BSM Probability notes for the online course
Spring 2020

Péter Bálint

March 21, 2020

Abstract

On March 12 2020, the Budapest Semesters in Mathematics has decided to continue teaching in the form of distance education, effective March 23. These notes discuss the material for the Probability course. The course textbook, Ross: A first course in probability, is frequently referred.

The Poisson distribution

This is the content of section 4.7 in the Ross book, which we have already started studying in Budapest. Let me recall, that given a parameter $\lambda > 0$, the random variable X is Poisson distributed with parameter λ (notation $X \sim Poi(\lambda)$) if

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}; \qquad k = 0, 1, 2, ...$$

Also, we have seen that the Poisson distribution arises in a particular asymptotic regime of the binomial distribution, when $n \to \infty$, $p = p(n) \to 0$ such that $np \to \lambda$. In words, the Poisson distribution is a good model for the number of successes when having *many independent* trials such that *the individual success rate is small.* This is a very frequent scenario – here are some Poisson distributed quantities:

- the number of calls received by the call center of a large bank within an hour. (There are many customers, but for each of them the chance of calling within that hour is small.)
- the number of accidents within a month at some busy junction. (There are many cars passing, but for each particular car the chance of a crash is very small.)
- the number of typos in a book chapter. (There are many characters, and for each character the chance of being misspelt is small.)

Here we compute the expected value of a variable $X \sim Poi(\lambda)$:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

where we changed the index of summation to $m = k - 1$. You may say this is obvious, as we obtained the Poisson as a limit of the binomial – the expected value of which is np, and we had $np \to \lambda$. Yet, what we see here is that the order of taking the limit and the infinite summation (or in a related context, the integration) can be swapped, which is a highly nontrivial issue. We may proceed to compute the variance by noting $\mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}X$, and

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda}\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = e^{-\lambda}\lambda^2 \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2$$

thus

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}X - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Example 1. *How many chocolate chips should you plan per muffin to ensure that no more than one percent of customers get upset?*

What I mean is that a customer gets upset if she/he finds no chocolate chips at all in her/his muffin. Let X denote the number of chocolate chips in one particular muffin. *Claim:* X is Poisson distributed.

Assuming that the claim holds, given many customers, by Bernoulli's Law of Large Numbers we have that

$$\frac{\#\{\text{customers who get no muffin}\}}{\text{total number of customers}} \longrightarrow \mathbb{P}(X = 0)$$

hence we want that $e^{-\lambda} = \mathbb{P}(X = 0) < 0.01$, which implies the following lower bound on the parameter of the Poisson distribution: $\lambda > \ln 100 \approx 4.6$.

Now let us argue why the claim holds. To bake the muffins, a large amount of dough is prepared. This will be later chopped into $M \gg 1$ portions of equal size, where each portion corresponds to one muffin. But before splitting it up, $N \gg 1$ chocolate chips are put evenly into the dough. For any particular chip, the chance of landing in the portion that corresponds to my muffin is $1/M$. Hence we have many (N) trials with a small individual success rate ($1/M$), and the number of successes (chocolate chips in my muffin) is Poisson distributed. Moreover, $\lambda \approx N/M$, the number of chocolate chips planned per muffin.

You may skip the somewhat lengthy Example 7d on the length of the longest run in the Ross book. However, the material of pages 155–157 (ninth edition) on the Poisson process is definitely relevant for us.

The Poisson process

In many of the examples of quantities that are Poisson distributed, there is a *time scale* involved. This opens the perspectives to an exciting branch of probability: the theory of stochastic processes, which studies random phenomena evolving in time.

For the particular case of the Poisson process, we have a *point process*, a countable random subset of the halfline $[0, +\infty)$. Here the halfline is typically interpreted as time. The points in the random set will be referred as *impact*s, which may correspond to the calls at the call center, the accidents at the junction etc.

The Poisson process is defined by three characteristic properties. To formulate these, we need some terminology, which will be useful for future reference as well.

- Consider a continuous function $f : [0, +\infty) \to \mathbb{R}$, which we study for small values $h \to 0+$. $f(h) = o(h)$ ("little o of h") if

$$f(h) = o(h) \quad \Longleftrightarrow \quad \lim_{h \to 0+} \frac{f(h)}{h} = 0.$$

  In particular $h^2 = o(h)$, but $0.01 \cdot h \neq o(h)$. Also, $\sqrt{h} \neq o(h)$, however, $h = o(\sqrt{h})$.

- Consider two discrete random variables X and Y, taking values $x_1, x_2, ...$ and $y_1, y_2, ...$, respectively. X and Y are independent if for any pair of values x_k and y_ℓ the events $\{X = x_k\}$ and $\{Y = y_\ell\}$ are independent. Given random variables $X_1, X_2, ...X_m$, their independence (as a collection) is defined analogously.

