CS 107, Probability, Spring 2019
Lecture 25

Michael Poghosyan

AUA

22 March 2019

Content

- Examples of Important Discrete R.V.s
- Examples of Important Continuous R.V.s

Do you think all digits are equal?

LZ

Do you think all digits are equal? I mean, in the sense of rights? $\ddot\smile$

LZ

Do you think all digits are equal? I mean, in the sense of rights? ☺

Benford's law

In many numerical datasets, the chance that the first digit will be $n \in \{1, 2, ..., 9\}$ is

$$p_n =$$

LZ

Do you think all digits are equal? I mean, in the sense of rights? $\ddot\smile$

Benford's law

In many numerical datasets, the chance that the first digit will be $n \in \{1, 2, ..., 9\}$ is

$$p_n = \log_{10}\left(1 + \frac{1}{n}\right).$$

LZ

Do you think all digits are equal? I mean, in the sense of rights? $\ddot\smile$

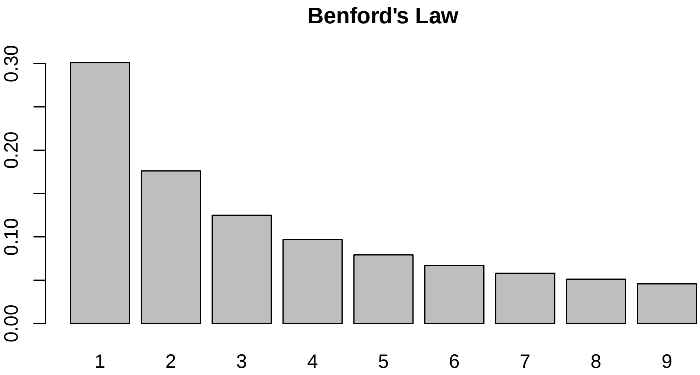
Benford's law

In many numerical datasets, the chance that the first digit will be $n \in \{1, 2, ..., 9\}$ is

$$p_n = \log_{10}\left(1 + \frac{1}{n}\right).$$

Say, the chance that our dataset number will start by 1 is 30.1%, by 2 is 17.6% etc.

LZ



LZ

LZ

This can be used in, say, Accounting fraud detection, Election data Fraud analysis,...

LZ

This can be used in, say, Accounting fraud detection, Election data Fraud analysis,...

See, e.g., `https://en.wikipedia.org/wiki/Benford's_law`

LZ

This can be used in, say, Accounting fraud detection, Election data Fraud analysis,...
See, e.g., https://en.wikipedia.org/wiki/Benford's_law
Here are some examples: http://testingbenfordslaw.com/

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?
- Plot the CDF of X

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?
- Plot the CDF of X
- Plot the PMF of X

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?
- Plot the CDF of X
- Plot the PMF of X
- In which case $X = \sqrt{2}$? When is $X = \sqrt{2}$?

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?
- Plot the CDF of X
- Plot the PMF of X
- In which case $X = \sqrt{2}$? When is $X = \sqrt{2}$? Gotcha $\ddot\smile$!

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?
- Plot the CDF of X
- Plot the PMF of X
- In which case $X = \sqrt{2}$? When is $X = \sqrt{2}$? Gotcha $\ddot\smile$!
- Give some possible result from one Experiment.

Prelude

Consider the r.v. X given by

Values of X	$\sqrt{2}$	0	10
$\mathbb{P}(X = x)$	0.6	0.1	0.3

- Describe the r.v. X: is it Discrete or Continuous?
- Which values X can take?
- What is the Range of X?
- What is the Probability $\mathbb{P}(X = 2)$?
- Plot the CDF of X
- Plot the PMF of X
- In which case $X = \sqrt{2}$? When is $X = \sqrt{2}$? Gotcha $\ddot\smile$!
- Give some possible result from one Experiment.
- Give some possible result from 10 Trials.

Prelude: Poisson Distribution

Let us recall the Poisson Distribution definition: $X \sim Poisson(\lambda)$ $(\lambda > 0)$, if the PMF of X is given by

$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

that is,

Values of X :	0	1	2	$\ldots$	k	$\ldots$
$\mathbb{P}(X = x)$:	$e^{-\lambda}$	$e^{-\lambda} \cdot \frac{\lambda^1}{1!}$	$e^{-\lambda} \cdot \frac{\lambda^2}{2!}$	$\ldots$	$e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	$\ldots$

Prelude: Poisson Distribution

Let us recall the Poisson Distribution definition: $X \sim Poisson(\lambda)$ $(\lambda > 0)$, if the PMF of X is given by

$$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \dots$$

that is,

Values of X :	0	1	2	$\dots$	k	$\dots$
$\mathbb{P}(X = x)$:	$e^{-\lambda}$	$e^{-\lambda} \cdot \frac{\lambda^1}{1!}$	$e^{-\lambda} \cdot \frac{\lambda^2}{2!}$	$\dots$	$e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	$\dots$

Do not read as **Poison** ! $\ddot\smile$

Poisson Distribution: Examples

Poisson Distribution: Examples

- Classic Poisson example: the data set of von Bortkiewicz (1898), for the chance of a Prussian cavalryman being killed by the kick of a horse. See, for example, `http://www.umass.edu/wsp/resources/poisson/index.html`

Poisson Distribution: Examples

- Classic Poisson example: the data set of von Bortkiewicz (1898), for the chance of a Prussian cavalryman being killed by the kick of a horse. See, for example, `http://www.umass.edu/wsp/resources/poisson/index.html`
- DeepMind Lecture on Reinforcement Learning!

