14. Poisson Convergence

We now turn our attention to one of the more ubiquitous discrete limiting distributions, the Poisson, beginning with the "law of rare events" (or "weak law of small numbers"). It is instructive to compare the following result and its corresponding proof with that of the Lindeberg-Feller theorem.

Theorem 14.1. *For each $n \in \mathbb{N}$, let $X_{n,1}, ..., X_{n,n}$ be independent with $P(X_{n,m} = 1) = p_{n,m}$ and $P(X_{n,m} = 0) = 1 - p_{n,m}$.*

Suppose that as $n \to \infty$,

(1) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$,
(2) $\max_{1 \le m \le n} p_{n,m} \to 0$.

If $S_n = X_{n,1} + ... + X_{n,m}$, then $S_n \Rightarrow W$ where $W \sim$ Poisson(λ).

Proof. The characteristic function of $X_{n,m}$ is

$$\varphi_{n,m}(t) = E\left[e^{itX_{n,m}}\right] = 1 - p_{n,m} + p_{n,m}e^{it},$$

so it follows from the independence assumption that S_n has ch.f.

$$\varphi_{S_n}(t) = E\left[e^{itS_n}\right] = \prod_{m=1}^{n}\left[1 + p_{n,m}\left(e^{it} - 1\right)\right].$$

Now, for $p \in [0, 1]$,

$$\left|\exp\left(p(e^{it} - 1)\right)\right| = \exp\left[\text{Re}\left(p(e^{it} - 1)\right)\right] = \exp\left[p\left(\cos(t) - 1\right)\right] \le 1$$

and $\left|1 + p(e^{it} - 1)\right| = \left|p \cdot e^{it} + (1 - p) \cdot 1\right| \le 1$ since it is on the line segment connecting 1 to e^{it}, which is a chord of the unit circle in $\mathbb{C}$.

Thus, taking $z_m = 1 + p_{n,m}\left(e^{it} - 1\right)$, $w_m = \exp\left(p_{n,m}(e^{it} - 1)\right)$ in Lemma 13.1, we have

$$\begin{aligned}\left|\prod_{m=1}^{n}\left(1 + p_{n,m}\left(e^{it} - 1\right)\right) - \exp\left(\sum_{m=1}^{n} p_{n,m}(e^{it} - 1)\right)\right| &= \left|\prod_{m=1}^{n}\left(1 + p_{n,m}\left(e^{it} - 1\right)\right) - \prod_{m=1}^{m}\exp\left(p_{n,m}(e^{it} - 1)\right)\right| \\ &\le \sum_{m=1}^{n}\left|\exp\left(p_{n,m}(e^{it} - 1)\right) - \left[1 + p_{n,m}\left(e^{it} - 1\right)\right]\right|.\end{aligned}$$

By assumption 2, we have $\max_{1 \le m \le n} p_{n,m} \le \frac{1}{2}$, and thus $\max_{1 \le m \le n}\left|p_{n,m}\left(e^{it} - 1\right)\right| \le 1$, for n sufficiently large.

Using Lemma 13.2, we conclude that

$$\begin{aligned}\left|\prod_{m=1}^{n}\left(1 + p_{n,m}\left(e^{it} - 1\right)\right) - \exp\left(\sum_{m=1}^{n} p_{n,m}(e^{it} - 1)\right)\right| &\le \sum_{m=1}^{n}\left|\exp\left(p_{n,m}(e^{it} - 1)\right) - \left[1 + p_{n,m}\left(e^{it} - 1\right)\right]\right| \\ &\le \sum_{m=1}^{n} p_{n,m}^2\left|\left(e^{it} - 1\right)\right|^2 \le 4\sum_{m=1}^{n} p_{n,m}^2 \\ &\le 4\max_{1 \le m \le n} p_{n,m}\sum_{m=1}^{n} p_{n,m} \to 0\end{aligned}$$

by assumptions 1 and 2.

