Chapter 5. Multiple Random Variables

5.5: Convolution

Slides (Google Drive)

Alex Tsun

Video (YouTube)

In section 4.4, we explained how to transform random variables (finding the density function of $g(X)$). In this section, we'll talk about how to find the distribution of the sum of two independent random variables, $X + Y$, using a technique called convolution. It will allow us to prove some statements we made earlier without proof (like sums of independent Binomials are Binomial, sums of indepenent, Poissons are Poisson), and also derive the density function of the Gamma distribution which we just stated.

5.5.1 Law of Total Probability for Random Variables

We did secretly use this in some previous examples, but let's formally define this!

Definition 5.5.1: Law of Total Probability for Random Variables

Discrete version: If X, Y are discrete random variables:

$$p_X(x) = \sum_y p_{X,Y}(x, y) = \sum_y p_{X|Y}(x \mid y) p_Y(y)$$

Continuous version: If X, Y are continuous random variables:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) dy$$

This should just remind of you of the LTP we learned in section 2.2, or the definition of marginal PMF/PDFs from earlier in the chapter! We'll use this LTP to help us derive the formulae for convolution.

5.5.2 Convolution

Convolution is a mathematical operation that allows to derive the distribution of a sum of two independent random variables. For example, suppose the amount of gold a company can mine is X tons per year in country A, and the amount of gold the company can mine is Y tons per year in country B, independently. You have some distribution to model each. What is the distribution of the total amount of gold you mine, $Z = X + Y$? Combining this with 4.4, if you know your profit is some function of $g(Z) = \sqrt{X + Y}$ of the total amount of gold, you can now find the density function of your profit!

I think this is best learned through examples:

Example(s)

Let $X, Y \sim \text{Unif}(1, 4)$ be independent rolls of a fair 4-sided die. What is the PMF of $Z = X + Y$?

Solution We know that for the range of Z we have the following, since it is the sum of two values each in the range $\{1, 2, 3, 4\}$:

$$\Omega_Z = \{2, 3, 4, 5, 6, 7, 8\}$$

Should the probabilities be uniform? That is, would you be equally likely to roll a 2 as a 5? No, because there is only one way to get a 2 (rolling $(1, 1)$), but many ways to get a 5.

If I wanted to compute the probability that $Z = 3$ for example, I could just sum over all possible values of X in $\Omega_X = \{1, 2, 3, 4\}$ to get:

$$\begin{aligned}
\mathbb{P}(Z = 3) &= \mathbb{P}(X = 1, Y = 2) + \mathbb{P}(X = 2, Y = 1) + \mathbb{P}(X = 3, Y = 0) + \mathbb{P}(X = 4, Y = -1) \\
&= \mathbb{P}(X = 1)\,\mathbb{P}(Y = 2) + \mathbb{P}(X = 2)\,\mathbb{P}(Y = 1) + \mathbb{P}(X = 3)\,\mathbb{P}(Y = 0) + \mathbb{P}(X = 4)\,\mathbb{P}(Y = -1) \\
&= \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{4} + \frac{1}{4}\cdot 0 + \frac{1}{4}\cdot 0 \\
&= \frac{2}{16}
\end{aligned}$$

where the first line is all ways to get a 3, and the second line uses independence. Note that is is not possible that $Y = 0$ or $Y = -1$, but we write this for completion. More generally, to find $p_Z(z) = \mathbb{P}(Z = z)$ for any value of z, we just write

$$\begin{aligned}
p_Z(z) &= \mathbb{P}(Z = z) \\
&= \sum_{x \in \Omega_X} \mathbb{P}(X = x, Y = z - x) \\
&= \sum_{x \in \Omega_X} \mathbb{P}(X = x)\,\mathbb{P}(Y = z - x) \\
&= \sum_{x \in \Omega_X} p_X(x) p_Y(z - x)
\end{aligned}$$

The intuition is that if we want $Z = z$, we sum over all possibilities of $X = x$ but require that $Y = z - x$ so that we get the desired sum of z. It is very possible that $p_Y(z - x) = 0$ as we saw above. □

It turns out that formula at the bottom was extremely general, and works for any sum of two independent discrete RVs. Now let's consider the continuous case. What if X and Y are continuous RVs and we define $Z = X + Y$; how can we solve for the probability *density* function for Z, $f_Z(z)$? It turns out the formula is extremely similar, just replacing p with f !

Theorem 5.5.1: Convolution

Let X, Y be *independent* RVs, and $Z = X + Y$.
Discrete version: If X, Y are discrete:

$$p_Z(z) = \sum_{x \in \Omega_X} p_X(x) p_Y(z - x)$$

Continuous version: If X, Y are continuous:

$$f_Z(z) = \int_{x \in \Omega_X} f_X(x) f_Y(z - x) dx$$

Note: You can swap the roles of X and Y. Note the similarity between the cases!

