MATH 460: Probability & Statistics — September 2018

Chapter 3: Expectation & Chapter 6: Estimation Theory

Class Notes — **Dr. Basilio**

Southern California Institute of Technology

Wednesday 9.26.2018

* * *

Chapter 4: Probability Distribution Functions

Poisson Distribution

Definition 1: Poisson-Distribution

For a Poisson distribution, we make the following assumptions:

- X is a discrete random variable, i.e. it takes values 0,1,2, ... So X = the number of successes of a Poisson random variable.
- the number of successes in two disjoint time intervals is independent
- the probability of success during a small time interval is proportional to the entire length of the time interval

Then, if μ is the mean of the number of successes in the given time interval, the probability that X is successful x times is given by

$$\boxed{P(X = x) = \frac{e^{-\mu} \cdot \mu^x}{x!}} \tag{1}$$

FACTS:

(a) If n is an interval of time, and p is the probability of success, then $\mu = np$.

(b) A Poisson Distribution is often used to approximate the Binomial Distribution, when n is "large" and p is "small" (general rule is $n \geq 30$ and $p < 0.05$)

(c) A Poisson distribution has $\sigma = \sqrt{\mu}$.

USING CALCULATOR TI83: poissonpdf(μ,x)

$\boxed{DIST}$ key in yellow ($\boxed{2nd}$ > $\boxed{VARS}$) > Scroll to A. "poissonpdf("

- Given mean μ and the number of successes x, the probability can be calculated by the "poisson distribution:"

$$\text{poissonpdf}(\mu, x) = P(X = x) \tag{2}$$

Activity 1: Poisson-Distribution

ACME Realty reports it sells 75 homes in 25 days. What is the probability that exactly 2 homes will be sold tomorrow? *(Note: this is problem 8 from our Midterm review)*

Activity 2: Poisson-Distribution

A Life Insurance (LI) salesman sells on average 3 LI policies per week. Assuming a Poisson Distribution, calculate the probability that in a given week she will sell:

(a) some policies

(b) 2 or more but less than 5 policies

(c) Assuming a five day workweek, what is the probability that in a given day, she will sell a policy?

Activity 3: Poisson-Distribution

A company makes electrical motors. The probability an electrical motor is defective is 0.01. What is the probability that a sample of 415 electrical motors will contain exactly five defective motors?

Activity 4: Visualizing-Poisson-Distribution

A 911 operator receives about six telephone calls between 8 a.m. and 10 a.m.

(a) What is the probability that she receives more than one call in the next 15 minutes?

(b) Plot the histogram for the probability $P(x) = P(X = x)$ for $x = 0, 1, 2, 3, \ldots$

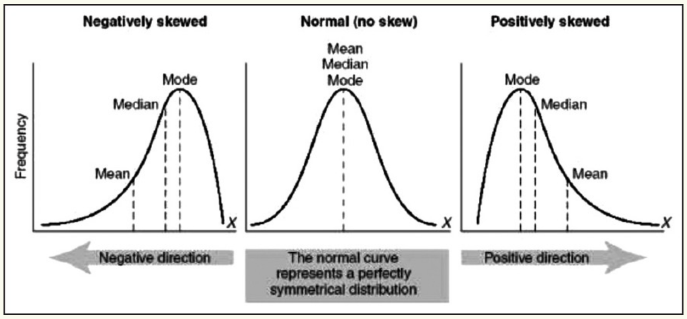
x	$P(x)$
$\vdots$	$\vdots$

Skewness

Definition 2: Skewness

Skewness is asymmetry in a statistical distribution, in which the histogram (or curve) appears distorted or skewed either to the left or to the right. Skewness can be quantified to define the extent to which a distribution differs from a normal distribution.

- **Positively Skewed:** the "tail" of the distribution is to the right of the mean, or the mean is greater than the median and mode ("hump")
- **Negatively Skewed:** the "tail" of the distribution is to the left of the mean, or the mean is smaller than the median and mode ("hump")



Activity 5: Poisson-Distribution-Skewness

Look back at your histogram from Activity 4 part (b). Is the Poisson Distribution positively or negatively skewed?

Activity 6: Frequency-Skewness

The following is a list of prices (in dollars) of birthday cards found in various drug stores:

1.45	2.20	0.75	1.23	1.25
1.25	3.09	1.99	2.00	0.78
1.32	2.25	3.15	3.85	0.52
0.99	1.38	1.75	1.22	1.75

(a) Organize this data with intervals of 50 cents (i.e. .50-0.99, 1.00-0.49, and so on) using create a frequency distribution table.

