Elementary Statistics

STAT 251

Lecture 18

Midterm Exam Review

Dr. Lasantha Premarathna

Midterm Exam Review

- Chapter 1: Summary and Display of Univariate Data
- Chapter 3: Sets and Probability
- Chapter 4: Random Variables and Distributions
- Chapter 5: Normal Distribution
- Chapter 6: Some Probability Models

Chapter 1: Summary and Display of Univariate Data

- Classification of Variables
- Descriptive vs. Inferential Statistics
- Summarizing data using tables and graphs
  - Frequency Table
  - Pie Chart
  - Bar Graphs
  - Dot plot
  - Stem-and-leaf plots
  - Histograms

    Describing a distribution
  - Box plot
- Measures of Center
  - Mean
  - Median
  - Comparing the Mean and Median

2

Chapter 1: Summary and Display of Univariate Data

- Measures of variability
  - Range
  - Variance
  - Standard deviation
  - Inter-quartile range ($IQR = Q_3 - Q_1$)
- Percentiles
- Identifying outliers
- How do outliers affect mean, median, variance, standard deviation, Q_1, Q_3, IQR, etc
- How do location/scale changes affect mean and variance

Chapter 3: Sets and Probability

- Random experiments
- Sample Space
- Event
- Equally likely outcomes
- Set theory for events using Venn Diagrams
  - Complement of an event
  - Intersection of events
  - Union of events
  - Disjoint or Mutually Exclusive Events
- Addition Rule
- Complement rule

Chapter 3: Sets and Probability

- Conditional Probability
- Multiplication Rule
- Independent Events Defined Using Conditional Probabilities
- Bayes Theorem
- Tree Diagram

Chapter 4: Random Variables and Distributions

- Notations
- Discrete Random Variables
- Continuous Random Variables
- Probability mass function (pmf)
- Probability density function (pdf)
- The Mean, the Variance and the Standard Deviation
- Cumulative Distribution Function (cdf)
- Max and Min of Independent Random Variables

Chapter 4: Random Variables and Distributions

- Random Variables
  - Discrete random variable
  - Continuous Random Variables

- Discrete random variable
  - Probability mass function (pmf) of a discrete random variable X, is defined as
    $f(x) = P(X = x)$ **for all possible values of** X

  - $f(x)$ gives the probability of each possible value x of the rv X. It has the following properties.
    1. $f(x) \geq 0$ for all x in X
    2. $\sum_x f(x) = 1$

Chapter 4: Random Variables and Distributions

- Discrete random variable
  - The cumulative distribution function (cdf) of a discrete random variable X with pmf $f(x)$ is defined as

$$F(x) = P(X \leq x) = \sum_{k \leq x} f(k), \qquad \text{for all real } x$$

  - Mean/Expectation /Expected value of a discrete random variable X with pmf $f(x)$ is

$$\mu = E(X) = \sum_{x} x f(x)$$

  - Expected value of some function $g(X)$ corresponding to the random variable X with pmf $f(x)$ is

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

Chapter 4: Random Variables and Distributions

- Consider the discrete random variable X with pmf $f(x)$
  - Variance of X is

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

  - standard deviation of X is

$$\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var(X)}}$$

  - Following formula is often used to calculate the variance

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

- Continuous random variable
  - Let X be a continuous random variable. Then the probability density function (pdf) of X is a function $f(x)$ such that for any two numbers a and b with $a < b$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

  - $f(x)$ to be a legitimate pdf, it must satisfy
    1. $f(x) \geq 0$ for all x
    2. $\int_{-\infty}^{\infty} f(x)dx = 1$

       = this is the area under the graph of $f(x)$

Chapter 4: Random Variables and Distributions

- Continuous random variable
  - The cumulative distribution function (cdf) of a continuous random variable X with pdf $f(x)$ is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \qquad \text{for all } x$$

  - Mean/Expectation /Expected value of a continuous random variable X with pdf $f(x)$ is

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

  - in general

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

$g(X)$ is a function of X

Chapter 4: Random Variables and Distributions

- Continuous random variable
  - Variance of X is

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

  - Standard deviation of X is

$$\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var(X)}}$$

** Following formula is often used to calculate the variance

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

Chapter 4: Random Variables and Distributions

- Uniform Distribution
  If X is a uniform random variable, X is uniformly distributed on the interval [a,b]

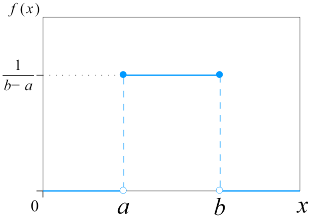
pdf of X is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & ; a \leq x \leq b \\ 0 & ; \text{otherwise} \end{cases}$$

Notation: $X \sim U(a, b)$

$$\mu = E(X) = \frac{a+b}{2}$$

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

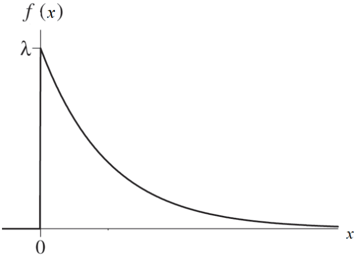


Chapter 4: Random Variables and Distributions

- Exponential Distribution

  Exponential random variables are often used to model the time until an even occur. If X is a exponential random variable with $\lambda > 0$ (rate parameter), then the pdf of X is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x \geq 0 \\ 0 & ; x < 0 \end{cases}$$



