Introduction to Data Science With Probability and Statistics

Lecture 8: Discrete Random Variables and their Distributions

CSCI 3022 - Summer 2020
Sourav Chakraborty
Dept. of Computer Science
University of Colorado Boulder

What will we learn today?

- Refresher
- Examples
- Negative Binomial Distribution.
- Geometric Distribution

Review from Last Time

A **discrete random variable** (r.v.) X is a function that maps the elements of the sample space Ω to a finite number of values $a_1, a_2, \dots, a_n$ or an infinite number of values $a_1, a_2, \dots$

A discrete random variable X has a **Bernoulli distribution** (X~Ber(p)), where $0 \leq p \leq 1$, if its probability mass function is given by

$$f(1) = p_X(1) = P(X = 1) = p \quad \text{and} \quad p_X(0) = P(X = 0) = 1 - p$$

A discrete random variable X has a **Binomial Distribution** (X~Bin(n, p)) with $n = 1, 2, \dots$ and $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n$$

Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. If the medication is given to a new patient with allergies, what is the probability that it is effective to that patient?

Distribution ?

Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. If the medication is given to a new patient with allergies, what is the probability that it is effective to that patient?

Distribution ?

- p : probability of success (relief with the medicine) = 0.8
- single patient can be thought of as a single trail.
- $X = \begin{cases} 1, & \text{if cured} \\ 0, & \text{if not} \end{cases}$

$\therefore \boxed{X \sim Ber(0.8)}$

Allergies Strikes Back

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. If the medication is given to 10 new patients with allergies, what is the probability that it is effective in exactly seven?

Distribution?

Allergies Strikes Back

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. If the medication is given to 10 new patients with allergies, what is the probability that it is effective in exactly seven?

Distribution?

Let $X_i = \begin{cases} 1, & \text{if } i\text{th patient is cured} \\ 0, & \text{if } i\text{th patient is NOT cured.} \end{cases}$

$\therefore X_i \sim Ber(0.8)$

Let X = Number of patients cured with 10 trials

$= \sum_{i=1}^{10} X_i$

$\therefore \boxed{X \sim Bin(10, 0.8)}$

$P(X=7)$

$= \binom{10}{7} (0.8)^7 (0.2)^3$

Geometric Distribution

Example: Suppose that you flip the same biased coin repeatedly. How many times do you flip the coin before you see your first Heads?

Hint: Get the distribution / PMF

Geometric Distribution

$P(\text{heads}) = p$

Example: Suppose that you flip the same biased coin repeatedly. How many times do you flip the coin before you see your first Heads?

Let X be the trial number when we recieve our first heads.

