Lecture 9

1. The geometric distribution, continued

1.1. An example. A couple has children until their first son is born. Suppose the sexes of their children are independent from one another [unrealistic], and the probability of girl is 0.6 every time [not too bad]. Let X denote the number of their children to find then that $X = \text{Geom}(0.4)$. In particular,

$$\begin{aligned}
\text{P}\{X \le 3\} &= f(1) + f(2) + f(3) \\
&= p + p(1-p) + p(1-p)^2 \\
&= p\left[1 + 1 - p + (1-p)^2\right] \\
&= p\left[3 - 3p + p^2\right] \\
&= 0.784.
\end{aligned}$$

1.2. The tail of the distribution. Now you may be wondering why these random variables are called "geometric." In order to answer this, consider the tail of the distribution of X (probability of large values). Namely, for all $n \ge 1$,

$$\begin{aligned}
\text{P}\{X \ge n\} &= \sum_{j=n}^{\infty} p(1-p)^{j-1} \\
&= p \sum_{k=n-1}^{\infty} (1-p)^k.
\end{aligned}$$

Let us recall an elementary fact from calculus.

Lemma 9.1 (Geometric series). *If $r \in (0\,,1)$, then for all $n \geq 0$,*

$$\sum_{j=n}^{\infty} r^j = \frac{r^n}{1-r}.$$

Proof. Let $s_n = r^n + r^{n+1} + \cdots = \sum_{j=n}^{\infty} r^j$. Then, we have two relations between s_n and s_{n+1}:

(1) $rs_n = \sum_{j=n+1}^{\infty} r^j = s_{n+1}$; and
(2) $s_{n+1} = s_n - r^n$.

Plug (2) into (1) to find that $rs_n = s_n - r^n$. Solve to obtain the lemma. □

Return to our geometric random variable X to find that

$$\mathrm{P}\{X \geq n\} = p\frac{(1-p)^{n-1}}{1-(1-p)} = (1-p)^{n-1}.$$

That is, $\mathrm{P}\{X \geq n\}$ vanishes geometrically fast as $n \to \infty$.

In the couples example (§1.1),

$$\mathrm{P}\{X \geq n\} = 0.6^{n-1} \qquad \text{for all } n \geq 1.$$

2. The negative binomial distribution

Suppose we are tossing a p-coin, where $p \in (0\,,1)$ is fixed, until we obtain r heads. Let X denote the number of tosses needed. Then, X is a discrete random variable with possible values $r, r+1, r+2, \ldots$. When $r = 1$, then X is Geom(p). In general,

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{if } x = r, r+1, r+2, \ldots\,, \\ 0 & \text{otherwise.} \end{cases}$$

This X is said to have a *negative binomial distribution with parameters r and p.* Note that our definition differs slightly from that of your text (p. 117).

3. The Poisson distribution

Choose and fix a number $\lambda > 0$. A random variable X is said to have the *Poisson distribution with parameter* λ (written Poiss(λ)) if its mass function is

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{if } x = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

In order to make sure that this makes sense, it suffices to prove that $\sum_x f(x) = 1$, but this is an immediate consequence of the Taylor expansion of e^λ, viz.,

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$$

3.1. Law of rare events. Is there a physical manner in which Poiss(λ) arises naturally? The answer is "yes." Let $X = \text{Bin}(n\,, \lambda/n)$. For instance, X could denote the total number of sampled people who have a rare disease (population percentage $= \lambda/n$) in a large sample of size n. Then, for all fixed integers $k = 0\,, \ldots, n$,

$$f_X(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}. \tag{8}$$

Poisson's "law of rare events" states that if n is large, then the distribution of X is approximately Poiss(λ). In order to deduce this we need two computational lemmas.

Lemma 9.2. *For all* $z \in \mathbf{R}$,

$$\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n = e^z.$$

Proof. Because the natural logarithm is continuous on $(0\,,\infty)$, it suffices to prove that

$$\lim_{n\to\infty} n \ln\left(1 + \frac{z}{n}\right) = z. \tag{9}$$

By Taylor's expansion,

$$\ln\left(1 + \frac{z}{n}\right) = \frac{z}{n} + \frac{\theta^2}{2},$$

where θ lies between 0 and z/n. Equivalently,

$$\frac{z}{n} \le \ln\left(1 + \frac{z}{n}\right) \le \frac{z}{n} + \frac{z^2}{2n^2}.$$

Multiply all sides by n and take limits to find (9), and thence the lemma. $\square$

Lemma 9.3. *If* $k \ge 0$ *is a fixed integer, then*

$$\binom{n}{k} \sim \frac{n^k}{k!} \qquad \text{as } n \to \infty.$$

where $a_n \sim b_n$ *means that* $\lim_{n\to\infty}(a_n/b_n) = 1$.

Proof. If $n \geq k$, then

$$\begin{aligned}\frac{n!}{n^k(n-k)!} &= \frac{n(n-1)\cdots(n-k+1)}{n^k}\\ &= \frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-k+1}{n}\\ &\to 1 \qquad \text{as } n \to \infty.\end{aligned}$$

The lemma follows upon writing out $\binom{n}{k}$ and applying the preceding to that expression. □

Thanks to Lemmas 9.2 and 9.3, and to (8),

$$f_X(k) \sim \frac{n^k}{k!}\frac{\lambda^k}{n^k}e^{-\lambda} = \frac{e^{-\lambda}\lambda^k}{k!}.$$

That is, when n is large, X behaves like a Poiss(λ), and this proves our assertion.