Therefore, since assumption 1 implies $\exp\left(\sum_{m=1}^{n} p_{n,m}(e^{it}-1)\right) \to e^{\lambda(e^{it}-1)}$,

$$\varphi_{S_n}(t) = \prod_{m=1}^{n}\left[1+p_{n,m}\left(e^{it}-1\right)\right] \to e^{\lambda(e^{it}-1)} = \varphi_W(t),$$

and the result follows from the continuity theorem. □

An easy consequence of Theorem 14.1 is

Theorem 14.2. *Let $X_{n,m}$, $1 \le m \le n$ be independent $\mathbb{N}_0$-valued random variables with $P(X_{n,m}=1) = p_{n,m}$ and $P(X_{n,m} \ge 2) = \varepsilon_{n,m}$ where*

(1) $\sum_{m=1}^{n} p_{n,m} \to \lambda$,
(2) $\max_{1\le m\le n} p_{n,m} \to 0$,
(3) $\sum_{m=1}^{n} \varepsilon_{n,m} \to 0$

as $n \to \infty$. Then $S_n = X_{n,1} + \ldots + X_{n,n}$ converges weakly to the Poisson(λ) distribution.

Proof. Let $Y_{n,m} = 1\{X_{n,m}=1\}$ and $T_n = \sum_{m=1}^{n} Y_{n,m}$. Theorem 14.1 implies $T_n \Rightarrow W \sim$ Poisson(λ), so, since $P(S_n \ne T_n) \le \sum_{m=1}^{n} \varepsilon_{n,m} \to 0$ (thus $S_n - T_n \to_p 0$), $S_n = T_n + (S_n - T_n) \Rightarrow W$ by Slutsky's theorem. □

It is worth mentioning that, just as in the normal case, independence is not a strictly necessary condition for Poisson convergence. To relax the assumption in general one needs to use a different proof strategy than convergence of characteristic functions. However, it is sometimes possible to give direct proofs by simple calculations.

Example 14.1 (Hat check, Lazy Secretary, etc...)**.**

Define $X_{n,m} = X_{n,m}(\pi) = 1\{\pi(m) = m\}$ where π is chosen from the uniform measure on S_n, the symmetric group on $\{1, ., , , n\}$.

Then $T_n = \sum_{m=1}^{n} X_{n,m}$ is the number of fixed points in a random permutation of length n.

Inclusion-exclusion gives the probability of at least one fixed point as

$$\begin{aligned} P(T_n > 0) = P\left(\bigcup_{m=1}^{n}\{X_{n,m}=1\}\right) &= \sum_{m=1}^{n} P(X_{n,m}=1) - \sum_{l<m} P(X_{n,l} = X_{n,m} = 1) \\ &+ \sum_{k<l<m} P(X_{n,k} = X_{n,l} = X_{n,m} = 1) - \ldots + (-1)^{n+1} P(X_{n,1} = \ldots = X_{n,n} = 1) \\ &= \sum_{k=1}^{n} (-1)^{k+1}\binom{n}{k}\frac{(n-k)!}{n!} = \sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!} \end{aligned}$$

since the number of permutations with k specified fixed points is $(n-k)!$.

It follows that the probability of a derangement is given by

$$P(T_n = 0) = 1 - P(T_n > 0) = 1 - \sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!} = \sum_{k=0}^{n}\frac{(-1)^k}{k!} \to e^{-1}.$$

To compute other values of the mass function for T_n, note that

$$P(T_n = m) = \binom{n}{m}\frac{(n-m)!}{n!}P(T_{n-m} = 0) = \frac{1}{m!}P(T_{n-m} = 0) \to \frac{1}{m!}e^{-1}.$$

Therefore, for every $x \in \mathbb{R}$, we have

$$P(T_n \le x) = \sum_{m\in\mathbb{N}_0:m\le x} P(T_n = m) \to \sum_{m\in\mathbb{N}_0:m\le x} \frac{1}{m!}e^{-1}$$

(as the above sums contain finitely many terms), so the number of fixed points in a permutation of length n converges weakly to $W \sim \text{Poisson}(1)$ as $n \to \infty$.

For most common discrete random variables, rather than memorize the p.m.f.s, one needs only to understand the stories that they tell and the probabilities follow easily from combinatorial considerations.

For example, if $X \sim \text{Binomial}(n, p)$, then the story is that X gives the number of heads in n independent flips of a coin with heads probability p. To compute the probability that $X = k$ for $k = 0, 1, ..., n$ we note that any sequence of k heads and $n-k$ tails has probability $p^k(1-p)^{n-k}$ by independence. Since the number of such sequences is determined by specifying where the heads occur, of which there are $\binom{n}{k}$ possibilities, the binomial p.m.f. is given by $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$, $k = 0, 1, ..., n$.

Similarly, if $Y \sim \text{Hypergeometric}(N, M, n)$, then the story is that we sample n items without replacement from a set of N items of which M are distinguished, and Y counts the number of distinguished items in our sample. For $k \le \min(n, M)$, there are $\binom{M}{k}$ ways to choose k distinguished items, $\binom{N-M}{n-k}$ ways to choose the remaining $n-k$ items, and $\binom{N}{n}$ possible samples, so $P(Y = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$.

