THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

# SDS 321: Introduction to Probability and Statistics

## Lecture 10: Expectation and Variance


Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

# The Geometric random variable

- The Bernoulli PMF describes the probability of success/failure in a single trial.
- The Binomial PMF describes the probability of $k$ successes out of $n$ trials.
- Sometimes we may also be interested in doing trials until we see a success.

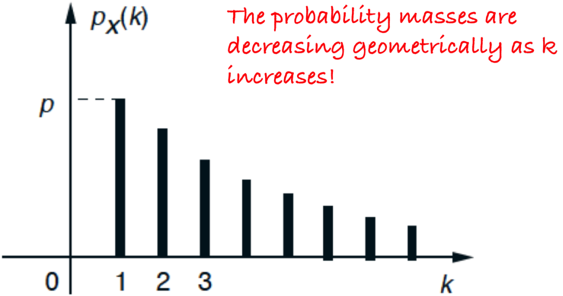
- Alice resolves to keep buying lottery tickets until he wins a hundred million dollars. She is interested in the random variable "number of lottery tickets bought until he wins the 100M$ lottery".
- Annie is trying to catch a taxi. How many occupied taxis will drive pass before she finds one that is taking passengers?

- The number of trials required to get a single success is a **Geometric Random Variable**

# The geometric random variable

We repeatedly toss a biased coin ($P(\{H\}) = p$). The geometric random variable is the number $X$ of tosses to get a head.

- $X$ can take any integral value.
- $P(X = k) = P(\{\underbrace{TT\ldots T}_{k-1}H\}) = (1-p)^{k-1}p$.
- $\sum_k P(X = k) = 1$ (why?)



# The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

# The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

- $P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}$

# The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

- $P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}$
- Intuitively, this is asking for the probability that the first $k-1$ tosses are tails.
- This probability is $P(X \geq k) = (1-p)^{k-1}$

# The geometric random variable

What is $P(X \geq k)$? What is $P(X > k)$?

- $P(X \geq k) = \sum_{i=k}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}$
- Intuitively, this is asking for the probability that the first $k-1$ tosses are tails.
- This probability is $P(X \geq k) = (1-p)^{k-1}$
- $X > k$ is the event that $X \geq k+1$, and so $P(X > k) = (1-p)^{k}$

# The memoryless property

What is $P(X = a + b | X > a)$?

# The memoryless property

What is $P(X = a + b | X > a)$?

- $$\begin{aligned} P(X = a + b | X > a) &= \frac{P(X = a + b)}{P(X > a)} \\ &= \frac{p(1-p)^{a+b-1}}{(1-p)^a} \\ &= p(1-p)^{b-1} = P(X = b) \end{aligned}$$
- You forgot about $X > a$ and started the clock afresh!

# The memoryless property

What is $P(X > a + b | X > a)$?

# The memoryless property

What is $P(X > a + b | X > a)$?

$$P(X > a + b | X > a) = \frac{P(X > a + b)}{P(X > a)}$$

- $$= \frac{(1-p)^{a+b}}{(1-p)^{a}} = (1-p)^{b}$$
  $$= P(X > b)$$
- You forgot about $X > a$ and started the clock afresh!

# The memoryless property

What is $P(X \leq a + b | X > a)$?

# The memoryless property

What is $P(X \leq a+b|X > a)$?

- $$\begin{aligned} P(X \leq a+b|X > a) &= \frac{P(a < X \leq a+b)}{P(X > a)} \\ &= \frac{P(X > a) - P(X > a+b)}{(1-p)^a} \\ &= \frac{(1-p)^a - (1-p)^{a+b}}{(1-p)^a} \\ &= 1 - (1-p)^b = P(X \leq b) \end{aligned}$$
- You forgot about $X > a$ and started the clock afresh!

# The Poisson random variable

I have a book with 10000 words. Probability that a word has a typo is 1/1000. I am interested in how many misprints can be there on average? So a Poisson often shows up when you have a Binomial random variable with very large n and very small p but $n \times p$ is moderate. Here $np = 10$.

Our random variable might be:

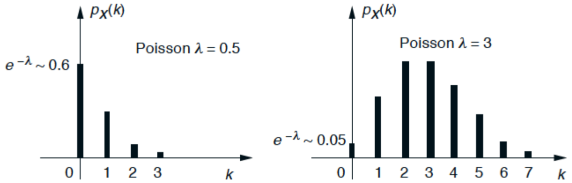
- The number of car crashes in a given day.
- The number of buses arriving within a given time period.
- The number of mutations on a strand of DNA.

We can describe such situations using a **Poisson random variable**.

# The Poisson random variable

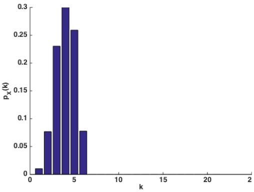
- A Poisson random variable takes non-negative integers as values. It has a nonnegative parameter $\lambda$.
- $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, for $k = 0, 1, 2 \ldots$.
- $\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots) = 1$. (Exponential series!)

The PMF is monotonically decreasing for λ=0.5

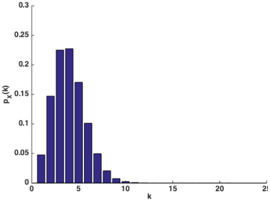



The PMF is increasing and then decreasing for λ=3

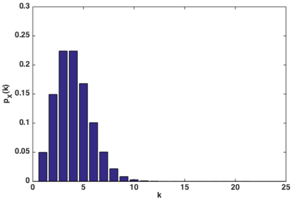
# Poisson random variable




Binomial(5,0.6)




Binomial(100,0.03)




Poisson(3)

- When $n$ is very large and $p$ is very small, a binomial random variable can be well approximated by a Poisson with $\lambda = np$.
- In the above figure we increased $n$ and decreased $p$ so that $np = 3$.
- See how close the PMF's of the Binomial(100,0.03) and Poisson(3) are!
- More formally, we see that $\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}$ when $n$ is large, $k$ is fixed, and $p$ is small and $\lambda = np$.

# Example

Assume that on a given day 1000 cars are out in Austin. On an average three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?

4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?

# Example

Assume that on a given day 1000 cars are out in Austin. On an average three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?

2. Use poisson approximation!

4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?

# Example

Assume that on a given day 1000 cars are out in Austin. On an average three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?
2. Use poisson approximation!
3. $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-3}(1 + 3) = 0.8$
4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?

# Example

Assume that on a given day 1000 cars are out in Austin. On an average three out of 1000 cars run into a traffic accident per day.

1. What is the probability that we see at least two accidents in a day?
2. Use poisson approximation!
3. $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-3}(1 + 3) = 0.8$
4. If you know there is at least one accident, what is the probability that the total number of accidents is at least two?
5. $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-3} = 0.950$.
   $P(X \geq 2 | X \geq 1) = P(X \geq 2)/P(X \geq 1) = 0.8/0.950 = 0.84$