CS 107, Probability, Spring 2019
Lecture 26

Michael Poghosyan

AUA

25 March 2019

Content

- End of the story about Important Discrete R.V.s
- Examples of Important Continuous R.V.s

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? $\ddot\smile$

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? $\ddot\smile$

Zipf's law and Pareto Principle

https://www.youtube.com/watch?v=fCn8zs912OE

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? ⌣̈

Zipf's law and Pareto Principle

https://www.youtube.com/watch?v=fCn8zs912OE

Some comments under the Video:

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? $\ddot\smile$

Zipf's law and Pareto Principle

`https://www.youtube.com/watch?v=fCn8zs912OE`

Some comments under the Video:

- The way one of my teachers explained the 80/20 thing: 80% of the noise in a classroom is caused by 20% of the students.

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? $\ddot\smile$

Zipf's law and Pareto Principle

https://www.youtube.com/watch?v=fCn8zs912OE

Some comments under the Video:

- The way one of my teachers explained the 80/20 thing: 80% of the noise in a classroom is caused by 20% of the students.
- 80% of comments are created by 20% of users $\ddot\smile$

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? ⌣̈

Zipf's law and Pareto Principle

https://www.youtube.com/watch?v=fCn8zs912OE

Some comments under the Video:

- The way one of my teachers explained the 80/20 thing: 80% of the noise in a classroom is caused by 20% of the students.
- 80% of comments are created by 20% of users ⌣̈
- My phone is at 20% 80% of the time ⌣̈

LZ

The of and to. A in is I. That it, for you, was with on. As have ... but be they.

Can you guess what this means? $\ddot\smile$

Zipf's law and Pareto Principle

https://www.youtube.com/watch?v=fCn8zs912OE

Some comments under the Video:

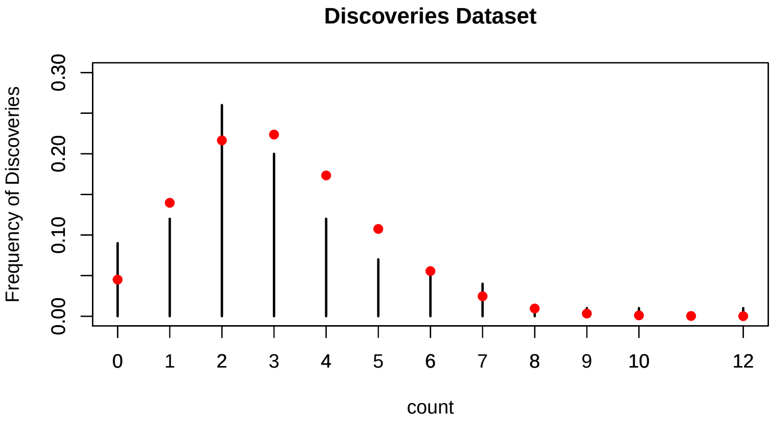
- The way one of my teachers explained the 80/20 thing: 80% of the noise in a classroom is caused by 20% of the students.
- 80% of comments are created by 20% of users $\ddot\smile$
- My phone is at 20% 80% of the time $\ddot\smile$
- Studied 20% of the material, get a 80% on the finals. Happened to no student ever. (MP: Exactly!)

Poisson Distribution: R Examples

R Code

```
help("discoveries")
disc <- discoveries
plot(disc)
hist(disc, breaks = seq(0,13,1)) #histogram
#Fitting the data by the Poisson
lambda = mean(disc)
table(disc)
plot(table(disc)/length(disc), xlim = c(0,max(disc)),
   ylim = c(0,0.3), main = "Discoveries Dataset",
   ylab = "Frequency of Discoveries", xlab = "count")
n = 0:max(disc)
m = dpois(n, lambda)
par(new = T) #To keep the previous graph
plot(n,m, xlim = c(0,max(disc)), ylim = c(0,0.3),
   pch = 19, col="red", main = "", xlab = "", ylab = "")
```