Because the Poisson distribution assigns positive mass to infinitely many outcomes, determining the p.m.f. is not such a simple matter of counting. Nonetheless, the preceding results do supply us with an appropriate story:

For $0 \le s < t$, let $N(s, t)$ denote the number of occurrences of a given type of event in the time interval $(s, t]$ - say, the number of arrivals at a restaurant between s and t minutes after it opens. Suppose that

(1) The number of occurrences in disjoint time intervals are independent.
(2) The distribution of $N(s, t)$ depends only on $t - s$ (*stationary increments*).
(3) $P(N(0, h) = 1) = \lambda h + o(h)$.
(4) $P(N(0, h) \ge 2) = o(h)$.

Theorem 14.3. *If properties* $1-4$ *hold, then* $N(0, t) \sim$ *Poisson*(λt).

Proof. Define $X_{n,m} = N\left(\frac{(m-1)t}{n}, \frac{mt}{n}\right)$. Property 1 shows that $X_{n,1}, ..., X_{n,n}$ are independent; properties 2 and 3 show that $p_{n,m} = P(X_{n,m} = 1) = P(X_{n,1} = 1) = \frac{\lambda t}{n} + o\left(\frac{t}{n}\right)$; and properties 2 and 4 show that $\varepsilon_{n,m} = P(X_{n,m} \ge 2) = P(X_{n,1} \ge 2) = o\left(\frac{t}{n}\right)$.

Since $\sum_{m=1}^{n} p_{n,m} = n\left(\frac{\lambda t}{n} + o\left(\frac{1}{n}\right)\right) \to \lambda t$ and $\sum_{m=1}^{n} \varepsilon_{n,m} = no\left(\frac{1}{n}\right) \to 0$ as $n \to \infty$, Theorem 14.2 implies that $X_{n,1} + ... + X_{n,n} \Rightarrow W \sim \text{Poisson}(\lambda t)$. The result follows by observing that $X_{n,1} + ... + X_{n,n} = N(0, t)$ for all n. □

The random variables $N(0,t)$ as t ranges over $[0,\infty)$ are an example of a continuous time stochastic process:

Definition. A family of random variables $\{N(t)\}_{t\geq 0}$ is called a *Poisson process with rate* λ if it satisfies:

(1) For any $0 = t_0 < t_1 < ... < t_n$, the random variables $N(t_k) - N(t_{k-1})$, $k = 1, ..., n$ are independent;
(2) $N(t) - N(s) \sim \text{Poisson}\,(\lambda(t-s))$.

To better understand the process $\{N(t)\}_{t\geq 0}$, it is useful to consider the following construction which explains our "arrivals story" and provides a bridge between the Poisson and exponential distributions:
Let $\xi_1, \xi_2, ...$ be i.i.d. exponentials with mean λ^{-1} - that is $P(\xi_i > t) = e^{-\lambda t}$ for $t \geq 0$.
Define $T_n = \sum_{i=1}^n \xi_i$ and $N(t) = \sup\{n : T_n \leq t\}$.
If we think of the $\xi_i's$ as interarrival times, then T_n gives the time of the nth arrival and $N(t)$ is the number of arrivals by time t.
Since a sum of n i.i.d. Exponential(λ) R.V.s has a $\Gamma(n, \lambda^{-1})$ distribution*, we see that T_n has density

$$f_{T_n}(s) = \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} \text{ for } s \geq 0.$$

Accordingly,

$$P(N(t) = 0) = P(T_1 > t) = e^{-\lambda t} = e^{-\lambda t}\frac{(\lambda t)^0}{0!}$$

and

$$\begin{aligned} P(N(t) = n) = P(T_n \leq t < T_{n+1}) &= \int_0^t f_{T_n}(s) P(\xi_{n+1} > t-s) ds \\ &= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} e^{\lambda(s-t)} ds = e^{-\lambda t}\frac{\lambda^n}{n!}\int_0^t n s^{n-1} ds = e^{-\lambda t}\frac{(\lambda t)^n}{n!} \end{aligned}$$

for $n \geq 1$, so $N(t) \sim \text{Poisson}(\lambda t)$.