Definition 2. *Fix some positive parameter λ. A point process on the positive halfline $[0, +\infty)$ is a* ***Poisson process of intensity*** *λ if*

P1 ***(Independence)*** *Consider $I_1, I_2, ...I_m$, an arbitrary finite collection of non-overlapping intervals in $[0, +\infty)$, and let $X_1, ..., X_m$ denote the number of impacts in the intervals $I_1, ...I_m$, respectively. The random variables $X_1, ..., X_m$ are independent.*

P2 ***(Homogeneity)*** *Let I be an (infinitesimally small) interval of length h. Then*

$$\mathbb{P}(\text{There is at least one impact in I}) = \lambda \cdot \text{h} + \text{o(h)}$$

P3 ***(No Accumulation)*** *Let I be an (infinitesimally small) interval of length h. Then*

$$\mathbb{P}(\text{There are at least two impacts in I}) = \text{o(h)}$$

Remark. Do not confuse λ, the intensity of the process with the various λ-s that appeared previously as parameters of a Poisson distribution. In particular, the intensity of the process has dimensions $\frac{1}{\text{time}}$. If time is measured in different units, λ has to be rescaled.

Proposition 3. *Consider a Poisson process of intensity λ, and let $I_t \subset [0, +\infty)$ be an interval of length t. Let $N(t)$ denote the number of impacts inside I_t. Then $N(t) \sim Poi(\lambda t)$.*

For the proof of this proposition we refer to the Ross book.

Example 4. *This is a variation on Example 7e from Ross. Let us assume that on a highway the average number of accidents is* 3 *per month. We start inspecting that highway on a particular day.*

(a) What is the chance that there are at least two accidents in the next twenty days?

(b) What is the chance that at least t days elapse until the first accident?

The number of accidents is a Poisson process of intensity $\lambda = 3\frac{1}{\text{month}} = 0.1\frac{1}{\text{day}}$. To solve part (a), let us measure time in days, then

$$N(20) \sim Poi\left(0.1\frac{1}{\text{day}} \cdot 20\text{days}\right) = Poi(2).$$

Hence

$$\begin{aligned}\mathbb{P}(N(20) > 2) &= 1 - \mathbb{P}(N(20) = 0) - \mathbb{P}(N(20) = 1) = 1 - e^{-2} - \frac{2}{2!} = \\ &= e^{-2}1 - 2e^{-2} \approx 1 - 2 \cdot 0.135 = 0.73.\end{aligned}$$

To solve part (b), let T denote the random variable that measures the time (in days) that elapses until the first accident. Then

$$\mathbb{P}(T > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t} = e^{-0.1 \cdot t}.$$

Further discrete distributions

Here we discuss the material of section 4.8 from the Ross book.

The geometric distribution

Let $p \in (0, 1)$, the individual success rate in a sequence of independent trials (for example, $p = \frac{1}{6}$ for subsequent rolls of a fair die). The random variable X is geometrically distributed with parameter p ($X \sim Geom(p)$) if it can take the values $k = 1, 2, ...$ and

$$\{X = k\} \iff \text{The first success is at the kth trial.}$$

For brevity, let us introduce $q = 1 - p$. Then, as $X = k$ means there are $k - 1$ failures in a row, followed by a success, the mass function is

$$\mathbb{P}(X = k) = q^{k-1}p.$$

Note

$$\sum_{k=1}^{\infty} q^{k-1}p = p\sum_{m=0}^{\infty} q^m = p \cdot \frac{1}{1-q} = 1$$

where we have summed up a geometric series, which is the reason for the name of this distribution. To compute expected value and variance, introduce the function $g(q) = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$. As $q \in (0,1)$, this power series converge. Hence, denoting differentiation w.r. to q by prime,

$$\begin{aligned}\sum_{k=1}^{\infty} kq^{k-1} &= g'(q) = \frac{1}{(1-q)^2},\\ \sum_{k=2}^{\infty} k(k-1)q^{k-2} &= g''(q) = \frac{2}{(1-q)^3}.\end{aligned}$$

Now

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kq^{k-1}p = pg'(q) = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

In particular, in subsequent rolls of a fair die, the *expected time* when the first 6 occurs is at the 6th roll. Also,

$$\mathbb{E}(X(X-1)) = \sum_{k=2}^{\infty} k(k-1)q^{k-1}p = pqg''(q) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2},$$

thus

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}X - \mathbb{E}X^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2},$$

where we have used $q + p = 1$. Then $\mathbb{D}(X) = \frac{\sqrt{q}}{p}$.

The geometric distribution has another remarkable feature, the **memoryless property**. Let k and n be arbitrary positive integers.

$$\begin{aligned}\mathbb{P}(X \geq k) &= q^{k-1}, \qquad \mathbb{P}(X > n) = q^n\\ \mathbb{P}(X \geq k+n | X > n) &= \frac{q^{k+n-1}}{q^n} = q^{k-1} = \mathbb{P}(X \geq k).\end{aligned}$$

In words: if there was no success in the first n trials, the probability that we have to make another k trials for the first success is independent of n. (What has happened previously does not change the distribution of the additional time needed for the first success.)

The negative binomial distribution

Fox two parameters, an integer $r \geq 1$ and a success rate $p \in (0,1)$. Consider again a sequence of independent trials. Then X has a negative binomial distribution with parameters r and p ($X \sim NegBinom(r,p)$) if

$$\{X = k\} \iff \text{The rth success is at the kth trial.}$$

Apparently, X can take values $r, r+1, ...$ Try to figure out the mass function!