So, $X=1$, $\{H\}$: $P(X=1) = p$

$X=2$, $\{TH\}$: $P(X=2) = (1-p)p$

$X=3$, $\{TTH\}$: $P(X=3) = (1-p)^2 p$

$X=4$, $\{TTTH\}$: $P(X=4) = (1-p)^3 p$

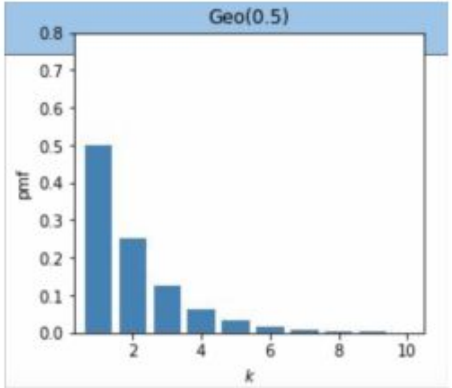
$\vdots$

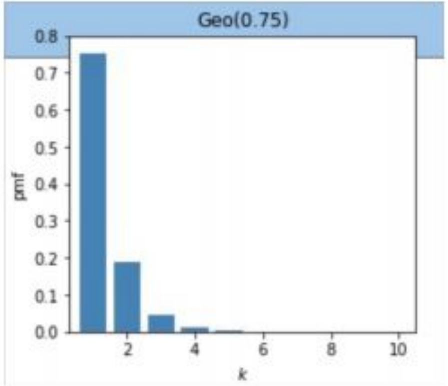
$$\boxed{P(X=x) = (1-p)^{x-1} p}$$

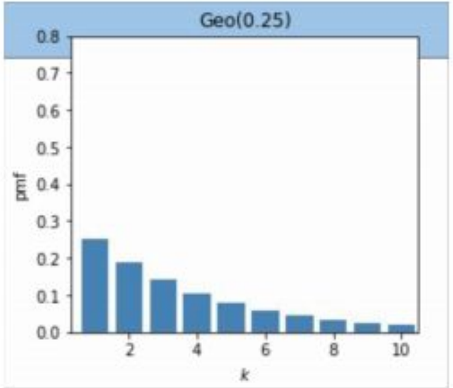
Geometric Distribution

A discrete random variable X has a **Geometric distribution** (X~Geo(p)) with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = (1-p)^{k-1} \cdot p \quad \text{for } k = 1, 2, 3, \ldots$$







Geometric Distribution

A discrete random variable X has a **Geometric distribution** (X~Geo(p)) with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by

$$p_X(k) = P(X = k) = (1 - p)^{k-1} \cdot p \quad \text{for } k = 1, 2, 3, \dots$$

Question: What assumptions did we implicitly make in deriving the Geometric distribution?

- Each trial is independent.
- Each trial is a Bernoulli random variable with probability of success p.

Return of the Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. How many patients with allergies do you need to give this medication to get the first case of relief/cure?

Distribution?

Return of the Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. How many patients with allergies do you need to give this medication to get the first case of relief/cure?

Distribution? Let X_i : $\begin{cases} 1, & \text{if ith patient is cured (success)} \\ 0, & \text{if ith patient is not cured. (failure)} \end{cases}$

Let X : The patient Number when we will get our first cure.

Since $X_i \sim Ber(0.8)$

p

$\boxed{X \sim Geo(0.8)}$

Negative Binomial Distribution

<u>**Example**</u>: Suppose that you flip the same biased coin repeatedly. How many times do you flip the coin before you see 3 Heads total?

Negative Binomial Distribution

Example: Suppose that you flip the same biased coin repeatedly. How many times do you flip the coin before you see 3 Heads total?

Suppose we get our 3rd heads in the trial/flip number k.

$\Rightarrow \underbrace{\text{first } (k-1) \text{ trials had exactly 2 heads}}_{\text{Binomial ?}}$.

$$P(X=k) = \underbrace{\binom{k-1}{2} p^2 (1-p)^{(k-1)-2}}_{\text{2 heads in } k-1 \text{ trials}} \cdot \underbrace{p}_{\text{heads in the } k\text{th trials.}}$$

Negative Binomial Distribution

A discrete random variable X has a **Negative Binomial distribution** (X~NB(r, p)) with parameters r and p, where $r > 1$ and $0 \leq p \leq 1$, if its probability mass function is given by