(b) Draw a Histogram of the data. State the skewness of the data.

Chapter 3: Expectation

Definition 3: Expectation

- A very important concept in probability and statistics is that of the mathematical expectation, expected value, or briefly the **expectation**, of a random variable. Expected value uses probability to tell us what outcomes to expect in the long run.
- For a discrete random variable X having the possible values $x_1, x_2, \ldots, x_n$, the expectation of X is defined as

$$E(X) = x_1 P(X = x_1) + x_n P(X = x_n) + \cdots + x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i) \tag{3}$$

- As a special case of (2), where the probabilities are all equal, we have

$$E(X) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{4}$$

  which is called the arithmetic mean, or simply, the mean of the $x_1, x_2, \ldots, x_n$.
- The expectation of X is very often called the mean of X and is denoted by μ_X, or simply μ, when the particular random variable is understood.
- The mean, or expectation, of X gives a single value that acts as a representative or average of the values of X, and for this reason it is often called a **measure of central tendency**.

Example 1: Expectation

(a) Alex is a basketball player who makes 50% of his 2-point shots and 20% of his 3-point shots. If $X =$ making a 2-point shot, then his expectation in the long term is $E(X) = 2 \cdot 0.5 = 1$ point since the outcome is $x = 2$ and the probability is $p = 0.5$. If $X =$ making a 3-point shot, then his expectation in the long term is $E(X) = 3 \cdot 0.2 = 0.6$ points since the outcome is $x = 3$ and the probability is $p = 0.2$. This is confusing to many people at first since you can't score 0.6 points! But this is because we are taking an average!

(b) If marks of five students is given to be 65, 76, 88, 34, and 90, then the expected value of mark for a random student is

$$E(X) = \frac{65 + 76 + 88 + 34 + 90}{5} = 70.6$$

Activity 7: Expectation

Suppose that a game is to be played with a single die assumed fair. In this game a player wins $20 if a 2 turns up, $40 if a 4 turns up; loses $30 if a 6 turns up; while the player neither wins nor loses if any other face turns up.

(a) State what the random variable X is

(b) Find all the outcomes $x_1, \ldots, x_6$

(c) Find all the probabilities for each respective outcome

(d) Find the expected sum of money to be won (or lost).

(e) In a fair game, what do you think is a reasonable buy-in is in order to play the game?

Activity 8: Expectation

A game is played where a player rolls a six sided die and if the result is an even number, they win 4 times the number in dollars, but if the result is odd, they lose 6 times the number in dollars. Find the expected winnings (or losings).

(a) Find the expected winnings (or losings).

(b) Even if the game is free, should you play?

Chapter 6: Estimation Theory

Confidence Intervals

Suppose you were trying to determine the mean rent of a two-bedroom apartment in your town. You might look in the classified section of the newspaper, write down several rents listed, and average them together. You would have obtained a point estimate of the true mean. If you are trying to determine the percentage of times you make a basket when shooting a basketball, you might count the number of shots you make and divide that by the number of shots you attempted. In this case, you would have obtained a point estimate for the true proportion.

Definition 4: Inferential-Statistics

- **POPULATION vs SAMPLE:** In statistics, we frequently want to understand data that comes from a particular group. For example, suppose I want to have a better understanding of all college student's in CA. The **population** is the entire group we are looking to study. In this case, our population is all students who are attending college in CA. Since it is generally not possible to collect data on an entire population, we collect data from a smaller group or subset taken from the population. This smaller group is called the **sample**.
- **PARAMETER vs STATISTIC:** A number that represents a characteristic of the population is called a **parameter**. A number that represents a characteristic of the sample is called a **statistic**.
- **INFERENTIAL STATISTICS:** We use **sample data** to make generalizations about an unknown population. This part of statistics is called **inferential statistics**. The sample data help us to make an estimate of a population parameter.
- **CONFIDENCE INTERVALS:** We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. After calculating point estimates, we construct interval estimates, called **confidence intervals**.
- **ASSUMPTION:** For $n \geq 30$ the sample can be assumed to be nearly a normal distribution.