Proof of Convolution.:

- Discrete case: Even though we proved this earlier, we'll do it again a different way (using the LTP/def of marginal):

$$\begin{aligned}
p_Z(z) &= \mathbb{P}(Z = z) \\
&= \sum_{x \in \Omega_X} \mathbb{P}(X = x, Z = z) && \text{[LTP/marginal]} \\
&= \sum_{x \in \Omega_X} \mathbb{P}(X = x, Y = z - x) && [(X = x, Z = z) \text{ equivalent to } (X = x, Y = z - x)] \\
&= \sum_{x \in \Omega_X} \mathbb{P}(X = x)\, \mathbb{P}(Y = z - x) && [X \text{ and } Y \text{ are independent}] \\
&= \sum_{x \in \Omega_X} p_X(x) p_Y(z - x)
\end{aligned}$$

- Continuous case: Since we should never work with densities as probabilities, let's start with the CDF and differentiate:

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \le z) \\
&= \mathbb{P}(X + Y \le z) && [\text{def of } Z] \\
&= \int_{x \in \Omega_X} \mathbb{P}(X + Y \le z \mid X = x) f_X(x) dx && [\text{LTP, conditioning on } X] \\
&= \int_{x \in \Omega_X} \mathbb{P}(x + Y \le z \mid X = x) f_X(x) dx && [\text{given } X = x] \\
&= \int_{x \in \Omega_X} \mathbb{P}(Y \le z - x \mid X = x) f_X(x) dx && [\text{algebra}] \\
&= \int_{x \in \Omega_X} \mathbb{P}(Y \le z - x) f_X(x) dx && [X \text{ and } Y \text{ are independent}] \\
&= \int_{x \in \Omega_X} F_Y(z - x) f_X(x) dx && [\text{def of CDF of } Y]
\end{aligned}$$

  Now we can take the derivative (with respect to z) of the CDF to get the density (F_Y becomes f_Y):

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{x \in \Omega_X} f_X(x) f_Y(z - x) dx$$

Note the striking similarity in the formulae! □

Example(s)

Suppose X and Y are two independent random variables such that $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$, and let $Z = X + Y$. Prove that $Z \sim \text{Poi}(\lambda_1 + \lambda_2)$.

The range of X, Y are $\Omega_X = \Omega_Y = \{0, 1, 2, \dots\}$, and so $\Omega_Z = \{0, 1, 2, \dots\}$ as well. For $n \in \Omega_Z$: Note that the convolution formula says:

$$p_Z(n) = \sum_{k \in \Omega_X} p_X(k) p_Y(n - k) = \sum_{k=0}^{\infty} p_X(k) p_Y(n - k)$$

However, if you blindly plug in the PMFs p_X and p_Y, you will get the wrong answer, and here's why. We only want to sum things that are non-zero (otherwise what's the point?), and if we want $p_X(k)p_Y(n-k) > 0$, we need BOTH to be nonzero. That means, k must be in the range of X AND $n-k$ must be in the range of Y. Remember the dice example (we had $p_Y(-1)$ at some point, which would be 0 and not 1/4). We are guaranteed $p_X(k) > 0$ because we are only summing over valid $k \in \Omega_X$, but we must have $n-k$ be a nonnegative integer (in the range $\Omega_Y = \{0, 1, 2, \dots\}$, so actually, we must have $k \leq n$. Now, we can just plug and chug:

$$\begin{aligned}
p_Z(n) &= \sum_{k=0}^{n} p_X(k)p_Y(n-k) && \text{[convolution formula]} \\
&= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} && \text{[plug in Poisson PMFs]} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{1}{k!(n-k)!}\lambda_1^k(1-\lambda_2)^{n-k} && \text{[algebra]} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!}\sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\lambda_1^k(1-\lambda_2)^{n-k} && \text{[multiply and divide by } n!\text{]} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{1}{n!}\sum_{k=0}^{n} \binom{n}{k}\lambda_1^k(1-\lambda_2)^{n-k} && \left[\binom{n}{k} = \frac{n!}{k!(n-k)!}\right] \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!} && \text{[binomial theorem]}
\end{aligned}$$

Thus, $Z \sim \text{Poi}(\lambda_1 + \lambda_2)$, as its PMF matches that of a Poisson distribution! Note we wouldn't have been able to do that last step if our sum was still $k = 0$ to n. You MUST watch out for this at the beginning, and after that, it's just algebra.

Example(s)

Suppose X, Y are independent and identically distributed (iid) continuous Unif$(0, 1)$ random variables. Let $Z = X + Y$. What is $f_Z(z)$?

Solution We always begin by calculating the range: we have $\Omega_Z = [0, 2]$. Again, we shouldn't expect Z to be uniform, since we should expect a number around 1, but not 0 or 2.

For a $U \sim \text{Unif}(0, 1)$ (continuous) random variable, we know $\Omega_U = [0, 1]$, and that

$$f_U(u) = \begin{cases} 1 & 0 \leq u \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Z(z) = \int_{x\in\Omega_X} f_X(x)f_Y(z-x)dx = \int_0^1 f_X(x)f_Y(z-x)dx = \int_0^1 f_Y(z-x)dx$$

where the last formula holds since $f_X(x) = 1$ for all $0 \leq x \leq 1$ as we saw above. Remember, we need to make sure $z - x \in \Omega_Y = [0, 1]$, otherwise the density will be 0.

