- 23

Stat 230 - Probability **Fall 2016**

Lecture 18 - 23: October 24 - November 4, 2016

Lecturer: Nagham Mohammad *Notes By: Harsh Mistry*

18.1 Expected Value and Variance

18.1.1 Summarizing Data on Random Variables

18.1.1.1 Types of Descriptive statistics

- Representing data using visual techniques
  - Tables
  - Graphs
- Numerical summary measures for data sets
  - Central Tendency
  - Variation

18.1.1.2 Pictorial Methods : Histogram

- The purpose of a histogram is to put numerical information into graphic form so it is easier to understand
- Histograms are good summaries of data because they show the variability in observed outcomes

Constructing a Histogram :

1. Determine the frequency and relative frequency of each x.

$$\text{Frequency of a value} = \text{number of times the value occurs}$$

$$\text{Relative frequency of a value} = \frac{\text{number of times the value occurs}}{\text{number of observations in teh data set}}$$

2. Mark possible x values on horizontal scale
3. Above each value of x, draw a rectangle whose height is the relative frequency or the frequency of that value.

Describing the shape of a Histogram :

- Symmetry : If one half is a mirror image of the other

- Unimodal : Single prominent peak (A local maximum in a chart)
- Bimodal : Two prominent peaks
- Multimodal : more than two prominent peaks
- Skewness : weather or not the data is pulled to one side

18.1.2 Measures of Location

1. Arithmetic Mean : The arithmetic average value of observations
2. Median : The middle value
   a) Can be found by arranging all the observations from lowest value to highest value and picking the middle one.
   b) If there are two middle numbers, then take their mean for the orders values

$$\widetilde{x} = \begin{cases} \text{The single middle value if n is odd} = (n+1)/2 \\ \text{The average of the two middle values of n is even} = \text{averge of } \{n/2\} \text{ and } \{(n/2)+1\} \end{cases}$$

18.1.2.1 The Sample Mode

The mode is the value that appears most often in a set of data.

- If there is 2 modes, then the data set may be said to be **bimodal**
- if there is more than 2 modes, then set may be described as **multimodal**
- If all items in the set equally repeat the same amount of times, then the set is not unique and said to be **uniform**

18.1.3 Variability

18.1.3.1 Sample Standard Deviation

- The standard deviation is used to describe the variation (measure of dispersion) around the mean
- To get the standard deviation of a sample of data :
  - Calculate the variance of S^2
  - Take the square root to get the standard deviation S
- The larger the S, the further the individual cases are from the mean
- The smaller the S, the closer the individual scores are to the mean.

Note:

- Although variance is a useful measure of spread, its units are units squared
- The standard deviation is more intuitive, because it has the same units as the raw data and the mean
- Like the mean, the s.d will be inflated by an outlier

18.1.3.2 Mean and Variance of Discrete Random Variables

- Used to summarize the probability distribution of a random variable X
- The expectation (also called the mean or the expected value) of a discrete random variable X with probability function $f(x)$
- The mean is a measure of the center of the probability distribution.
- the expectation of X is also often denoted by the Greek letter μ
- The mean of the discrete random variable X is weighted average of the possible values of X, with weights equal to the probabilities

If X is a discrete Random Value, then

- The expected value or mean of X is

$$\mu_x = E(X) = \sum_x x f(x)$$

- The variance of X, denoted as σ^2 or $V(x)$, is

$$\begin{aligned}\sigma^2 = V(X) = E(X-\mu)^2 &= \sum_x (x-\mu)^2 f(x) = \sum x^2 f(x) - \mu^2 \\ &= E(X^2) - [E(X)]^2\end{aligned}$$

- The standard deviation of X is $\sqrt{\sigma^2} = \sigma$
- Note : $Var(X) = E(X^2) - \mu^2$

18.1.3.3 The Expected Value of a Function

If x is a discrete random value with set of possible values D and probability function f(x) then the expected value of any function g(x), denoted by E[g(x)] or $\mu_{g(x)}$, is computed as

$$E[g(X)] = \sum_{\text{all x}} g(x) f(x)$$

Rule of Expected Value :

$$(E(aX+b) = aE(X) + b$$

Properties of Expectation :