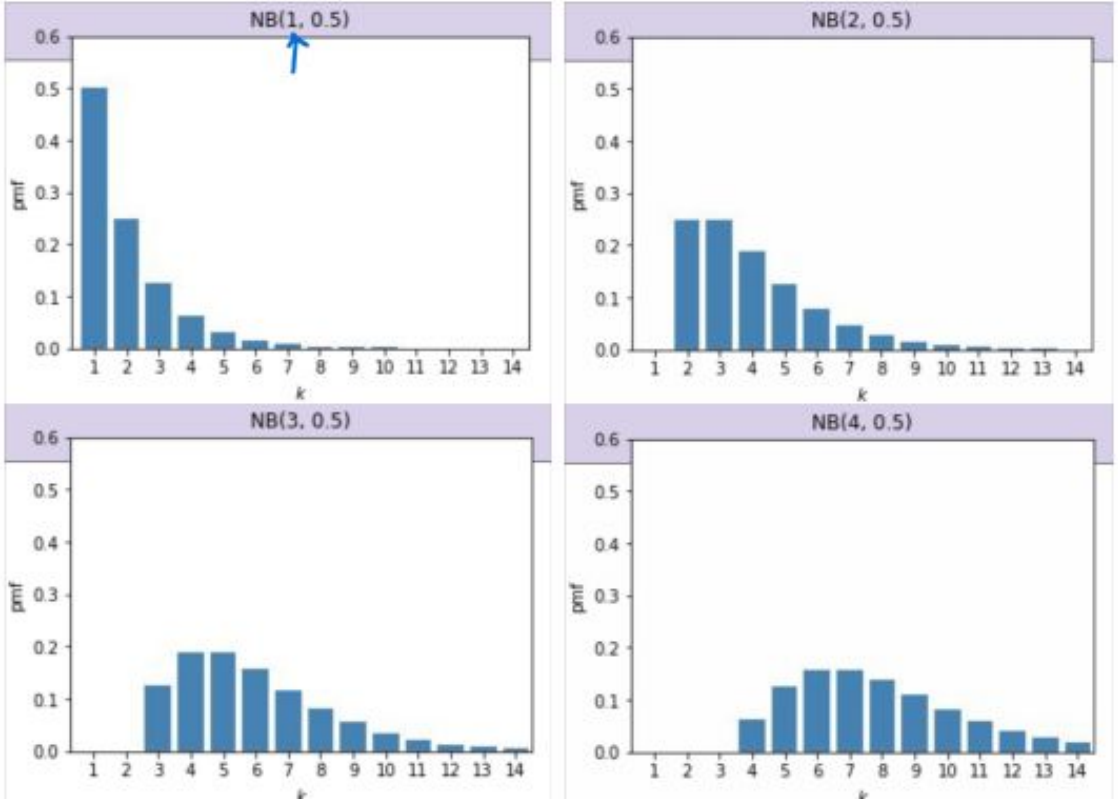
$$p_X(k) = P(X = k) = \binom{k-1}{r-1} \mathrm{p}^{\mathrm{r}} (1 - \mathrm{p})^{k-r}$$

p = probability of success for each trial

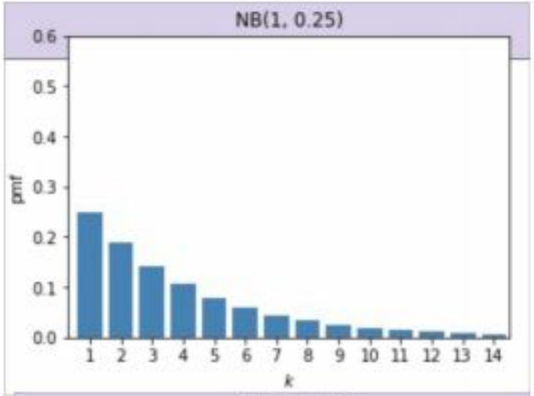
r = number of successes we want to observe

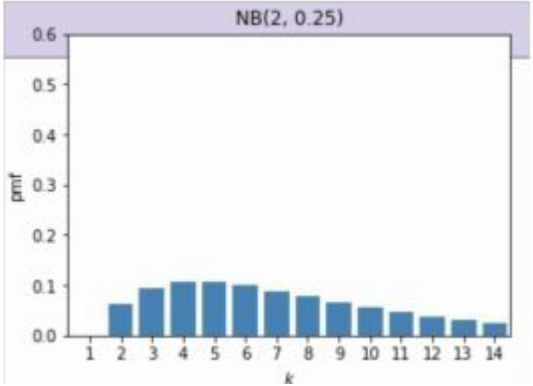
X = number of trials needed before we observe r successes

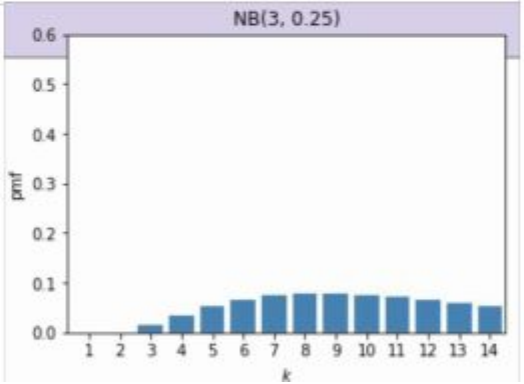
Negative Binomial Distribution



Negative Binomial Distribution







Question: What assumptions did we implicitly make in deriving the Negative Binomial distribution?

