Chapter 12

Negative Binomial, Poisson, and the Poisson Process

Lecture 12 | Devore 9e: 3.5–3.6

So far the discrete families are geometric, binomial, hypergeometric. Today we add the last two, and then collect everything in one table.

12.1 The negative binomial distribution

Definition 12.1 (Bernoulli random variable)**.** A single trial with two outcomes S, F and $\mathbb{P}(S) = p$, $\mathbb{P}(F) = 1 - p$. Coded as a random variable, $X \sim \text{Bernoulli}(p)$ takes the value 1 on S and 0 on F.

Definition 12.2 (Negative binomial distribution)**.** Take a sequence of i.i.d. (*independent and identically distributed*) Bernoulli trials and repeat until a total of r successes has occurred. Let

$$X := \text{the number of failures before the } r\text{-th success.}$$

Then $X \sim \text{NegBinomial}(r, p)$.

Theorem 12.3.

$$\mathbb{P}\{X = x\} = \binom{x+r-1}{x}(1-p)^x p^r, \qquad x = 0, 1, 2, \ldots$$

Proof. The experiment stops at the r-th success, so the *last* trial is forced to be S. Before it there are $x + r - 1$ trials containing exactly x failures and $r - 1$ successes, in any order:

$$\underbrace{\cdots}_{x\ F\text{'s and } (r-1)\ S\text{'s, any order}} \qquad \boxed{S}$$

There are $\binom{x+r-1}{x}$ such arrangements, each of probability $(1-p)^x p^{r-1}$, and then the final S contributes another factor p. $\square$

Remark 12.4 (Consistency check)**.** Set $r = 1$: $\mathbb{P}\{X = x\} = (1-p)^x p$, the geometric distribution counted by *failures*. So the geometric is NegBinomial$(1, p)$, up to the usual shift between "number of trials" and "number of failures".

Theorem 12.5.

$$\mathbb{E}[X] = r\Big(\frac{1}{p} - 1\Big) = \frac{r(1-p)}{p}, \qquad \mathrm{Var}(X) = \frac{r\big(\frac{1}{p} - 1\big)}{p} = \frac{r(1-p)}{p^2}.$$

The mean is worth seeing intuitively. Reaching one success takes $1/p$ trials on average, of which one is the success itself, so $\frac{1}{p} - 1$ are failures. Reaching r successes just repeats that r times.

12.2 The Poisson distribution

Recall $b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$. Now suppose n is very large while the mean $np = \mu$ stays moderate — so p is tiny. Many trials, each almost certainly a failure. What happens to the pmf?

Theorem 12.6 (Poisson limit)**.** *If $n \to \infty$ and $p \to 0$ with $np = \mu$ fixed, then for each $x = 0, 1, 2, \dots$*

$$\binom{n}{x} p^x (1-p)^{n-x} \longrightarrow \frac{\mu^x}{x!} e^{-\mu}.$$

Sketch (this is Homework 5). Write $p = \mu/n$ and expand:

$$\binom{n}{x}\Big(\frac{\mu}{n}\Big)^x\Big(1 - \frac{\mu}{n}\Big)^{n-x} = \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1} \cdot \frac{\mu^x}{x!} \cdot \underbrace{\Big(1 - \frac{\mu}{n}\Big)^n}_{\to e^{-\mu}} \cdot \underbrace{\Big(1 - \frac{\mu}{n}\Big)^{-x}}_{\to 1},$$

using $\lim_{n\to\infty}(1 + \frac{a}{n})^n = e^a$. □

Definition 12.7 (Poisson distribution)**.** $X \sim \mathrm{Poisson}(\mu)$ if

$$\mathbb{P}\{X = x\} = \frac{\mu^x}{x!} e^{-\mu}, \qquad x = 0, 1, 2, \dots$$

Remark 12.8 (Sanity check)**.** $\displaystyle\sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{-\mu} e^{\mu} = 1.$ ✓

Theorem 12.9. *If $X \sim \mathrm{Poisson}(\mu)$ then $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \mu$.*

Proof. $\mathbb{E}[X] = \sum_{x\ge1} x \frac{\mu^x}{x!} e^{-\mu} = \mu e^{-\mu} \sum_{x\ge1} \frac{\mu^{x-1}}{(x-1)!} = \mu e^{-\mu} e^{\mu} = \mu$ (the $x = 0$ term contributes nothing, and $x/x! = 1/(x-1)!$). The same trick twice gives

$$\mathbb{E}[X(X-1)] = \sum_{x\ge2} x(x-1) \frac{\mu^x}{x!} e^{-\mu} = \mu^2 e^{-\mu} \sum_{x\ge2} \frac{\mu^{x-2}}{(x-2)!} = \mu^2 e^{-\mu} e^{\mu} = \mu^2,$$

so $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \mu^2 + \mu$ and $\mathrm{Var}(X) = \mu^2 + \mu - \mu^2 = \mu$. □

Remark 12.10. Mean = variance is the fingerprint of the Poisson. Count data whose variance noticeably exceeds its mean ("overdispersed") is telling you the Poisson model is wrong.

