CIS 490 Machine Learning

Lecture 4

Instructor: (Julia) Hua Fang

1

2

Last time

- Review: Probability Theory (II)
  - Normal vs. Standard Normal distribution
  - Covariance vs. correlation
  - Descriptive Statistics
  - Central limited theory (I)
- Reminder: Study group work agreement, due 1/31.

Adapted from Jeff Howbert, Greg Shakhnarovich

2

3

Answers to Quick quiz in LS3

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

3

4

Answers to Quick quiz in LS3

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

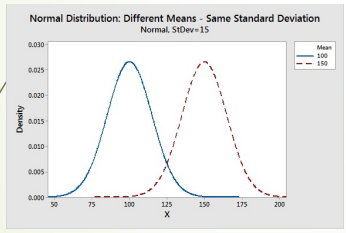
4

5

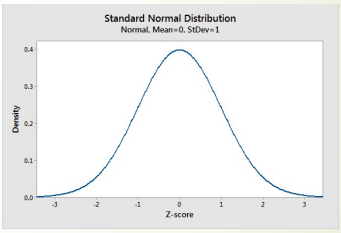
Answers to Quick quiz in LS3

Which graph refers to normal distributions? a

Which graph is more likely to be standard normal distribution? b



a.



b.

5

6

Answers to Quick quiz in LS3

1. Is covariance matrix symmetric? Y/N
2. What is the range of elements in covariance matrix?

   $(-\infty, +\infty)$

3. Is correlation matrix symmetric? Y/N
4. What is the range of elements in correlation matrix?

   $[-1, 1]$

5. Are both Covariance and correlation matrices square matrices? Y/N
6. What are the values called on the diagonal in the covariance matrix? Variance What are the values on the diagonal in the correlation matrix? 1
7. What is the relationship between covariance matrix and correlation matrix?

Answer: Correlation is seen as the covariance matrix of the standardized random variables, ranging from -1 to 1. The correlation measures both the strength and direction of the linear relationship between two variables. Covariance values are not standardized. Therefore, the covariance can range from negative infinity to positive infinity.

6

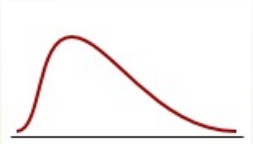
7

Answers to Quick quiz in LS3

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



Skewness > 0, right-skewed, Positively skewed.

7

8

Outline

- Review: Probability Theory (III)
  - Central limited theory (II)
  - Multivariate distribution
    - Joint probability
    - Marginal probability
    - Conditional probability

Adapted from Jeff Howbert, Greg Shakhnarovich

8

9

Recall: Central Limit Theorem

- Sampling distribution of the means
- CLT applies to most probability distributions of independent and identically distributed (iid.) variables with finite variance.

  (Suppl.: Can't apply to Cauchy distribution because it has an infinite variance)
- 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 $(\bar{X})$ = $\mu_{population}$
  - Standard deviation of sample means (s)= $\sigma_{population}/\sqrt{n}$
  - n = the sample size
- Rule of thumb: In order for the result of the CLT to hold, the sample typically should be sufficiently large ($n \geq 30$), except normal distribution.

9

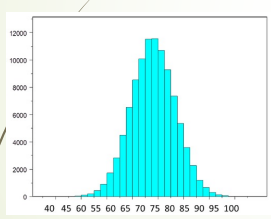
10

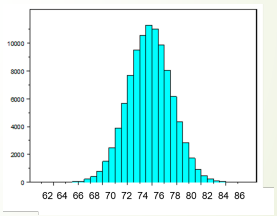
Central Limit Theorem (CLT): in class exercise

- **CLT with a Normal Population:** CLT holds true even for samples n < 30.

(a) True population mean (μ)= 75; standard deviation (σ) of the population =8.

(b) Randomly taking n=10, per CLT, the mean of the sample means = ?; standard deviation of the sample means =?.





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

10

Step for solution: Normal

11

Given

mean of sample means $(\bar{X})$ = $\mu_{population}$,
standard deviation (SD) of sample means (s)= $\sigma_{population}/\sqrt{n}$
per CLT,

The mean of the sample means for this example is

$\bar{X} = \mu_{population} = 75$

The standard deviation (SD) of the sample means for this example is:

$SD = \sigma_{population}/\sqrt{n} = 8/\sqrt{10}$
$=2.53$

11

Central Limit Theorem (CLT): in class exercise

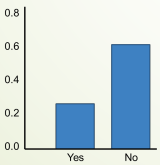
12

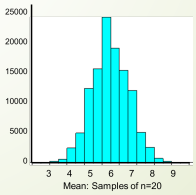
- **CLT with a Binomial Population:** CLT holds true provided that the minimum of np and n(1-p) is at least 5, where n= sample size; p = the prob. of success on any given trial.

