(X + Y)^2] - [\mathbb{E}(X + Y)]^2 \\
&= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - [\mathbb{E}(X)]^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - [\mathbb{E}(Y)]^2 \\
&= \Big(\mathbb{E}(X^2) - [\mathbb{E}(X)]^2\Big) + \Big(\mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2\Big) + 2 \cdot [\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)]
\end{aligned}$$

Note that $\mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \text{var}(X)$ and $\mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 = \text{var}(Y)$ by definition. In addition, $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ is called the ***covariance*** of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Don't worry about the covariance right now, because we are going to talk about it again when we start to look at bivariate data in the later section of the course.

Therefore, the above equation becomes:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) - 2 \cdot [\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] \qquad (24)$$

We need to figure out what $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ is. We can expand it by the definition of the expectation:

$$\begin{aligned}
\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) &= \sum_{x,y} xy \cdot p_{\boldsymbol{X},\boldsymbol{Y}}(x,y) - \sum_x x p_{\boldsymbol{X}}(x) \sum_y y p_{\boldsymbol{Y}}(y) \\
&= \sum_{x,y} xy \cdot P(X = x, Y = y) - \sum_x x p_{\boldsymbol{X}}(x) \sum_y y p_{\boldsymbol{Y}}(y) \qquad (25)
\end{aligned}$$

If X and Y are independent, then:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) = p_{\boldsymbol{X}}(x) p_{\boldsymbol{Y}}(y) \qquad (26)$$

Put equation (26) into equation (25):

$$\begin{aligned}
\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) &= \sum_{x,y} xy \cdot p_{\boldsymbol{X}}(x) p_{\boldsymbol{Y}}(y) - \sum_x x p_{\boldsymbol{X}}(x) \sum_y y p_{\boldsymbol{Y}}(y) \\
&= 0 \qquad (27)
\end{aligned}$$

Put equation (27) into equation (24), we have:

$$\begin{aligned}
\mathrm{var}(X + Y) &= \mathrm{var}(X) + \mathrm{var}(Y) - 2 \cdot 0 \\
&= \mathrm{var}(X) + \mathrm{var}(Y) \qquad (28)
\end{aligned}$$

Therefore, we see that the variance of the sum of two **independent** random variables are the sum of their variances. It can be easily extended to n **independent** random variables.

7 The PMF or PDF of The Sum of Independent Random Variables

7.1 The Sum of Independent Random Variables In General

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. I'm going to say it again: **whenever we start to do something new, *always, always* start with something simple to get an intuition**.

For a start, we can look at the simplest case: the sum of two independent 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 $\boldsymbol{Y}$.

7.1.1 The Discrete Case

Again, let's begin with simpler random variables, that is, the discrete random variables.

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are discrete random variables, the situation is straightforward. We have:

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

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:

$$\begin{aligned} p_{\boldsymbol{Z}}(z) &= \sum_{\{(x,y)\,|\,x+y=z\}} P(\boldsymbol{X} = x, \boldsymbol{Y} = y) = \sum_x P(\boldsymbol{X} = x, \boldsymbol{Y} = z - x) \\ &= \sum_x p_{\boldsymbol{X}}(x) p_{\boldsymbol{Y}}(z - x) \end{aligned} \tag{29}$$

Equation (29) $p_{\boldsymbol{Z}}(z) = \sum_x p_{\boldsymbol{X}}(x) p_{\boldsymbol{Y}}(z - x)$ is called the **convolution** formula.

7.1.2 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.

7.2 The Sum of Independent Normal Random Variables

Since the normal random variables are quite common and useful, we are often facing problems where we need to compute the probability of the sum of different normal random variables.

Let's just start with simplest case: the sum of two normal random variables. Let $\boldsymbol{X} \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $\boldsymbol{Y} \sim \mathcal{N}(\mu_y, \sigma_y^2)$ be two independent normal random variables. We want to derive the PDF of $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$.

First, using the linearity of expectation, we have:

$$\mathbb{E}[Z] = \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \mu_x + \mu_y$$

Using equation (28) from **Section 6**, we also have:

$$\text{var}(Z) = \text{var}(X+Y) = \text{var}(X) + \text{var}(Y) = \sigma_x^2 + \sigma_y^2$$

We have easily derived the mean and the variance of $\boldsymbol{Z}$. Now we need to figure out the shape of $\boldsymbol{Z}$. What is our best guess? Well ... intuitively, $\boldsymbol{Z}$ should also be a normal random variable. Let's see if we could justify our guess.

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