Week 12 Recap

Monday November 6 and Wednesday November 8: Sums of independent random variables.

This is one of the most important week for this course. The material is in Chapter 7 in the book.

Given a vector $X = (X_1, \ldots, X_k)$ we want to learn how to find the distribution of $S = X_1 + X + 2 + \cdots + X_k$. This is an example of a function of a multivariate random variable and the usual techniques apply.

Now, we are interested in the case when the X_i are mutually independent. Let start with two discrete random variables, X, Y that are independent, take values in $\mathbb{Z}$, and have respective mass probability function $P(X = x) = p_x, x \in \mathbb{Z}$, and $P(Y = y) = q_y, y \in \mathbb{Z}$. We want to find the mass probability function of $Z = X + Y$ (which, obviously, take values in $\mathbb{Z}$). We have \P(Z=z)= \sum_{x\in \mathbb Z} p_xq_{z-x}=\sum _{y\in \mathbb Z} p_{z-y} q_y,\;\;\;z\in \mathbb Z.\]

Example: Sum of two independent geometric with parameter p. A geometric random variable X with parameter p has distribution $P(X = k) = p(1-p)^{k-1}, k = 1, 2, \ldots$. If X and Y are independent geometric random variables with parameter p, the distribution of their sum, $Z = X + Y$, is given by

$$\begin{aligned}
P(Z = z) &= \sum_{x=1}^{\infty} P(X = x)P(Y = z - x) \\
&= \sum_{x=1,\;\; z-x\geq 1}^{\infty} p(1-p)^{x-1}p(1-p)^{z-x-1} \\
&= p^2 \sum_{x=1}^{z-1} (1-p)^{z-2} = p^2(z-1)(1-p)^{z-2} \\
&= \binom{z-1}{1} p^2(1-p)^{z-2}.
\end{aligned}$$

Can you recognize this mass distribution function? Suppose we consider a sequence of independent identical experiments
with probability p of success. We think of the experiment occurring every minute, the first at minute 1, the second at minute 2, etc. What is the probability that the second success occurs at time z? The answer is $\binom{z-1}{1} p^2(1-p)^{z-2}$. (You should be able to explain why).

Our computation in this example shows that the sum of two independent geometric random variables with parameter p is a {\em negative binomial} with parameters $r = 2$ and p, the same distribution that the distribution of the time at which a second success occurs in a series of independent identical experiments with probability p of success. Let us explain why this is true, without using any computation at all. Imagining this sequence of experiments, let X be the time at which the first success occurs, let Y be the *additional* time needed for the second success to occur. Let Z be the time at which the second success occurs. If the first success is at time m and the second at time n then $X = m$ and $Y = n - m$, and $Z = n = m + n - m = X + Y$. Now explain why, in this situation, X is geometric with parameter p; Y is also a geometric with parameter p; and X and Y are independent.

So, using this thought experiment, we see that the sum of two independent geometric random variables with parameter p is a negative binomial with parameter $r = 2$ and p. We should probably ask ourselves:

What is the sum of r independent geometric random variable with parameter p?

The continuous RVs case:

Let X and Y be continuous independent random variables taking values in $\mathbb{R}$ with respective density functions f_X and f_Y.

Then their sum Z is a continuous random variable with density function f_Z given by

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)dx = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)dy, \quad z \in \mathbb{R}.$$

Example: Sum of two independent normal $\mathcal{N}(0,1)$. Assume that X and Y are independent normal $\mathcal{N}(0,1)$. Then their sum Z had density function

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(z-x)^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(x^2+(z-x)^2)/2} dx \end{aligned}$$

We write

$$x^2 + (z-x)^2 = 2x^2 - 2xz + z^2 = 2\left(x - \frac{1}{2}z\right)^2 + \frac{1}{2}z^2$$

and use this expression in the integral to get

$$\begin{aligned} f_Z(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/4} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-\frac{1}{2}z)^2)} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-z^2/4} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2} du = \frac{1}{\sqrt{4\pi}} e^{-z^2/4}. \end{aligned}$$

The first equality in the last line is obtained by using the change of variable $u = x - z/2$ and the last equality by using the change of variable $v = \sqrt{2}\,u$ which shows that

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2}\,du = \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-v^2/2}\,dv = \frac{1}{\sqrt{2}}.$$

The random variable Z is a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 2$. Indeed, the density of a normal $\mathcal{N}(\mu, \sigma^2)$ is $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$.

MGF and sums of independent RV

Let X and Y be independent random variables with generating functions M_X, M_Y. Then their sum $Z = X + Y$ has generating function M_Z given by

$$M_Z(t) = M_X(t) M_Y(t)$$

for all $t \in \mathbb{R}$ such that both M_X and M_Y are defined.

Example: Sum of two independent negative binomial X, Y with parameters r, p and s, p:

The sum Z of two independent negative binomial random variables X, Y with parameters r, p and s, p is a negative binomial random variable with parameters $r + s$ and p.

We give two proofs.

Proof I: Consider a sequence of repeated identical independent experiments with probability p of success.
Think of X as being the number of experiments until the r-th success. Similarly think of Y as the number of experiments required to see s successes after the r-th success. By construction, these two random variables are independent and $Z = X + Y$ counts the number of experiments needed to see $r + s$ successes. So, Z is a negative binomial with parameters $r + s, p$.

FIGURE 2. A sequence of experiments used to represent X and Y, and the sum $X + Y$. Here $r = 3$, $s = 4$, $X = 9$ and $Y = (17 - 9) = 8$

Proof II: The moment generating function of X and Y are $M_X(t) = \left(\frac{pe^t}{1+(1-p)e^t}\right)^r$ and $M_Y(t) = \left(\frac{pe^t}{1+(1-p)e^t}\right)^s$. It follows that

$$M_Z(t) = M_X(t)M_Y(t) = \left(\frac{pe^t}{1+(1-p)e^t}\right)^{r+s}$$

and this shows that Z has a negative binomial distribution with parameters $r+s, p$.

Remark: A negative binomial random variable with parameter $r = 1$ is really a geometric random variable. So we recover the fact that the sum of two independent geometric random variables with the same parameter p is a negative binomial with parameters $r = 2$ and p. A simple induction argument then tells us that the sum of r mutually independent geometric random variables with identical parameter p is a negative binomial with parameters r, p. Can you explain this fact without computation by using a sequence of independent identical tries?