1. For constants a and b,
$$E[ag(X)+b] = aE[g(x)] + b$$
2. For constants a and b and functions g_1 and g_2, it is also easy to
$$E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$$

18.1.3.4 Rule of Variance

$$V(aX + b) = a^2\sigma_x^2$$
$$\sigma_{ax+b} = \mid a \mid \sigma_x$$

18.1.4 The mean and variance of X

18.1.4.1 Binomial Distribution

If x follows a Binomial distribution with parameters n and p: X Binomial(n,p), then

$$\mu_x = E(X) = np$$
$$\sigma_x^2 = Var(X) = np(1-p)$$
$$\sigma_x = SD(X) = \sqrt{np(1-p)}$$

18.1.4.2 Hyper geometric Distribution

$$E(X) = n(r/N) = np$$
$$V(X) = \frac{N-n}{N-1} \cdot n \cdot p \cdot (1-p)$$

18.1.4.3 Poisson Distribution

- Mean : $\mu = \lambda t$
- Variance : $\sigma^2 = \lambda t$
- Standard Deviation : $\sigma = \sqrt{\lambda t}$

18.1.4.4 Uniform Distribution

$$E(X) = \frac{a+b}{2}$$
$$Var(X) = \frac{(b-a+1)^2 - 1}{12}$$

18.2 Continuous Probability Distributions

18.2.1 Continuous Random Variables and Probability Density Functions

A random variable that can assume any value in an entire interval of real numbers is said to be continuous, that is for some $a < b$, any number x between a and b is possible, and P(X = x) = 0 for each x.

A probability density function (p.d.f) of a continuous random variable X is a function f(x), such that for any two numbers $a \leq b$

$$p[a \leq X \leq b] = \int_a^b f(x)dx$$

which has the following properties :

1. It is non-negative
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

18.2.2 The Cumulative Distribution Function (c.d.f)

The cumulative distribution function, F(x) for a continuous r.v. X is defined for every number $x \in R$ by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

For each, F(x) is the area under the density curve to the left of x.

18.2.2.1 Properties

- The function F is non-negative : $F(x) \geq 0$
- The function F is non-decreasing : If $b \geq a$, then $F(b) \geq F(a)$
- The function F is continuous for all $x \in R$

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(X) = 1$$

- If X is a purely discrete random variable, then it attains values$x_1, \ldots,$ with probability $p_i = P(x_i)$ and the CDF of X will be discontinuous at the points x_i and constant in between
- The cumulative distribution function of a continuous probability distribution
- The cumulative distribution function which has both a continuous part and discrete part.

18.2.2.2 Using F(x) to Compute Probabilities

- Let X be a continuous r.v with p.d.f, f(x) and c.d.f. F(x). Then for any number a
- $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$
- $P(a \leq x \leq b) = F(b) = F(a)$

18.2.2.3 Obtaining f(x) from F(x)

- Suppose X is a continuous random variable with cumulative distribution function F(x).
- The **probability density function (p.d.f)** of X is defined as

$$F'(x) = f(x)$$

$$f(x) = \frac{d}{dx}F(x)$$

18.2.3 Defined Variables or Change of Variable

When we know the p.d.f. or c.d.f. for a continuous random variable X we sometimes want to find the p.d.f. or c.d.f. for some other random variable Y which is a function of X

1. Write the c.d.f of Y as a function of X
2. Use $F_X(x)$ to find $F_Y(y)$. Then if you want the p.d.f. $f_Y(y)$ you can differentiate the expression $F_Y(y)$
3. Find the range of values of y

18.2.4 Expected value for the Continuous Random Variables

The expected or mean of continuous r.v. X with p.d.ff(x) is

$$\mu_x = E(X)$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

18.2.5 The Variance of a Continuous Random Variable

If the random variable X is continuous with probability density function f(x), then the variance is given by

$$Var(X) = E((X - \mu)^2)$$

$$Var(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2$$

where μ is the expected value.