(a) True Binomial Population : success of a medical procedure, yes/no, p =0.3

(b) If randomly taking samples of size n =20; per CLT, the mean of the sample means = ?; standard deviation of the sample means =?

Hint: first recall the formula for population mean and variance of a binomial distribution





12

Steps for solution: Binomial

13

Step 1: Check if CLT holds true for this example in the previous slide. Remember CLT holds true provided that the minimum of np and n(1-p) is at least 5 for binomial distribution.

Np= 20*.3=6; n(1-p) =20*.7 = 14, min (np, n(1-p)) = min (6, 14) = 6

Yes, since the min. is larger than 5, CLT holds true for this example

Step 2: Recall Binomial distribution, $\mu_{population} = np$; $\sigma^2_{population} = npq = np(1-p)$.

Given

mean of sample means $(\bar{X}) = \mu_{population}$,

standard deviation (SD) of sample means (s)= $\sigma_{population}/\sqrt{n}$

per CLT,

The mean of the sample means for this example is

$\bar{X} = \mu_{population} = np = 20 * .3 = 6$

The standard deviation (SD) of the sample means for this example is:

$SD = \sigma_{population}/\sqrt{n} = sqrt(\sigma^2_{population})/\sqrt{n}$

$= sqrt[(np(1-p)]/\sqrt{n}$

$= sqrt(20*.3*.7) / sqrt(20) = 0.46$

13

Central Limit Theorem (CLT): in class exercise

14

- Central Limit Theorem with a Skewed Distribution

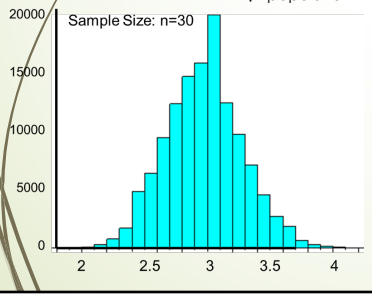
Let's assume the number of spam follows the Poisson distribution

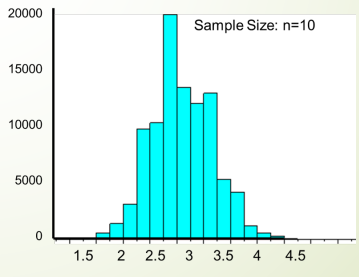
$$p(x,\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

and randomly draw samples of size =30 and size =10 with the mean of sample means ($\bar{X}$) of 3.

Which graph shows CLT seem to apply? Estimate the population mean (μ) and population standard deviation (σ) based on CLT?

$\mu_{population} = ?$; $\sigma_{population} = ?$





14

Steps for solution: Poisson

15

Step 1: Rule of thumb: In order for the result of the CLT to hold, the sample must be sufficiently large ($n \geq 30$), except normal distribution.
So, when sample size n =30, it shows that CLT seems to apply

Step 2: Given
mean of sample means ($\bar{X}$) = $\mu_{population}$,
standard deviation (SD) of sample means (s)= $\sigma_{population}/\sqrt{n}$
per CLT,

The population mean for this example can be estimated as

$$\mu_{population} = \bar{X} = \lambda = 3$$

For this Poisson example, we want to estimate $\sigma_{population}$ = ?, and we know the mean of sample means ($\bar{X}$) = 3, so the variance of sample means = 3 per the formula for calculating the variance for Poisson Distribution. Plug these values in SD = $\sigma_{population}/\sqrt{n}$ per CLT, then

sqrt (λ=3) = $\sigma_{population}/\sqrt{30}$ → $\sigma_{population}$ = sqrt (90).

15

16

Conclusion on Central Limit Theorem (CLT):

- Understanding the central limit theorem is crucial when it comes to trusting the validity of your results and assessing the precision of your estimates.
- Use large sample sizes to satisfy the normality assumption even when your data are nonnormally distributed and to obtain more precise estimates!

