CIS 490 Machine Learning

Lecture 3

Instructor: (Julia) Hua Fang

1

2

Reminders:

- Task 1: Study groups (Self-assembled study groups) due Jan 25, no more than 3. **If you prefer to work on your own, please do submit your intention here as well.**

  Random group assignment by Jan 28: **If you don't want to work on your own** but haven't found a partner, we will randomly assign you to a group that can accept a member.

  Since you chose to work as a group, each member must sign on the group work agreement, no matter if you chose our own members or you were randomly assigned to a group by Jan 31, posted in the "Course Logistics" folder at MyCourses.

- LA 1: Due Jan 26

2

3

Answers to class exercises in LS 2: Common Continuous Distributions

- Uniform
- Gaussian/Normal
- Student t
- Chi-squared
- Gamma
- Beta
- Pareto

(Q: Which ones belong to exponential family distributions?):
Normal/Gaussian, Chi-squared, Gamma, Beta

Note: Review online if you forget these common Continuous Distributions

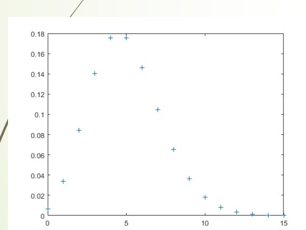
3

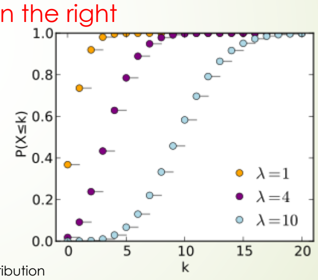
4

Answers to Quick Quiz 1 in LS2

- Which statistical distribution do these two graphs show?
Poisson
- Are they continuous or discrete distributions? Discrete
- Which one shows the pmf of this distribution? The graph on the left

  Which one shows cdf? The graph on the right





https://en.wikipedia.org/wiki/Poisson_distribution

4

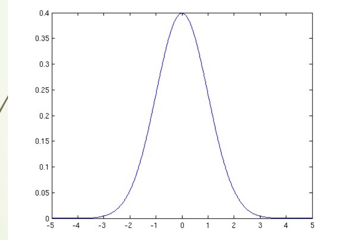
5

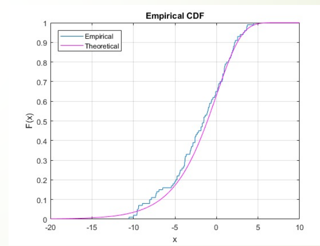
Answers to Class exercise 2 in LS2

- Which statistical distribution do these two graphs show?

Normal/Gaussian

- Are they continuous or discrete distributions? Continuous
- Which one shows the pdf of this distribution? The graph on the left Which one shows cdf? The graph on the right





Refer to: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Normal.html

5

6

Quick quiz 1

If we have a PDF expressed as $\frac{1}{\sqrt{2\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

identify which probability distribution does this PDF describe?

a. Poisson
b. Normal
c. Uniform
d. Gamma

6

7

Quick quiz 2

If we have a PMF expressed as $p\,(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

for x = 0, 1,2,… where λ is the shape parameter which indicates the average number of events in the given time interval, which probability distribution has this PMF?

a. Poisson
b. Normal
c. Uniform
d. Gamma

7

8

Lecture outline

- Review: Probability Theory (II)
  - Normal vs. Standard Normal distribution
  - Covariance vs. correlation
  - Descriptive Statistics
  - Central limited theory (I)

Adapted from Jeff Howbert, Greg Shakhnarovich

8

9

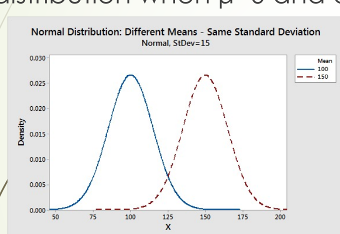
Normal vs. Standard Normal distribution

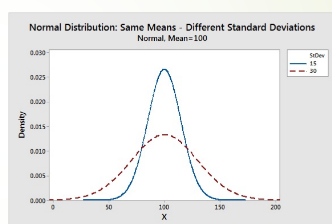
9

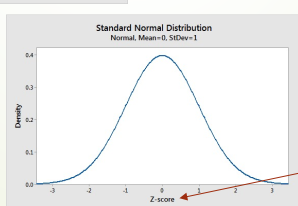
10

Normal vs. standard normal distribution

- **<u>Standard</u>** normal distribution is a special case of normal distribution when $\mu=0$ and $\sigma^2=1$.