Notation: $X \sim Exponental(\lambda)$

or $X \sim Exp(\lambda)$

$$\mu = E(X) = \frac{1}{\lambda}$$

$$\sigma^2 = Var(X) = \frac{1}{\lambda^2}$$

Chapter 4: Random Variables and Distributions

- Properties of the Mean and Variance

  1. $E(aX + b) = aE(X) + b$ for all constants a and b
  2. $E(X + Y) = E(X) + E(Y)$ for all pairs of random variables X and Y.
  3. $E(XY) = E(X)E(Y)$ for all pairs of **independent random variables** X and Y
  4. $Var(aX + b) = a^2 Var(X)$ for all constants a and b
  5. If X and Y are **independent random variables**

     $$Var(X + Y) = Var(X) + Var(Y)$$
     $$Var(X - Y) = Var(X) + Var(Y)$$

Chapter 4: Random Variables and Distributions

- Covariance
  - In bivariate setting involving random variables X and Y, covariance is given by
$$Cov(X,Y) = E[XY] - E[X]E[Y]$$
  - If X and Y are **independent random variables**
$$Cov(X,Y) = 0$$
  - $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
  - More generally
$$Var(aX + bY + c) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$
  - if X and Y are **independent random variables**
$$Var(aX + bY + c) = a^2Var(X) + b^2Var(Y)$$

Chapter 4: Random Variables and Distributions

- Linear Combination of independent rvs
- Average of iid rvs with mean $\mu$and variance σ^2
- Max and Min of Independent Random Variables

Chapter 5: Normal Distribution

- Normal Distribution, $X \sim N(\mu, \sigma^2)$
- The 68-95-99.7 rule (Empirical rule)
- Z- Score
- The Standard Normal Distribution, $Z \sim N(0, 1)$
- Calculate normal probabilities using standard normal distribution
- linear combination of Normal random variables

Chapter 6: Some Probability Models

- Bernoulli Distribution

$$X \sim \text{Bernulli}(p)$$

When the random variable X follows a Bernoulli distribution with probability of success p, then it's has the following pmf

$$P(X = x) = p^x (1-p)^{1-x}\ , \qquad x = 0, 1$$

  - Mean and Variance of Bernoulli random Variable

$$\text{mean} = \mu = E(X) = p$$

$$\text{variance} = \sigma^2 = Var(X) = p(1-p)$$

Chapter 6: Some Probability Models

- Binomial Distribution
  - Binomial random variable is the number of success for n independent trials.
  - $X \sim \text{Bin}(n, p)$ and it has the pmf

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \;, \qquad x = 0, 1, 2, 3, \;\cdots\;, n$$

    where n is the number of trials and p is the probability of success
  - Mean and Variance of Binomial random Variable

$$\text{mean} = \mu = E(X) = np$$

$$\text{variance} = \sigma^2 = Var(X) = np(1-p)$$

Chapter 6: Some Probability Models

- Geometric Distribution
  - A Geometric random variable X counts the number of independent trials to get the first success

$$X \sim Geo(p) \quad ; p \text{ is the probability of success}$$

$$\underline{pmf}: \quad P(X = x) = p(1-p)^{x-1} \qquad ; x = 1, 2, 3, \cdots$$

$$\text{mean} = \mu = E(X) = \frac{1}{p}$$

$$\text{variance} = \sigma^2 = Var(X) = \frac{1-p}{p^2}$$

Chapter 6: Some Probability Models

- Poisson Distribution
  - A random Variable X: the number of occurences in a given interval/space has a poison distribution with $\lambda > 0$, if it has the pmf
    $$X \sim Poisson(\lambda)$$
    $$\underline{pmf} : P(X = x) = \frac{\lambda^x e^{-\lambda}}{x\,!} \qquad ; x = 0, 1, 2, 3, \cdots$$
    where λ is the rate of occurrences
  - The Poisson distribution is especially good at modeling rare events.

$$\text{mean} = \mu = E(X) = \lambda$$

$$\text{variance} = \sigma^2 = Var(X) = \lambda$$

Chapter 6: Some Probability Models

- Poisson process and associated random variables
  - Let the random Variable X: the number of occurrences in a given interval has a Poisson distribution

$$X \sim Poisson(\lambda)$$

  - Then let T be the time between two consecutive occurrences of events. We also can consider T as the waiting time until the first event. Then T is a continuous random variable and it has exponential density.

$$T \sim Exp(\lambda)$$

pdf of T: $f(t) = \lambda e^{-\lambda t} \qquad ; t \geq 0$

cdf of T: $F(t) = 1 - e^{-\lambda t} \qquad ; t \geq 0$

$$E(T) = \frac{1}{\lambda} \qquad Var(T) = \frac{1}{\lambda^2}$$

Next class ...

Visit the course website at canvas.ubc.ca

- Next class:
  - **Midterm Exam: Wednesday, October 26 (8:00am-8:50am)**