Discoveries Dataset Model Result

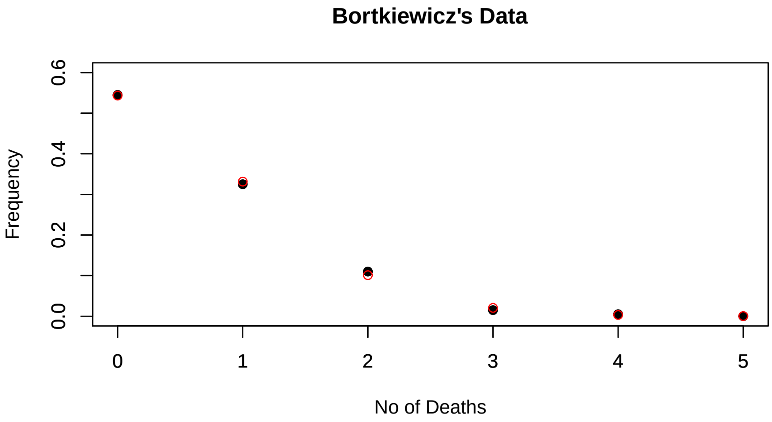


Poisson Distribution: R Examples

R Code

```
#Bortkiewicz's Data
#Deaths by horse-kick in Prussian Army cavalry corps
#(Bortkiewicz 1898). N is the number of units
#corresponding to each number of deaths.
hk<-data.frame(D = c(0,1,2,3,4,5),
   N = c(109,65,22,3,1,0))
plot(hk)
lambda <- sum(hk$D*hk$N)/sum(hk$N)
n = 0:5
m = dpois(n, lambda)
plot(n,m, xlim = c(0,5), ylim = c(0,0.6),
    col="red", type = "l", lwd = 3)
par(new = T)
plot(hk$D, hk$N/sum(hk$N) , xlim = c(0,5),
    ylim = c(0,0.6), pch=19)
```

Bortkiewicz Dataset Model Result



Approximation of Binomial Distribution through Poisson

Now, let us talk about one application of the Poisson Distribution for calculation of Binomial Probabilities.

Approximation of Binomial Distribution through Poisson

Now, let us talk about one application of the Poisson Distribution for calculation of Binomial Probabilities. Recall that when doing n independent Trials with Probability p of Success in each Trial, then the Probability of having exactly k Successes is (X is the number of Successes):

$$\mathbb{P}(X = k)$$

Approximation of Binomial Distribution through Poisson

Now, let us talk about one application of the Poisson Distribution for calculation of Binomial Probabilities. Recall that when doing n independent Trials with Probability p of Success in each Trial, then the Probability of having exactly k Successes is (X is the number of Successes):

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Approximation of Binomial Distribution through Poisson

Now, let us talk about one application of the Poisson Distribution for calculation of Binomial Probabilities. Recall that when doing n independent Trials with Probability p of Success in each Trial, then the Probability of having exactly k Successes is (X is the number of Successes):

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Now, assume n is large enough, but also p is small enough, and $\lambda = n \cdot p$.

Approximation of Binomial Distribution through Poisson

Now, let us talk about one application of the Poisson Distribution for calculation of Binomial Probabilities. Recall that when doing n independent Trials with Probability p of Success in each Trial, then the Probability of having exactly k Successes is (X is the number of Successes):

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Now, assume n is large enough, but also p is small enough, and $\lambda = n \cdot p$. Then,

$$\mathbb{P}(X = k)$$

Approximation of Binomial Distribution through Poisson

Now, let us talk about one application of the Poisson Distribution for calculation of Binomial Probabilities. Recall that when doing n independent Trials with Probability p of Success in each Trial, then the Probability of having exactly k Successes is (X is the number of Successes):

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Now, assume n is large enough, but also p is small enough, and $\lambda = n \cdot p$. Then,

