Lecture 13. The Poisson distribution and approximation

Martin Klazar

January 5, 2021

The Poisson distribution

We begin with a definition.

Definition (Poisson random variable). *The Poisson random variable, or the RV with Poisson distribution, with mean (or parameter) $\mu \ge 0$ is any discrete random variable X on a probability space $(\Omega, \Sigma, \mathrm{Pr})$ such that $X(\Omega) = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ and for every $j \in \mathbb{N}_0$,*

$$\mathrm{Pr}(X = j) = \frac{e^{-\mu}\mu^j}{j!}\ .$$

$\square$

In particular, $\mathrm{Pr}(X = 0) = e^{-\mu}$. It is easy to see that it is a correct definition and such discrete RV exists because the probabilities of all values sum up to 1, see Lecture 9. Indeed,

$$\sum_{j=0}^{\infty} \frac{e^{-\mu}\mu^j}{j!} = \frac{1}{e^{\mu}} \sum_{j=0}^{\infty} \frac{\mu^j}{j!} = \frac{e^{\mu}}{e^{\mu}} = 1\ ,$$

by the Taylor series of the exponential function. Also, the mean (expectation) of any Poisson random variable with mean μ is indeed μ,

$$\mathbb{E}X = \sum_{j=0}^{\infty} j \cdot \frac{e^{-\mu}\mu^j}{j!} = \frac{\mu}{e^{\mu}} \sum_{j=1}^{\infty} \frac{\mu^{j-1}}{(j-1)!} = \frac{\mu e^{\mu}}{e^{\mu}} = \mu\ .$$

A *Poisson random variable* or a random variable with *Poisson distribution* is a Poisson random variable with some mean μ. We have the next nice property for these RVs.

Proposition (sum of Poissons). *Suppose that $X_1, X_2, \dots, X_n$ are independent Poisson random variables. Then $X := X_1 + X_2 + \cdots + X_n$*

is a Poisson random variable. Moreover, the mean of X is the sum of the means of the X_i.

Proof. Suppose that X_i has mean $\mu_i \in \mathbb{R}$ and that $j \in \mathbb{N}_0$. Then $X(\Omega) = \mathbb{N}_0$ and

$$\begin{aligned}
\Pr(X = j) &\overset{\text{def. of } X}{=} \sum_{\substack{j_i \in \mathbb{N}_0 \\ j_1+\cdots+j_n=j}} \Pr(X_1 = j_1 \wedge \cdots \wedge X_n = j_n) \\
&\overset{\text{indep. of } X_i}{=} \sum_{\substack{j_i \in \mathbb{N}_0 \\ j_1+\cdots+j_n=j}} \prod_{i=1}^{n} \frac{e^{-\mu_i}\mu_i^{j_i}}{j_i!} \\
&\overset{\text{algebra}}{=} \frac{1}{e^{\mu_1+\cdots+\mu_n} j!} \sum_{\substack{j_i \in \mathbb{N}_0 \\ j_1+\cdots+j_n=j}} \binom{j}{j_1,\, \ldots,\, j_n} \prod_{i=1}^{n} \mu_i^{j_i} \\
&\overset{\text{multin. thm.}}{=} \frac{e^{-\mu_1-\cdots-\mu_n}(\mu_1+\cdots+\mu_n)^j}{j!} .
\end{aligned}$$

The last line follows by the multinomial theorem. The theorem says that for any variables $x_1, \ldots, x_m$ and any $n \in \mathbb{N}_0$,

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{\substack{k_i \in \mathbb{N}_0 \\ k_1+\cdots+k_m=n}} \binom{n}{k_1,\, k_2,\, \ldots,\, k_m} x_1^{k_1} x_2^{k_2} \ldots x_m^{k_m}$$

where $\binom{n}{k_1,k_2,\ldots,k_m} = \frac{n!}{k_1!k_2!\ldots k_m!}$ are multinomial coefficients. □

• *Siméon D. Poisson (1781–1840)* was a French mathematician, engineer, and physicist. An interesting account on his life and career is given at `https://mathshistory.st-andrews.ac.uk/Biographies/Poisson/`