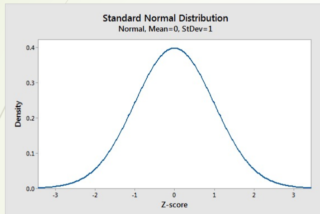


Z-score: standard score, the value on the standard normal distribution

10

11

Normal vs. **Standard** normal distribution: How to compute z-score (in class exercise)



$$Z = \frac{X - \mu}{\sigma}$$

X: ?
μ (Mu): ?
σ (Sigma): ?

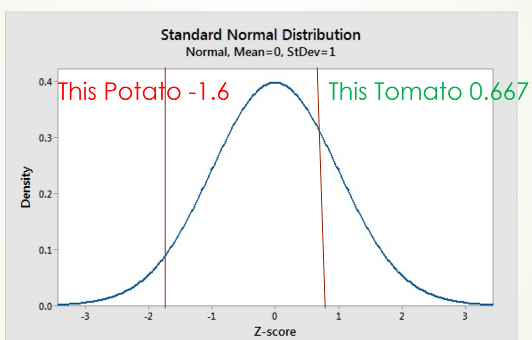
Class exercise: A tomato weighs 110 grams and a potato weighs 100 grams. Let's calculate the z-core of this tomato and z-score of this Potato and compare their standard scores!

	Tomato Population	Potato Population
μ	100	140
σ	15	25

11

12

in class exercise: continued



12

13

Covariance vs Correlation

13

14

Covariance matrix vs Correlation matrix (1)

- Covariance Matrix (**C**): also called variance-covariance matrix. For a sample data (n*d)

$$C = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T \qquad C \in \mathbb{R}^{d \times d} \qquad X \in \mathbb{R}^{n \times d}$$

where ***n*** is # of cases and ***d*** is dimensions or # of random variables.

*Q: Is **C** symmetric; what is the range of elements in **C**?*

```
> correlation <- cor(data)
> correlation
             x1          x2          x3
x1  1.000000000 -0.02348131 0.004810578
x2 -0.023481314  1.00000000 0.140799229
x3  0.004810578  0.14079923 1.000000000
> covariance <- cov(data)
> covariance
            x1          x2         x3
x1  2.99794854 -0.06022025 0.01942999
x2 -0.06022025  2.19389664 0.48648745
x3  0.01942999  0.48648745 5.44159936
```

14

15

Covariance matrix vs Correlation matrix (2)

- Correlation matrix :

Seen as the covariance matrix of the standardized random variables $X_i/\sigma(X_i)$ for $i = 1, \ldots, n$.

and the correlation coefficient of two RVs is $r_{xixj} = \frac{Cov(xi,xj)}{\sigma_{xi}\sigma_{xj}}$ where σ (sigma): standard deviation.

Q: Is Correlation Matrix symmetric; what is the range of elements in this matrix?

```
> correlation <- cor(data)
> correlation
             x1          x2          x3
x1  1.000000000 -0.02348131 0.004810578
x2 -0.023481314  1.00000000 0.140799229
x3  0.004810578  0.14079923 1.000000000
> covariance <- cov(data)
> covariance
            x1          x2         x3
x1  2.99794854 -0.06022025 0.01942999
x2 -0.06022025  2.19389664 0.48648745
x3  0.01942999  0.48648745 5.44159936
```

15

16

Using R: Covariance matrix vs Correlation matrix

- Using R, generate covariance and correlation matrices.

cor( ) returns the correlation matrix

cov() returns the covariance matrix

```
> correlation <- cor(data)
> correlation
             x1          x2          x3
x1  1.000000000 -0.02348131 0.004810578
x2 -0.023481314  1.00000000 0.140799229
x3  0.004810578  0.14079923 1.000000000
> covariance <- cov(data)
> covariance
            x1          x2         x3
x1  2.99794854 -0.06022025 0.01942999
x2 -0.06022025  2.19389664 0.48648745
x3  0.01942999  0.48648745 5.44159936
```

cov2cor(V): convert covariance matrix to correlation matrix

16

17

Descriptive Statistics

17

18

Descriptive Statistics

- Summarize a sample, training or testing/validation data
- Measure central tendency or variability/dispersion of data

18

19

Descriptive Statistics: Measures of central tendency/location

- Measures of central tendency/location:
  e.g., Mean, Median, Mode.

