Today: Exponential distribution

Next: ASV 5.1

Week 7:

- homework 5 (due Friday, November 11, 11:59PM)
- homework 3 regrades open until Sunday, Nov 13
- Midterm 2 on Wednesday, November 16 (lectures 12-20)
Approximating probabilities of rare events

Poisson distribution is used to model the occurrences of rare events. Examples:

- Customers arriving in a store: all potential customers decide independently to come or not.
- Number of emergency calls: all people in the city have an emergency or not "independently" of each other.
- Number of car accidents: all drivers in the county have accidents (or not) "independently".
- Number of goals scored in a hockey game: a lot of "independent" shots.

If random variable $X$ counts the occurrences of rare events that are not strongly dependent, then $P(X = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}$ with $\lambda = E(X)$. 
**Example**

Number of phone calls in a day can be modeled by Poisson random variable. We know that on average 0.5% of the time the call center receives no calls at all. What is the average number of calls per day?

Let $X$ be the number of calls per day.

$X \sim \text{Poisson}(\lambda)$, $E(X) = \lambda$

It is given that $P(X = 0) = 0.005 = e^{-\lambda} = \frac{1}{200}$

Therefore, $\log e^{-\lambda} = -\lambda = \log \frac{1}{200} = -\log 200$

$\lambda = \log 200 \approx 5.298$

$E(X) \approx 5.298$
Exercise

10% of households earn $> 80000$

0.25% of households earn $> 450000$

Choose 400 households at random. Denote

$X = \#$ households $> 80000$ , $Y = \#$ households $> 450000$

Estimate $P(X \geq 48)$ and $P(Y \geq 2)$

1) For $X$, $n = 400$, $p_x = 0.1$, $np_x = 40$, $np_x(1-p_x) = 36 > 10$

$P(X \geq 48) = P\left( \frac{X - 40}{6} \geq \frac{48 - 40}{6} \right) \approx 1 - \Phi(1.33)$

2) For $Y$, $n = 400$, $p_y = \frac{1}{400}$, $np_y = 1$, $np_y^2 = \frac{1}{400}$ — use Poisson approximation

$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) \approx 1 - e^{-1} - 1 \cdot e^{-1} = 1 - \frac{2}{e} \approx 0.2642$
A fair 20-sided die is tossed 400 times. We want to calculate the probability that a 13 came up at least 25 times. We should use:

(a) Poisson approximation
(b) Normal approximation
(c) Neither
(d) Both

Poisson: $X \sim \text{Poisson (20)}$, $P(X \geq 25) \approx 15.68\%$

Normal: $Y \sim N(20, 19)$, $P(Y \geq 25) \approx 12.57\%$

True value: $S_n = \text{Bin}(400, \frac{1}{20})$, $P(S_n \geq 25) \approx 15.1\%$
Waiting for a customer

Suppose that customers arrive in a store with rate $\lambda$ customers per hour. How can we model the time until the first (or next) customer arrives?

Let $X =$ time when the first customer arrives

Additional assumptions: if the intervals are small enough, then

- only one customer can arrive per interval
- customers arrive/do not arrive for each interval independently
- $P(\text{customer arrives during } T_k) = \frac{\lambda t}{n}$

Q: What is $P(X > t) = P(\text{no customers in each } T_k)$
**Exponential distribution**

\[ P(X > t) = P(\text{no customer in each } T_k) \]

\[ = \prod_{k=1}^{n} P(\text{no customer in } T_k) = \prod_{k=1}^{n} \left(1 - \frac{\lambda t}{n}\right) \]

\[ = \left(1 - \frac{\lambda t}{n}\right)^n \xrightarrow{n \to \infty} e^{-\lambda t} \]

\text{CDF: } P(X \leq t) = 1 - e^{-\lambda t}, \quad t \geq 0

\text{PDF: } f_X(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0 \\ 0, & \text{otherwise} \end{cases}

\text{Def. Let } \lambda > 0. \text{ We say that random variable } X \text{ has exponential distribution with rate parameter } \lambda, \text{ if}

\[ f_X(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0 \\ 0, & t \leq 0 \end{cases}, \text{ denote } X \sim \text{Exp}(\lambda) \]
Exponential distribution

Let $X \sim \text{Exp}(\lambda)$. Then

- $E(X) = \int_0^\infty t \lambda e^{-\lambda t} dt = \int_0^\infty (-e^{-\lambda t})' dt = -t \lambda e^{-\lambda t} \bigg|_0^\infty + \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda}$

- $E(X^2) = \frac{2}{\lambda^2}$, so $\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

- $P(X > t) = e^{-\lambda t}$

Exponential distribution is used to model waiting times.

Example: Length of a phone call is modeled by exponential random variable with mean 10 (minutes). What is the probability that the call takes > 8 minutes? Between 8 and 22?

$X \sim \text{Exp}(\lambda)$, $E(X) = \frac{1}{\lambda} = 10$, so $\lambda = \frac{1}{10}$. Then $P(X > 8) = e^{-\frac{8}{10}} = e^{-0.8}$

$P(8 < X \leq 22) = P(X > 8) - P(X > 22) = e^{-0.8} - e^{-2.2} \approx 0.4493$
Memoryless property

Proposition. Let $X \sim \text{Exp}(\lambda)$, $\lambda > 0$. Then for any $s, t > 0$

Proof. $P(X > s + t | X > s)$

Exp($\lambda$) is the only continuous distribution with memoryless property.

Remark. If $N \sim \text{Geom}(p)$, then $P(N > k) = (1-p)^k$, and

$$P(N > k + l | N > k) = \frac{P(N > k + l)}{P(N > k)} = \frac{(1-p)^{k+l}}{(1-p)^k} = (1-p)^l = P(N > l)$$