§ 3.5 Poisson distribution

When modeling with the binomial distribution, we require a fixed number of trials. However, consider counting:

- number of babies born in a hospital in 1 day
- number of emails received in 1 hour
- number of highway accidents in 1 year

These scenarios involve counting during a fixed time period but the notion of independent trials doesn't make sense.

Instead, we model them using the Poisson distribution.

Def A random variable X has the Poisson distribution with parameter $\lambda > 0$ if its range is $\{0, 1, 2, \dots\}$ and

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \dots .$$

Shorthand: $X \sim \mathrm{Pois}(\lambda)$.

Remark: the parameter λ is the average count (or we often use the term "number of arrivals" instead of "count") over the fixed time period

Example Suppose someone receives on average 10 emails per hour. We model the number of emails in 1 hour as a random variable X with the Poisson distribution.

① Find the probability they receive exactly 9 emails in 1 hour.

$X \sim \mathrm{Pois}(10)$ and $P(X=9) = e^{-10} \dfrac{10^9}{9!}$

② Find the probability they receive exactly 21 calls in 2 hours. Let $Y \sim \mathrm{Pois}(20)$, $P(Y=21) = e^{-20} \dfrac{20^{21}}{21!}$

Example Suppose someone receives 2 calls per hour on average. What is the probability that over 24 hours they receive exactly 50 calls? at least 50 calls?

Let $X \sim \mathrm{Pois}(48)$ count the number of calls received over 24 hours. Then

$$P(X=50) = e^{-48} \frac{48^{50}}{50!}$$

$$P(X \geq 50) = 1 - P(X \leq 49)$$

compute this with R

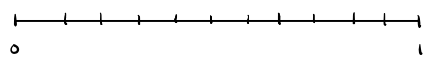
R commands if $X \sim \mathrm{Pois}(\lambda)$, then

- $P(X=k)$ can be computed in R with dpois(k, λ)
- $P(X \leq k)$ with ppois(k,λ)
- we can get an i.i.d. sample of size k with rpois(k, λ).

Derivation of the Poisson distribution

Suppose on average λ babies are born over 1 day.

Let's break up the day into n pieces of length $\frac{1}{n}$.



- On average, in any of these subintervals $\frac{\lambda}{n}$ babies are born.
- Moreover, if $X_1, \ldots, X_n$ denotes the number of babies born in the first, second, etc. subinterval, then when n is big enough each of these X_k could only take values in $\{0, 1\}$.

  (ie. it's very unlikely to have 2 or more babies born simultaneously)
- So $X_1, \ldots, X_n$ are i.i.d. $Ber(p)$ with $p = \frac{\lambda}{n}$.
- Since $X = \lim_{n \to \infty} X_1 + \cdots + X_n$ and $X_1 + \cdots + X_n \sim Bin(n, \frac{\lambda}{n})$

$$P(X = k) = \lim_{n \to \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{(using calculus)}$$

$$
\begin{aligned}
\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} &= \lim_{n\to\infty} \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1-\frac{\lambda}{n}\right)^n \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty} \underbrace{\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}}_{\to 1} \left(1-\frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \\
&= e^{-\lambda} \frac{\lambda^k}{k!}
\end{aligned}
$$

Problem 1. Customers arrive at a checkout counter in a department store according to a Poisson distribution at an average of 7 per hour.

a. Find the probability

1. exactly five customers arrive in the next hour
2. no more than three customers arrive in the next hour
3. at least two customers arrive in the next hour

b. Answer the same questions but change "next hour" to "next 30 minutes."

Part a:

```{r}
dpois(5, 7)
ppois(3, 7)
1-ppois(1, 7)
```

```
[1] 0.1277167
[1] 0.08176542
[1] 0.9927049
```

Part b:

```{r}
dpois(5, 3.5)
ppois(3, 3.5)
1-ppois(1, 3.5)
```

```
[1] 0.1321686
[1] 0.5366327
[1] 0.8641118
```

(a) $\underline{X} \sim \text{Pois}(7)$

$P(\underline{X}=5) = e^{-7}\frac{7^5}{5!} \approx 0.1277167$

$P(\underline{X} \leq 3) \approx 0.08176542$

$P(\underline{X} \geq 2) \approx 0.9927049$

(b) $\underline{X} \sim \text{Pois}(3.5)$, see right for values

Problem 2. A telemarketing company has found that the probability of making a sale when calling someone on the telephone is approximately 0.0002. If the salesperson contacts 10000 prospects, what is the probability of making between 2 and 5 sales?

a. Calculate this using the binomial distribution. What are n and p?

b. Let $Y \sim \text{Pois}(np)$. Compute $P(2 \leq Y \leq 5)$ and compare with your previous answer.

Problem 2

Part a:

```{r}
pbinom(5, 10000, 0.0002) - pbinom(1, 10000, 0.0002)
```

```
[1] 0.5774684
```

Part b:

```{r}
ppois(5, 10000*0.0002) - ppois(1, 10000*0.0002)
```

```
[1] 0.5774305
```

(a) $\underline{X} \sim \text{Bin}(10000, 0.0002)$

$P(2 \leq \underline{X} \leq 5) \approx 0.5774684$

(b) $\underline{Y} \sim \text{Pois}(2)$

$P(2 \leq \underline{Y} \leq 5) \approx 0.5774305$