Quick review

Given a data set, 10,10,20,40,70, find mean, median and mode.

NOTE: Review mean and variance formula for common discrete and continuous distributions mentioned in LS2.

e.g. How to calculate mean and variance for binomial data?

19

20

Descriptive Statistics: Measures of variability (1)

- Measures of variability:
  e.g., Variance, Standard Deviation, *Kurtosis, skewness*

Variance: $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$ or $\sigma^2 = \frac{\sum X^2}{N} - \mu^2$

Standard Deviation: σ

Quick review:
1) Given a data set, 3, 4, 4, 5, 6, 8, find variance and Standard Deviation

NOTE: You are expected to know and review mean and variance formula for common discrete and continuous distributions mentioned in LS2.

e.g. How to calculate mean and variance for binomial data?
https://www.sciencebuddies.org/science-fair-projects/science-fair/variance-and-standard-deviation
https://en.wikipedia.org/wiki/Binomial_distribution

20

21

Descriptive Statistics: Measures of variability (2)- Kurtosis

- Kurtosis: a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.
  - ✓ A Gaussian distribution has a kurtosis of 0.
  - ✓ A flatter distribution has a negative kurtosis,
  - ✓ A distribution more peaked than a Gaussian distribution has a positive kurtosis.
  - ✓ Kurtosis has no units.

21

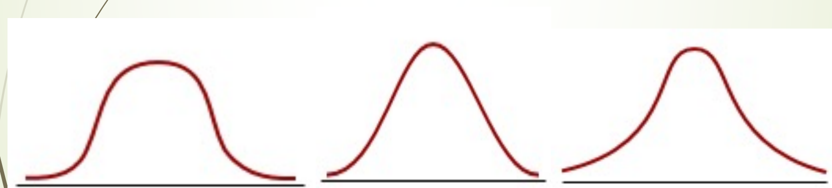
22

Descriptive Statistics: Kurtosis in-class exercises

- Kurtosis example:

Q: 1) Which one is more like a normal distribution?

2) Which one shows Kurtosis< 0 ? Kurtosis =0 ? Kurtosis > 0?



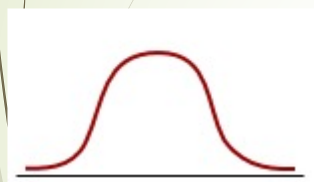
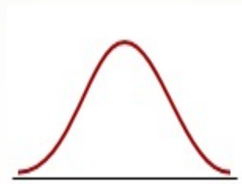
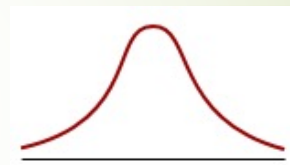
Westfall PH (2014) Kurtosis as Peakedness, 1905 - 2014. R.I.P. The American Statistician 68:191-195.

22

23

Descriptive Statistics: Kurtosis in-class exercises

- Kurtosis example: answer key

Thinner/lighter tails
Kurtosis< 0

(Normal/Gaussian Distribution)
Kurtosis =0

Fatter/Heavier tails
Kurtosis > 0

Westfall PH (2014) Kurtosis as Peakedness, 1905 - 2014. R.I.P. The American Statistician 68:191-195.

https://www.medcalc.org/manual/skewnesskurtosis.php

23

24

Descriptive Statistics: Measures of variability (3)
Skewness in-class exercises

Skewness

- The coefficient of Skewness is a measure for the degree of symmetry in the variable distribution

Q: 1) Which one is more like normal distribution
2) Which one shows Skewness < 0 ? Skewness =0 ? Skewness > 0?