We describe a situation leading non-rigorously to a Poisson distribution. Imagine three-dimensional box (container) $K = [0,1]^3$ that has unit volume and contains $N \approx 10^{20}$ particles moving in various directions. We assume that all particles have the same speed, do not interact among themselves, and bounce perfectly elastically from the walls of K. The particles move "randomly": if $L \subset K$ is any small box inside the box K with volume $\mathrm{vol}(L)$ (or L is a more general set with defined volume), then any of the particles is found in L with "probability" $\frac{\mathrm{vol}(L)}{\mathrm{vol}(K)} = \mathrm{vol}(L)$. This means that if we watch any particle and its movement in K in time interval $[0,t]$, then — if t_L denotes the amount of time the particle spends in L — $\lim_{t\to+\infty} t_L/t = \mathrm{vol}(L)$. We further assume that all particles move mutually independently: the "events" of their occurrences in L are independent. Suppose that L has volume proportional to $\frac{1}{N}$, where N is the number of particles. Given $k \in \mathbb{N}_0$, what is the "probability" that we find in L exactly k particles? We again interpret this "probability" as the limit ratio of the amount of time during which the evolving system has in L exactly k particles, to the total time. We compute the "probability" with the help of the next properties of the exponential function.

Lemma. *The following hold.*

1. *For every $x \in \mathbb{R}$,*
$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x\ .$$
2. *More generally, if $x \in \mathbb{R}$ and $(x_n) \subset \mathbb{R}$ is a sequence such that $\lim_{n\to\infty} x_n = x$, then again*
$$\lim_{n\to\infty}\left(1+\frac{x_n}{n}\right)^n = e^x\ .$$

Proof. Part 1 is a well known limit. You can deduce part 2 from part 1 as an exercise. □

Proposition (on particles). *Suppose that $\lambda > 0$ is a real number, $L \subset K$ is a small box inside the unit box K with volume* $\mathrm{vol}(L) = \lambda/N$, *where $N \in \mathbb{N}$ is the (large) number of particles, and $k \in \mathbb{N}_0$. Then in the above described situation,*
$$\lim_{N\to\infty} \text{“}\mathrm{Pr}(L \textit{ contains exactly } k \textit{ particles})\text{”} = \frac{e^{-\lambda}\lambda^k}{k!}\ .$$
Thus the number of particles in L follows in limit the Poisson distribution with mean λ.

Proof. The above displayed "probability" is exactly
$$\binom{N}{k}\cdot(\lambda/N)^k\cdot(1-\lambda/N)^{N-k}\ .$$

The binomial coefficient counts ways to select an unordered k-tuple of particles of all N particles, the second factor is the “probability” that the k-tuple lies in L, and the third factor is the “probability” that none of the remaining $N-k$ particles lies in L. We use independence of occurrences of a particle in L, as well as the fact that the “probability” of the union of disjoint “events” equals to the sum of their “probabilities”. We rewrite the above displayed expression as

$$\frac{N(N-1)\dots(N-k+1)}{N^k}\cdot(1-\lambda/N)^{-k}\cdot(1-\lambda/N)^N\cdot\frac{\lambda^k}{k!}$$

where for $k=0$ the descending product is set to 1. For $N\to\infty$ the first two factors go to 1. By part 1 of the previous lemma, the third factor goes to $e^{-\lambda}$ and the limit follows. □

The above quotation marks do not look nicely but we have to write them if we want to be consistent. Given initial positions and directions of movements of the particles, the evolution of the system is completely deterministic and involves no randomness. We proved that for $N\to\infty$ the above expression has limit $e^{-\lambda}\lambda^k/k!$ but there is no probability space and no random variable.

J. Beck (see Lecture 9) proved in Theorem 1 in his article Deterministic approach to the kinetic theory of gases, *J. Stat. Phys.* **138** (2010), 160–269 (the statement of the theorem takes over one page and its proof more than fifty pages) that an overwhelming majority — in the sense of initial positions and initial velocities[1] of the particles — of systems of N particles in the box K behave for large enough N and sufficiently long time t of evolution in a very precise accord with a Poisson distribution. All estimates in his Theorem 1 are (unlike our Proposition above) explicit, there are no limits with unquantified speed of convergence. From another Theorem 2 in the same article he deduces (see pp. 181–182) as an illustration the next result on the law of large numbers (closeness of a random variable to its mean) for systems of particles.