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

Poisson Distribution: R Examples

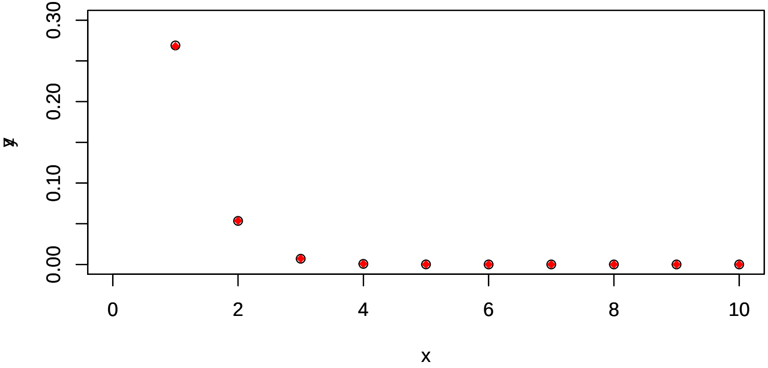
R Code

```
# Approximation of Binomial by Poisson
n <- 100
p <- 0.004
lambda <- n*p

x <- 0:20
y <- dbinom(x, size = n, prob = p)
plot(x,y, ylim = c(0,0.4))
z <- dpois(x, lambda = lambda)
par(new=T)
plot(x,z, pch = 18, col = "red", ylim = c(0,0.4))

#Plotting the difference
plot(x, abs(y-z))
```

Poisson approximation of Binomial Distribution



Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?

Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000.

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?
Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim$

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?

Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim Binom(20000, 0.0001)$.

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?
Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim Binom(20000, 0.0001)$. The actual probability is:

$$\mathbb{P}(X = 3) =$$

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?

Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim Binom(20000, 0.0001)$. The actual probability is:

$$\mathbb{P}(X = 3) = \binom{20000}{3} \cdot p^3 \cdot (1-p)^{20000-3}$$

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?
Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim Binom(20000, 0.0001)$. The actual probability is:

$$\mathbb{P}(X = 3) = \binom{20000}{3} \cdot p^3 \cdot (1-p)^{20000-3}$$

Instead of calculating this, we can use the Poisson approximation: we have $n = 20000$, $p = 0.0001$.

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?

Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim Binom(20000, 0.0001)$. The actual probability is:

$$\mathbb{P}(X=3) = \binom{20000}{3} \cdot p^3 \cdot (1-p)^{20000-3}$$

Instead of calculating this, we can use the Poisson approximation: we have $n = 20000$, $p = 0.0001$. Now take $\lambda = n \cdot p = 2$, and use (since n is large)

$$\mathbb{P}(X=3) = \binom{20000}{3} \cdot p^3 \cdot (1-p)^{20000-3} \approx$$

Example:

Problem: Assume the chance of being ill with some (rare) disease is $p = 0.0001$. What is the probability that among 20.000 persons exactly 3 persons will be ill with that disease?
Solution: Here we are dealing with the Binomial Model: let X be the number of ill persons among that 20.000. Clearly, $X \sim Binom(20000, 0.0001)$. The actual probability is:

$$\mathbb{P}(X = 3) = \binom{20000}{3} \cdot p^3 \cdot (1 - p)^{20000-3}$$

Instead of calculating this, we can use the Poisson approximation: we have $n = 20000$, $p = 0.0001$. Now take $\lambda = n \cdot p = 2$, and use (since n is large)

$$\mathbb{P}(X = 3) = \binom{20000}{3} \cdot p^3 \cdot (1 - p)^{20000-3} \approx e^{-\lambda} \cdot \frac{\lambda^3}{3!} = e^{-2} \cdot \frac{2^3}{3!} \approx 0.18.$$

Note about the Poisson Distribution

Assume the average number of downloads in a day for some software (or some mobile app) is λ. So if we want to model the number of daily downloads X of that software (X is a r.v.), we can model that as $X \sim Poisson(\lambda)$.

Note about the Poisson Distribution

Assume the average number of downloads in a day for some software (or some mobile app) is λ. So if we want to model the number of daily downloads X of that software (X is a r.v.), we can model that as $X \sim Poisson(\lambda)$.
Now, we want to model the number Y of downloads for 2 days.