12.3 The Poisson process

The Poisson distribution arose above as a limit of counting. It also arises directly, from a short list of assumptions about events in *time*.

Think of calls arriving at a call center. Suppose:

- In a short interval of length Δt, the probability that *exactly one* call arrives is
$$\alpha\,\Delta t + o(\Delta t),$$
where $o(\Delta t)$ denotes any quantity with $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$;
- the probability that *two or more* calls arrive in Δt is $o(\Delta t)$;
- the number of calls arriving in Δt is independent of the number that arrived before this interval.

Definition 12.11 (Poisson process)**.** A stream of events satisfying the three conditions above is a **Poisson process** with rate α.

Theorem 12.12. *Let $P_k(t)$ be the probability that exactly k events occur in an interval of length t. Then*
$$P_k(t) = \frac{(\alpha t)^k}{k!} e^{-\alpha t},$$
i.e. the count over an interval of length t is Poisson(μ) *with $\mu = \alpha t$.*

Remark 12.13. So α is the expected number of events *per unit time*, and $\mu = \alpha t$ is the expected number in a window of length t. The two roles of the parameter are the usual source of confusion: "4 calls per hour" is $\alpha = 4$; over 15 minutes the count is Poisson(1), not Poisson(4).

Example 12.14. A call center receives calls as a Poisson process with $\alpha = 4$ per hour.
$$\mathbb{P}\{\text{no calls in 30 minutes}\} = P_0(0.5) = e^{-2} \approx 0.135,$$
$$\mathbb{P}\{\text{at least 2 calls in an hour}\} = 1 - P_0(1) - P_1(1) = 1 - e^{-4} - 4e^{-4} \approx 0.908.$$

Remark 12.15. The waiting *time* between consecutive events of a Poisson process is exponentially distributed — we prove this in Lecture 15, and it is the cleanest link between the discrete and continuous halves of the course.

12.4 Summary: the discrete families

Distribution	**Scenario**	**Support**	**pmf** $\mathbb{P}\{X = k\}$	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
Bernoulli(p)	single trial, success rate p	$k \in \{0, 1\}$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial(n, p)	number of S in n i.i.d. Bernoulli trials	$k = 0, \ldots, n$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Hypergeometric(n, M, N)	sampling *without* replacement: N total, M good, n drawn	k feasible	$\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$n\dfrac{M}{N}$	$\dfrac{N-n}{N-1}\, n\dfrac{M}{N}\Big(1 - \dfrac{M}{N}\Big)$
Geometric(p)	number of trials until the 1st S	$k = 1, 2, \ldots$	$(1-p)^{k-1}p$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
NegBinomial(r, p)	number of F before the r-th S	$k = 0, 1, 2, \ldots$	$\binom{k+r-1}{k} p^r (1-p)^k$	$r\Big(\dfrac{1}{p} - 1\Big)$	$\dfrac{r(1-p)}{p^2}$
Poisson(μ)	number of occurrences in a window of a Poisson process	$k = 0, 1, 2, \ldots$	$\dfrac{\mu^k}{k!} e^{-\mu}$	μ	μ

Remark 12.16 (How they connect)**.**

$$\text{Bernoulli}(p) \xrightarrow[\text{count } S]{n \text{ trials}} \text{Binomial}(n, p) \xrightarrow[np=\mu]{n\to\infty,\ p\to 0} \text{Poisson}(\mu), \qquad \text{Binomial}(n, p) \xleftarrow[n \ll N]{} \text{Hypergeometric}(n, M, N$$

$$\text{Geometric}(p) = \text{NegBinomial}(1, p) \text{ (up to the shift)}, \qquad \text{NegBinomial}(r, p) = \text{sum of } r \text{ independent geometrics (co}$$

Everything on this page comes from one experiment — a coin with bias p — and differs only in what is held fixed and what is counted.