A cubic container with side 1 meter contains a system of $N=10^{27}$ point particles (molecules) moving in various di-

[1]Physical terminology: “speed” is a scalar quantity, measured in $\mathrm{m\,s^{-1}}$, “velocity” is a vector quantity, here an element of $\mathbb{R}^3$.

rections with the same speed 10^3 ms^{-1}. The particles do not mutually interact, there are no collisions between them (Beck remarks that this is unrealistic, in reality mutual collisions are frequent), and they elastically reflect from the walls of the container. One selects in the container a subset A with volume 0.5 m^3. In the given moment the system of particles is *balanced*, if the number of particles in A equals

$$\frac{(1 \pm 0.001)N}{2}\ ,$$

i.e. deviates from the expected value $\frac{N}{2}$ with the relative error at most one tenth of a percent. We let the system evolve for 100 years ($\approx 3 \cdot 10^9$ seconds). For how long during the century will the system be unbalanced? It follows from the estimates in Theorem 2 that more than 99.99% of such systems (in the sense of initial positions and initial velocities of the particles) will be unbalanced during the century for only less than 10 seconds!

We cast the previous Proposition more generally and more rigorously. Recall (from Lecture 9) that a random variable X has *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0,1]$ if $X(\Omega) = \{0, 1, \ldots, n\}$ and for every value k one has that $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$; so that X counts heads in a series of n independent tosses of a general coin with $\Pr(\text{head}) = p$.

Theorem (binomial → Poisson). *Suppose that random variables X_n, $n \in \mathbb{N}$, have binomial distributions with parameters $n \in \mathbb{N}$ and $p_n = \frac{\lambda_n}{n} \in [0,1]$ where $\lambda_n \to \lambda$ for $n \to \infty$ and a fixed real $\lambda \ge 0$. Then for every $k \in \mathbb{N}_0$,*

$$\lim_{n \to \infty} \Pr(X_n = k) = \frac{e^{-\lambda} \lambda^k}{k!}\ .$$

In this sense binomial distributions converge to a Poisson distribution.

Proof. As in the previous Proposition, for $n \to \infty$ in

$$\Pr(X_n = k) = \binom{n}{k}\left(\frac{\lambda_n}{n}\right)^k\left(1 - \frac{\lambda_n}{n}\right)^{n-k}$$
$$= \frac{n(n-1)\dots(n-k+1)}{n^k} \cdot \left(1 - \frac{\lambda_n}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda_n}{n}\right)^n \cdot \frac{\lambda_n^k}{k!}$$

the first two factors go to 1, the third one goes to $e^{-\lambda}$ by part 2 of the previous Lemma, and the fourth factor goes to $\lambda^k/k!$. The limit follows. □

This is Theorem 5.5 in Mitzenmacher and Upfal where the proof is more complicated. The previous theorem applies to the balls and bins model with m balls and n bins: if $n \to \infty$ and $\frac{m}{n} \to \lambda \geq 0$ then the number of balls in a given bin has in limit the Poisson distribution with mean λ.

Janson inequalities

We explain so called *Janson inequalities* (Alon and Spencer, pp. 115–117). They are related to LLL and strengthen it in a way, they show that if some events $B_1, \dots, B_n$ are only mildly dependent then

$$\Pr(\textstyle\bigcap_{i=1}^n \overline{B_i}) \approx \prod_{i=1}^n (1 - \Pr(B_i))$$

— equality would hold here if B_i were independent. Before stating the inequalities — we will not prove them, for proofs see pp. 117–119 in Alon and Spencer — we introduce the formal setup of random subsets, Alon and Spencer give it only in intuitive terms.

Definition (random subsets). Let $U \neq \emptyset$ be a finite set and

$$\overline{p} = (p_r \in [0,\,1] \mid r \in U) \in [0,\,1]^U$$

be a $|U|$-tuple of constants labeled by elements in U. We say that a pair (P, X) is a *$\overline{p}$-random subset of U* if $P = (\Omega, \Sigma, \Pr)$ is a probability space and X is a random variable on P such that the following hold.

1. The values of X are

$$X(\Omega) = \{0,\,1\}^U = \{\overline{a} = (a_r \mid r \in U) \mid a_r \in \{0,\,1\}\}\ ,$$

binary $|U|$-tuples indexed by elements in U, and

$$\Pr(X_r = a_r = 1) = p_r$$

for every $r \in U$.

2. For any $V \subset U$, the events $\{(X_r = 1) \mid r \in V\}$ are independent.

$\square$

We think of X as the random subset $R \subset U$ obtained by putting elements $r \in U$ mutually independently in R with probabilities p_r. As in the lemmas in Lectures 8 and 9, part 2 is equivalent to independence of any collection of events $\{(X_r = b_r) \mid r \in V\}$ where $V \subset U$ and $b_r \in \{0, 1\}$.