Sheskin DJ (2011) Handbook of parametric and nonparametric statistical procedures. 5th ed. Boca Raton: Chapman & Hall /CRC

24

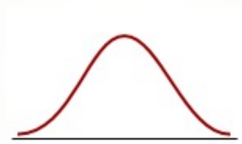
25

Descriptive Statistics

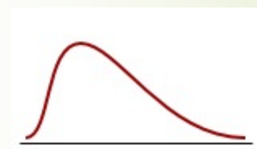
Skewness

- Answer key

Negatively skewed distribution or Skewed to the left Skewness <0



Normal distribution Symmetrical Skewness = 0



Positively skewed distribution or Skewed to the right Skewness > 0

Sheskin DJ (2011) Handbook of parametric and nonparametric statistical procedures. 5th ed. Boca Raton: Chapman & Hall /CRC

25

26

Central Limit Theorem (CLT)

- What is the sampling distribution of the mean?
- What does this CLT definition entail?

26

27

Central Limit Theorem (CLT):

- Given a population with mean μ and standard deviation σ, take sufficiently large random samples from the population with replacement, then the distribution of the sample means (called "the sampling distribution of means") will be approximately normally distributed:

http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability12.html

27

28

Central Limit Theorem (CLT): cont.

- The definition for CLT also refers to "the sampling distribution of the mean."

Example: perform a NIH study once (very costly!), and you might calculate the mean of that one sample.

Suppose that you can repeat the study many times and collect the same sample size for each one.

Well, calculate the mean for each sample and graph them on a histogram.

The histogram displays the distribution of sample means, that is, the sampling distribution of the mean.

28

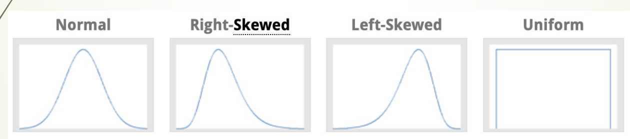
29

Central Limit Theorem (CLT): cont.

- CLT applies to most probability distributions of independent and identically distributed (aka. **iid**) variables with finite variance.

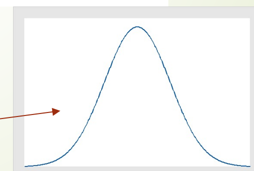
  (Suppl.: Can't apply to Cauchy distribution because it has an infinite variance)

  These population distributions range from normal, skewed, and uniform, etc. (examples below)



What does this CLT definition entail?

Answer: If you draw random samples from a variety of distributions, the sampling distribution of means per CLT is NORMAL!



Cool!

29

30

Central Limit Theorem (CLT): cont.

- **Properties** of CLT
  - The normality features of CLT.

    As the sample size increases, the sampling distribution (of sample means) converges on a normal distribution where
    - Mean of sample means (X) = $\mu_{population}$
    - Standard deviation of sample means (s) = $\sigma_{population}/\sqrt{n}$
    - n = the sample size

Sample size increases, then the sampling distribution of the sample means more closely approximates the normal distribution, and the spread of that distribution tightens.

http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability12.html

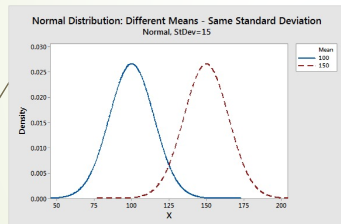
30

31

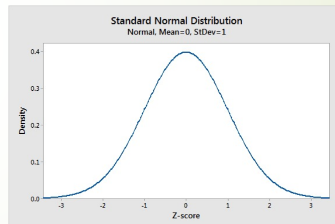
Quick Quiz 1:

Which graph refers to normal distributions?

Which graph is more likely to be standard normal distribution?



a.



b.

31

32

Quick Quiz 2:

1. Is covariance matrix symmetric? Y/N
2. What is the range of elements in covariance matrix?
3. Is correlation matrix symmetric? Y/N
4. What is the range of elements in correlation matrix?
5. Are both Covariance and correlation matrices square matrices? Y/N
6. What are the values called on the diagonal in the covariance matrix? What are the values on the diagonal in the correlation matrix?
7. What is the relationship between covariance matrix and correlation matrix?

32

33

Quick Quiz 3:

1. See the distribution below. Kurtosis< 0 ? Kurtosis =0 ? Kurtosis > 0?

2. See the distribution below. Is it Skewness < 0, Skewness =0 or Skewness > 0? Is it Left- or right-skewed? Is it Positively or negatively skewed?

33