16

17

Multivariate probability distributions

17

18

Multivariate probability distributions

- Scenario
  - Several random processes occur (doesn't matter whether in parallel or in sequence)
  - Want to know probabilities for each possible combination of outcomes
- Types:
  - Joint probability
  - Marginal probability
  - Conditional probability

18

Multivariate probability distributions: Joint probability

19

- Joint probability of several random variables

e.g.: two processes whose outcomes are represented by random variables X and Y. The probability that process X has outcome x and process Y has outcome y is denoted as:

$$p(X = x, Y = y)$$

Case study:

Probability Table		X= Model type			
		Sedan	Minivan	SUV	Sport
Y= Manufacturer	European	n/a	0.1481	n/a	n/a
	Asian	n/a	0.1111	n/a	n/a
	American	n/a	0.0741	n/a	n/a

Assuming calculating the probability of selling a high volume of cars

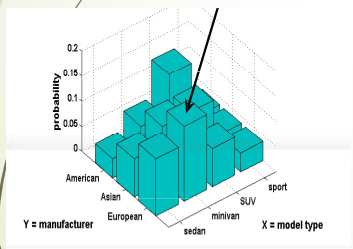
19

Example of Joint Probability: In class exercise

20

Example: Probability Table		X= Model type			
		Sedan	Minivan	SUV	Sport
Y= Manufacturer	European	n/a	0.1481	n/a	n/a
	Asian	n/a	0.1111	n/a	n/a
	American	n/a	0.0741	n/a	n/a

Assuming calculating the probability of selling a high volume of cars



Joint probability:
$p(X = \text{minivan}, Y = \text{European}) = ?$

20

Multivariate probability distributions: *Marginal* probability

21

- Marginal probability
  - Probability distribution of a single variable in a joint distribution

    e.g. two random variables X and Y:

    $p(X = x) = \sum_{y=\text{all values of } Y} p(X = x, Y = y)$

Example: Probability Table		X= Model type			
		Sedan	Minivan	SUV	Sport
Y= Manufacturer	European	n/a	0.1481	n/a	n/a
	Asian	n/a	0.1111	n/a	n/a
	American	n/a	0.0741	n/a	n/a
		Marginal	?		

Assuming calculating the probability of selling a high volume of cars

Marginal probability: *p(X* = minivan) =?

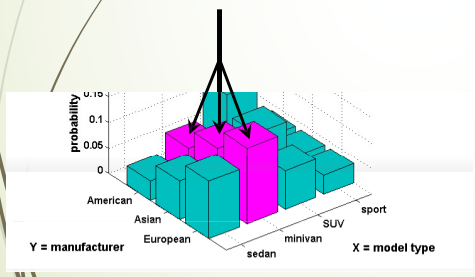
21

Example of Marginal Probability: In class exercise

22

Example: Probability Table		X= Model type			
		Sedan	Minivan	SUV	Sport
Y= Manufacturer	European	n/a	0.1481	n/a	n/a
	Asian	n/a	0.1111	n/a	n/a
	American	n/a	0.0741	n/a	n/a
		Marginal	?		

Assuming calculating the probability of selling a high volume of cars



Marginal probability:
p(X = minivan) =?

22

Multivariate probability distributions: *Conditional* probability

23

- Conditional probability
  - Probability distribution of one variable *given* that another variable takes a certain value

    e.g. two random variables *X* and *Y*:

    $p(X = x \mid Y = y) = p(X = x, Y = y) / p(Y = y)$

Example: Probability Table

Y= Manufacturer	X= Model type				
		Sedan	Minivan	SUV	Sport
	European	n/a	0.1481	n/a	n/a
	Asian	n/a	0.1111	n/a	n/a
	American	n/a	0.0741	n/a	n/a

Assuming calculating the probability of selling a high volume of cars

Conditional probability: *p*(*Y* = European | *X* = minivan) = ?

23

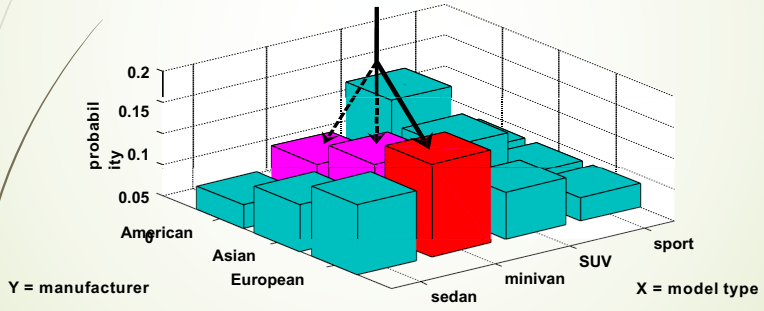
Example of Conditional Probability: In class exercise

24

Example: Probability Table

Y= Manufacturer	X= Model type				
		Sedan	Minivan	SUV	Sport
	European	n/a	0.1481	n/a	n/a
	Asian	n/a	0.1111	n/a	n/a
	American	n/a	0.0741	n/a	n/a
		Marginal	0.3333		