- Each trial is a Bernoulli r.v. with probability of success *p*
- Each trial is independent

Revenge of the Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. How many patients should I give the medication to make sure that exactly 5 people get relief/cured?

Revenge of the Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. How many patients should I give the medication to make sure that exactly 5 people get relief/cured?

Let X : trial number you will get 5th relief.

$$\therefore P(X=K) = \binom{K-1}{4} p^4 (1-p)^{(K-1)-4} \cdot p$$

$$\sim NB(\underset{r}{5}, \underset{p}{0.8})$$

Binomial vs. Negative Binomial Distributions

Relation?

Binomial vs. Negative Binomial Distributions

Relation?

Bin (n, p)	NB (r, p)
Number (variable) of successes in a fixed Number of trials (n)	Number (variable) of trials needed to get a fixed number of successes. (r)

Binomial-like Distributions

Example: You are doing an exit poll outside of a voting station on Election Day. As people exit, you ask them questions about their political affiliation, who they voted for, etc.

In particular, you are interested in how registered Independents voted. You know that about 20% of registered voters are registered as Independents.

Suppose you interview 100 people. ***Let X be a random variable describing the number of actual Independents you encounter.***

What distribution does X have?

Binomial-like Distributions

Example: You are doing an exit poll outside of a voting station on Election Day. As people exit, you ask them questions about their political affiliation, who they voted for, etc.

In particular, you are interested in how registered Independents voted. You know that about 20% of registered voters are registered as Independents.

Suppose you interview 100 people. ***Let X be a random variable describing the number of actual Independents you encounter.***

What distribution does X have?

$X \sim \mathrm{Bin}(100, 0.2)$

Binomial-like Distributions

Example: You are doing an exit poll outside of a voting station on Election Day. As people exit, you ask them questions about their political affiliation, who they voted for, etc.

In particular, you are interested in how registered Independents voted. You know that about 20% of registered voters are registered as Independents.

Suppose you talk to a lot of registered Republicans and Democrats, but haven't found an Independent yet. ***Let X be a random variable describing the number of people you have interviewed up to and including your first registered Independent voter.***

What distribution does X have?

Binomial-like Distributions

<u>Example</u>: You are doing an exit poll outside of a voting station on Election Day. As people exit, you ask them questions about their political affiliation, who they voted for, etc.

In particular, you are interested in how registered Independents voted. You know that about 20% of registered voters are registered as Independents.

Suppose you talk to a lot of registered Republicans and Democrats, but haven't found an Independent yet. ***Let X be a random variable describing the number of people you have interviewed up to and including your first registered Independent voter.***

What distribution does X have?

$X \sim Geo(0.2)$

Binomial-like Distributions

Example: You are doing an exit poll outside of a voting station on Election Day. As people exit, you ask them questions about their political affiliation, who they voted for, etc.

In particular, you are interested in how registered Independents voted. You know that about 20% of registered voters are registered as Independents.

Suppose you are really interested in talking to a lot of Independents. ***Let X be a random variable describing the number of people you have to talk to in order to interview exactly 100 registered Independents.***

What distribution does X have?

Binomial-like Distributions

Example: You are doing an exit poll outside of a voting station on Election Day. As people exit, you ask them questions about their political affiliation, who they voted for, etc.

In particular, you are interested in how registered Independents voted. You know that about 20% of registered voters are registered as Independents.

Suppose you are really interested in talking to a lot of Independents. ***Let X be a random variable describing the number of people you have to talk to in order to interview exactly 100 registered Independents.***

What distribution does X have?

$X \sim NB(100, 0.2)$

Next Time:

- More distributions!