For $f_Y(z-x) > 0$, we need $0 \leq z - x \leq 1$. We'll split into two cases depending on whether $z \in [0, 1]$ or $z \in [0, 2]$, which compose its range $\Omega_Z = [0, 2]$.

- If $z \in [0,1]$, we already have $z - x \leq 1$ since $z \leq 1$ (and $x \in [0,1]$). We also need $z - x \geq 0$ for the density to be nonzero: $x \leq z$. Hence, our integral becomes:

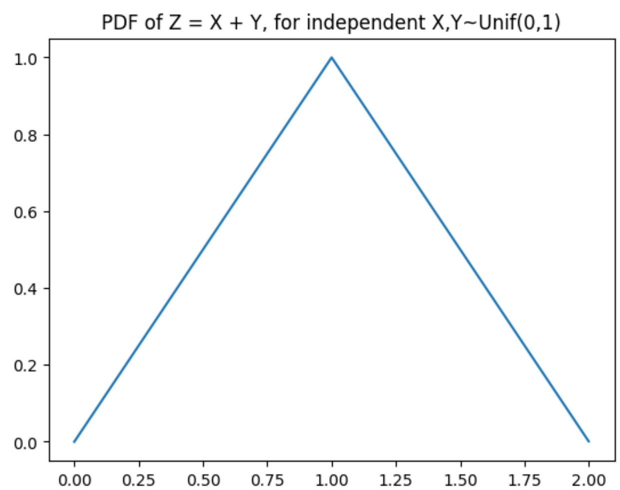
$$\begin{aligned} f_Z(z) &= \int_0^z f_Y(z-x)dx + \int_z^1 f_Y(z-x)dx \\ &= \int_0^z 1dx + 0 = [x]_0^z = z \end{aligned}$$

- If $z \in [1,2]$, we already have $z - x \geq 0$ since $z \geq 1$ (and $x \in [0,1]$). We now need the other condition $z - x \leq 1$ for the density to be nonzero: $x \geq z - 1$. Hence, our integral becomes:

$$\begin{aligned} f_Z(z) &= \int_0^{z-1} f_Y(z-x)dx + \int_{z-1}^1 f_Y(z-x)dx \\ &= 0 + \int_{z-1}^1 1dx = [x]_{z-1}^1 = 2 - z \end{aligned}$$

Thus, putting these two cases together gives:

$$f_Z(z) = \begin{cases} z & 0 \leq z \leq 1 \\ 2-z & 1 \leq z \leq 2 \\ 0 & \text{otherwise} \end{cases}$$



This makes sense because there are "more ways" to get a value of 1 for example than any other point. Whereas to get a value of 2, there's only one way - we need both X, Y to be equal to 1. □

Example(s)

Mitchell and Alex are competing together in a 2-mile relay race. The time Mitchell takes to finish (in hours) is $X \sim \text{Exp}(2)$ and the time Alex takes to finish his mile (in hours) is continuous $Y \sim \text{Unif}(0,1)$. Alex starts immediately after Mitchell finishes his mile, and their performances are independent. What is the distribution of $Z = X + Y$, the total time they take to finish the race?

Solution First, we know that $\Omega_X = [0, \infty)$ and $\Omega_Y = [0, 1]$, so $\Omega_Z = [0, \infty)$. We know from our distribution chart that

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0 \qquad \text{and} \qquad f_Y(y) = 1, 0 \leq y \leq 1$$

Let $z \in \Omega_Z$. We'll use the convolution formula, but this time over the range of Y (you could also do over X too!). We can do this because $X + Y = Y + X$, and there was no reason why we had to condition on X first.

$$f_Z(z) = \int_{\Omega_Y} f_Y(y) f_X(z - y) dy = \int_0^1 f_Y(y) f_X(z - y) dy$$

Since we are integrating over y, we don't need to worry about $f_Y(y)$ being 0, but we do need to make sure $f_X(z - y) > 0$. There are two cases again:

- If $z \in [0, 1]$, then since we need $z - y \geq 0$, we need $y \geq z$:

$$f_Z(z) = \int_0^z f_Y(y) f_X(z - y) dy = \int_0^z 1 \cdot \lambda e^{-\lambda(z-y)} dy = 1 - e^{-\lambda z}$$

- if $z \in (1, \infty)$, then $y \leq z$ always (since $y \in [0, 1]$), so

$$f_Z(z) = \int_0^1 f_Y(y) f_X(z - y) dy = (e^{\lambda} - 1)e^{-\lambda z}$$

Note this tiny difference in the upper limit of the integral made a huge difference! Our final result is

$$f_Z(z) = \begin{cases} 1 - e^{-\lambda z} & z \in [0, 1] \\ (e^{\lambda} - 1)e^{-\lambda z} & z \in (1, \infty) \\ 0 & \text{otherwise} \end{cases}$$

$\square$

The moral of the story is: always watch out for the ranges, otherwise you might not get what you expect! The range of the random variable exists for a reason, so be careful!