Assuming calculating the probability of selling a high volume of cars



conditional probability: $p(Y = \text{European} \mid X = \text{minivan}) = ?$

24

25

Continuous multivariate distribution

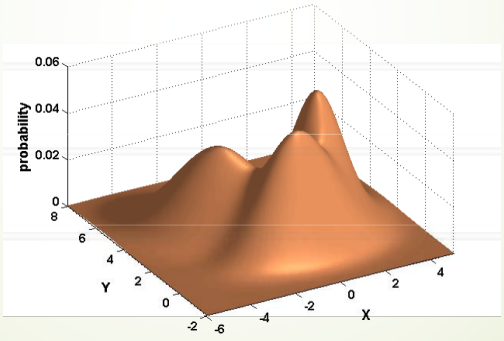
25

Continuous multivariate distribution

26

- Same concepts of joint, marginal, and conditional probabilities apply (except using integrals)

Example: three-component Gaussian mixture in two dimensions



26

27

Homework 1: Due Feb 4th, Posted on MyCourses

Example: use a specified mean vector and a covariance matrix to randomly generate 5 rows (**n=5**) for each of three random variables (RVs), x1, x2, and x3, using a random seed = **10**, from a multivariate **Gaussian** distribution, with a population mean vector of [1, 2, 3] (ie., mean of X1= 4, mean of X2= 9, mean of X3= 25) and a population covariance matrix of these RVs (shown below):

	X1	X2	X3
X1	4	2	3
X2	2	25	1
X3	3	1	9

27

28

Example: continued

Recall: In this example, a population mean vector of [4, 9, 25] and covariance matrix is

	X1	X2	X3
X1	4	2	3
X2	2	25	1
X3	3	1	9

```
> library(MASS) # must install and load this package
> library(moments) # must install and load this package
> mu = c(4,9,25); # mean vector
> mu
> sigma = matrix(c(4, 2, 3, 2, 25, 1, 3, 1, 9), 3, 3); #assign values to 3*3 covariance matrix
>sigma
>  set.seed(10);  # For reproducibility; do not worry about possible different values across different OS.
> rannum <- mvrnorm(n= 5, mu, sigma); # randomly generated values ;
> View(rannum); # view and replicate generated values with specificed mu and cov
```

Alternative Reference: https://astrostatistics.psu.edu/su07/R/html/MASS/html/mvrnorm.html

28

Example: continued

29

Recall: In this example, a population mean vector of [1, 2, 3] and covariance matrix is

	X1	X2	X3
X1	4	2	3
X2	2	25	1
X3	3	1	9

```
>summary(rannum);#get descriptive stats
#or you do like this
> X1 = rannum[,1];
> X2 = rannum[,2];
> X3 = rannum[,3];
> Means = c(mean(X1), mean(X2), mean(X3));
> Means
> Median = c(median(X1), median(X2), median(X3));
> Median
> Skewness = c(skewness(X1), skewness(X2), skewness(X3));
> Skewness
> Kurtosis =c(kurtosis(X1), kurtosis(X2), kurtosis(X3)) ;
> Kurtosis
> cov(rannum); #get cov. Matrix
> cor(rannum); #get cor. Matrix
```

29

Suppl.: Expected value (1)

- E[X]: the **expected value** or the **expectation** of X, or the arithmetic **mean** of the distribution for X.
- ✓ For any discrete random variable,

$$E(X) = \sum_{allx} xp_X(x)$$

Provided this number exists.

- ✓ For any continuous random variable, the expected value of X is the number E [X] defined by

$$E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$$

provided this number exists.

30

Suppl.: Expected value (2)

- Properties of the arithmetic **mean** E[X]

$E[mX+b] = mE[X]+b$, for all constants, m, b

$E[X+Y] = E[X]+E[Y].$

$E[XY] = E[X] \bullet E[Y]$ If X and Y are independent

31

Suppl.: Variance (3)

- For any random variable X, the variance of X is the number Var (X) given by

$$Var(X) = E[(X-\mu_x)^2]$$

- And the standard deviation of X is the number σ_x given by

$$\sigma_x = \sqrt{E[(x-\mu_x)^2]}$$

Where μ_x = E(X).

32

Suppl.: Variance (4)

- Properties of Variance

$Var(c) = 0$ For all constants c.

$Var(mX + b) = m^2 Var(X)$

$Var(X + Y) = Var(X) + Var(Y)$ If X and Y are independent

$Var(X + Y) = Var(X) + 2Cov(X, Y) + Var(Y)$

33