Definition 5: Confidence-Intervals

- **ASSUMPTION:** For $n \geq 30$ the sample can be assumed to be nearly a normal distribution.
- **POPULATION vs SAMPLE:** When you conduct a survey and calculate the sample mean, $\bar{x}$, and the **sample standard deviation**, S. You would use $\bar{x}$ to estimate the population mean and S to estimate the population standard deviation. The sample mean, $\bar{x}$, is the point estimate for the population mean, μ. The sample standard deviation, s, is the point estimate for the population standard deviation, σ.
- **STANDARD ERROR OF MEAN:** The **standard error of the mean** is given by $\frac{s}{\sqrt{n}}$.
- **CONFIDENCE:** Let C be a number from 0 to 1 and $C \cdot 100\%$ be a percentage between 0 and 100%.
- **z-SCORE:** Let z_C be the z-score such that the area between the interval $[-z_C, z_C]$ is C.
  To do this, you can use the inverse Normal distribution to find z_C but it is not simply C because we want the middle area. The formula is:

$$\boxed{z_C = \text{invNorm}\left(\frac{1+C}{2}\right)}$$

Table 6-1

Confidence Level	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

- **CONFIDENCE INTERVAL:** the interval that contains the mean of the population μ with a confidence of $C \cdot 100\%$

$$\boxed{\bar{x} \pm z_c \cdot \frac{s}{\sqrt{n}}} \text{ or } \left[\bar{x} - z_c \cdot \frac{s}{\sqrt{n}}, \quad \bar{x} + z_c \cdot \frac{s}{\sqrt{n}}\right] \tag{5}$$

- In general:
  - The more confident (larger C) we are that we know where the population mean, μ is, then the bigger the interval will be!
  - The shorter the interval predicting where the population mean μ is, then we will be less confident (smaller C)!

Example 2: Confidence Intervals

If you worked in the marketing department of an entertainment company, you might be interested in the mean number of songs a consumer downloads a month from iTunes.
Suppose we do not know the population mean μ, but we do know that the population standard deviation is $\sigma = 1$ and our sample size is 100.
The standard deviation for the sample mean is $\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 0.1$.
The empirical rule, which applies to bell-shaped distributions, says that in approximately 95% of the samples, the sample mean, $\bar{x}$, will be within two standard deviations of the population mean μ. So $z_C = 2$. Two standard deviations is $2 \cdot 0.1 = 0.2$. The sample mean $\bar{x}$ is likely to be within 0.2 units of μ.
Because $\bar{x}$ is within 0.2 units of μ, which is unknown, then μ is likely to be within 0.2 units of $\bar{x}$ in 95% of the samples. The population mean μ is contained in an interval whose lower number is calculated by taking the sample mean and subtracting two standard deviations and whose upper number is calculated by taking the sample mean and adding two standard deviations. In other words, μ is between $\bar{x} - 0.2$ and $\bar{x} + 0.2$ in 95% of all the samples.
Suppose that a sample produced a sample mean $\bar{x} = 2$. Then the unknown population mean μ is between $\bar{x} - 0.2 = 2?0.2 = 1.8$ and $\bar{x} + 0.2 = 2 + 0.2 = 2.2$.
We say that we are 95% confident that the unknown population mean number of songs downloaded from iTunes per month is between 1.8 and 2.2. The 95% confidence interval is $[1.8, 2.2]$.
The 95% confidence interval implies two possibilities. Either the interval (1.8, 2.2) contains the true mean μ or our sample produced an $\bar{x}$ that is not within 0.2 units of the true mean μ. The second possibility happens for only 5% of all the samples (95-100%).

Activity 9: Confidence-Interval

Find a $C \cdot 100\%$ confidence interval for μ for the given values:

(a) $\bar{x} = 75$, $s = 13.2$, and $n = 57$

(b) $\bar{x} = 315$, $s = 63$, and $n = 100$

Activity 10: Confidence-Interval

Below are the number of times per year 38 randomly selected employees for a large company feel overworked.

$$15, 23, 50, 31, 5, 27, 47, 43, 135, 164, 80, 123, 20, 34, 45, 56, 7, 12, 15,$$
$$16, 18, 64, 79, 84, 19, 32, 34, 56, 200, 0, 16, 61, 31, 52, 61, 70, 365, 105$$

(a) Find a 85% confidence interval for μ for the population mean of this data.

(b) Find a 90% confidence interval for the population mean of this data.