Note about the Poisson Distribution

Assume the average number of downloads in a day for some software (or some mobile app) is λ. So if we want to model the number of daily downloads X of that software (X is a r.v.), we can model that as $X \sim Poisson(\lambda)$.
Now, we want to model the number Y of downloads for 2 days. Since the average number of downloads in 2 days for the same software will be 2λ, then we can use $Y \sim Poisson(2\lambda)$.

Other important Discrete Distributions

Please read in textbooks about the following Discrete Distributions:

Other important Discrete Distributions

Please read in textbooks about the following Discrete Distributions:

- The HyperGeometric Distribution;

Other important Discrete Distributions

Please read in textbooks about the following Discrete Distributions:

- The HyperGeometric Distribution;
- Negative Binomial Distribution.

Important Continuous Distributions

Intro to Continuous Distributions

Recall that if X is a Continuous r.v., then we describe it by its

Intro to Continuous Distributions

Recall that if X is a Continuous r.v., then we describe it by its

- CDF $F(x) = \mathbb{P}(X \leq x)$;

Intro to Continuous Distributions

Recall that if X is a Continuous r.v., then we describe it by its

- CDF $F(x) = \mathbb{P}(X \leq x)$;
- PDF $f(x)$, where $F(x) = \int_{-\infty}^{x} f(t)dt$.

Recall that if X is a Continuous r.v., then we describe it by its

- CDF $F(x) = \mathbb{P}(X \leq x)$;
- PDF $f(x)$, where $F(x) = \int_{-\infty}^{x} f(t)dt$.

For example,

$$\mathbb{P}(X \in [a, b]) =$$

Intro to Continuous Distributions

Recall that if X is a Continuous r.v., then we describe it by its

- CDF $F(x) = \mathbb{P}(X \leq x)$;
- PDF $f(x)$, where $F(x) = \int_{-\infty}^{x} f(t)dt$.

For example,

$$\mathbb{P}(X \in [a, b]) = F(b) - F(a) =$$

Intro to Continuous Distributions

Recall that if X is a Continuous r.v., then we describe it by its

- CDF $F(x) = \mathbb{P}(X \leq x)$;
- PDF $f(x)$, where $F(x) = \int_{-\infty}^{x} f(t)dt$.

For example,

$$\mathbb{P}(X \in [a, b]) = F(b) - F(a) = \int_a^b f(x)dx.$$

Uniform Distribution

Uniform Distribution on $[a, b]$

We say that the r.v. X has a uniform distribution in $[a, b]$, and we will write $X \sim Unif[a, b]$ if its PDF is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b]; \\ 0, & \text{otherwise} \end{cases}$$

Uniform Distribution

Uniform Distribution on $[a, b]$

We say that the r.v. X has a uniform distribution in $[a, b]$, and we will write $X \sim Unif[a, b]$ if its PDF is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b]; \\ 0, & \text{otherwise} \end{cases}$$

Exercise: Check that this function is a PDF of some r.v.

Uniform Distribution

Uniform Distribution on $[a, b]$

We say that the r.v. X has a uniform distribution in $[a, b]$, and we will write $X \sim Unif[a, b]$ if its PDF is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b]; \\ 0, & \text{otherwise} \end{cases}$$

Exercise: Check that this function is a PDF of some r.v.

The Graph of the PDF: On the board.

The CDF of the Uniform Distribution

Fact: The CDF of $X \sim Unif[a, b]$ is given by:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

The CDF of the Uniform Distribution

Fact: The CDF of $X \sim Unif[a, b]$ is given by:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

Graph on the Board

Idea of Uniform Distribution

Uniform Distribution Models situations, where our r.v. X can take any value from $[a, b]$, and we want to have all values equiprobable.

Idea of Uniform Distribution

Uniform Distribution Models situations, where our r.v. X can take any value from $[a, b]$, and we want to have all values equiprobable. In fact, the probability of each value is 0, so we describe equiprobability in the following way: if we will take two subintervals in $[a, b]$ of the same length, then the probability that X is in each subinterval is the same.