Proposition (random subsets exist). *For every U and $\overline{p}$ as in the previous definition, there exists a $\overline{p}$-random subset of U.*

Proof. 1. We consider the finite probability space

$$P = (\Omega,\, \Sigma,\, \Pr) = \left(\{0,\, 1\}^U,\, \mathcal{P}\left(\{0,\, 1\}^U\right),\, \Pr\right)$$

where any atom $\{\overline{a}\}$ has probability

$$\Pr(\{\overline{a}\}) = \Pr\big(\{(a_r \mid r \in U)\}\big) := \prod_{r \in U,\, a_r = 1} p_r \prod_{r \in U,\, a_r = 0} (1 - p_r)\ .$$

From the identity

$$1 = \prod_{r \in U} (p_r + (1 - p_r)) = \sum_{\overline{a} \in \Omega} \Pr(\{\overline{a}\})$$

it follows that this is a correct definition of a probability space, probabilities of the atoms are nonnegative and sum up to 1. We define the random variable X as the identity map, $X(\overline{a}) = \overline{a}$. Let an $r \in U$ be given. We set $U' := U \setminus \{r\}$ and get that

$$\begin{aligned}
\Pr(X_r = 1) &= \sum_{\overline{a} \in \Omega,\, a_r = 1} \Pr(\{\overline{a}\}) = p_r \sum_{\substack{\overline{a} \in \Omega \\ a_r = 1}} \prod_{\substack{s \in U' \\ a_s = 1}} p_s \prod_{\substack{s \in U' \\ a_s = 0}} (1 - p_s) \\
&= p_r \prod_{s \in U'} (p_s + (1 - p_s)) = p_r\ .
\end{aligned}$$

2. Let $V \subset U$ be any subset. We set $U' := U \setminus V$ and compute that

$$\begin{aligned}
\Pr(\bigcap_{r\in V}(X_r = 1)) &= \sum_{\substack{\overline{a}\in\Omega \\ r\in V \Rightarrow a_r = 1}} \Pr(\{\overline{a}\}) \\
&= \prod_{r\in V} p_r \sum_{\substack{\overline{a}\in\Omega \\ r\in V \Rightarrow a_r = 1}} \prod_{\substack{r\in U' \\ a_r = 1}} p_r \prod_{\substack{r\in U' \\ a_r = 0}} (1 - p_r) \\
&= \prod_{r\in V} p_r \prod_{r\in U'} (p_r + (1 - p_r)) \\
&= \prod_{r\in V} \Pr(X_r = 1)\ .
\end{aligned}$$

□

Let U and $\overline{p}$ be as above and R (i.e. (P, X)) be a $\overline{p}$-random subset of U. Let $I \neq \emptyset$ be a finite set, $\{A_i \mid i \in I\} \subset \mathcal{P}(U)$ be a collection of subsets of U, and $\{B_i \mid i \in I\}$ be the corresponding events in P that $A_i \subset R$, so that B_i is the event

$$\bigwedge_{r\in A_i}(X_r = 1) = \bigcap_{r\in A_i}(X_r = 1)\ .$$

For $i, j \in I$ we write $i \sim j$ if $i \neq j$ and $A_i \cap A_j \neq \emptyset$.

Theorem (the Janson inequality). *Let an $\varepsilon > 0$ be given, U and $\overline{p}$ be as above, (P, X) be a $\overline{p}$-random subset of U, $\{A_i \mid i \in I\}$ be some subsets of U, and $\{B_i \mid i \in I\}$ be the corresponding events in P as above. We define*

$$\Delta := \sum_{\substack{i,j\in I \\ i\sim j}} \Pr(B_i \cap B_j),\ \ M := \prod_{i\in I} \Pr(\overline{B_i}) \ \text{ and } \ \mu := \sum_{i\in I} \Pr(B_i)\ .$$

In Δ we sum over $I \times I$, M would be the probability that none of the events B_i occurs if all B_i were independent, and $\mu = \mathbb{E}X$ where X is the sum of the indicator variables[2] *for the events B_i.*

If $\Pr(B_i) \leq \varepsilon$ for every $i \in I$ then

$$M \leq \Pr(\bigcap_{i\in I} \overline{B_i}) \leq M e^{\Delta/2(1-\varepsilon)} \ \text{ and } \ \Pr(\bigcap_{i\in I} \overline{B_i}) \leq e^{-\mu+\Delta/2}\ .$$

Here $\Pr(\bigcap_{i\in I} \overline{B_i}) = \Pr(X = 0)$.

