Lecture 8

1 Common Distributions

1.1 Poisson Distribution – $\mathrm{Poi}(\lambda)$

The Poisson distribution, parametrized with a real number $\lambda > 0$, is a distribution μ over $\mathbb{N} \cup \{0\}$, which is defined as

$$\mu(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!},$$

for every $k \in \mathbb{N} \cup \{0\}$. This is indeed a distribution as the support of μ is the countably infinite set $\mathbb{N} \cup \{0\}$, for every $k \in \mathbb{N} \cup \{0\}$ we have $\mu(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \geq 0$, and

$$\sum_k \mu(k) = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1,$$

where the penultimate equality holds by the Taylor series of the exponential function.

We have already seen this distribution (in Lecture 3): Let $\pi \in S_n$ be a permutation chosen uniformly at random and let X be the number of fixed points in π (i.e., the number of elements $1 \leq i \leq n$ such that $\pi(i) = i$). We proved that

$$\lim_{n\to\infty} \mathbb{P}(X = k) = e^{-1} \cdot \frac{1}{k!},$$

holds for every $k \geq 0$. This is indeed a Poisson distribution with parameter $\lambda = 1$. This is an example of a general phenomenon: for any sequence of independent (or "almost independent") events, each with a small probability to occur, such that we expect a constant number λ of events to hold, the number of events that will occur converges to the Poisson distribution $\mathrm{Poi}(\lambda)$ as the number of events tends to infinity. Formally:

Theorem 1.1 (Poisson limit theorem)**.** *Let $\{p_n\}_{n=1}^{\infty}$ be a sequence of probabilities (i.e., real numbers between 0 and 1), satisfying $\lim_{n\to\infty} np_n = \lambda$, for some real number $\lambda > 0$. Then the distribution $\mathrm{Bin}(n, p_n)$ converges to the distribution $\mathrm{Poi}(\lambda)$ as n tends to infinity, that is, for every integer $k \geq 0$ it holds that*

$$\lim_{n\to\infty} \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Proof. Fix an integer $k \geq 0$. Then

$$\binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = \frac{n(n-1) \cdot \ldots \cdot (n-k+1)}{k!} \cdot p_n^k \cdot (1 - p_n)^{n-k}.$$

Observe that

$$\begin{aligned}
\lim_{n\to\infty} n(n-1)\cdot\ldots\cdot(n-k+1)\cdot p_n^k &= \lim_{n\to\infty}(np_n)^k \cdot \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{n^k} \\
&= \lim_{n\to\infty}(np_n)^k \cdot \lim_{n\to\infty} 1\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right) \\
&= \lambda^k \cdot \prod_{i=0}^{k-1}\lim_{n\to\infty}\left(1-\frac{i}{n}\right) \\
&= \lambda^k.
\end{aligned}$$

We next show that $\lim_{n\to\infty}(1-p_n)^{n-k} = e^{-\lambda}$. In order to do so, we will state and prove a simple fact from calculus.

Lemma 1.2. *Let $0 \le p \le 1$ be a real number. Then*

(a) $1 - p \le e^{-p}$.

(b) If, moreover, $p \le 1/2$, then $1 - p \ge e^{-p-2p^2}$.

Proof. It follows by the Taylor Polynomial Approximation Theorem that

$$e^{-p} = 1 - p + \frac{e^{-c}}{2}\cdot p^2,$$

where $0 \le c \le p$ is some real number. Since, clearly, $\frac{e^{-c}}{2}\cdot p^2 \ge 0$, this proves (a).

Similarly

$$e^{-p-2p^2} = 1 - (p+2p^2) + \frac{1}{2}\cdot(p+2p^2)^2 - \frac{e^{-c}}{3!}\cdot(p+2p^2)^3,$$

where $0 \le c \le p$ is some real number. Since, clearly, $\frac{e^{-c}}{3!}\cdot(p+2p^2)^3 \ge 0$, it follows that

$$e^{-p-2p^2} \le 1 - (p+2p^2) + \frac{1}{2}\cdot(p+2p^2)^2 = 1 - p - 1.5p^2 + 2p^3 + 2p^4 \le 1 - p,$$

where the last inequality holds for $p \le 1/2$. This proves (b). □

We can now return to the proof of Theorem 1.1. It follows by Lemma 1.2(a) that

$$\lim_{n\to\infty}(1-p_n)^{n-k} \le \lim_{n\to\infty} e^{-np_n+kp_n} = e^{-\lambda},$$

where the last equality holds since $\lim_{n\to\infty} np_n = \lambda$ is finite and thus $\lim_{n\to\infty} p_n = 0$.