Idea of Uniform Distribution

Uniform Distribution Models situations, where our r.v. X can take any value from $[a, b]$, and we want to have all values equiprobable. In fact, the probability of each value is 0, so we describe equiprobability in the following way: if we will take two subintervals in $[a, b]$ of the same length, then the probability that X is in each subinterval is the same. Say, the probability that the value of X is in $[a, \frac{a+b}{2}]$ is $\frac{1}{2}$, and the probability that X is in $[\frac{a+b}{2}, b]$ is again $\frac{1}{2}$.

Uniform Distribution Examples

- Recall the experiments of picking a random number from some interval - we were talking, actually, about the Uniform Distribution. So if we are picking from $[a, b]$, and if we will denote the randomly chosen number by X, then $X \sim Unif[a, b]$ (if, of course, we are choosing *uniformly* $\ddot\smile$).

Uniform Distribution Examples

- Recall the experiments of picking a random number from some interval - we were talking, actually, about the Uniform Distribution. So if we are picking from $[a, b]$, and if we will denote the randomly chosen number by X, then $X \sim Unif[a, b]$ (if, of course, we are choosing *uniformly* $\ddot\smile$).
- If the Baghramyan Metro Station trains arrive in 10 min intervals, and I am visiting that station at some random time instant, then my waiting time X (in minutes) can be modeled by $X \sim Unif[0, 10]$.

Uniform Distribution Examples

- Recall the experiments of picking a random number from some interval - we were talking, actually, about the Uniform Distribution. So if we are picking from $[a, b]$, and if we will denote the randomly chosen number by X, then $X \sim Unif[a, b]$ (if, of course, we are choosing *uniformly* $\ddot\smile$).
- If the Baghramyan Metro Station trains arrive in 10 min intervals, and I am visiting that station at some random time instant, then my waiting time X (in minutes) can be modeled by $X \sim Unif[0, 10]$.
- Will the age distribution (exact age, not rounded to years) in AUA be Uniform ?

Uniform Distributions: Example

Example: Assume $X \sim \mathit{Unif}[1,4]$. Calculate $\mathbb{P}(X = 2)$, $\mathbb{P}(X \leq 2)$, $\mathbb{P}(2 \leq X \leq 3.5)$.

Uniform Distribution: R Examples

R Code

```
a <- 0
b <- 2
x <- seq(from = a-1, to = b+1, by = 0.01)
y <- dunif(x, min = a, max = b)
par(mfrow = c(1,2))
#plot the PDF
plot(x,y, type = "l", lwd = 3,
      main = "Uniform Distribution PDF")

z <- punif(x, min = a, max = b)
#plot the CDF
plot(x,z, type = "l", lwd = 3,
      main = "Uniform Distribution CDF")
```

Uniform Distribution: R Examples

R Code

```
#generating 10 uniformly distributed numbers in [0,1]
smpl <- runif(10) #equivalent to runif(10, min=0, max=1)
smpl

#generating 10000 uniformly distributed numbers in [0,1]
smpl <- runif(10000)
hist(smpl) # the histogram
```

Exponential Distribution

Exponential Distribution

We say that the r.v. X has an Exponential Distribution with the parameter $\lambda > 0$ (rate), and we write $X \sim Exp(\lambda)$, if its PDF is given by

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

Exponential Distribution

Exponential Distribution

We say that the r.v. X has an Exponential Distribution with the parameter $\lambda > 0$ (rate), and we write $X \sim Exp(\lambda)$, if its PDF is given by

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

Exercise: Check that this function is a PDF of some r.v.

Exponential Distribution

Exponential Distribution

We say that the r.v. X has an Exponential Distribution with the parameter $\lambda > 0$ (rate), and we write $X \sim Exp(\lambda)$, if its PDF is given by

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

Exercise: Check that this function is a PDF of some r.v.
Exercise: Prove that the CDF of $X \sim Exp(\lambda)$ is given by:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$