[2] Recall that the indicator RV $X = X_A$ of an event $A \in \Sigma$ in a probability space $(\Omega, \Sigma, \Pr)$ is defined by setting $X(\omega) = 1$ if $\omega \in A$, and $X(\omega) = 0$ if $\omega \in \Omega \setminus A$.

Theorem (the extended Janson inequality). *Under the assumptions of the previous theorem and the assumption that* $\Delta \geq \mu$,

$$\Pr(\bigcap_{i \in I} \overline{B_i}) \leq e^{-\mu^2/2\Delta} .$$

• *Svante Janson* is a Swedish mathematician, member of the Royal Swedish Academy of Sciences since 1994, and professor of mathematics at Uppsala University since 1987. As a child prodigy he matriculated at Uppsala University at age 13 in 1968.

As an application of the first Janson inequality (estimate)

$$M \leq \Pr(\bigcap_{i \in I} \overline{B_i}) \leq M e^{\Delta/2(1-\varepsilon)}$$

we refine the theorem in (Martin Tancer's) Lecture 6 that $p = p(n) = \frac{1}{n}$ is a threshold function for triangle containment in the random graph $G(n, p)$. In our terminology, for any $p \in [0, 1]$ the random graph $G(n, p)$ is the $\overline{p}$-random subset of U where $U = \binom{[n]}{2}$ ($[n] = \{1, 2, \dots, n\}$) and the vector $\overline{p} = (p, p, \dots, p)$ has length $\binom{n}{2}$.

Theorem (inside the threshold). *For any real number* $c > 0$*, if* $p = p(n) = c/n$ *then*

$$\lim_{n \to \infty} \Pr(G(n,\, p) \not\supset \Delta) = e^{-c^3/6} .$$

Proof. To use the first Janson inequality, we set (as above) $U = \binom{[n]}{2}$, $\overline{p} = \overline{p}(n) = (c/n, c/n, \dots, c/n)$, $I = \binom{[n]}{3}$, and for any $S \in I$ we consider the triangle $A_S := \binom{S}{2} \subset U$. Thus the corresponding event B_S, the event that $G(n, p) \supset A_S$, has probability $\Pr(B_S) = p^3$, and we may set $\varepsilon := p^3 = O(1/n^3)$. The event $\bigcap_{S \in I} \overline{B_S}$ is the event that $G(n, p) \not\supset \Delta$. For $S, T \in I$, the relation $S \sim T$ means that the triangles A_S and A_T are distinct but share an edge e. Then $\Pr(B_S \cap B_T) = p^5$ and (for $n \geq 3$)

$$\Delta = \sum_{\substack{S, T \in I \\ S \sim T}} \Pr(B_S \cap B_T) = \underbrace{3\binom{n}{3}(n-3)}_{e + S + (T \setminus e)}(c/n)^5 = O(1/n) .$$

Also,

$$\begin{aligned}
\lim_{n\to\infty} M &= \lim_{n\to\infty} \prod_{S\in I} (1 - \Pr(B_S)) \\
&= \lim_{n\to\infty} \left(1 - (c/n)^3\right)^{n(n-1)(n-2)/6} \\
&= e^{-c^3/6} \,.
\end{aligned}$$

The last limit follows by writing for $x \in (0, \frac{1}{2})$ the difference $1 - x$ as $e^{\log(1-x)} = e^{-x+O(x^2)}$. Since $e^{\Delta/2(1-\varepsilon)} = e^{O(1/n)} \to 1$ for $n \to \infty$, the first Janson inequality yields that

$$\lim_{n\to\infty} \Pr(G(n,\, c/n) \not\supset \Delta) = \lim_{n\to\infty} \Pr(\textstyle\bigcap_{S\in I} \overline{B_S}) = \lim_{n\to\infty} M = e^{-c^3/6} \,.$$

□

For $c \to 0$, respectively $c \to +\infty$, the probability goes indeed to 1, respectively to 0, as it should since $1/n$ is threshold edge probability for the random graph to contain a triangle. This is taken from Alon and Spencer, pp. 156–157.

Thank you!

(final version of January 13, 2021)