Now, observe that

$$\lim_{n\to\infty} np_n^2 = 0.$$

Hence, it follows by Lemma 1.2(b) that

$$\lim_{n\to\infty}(1-p_n)^{n-k} \ge \lim_{n\to\infty} e^{-np_n-2np_n^2+kp_n+2kp_n^2} = e^{-\lambda}.$$

It follows by the Squeeze Theorem (a.k.a. the Sandwich Theorem) that

$$\lim_{n\to\infty}(1-p_n)^{n-k} = e^{-\lambda}.$$

We conclude that

$$\begin{aligned}\lim_{n\to\infty}\binom{n}{k}\cdot p_n^k\cdot(1-p_n)^{n-k} &= \frac{1}{k!}\cdot\lim_{n\to\infty} n(n-1)\cdot\ldots\cdot(n-k+1)\cdot p_n^k\cdot\lim_{n\to\infty}(1-p_n)^{n-k}\\ &= e^{-\lambda}\cdot\frac{\lambda^k}{k!},\end{aligned}$$

as claimed. □

1.1.1 Poisson distribution in nature

Consider some company that receives phone calls in a given hour. Suppose we know that, on average, the company receives λ calls per hour. We would like to know how the number of phone calls the company receives is distributed. We can try to use the following modeling: divide the hour into 3600 seconds, where in every second the company will receive a call with probability $\frac{\lambda}{3600}$, independently of all other seconds, leading to a binomial distribution with parameters 3600 and $\lambda/3600$. This modeling has the following two flaws:

1. It does not take into account the fact that the company can receive more than one call in one second.
2. The number 3600 was chosen arbitrarily and there is no guarantee that it will suffice in practice.

We could try to overcome both problems by dividing the hour into more "blocks", say n blocks, where every "block" has a probability of $\frac{\lambda}{n}$ to get a call, independently of all other blocks. However, one could argue that the two problems were not actually solved, though they were somewhat reduced. It seems that, taking the limit as n tends to infinity, will solve the two aforementioned problems. Observe that any partition of the hour into n blocks entails that the number of calls received by the company during this hour, is distributed binomially with parameters n and λ/n. Hence, we could apply the Poisson Limit Theorem to infer that, as n tends to infinity, this distribution tends to $\mathrm{Poi}(\lambda)$.

1.2 Negative-Binomial Distribution – $\mathrm{NB}(r,p)$

The Negative-Binomial distribution, parametrized with a positive integer r and a real number $0<p<1$, is a distribution μ over $\{r, r+1, \ldots\}$, which is defined as

$$\mu(n) = \binom{n-1}{r-1}\cdot p^r\cdot(1-p)^{n-r},$$

for every integer $n \geq r$. We claim that this is indeed a distribution. The support of μ is countably infinite and $\mu(n) \geq 0$ for every integer $n \geq r$. Hence, it suffices to prove that $\sum_n \mu(n) = 1$.

Theorem 1.3. *Let $r \in \mathbb{N}$ and let $0 < p < 1$, be a real number. Then*

$$\sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1.$$

Proof. Note that

$$\sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = p^r \cdot \sum_{n=r}^{\infty} \binom{n-1}{r-1} (1-p)^{n-r}$$

and thus it suffices to prove that

$$\sum_{n=r}^{\infty} \binom{n-1}{r-1} (1-p)^{n-r} = p^{-r}.$$

We will do so by induction on r. For $r = 1$ we have

$$\sum_{n=1}^{\infty} \binom{n-1}{1-1} (1-p)^{n-1} = \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{1-(1-p)} = p^{-1}.$$

We assume the required equality holds for r and prove it for $r+1$. It suffices to prove that

$$p \cdot \sum_{n=r+1}^{\infty} \binom{n-1}{r} (1-p)^{n-1-r} = p^{-r}.$$

Indeed, we have

$$\begin{aligned}
p \cdot \sum_{n=r+1}^{\infty} \binom{n-1}{r} (1-p)^{n-1-r} &= [1-(1-p)] \cdot \sum_{n=r+1}^{\infty} \binom{n-1}{r} (1-p)^{n-1-r} \\
&= \sum_{n=r+1}^{\infty} \binom{n-1}{r} (1-p)^{n-1-r} - \sum_{n=r+1}^{\infty} \binom{n-1}{r} (1-p)^{n-r} \\
&= 1 + \sum_{n=r+2}^{\infty} \binom{n-1}{r} (1-p)^{n-1-r} - \sum_{t=r+2}^{\infty} \binom{t-2}{r} (1-p)^{t-1-r} \\
&= 1 + \sum_{n=r+2}^{\infty} \left[\binom{n-1}{r} - \binom{n-2}{r} \right] (1-p)^{n-1-r} \\
&= 1 + \sum_{n=r+2}^{\infty} \binom{n-2}{r-1} (1-p)^{n-1-r} \\
&= 1 + \sum_{m=r+1}^{\infty} \binom{m-1}{r-1} (1-p)^{m-r} \\
&= \sum_{m=r}^{\infty} \binom{m-1}{r-1} (1-p)^{m-r} = p^{-r}
\end{aligned}$$