Poisson Distribution: Examples

- Classic Poisson example: the data set of von Bortkiewicz (1898), for the chance of a Prussian cavalryman being killed by the kick of a horse. See, for example, `http://www.umass.edu/wsp/resources/poisson/index.html`
- DeepMind Lecture on Reinforcement Learning!
- Some other examples: the number of visitors to a web site per minute; the number of typographical errors found in a book; the number of network failures per day; the number of my OH visitors (maybe no? this changes close to midterms ?); the number of hungry professors visiting AUA cafeteria between 12:30PM and 14:30PM on MWF;

Poisson Distribution: Examples

- Classic Poisson example: the data set of von Bortkiewicz (1898), for the chance of a Prussian cavalryman being killed by the kick of a horse. See, for example, `http://www.umass.edu/wsp/resources/poisson/index.html`
- DeepMind Lecture on Reinforcement Learning!
- Some other examples: the number of visitors to a web site per minute; the number of typographical errors found in a book; the number of network failures per day; the number of my OH visitors (maybe no? this changes close to midterms ?); the number of hungry professors visiting AUA cafeteria between 12:30PM and 14:30PM on MWF;
- Your turn!

Poisson Distribution: R Examples

R Code

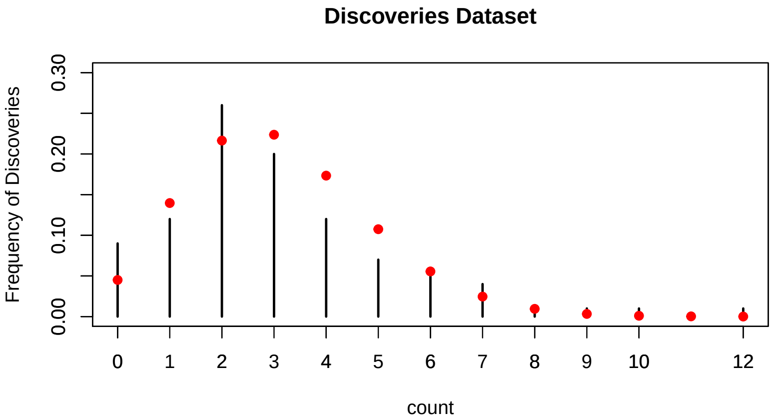
```
#Poisson Distribution
lambda <- 1.5
n <- 10
x <- 0:n
x.prob <- dpois(x,lambda) #PMF
#To have 2 plots side by side
par(mfrow=c(1,2))
#Plotting the PMF
plot(x, x.prob, type = "h", lwd = 5,
   main = "Poisson Distribution PMF")
#Now plotting the CDF
t <- seq(from = -1, to = n+1, by = 0.01)
y <- ppois(t,lambda)
plot(t,y, type = "s", lwd = 3,
   main = "Poisson Distribution CDF")
```

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

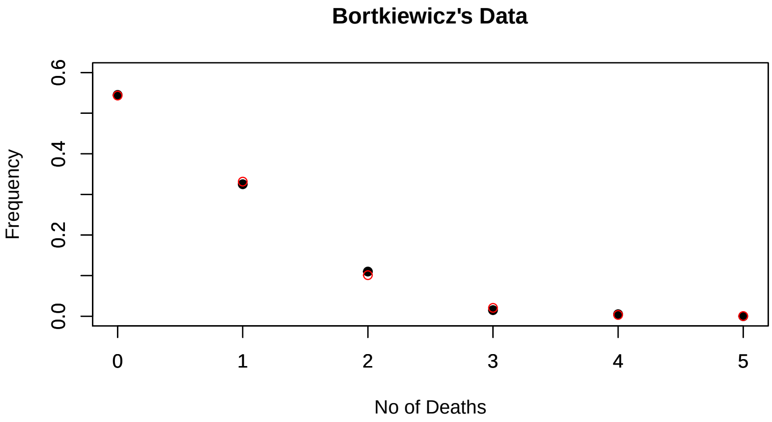


Poisson Distribution: R Examples

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

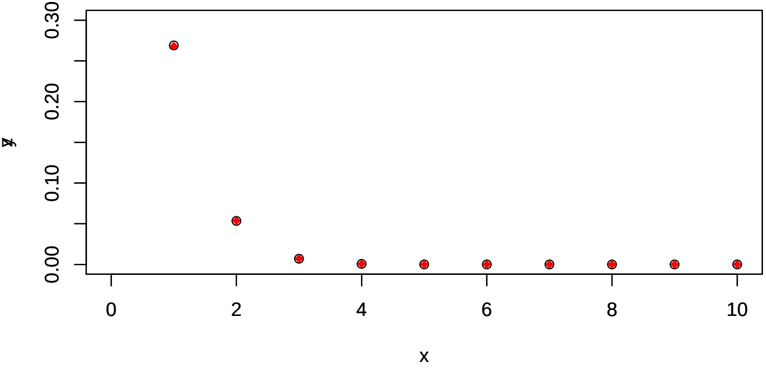
R Code

```
# 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



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.