To check that the number of arrivals in disjoint intervals is independent, we note that for all $n \in \mathbb{N}$ and all $u > t > 0$,

$$\begin{aligned} P(T_{n+1} \geq u, N(t) = n) = P(T_{n+1} \geq u, T_n \leq t) &= \int_0^t f_{T_n}(s) P(\xi_{n+1} \geq u-s) ds \\ &= \int_0^t \frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s} e^{\lambda(s-u)} ds = e^{-\lambda u}\frac{(\lambda t)^n}{n!} = e^{-\lambda(u-t)} e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda(u-t)} P(N(t) = n), \end{aligned}$$

and thus

$$P(T_{n+1} \geq u | N(t) = n) = \frac{P(T_{n+1} \geq u, N(t) = n)}{P(N(t) = n)} = e^{-\lambda(u-t)}.$$

Writing $s = u - t$, the above is equivalent to

$$P(T_{n+1} - t \geq s | N(t) = n) = e^{-\lambda s}$$

for all $n \in \mathbb{N}$, $s, t > 0$.

It follows that $T_1' = T_{N(t)+1} - t$ is independent of $N(t)$ and

$$P(T_1' \geq s) = \sum_{n=0}^{\infty} P(T_{n+1} - t \geq s | N(t) = n) P(N(t) = n) = e^{-\lambda s},$$

hence $T_1' \sim \text{Exponential}(\lambda)$.

Setting $T'_k = T_{N(t)+k} - T_{N(t)+k-1} = \xi_{N(t)}$ for $k \geq 2$ and observing that

$$\begin{aligned} P(N(t) = n, T'_1 \geq u - t, T'_k \geq v_k \text{ for } k = 2, ..., K) \\ = P(T_n \leq t, T_{n+1} \geq u, T_{n+k} - T_{n+k-1} \geq v_k \text{ for } k = 2, ..., K) \\ = P(T_n \leq t, T_{n+1} \geq u) \prod_{k=2}^{K} P(\xi_{n+k} \geq v_k), \end{aligned}$$

we see that $T'_1, T'_2, ...$ are i.i.d. Exponential(λ) and independent of $N(t)$.

In other words, the arrivals after time t are independent of $N(t)$ and have the same distribution as the original arrival sequence.

(Essentially, this is due to the "memorylessness property" of the exponential.)

It follows that for any $0 = t_0 < t_1 < ... < t_n$, $N(t_1) - N(t_0), ..., N(t_n) - N(t_{n-1})$ are independent Poissons.

This is because the vector $(N(t_2) - N(t_1), ..., N(t_n) - N(t_{n-1}))$ is measurable with respect to $\sigma(T'_1, T'_2, ...)$ (where the $T'_i s$ are constructed as above with $t = t_1$) and so is independent of $N(t_1)$.

Then an induction argument gives

$$P(N(t_1) - N(t_0) = k_1, ..., N(t_n) - N(t_{n-1}) = k_n) = \prod_{i=1}^{n} e^{-\lambda(t_i - t_{i-1})} \frac{(\lambda(t_i - t_{i-1}))^{k_i}}{k_i!}.$$

* To keep the discussion self-contained, we show that sums of independent exponentials are gammas.

This is a situation where convolution is more convenient than characteristic functions.

First, recall that for $\alpha, \beta > 0$, $X \sim \Gamma(\alpha, \beta)$ has density $f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$, $x > 0$, where the gamma function Γ satisfies $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$.

Also, for $\lambda > 0$, $W \sim \text{Exp}(\lambda)$ has density $f_W(w) = \lambda e^{-\lambda x}$, $x > 0$. Thus $W \sim \Gamma(1, \lambda^{-1})$.

Suppose that $Y \sim \Gamma(n, \lambda^{-1})$ and $W \sim \text{Exp}(\lambda)$ are independent and set $Z = W + Y$.

Then Z has positive support and Theorem 6.6 shows that for $z > 0$,

$$\begin{aligned} f_Z(z) = f_W * f_Y(z) &= \int_{-\infty}^{\infty} f_W(z-y) f_Y(y) dy \\ &= \int_0^z \lambda e^{-\lambda(z-y)} \frac{\lambda^n}{(n-1)!} y^{n-1} e^{-\lambda y} dy \\ &= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda z} \int_0^z y^{n-1} dy = \frac{\lambda^{n+1}}{\Gamma(n+1)} z^{(n+1)-1} e^{-\lambda z}. \end{aligned}$$

It follows by induction that if $X_1, \ldots, X_n$ are i.i.d. $\text{Exp}(\lambda)$, then $X_1 + \ldots + X_n \sim \Gamma(n, \lambda^{-1})$.