where the third equality holds by the substitution $t = n+1$ in the second sum, the fifth equality holds by Pascal's identity $\binom{a-1}{b} + \binom{a-1}{b-1} = \binom{a}{b}$, the sixth equality holds by the substitution $m = n-1$, and the last equality holds by the induction hypothesis. $\square$

Example 1: Toss a coin with probability p for heads until the rth time the outcome is heads, all coin tosses being mutually independent. Let X be the total number of tosses. Then $X \sim \mathrm{NB}(r,p)$. Indeed, if $X = k$, then the outcome of the kth toss is heads. Moreover, of the first $k-1$ tosses, exactly $r-1$ are heads and the rest are tails. Since all coin tosses are mutually independent, this happens with probability $\binom{k-1}{r-1} p^r (1-p)^{k-r}$. Note that we can view X as the sum of r independent geometric random variables $X_i \sim \mathrm{Geom}(p)$, where X_i represents the number of coin tosses from the $(i-1)$th time the outcome was heads (not including that toss) until the ith time the outcome was heads (including that toss). This will be made precise in the following result.

Theorem 1.4. *Let $r \in \mathbb{N}$ and let $0 < p < 1$ be a real number. Let $X_1, X_2, \ldots, X_r \sim \mathrm{Geom}(p)$ be independent random variables, and let $X = X_1 + X_2 + \ldots + X_r$. Then $X \sim \mathrm{NB}(r,p)$.*

Proof. Since every X_i takes values in $\{1, 2, \ldots\}$, it follows that X takes values in $\{r, r+1, \ldots\}$. For every r-tuple of natural numbers $\mathcal{I} = (x_1, \ldots, x_r) \in \mathbb{N}^r$ define the event

$$A_{\mathcal{I}} = \bigcap_{i=1}^{r} \{X_i = x_i\} .$$

Then, for every $n \in \{r, r+1, \ldots\}$ it holds that

$$\{X = n\} = \bigcup_{\substack{\mathcal{I}=(x_1,\ldots,x_r): \\ x_1+\ldots+x_r=n}} A_{\mathcal{I}}.$$

Observe that the number of such choices for $\mathcal{I}$ is the number of integer solutions to the equation $x_1 + \ldots + x_r = n$, where $x_i \geq 1$ for every $1 \leq i \leq r$, which is $\binom{n-1}{r-1}$. Since these events are pairwise disjoint, we infer that

$$\begin{aligned}
\mathbb{P}\left(X=n\right) &= \mathbb{P}\left(\bigcup_{\substack{\mathcal{I}=(x_1,\ldots,x_r):\\ x_1+\ldots+x_r=n}} A_{\mathcal{I}}\right) \\
&= \sum_{\substack{\mathcal{I}=(x_1,\ldots,x_r):\\ x_1+\ldots+x_r=n}} \mathbb{P}\left(A_{\mathcal{I}}\right) \\
&= \sum_{\substack{\mathcal{I}=(x_1,\ldots,x_r):\\ x_1+\ldots+x_r=n}} \prod_{i=1}^{r} \mathbb{P}\left(X_i = x_i\right) \\
&= \sum_{\substack{\mathcal{I}=(x_1,\ldots,x_r):\\ x_1+\ldots+x_r=n}} \prod_{i=1}^{r} (1-p)^{x_i-1} \cdot p \\
&= \sum_{\substack{\mathcal{I}=(x_1,\ldots,x_r):\\ x_1+\ldots+x_r=n}} p^r \cdot (1-p)^{x_1+\ldots+x_r-r} \\
&= \sum_{\substack{\mathcal{I}=(x_1,\ldots,x_r):\\ x_1+\ldots+x_r=n}} p^r \cdot (1-p)^{n-r} \\
&= \binom{n-1}{r-1} \cdot p^r \cdot (1-p)^{n-r},
\end{aligned}$$

where the third equality holds since $X_1, \ldots, X_r$ are mutually independent and the fourth equality holds by the definition of the Geometric distribution. $\square$