Midterm Review

Midterm Exam Rules

- **Zoom meeting: go to your TA section**
- April 29, Thursday 4 pm
- 110 minutes, starts promptly
- Calculators
  - No cell phones, computers, pagers
- Reference sheet:
  - Letter size, 1-side handwritten only
  - no problems or examples
- Turn in both reference sheet & exam: No submission of reference sheet imposes a penalty

Midterm Exam Rules

- Write and sign Honor Code
- Show your desk area using webcam in the beginning of the exam
- No speaking nor leaving the desk before completing exam/submission
- You can ask questions using chat window
- Submit both solutions and reference sheet
- Early submission is fine. Leave the zoom meeting after the submission is complete and TA's confirmation
- Internet disconnection: record locally until rejoining the zoom meeting and submit it with time stamp after the exam*
- Proctoring is recorded incl the submission process

Midterm Review

- Descriptive Statistics
- Probability reasoning questions
  - Unions and Intersections, Complements
  - Independence
  - Law of Total probability
- Conditional Probability and Bayes' Theorem
- Discrete Random Variables and Probability Distribution
  - Binomial Distribution

Measure of Location

- **Mean**

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- **Median**
  - Order the n data points from smallest to largest
    - the number in position $\frac{n+1}{2}$ if n is odd
    - the average of the numbers in positions $\frac{n}{2}$ and $\frac{n}{2}+1$ if n is even
- **Mode**
  - The value that has the highest frequency

Measure of Variability

- Sample Variance

$$\mathrm{Var}[x] = s^2$$

$$= \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$= \frac{S_{xx}}{n-1}$$

- Sample Standard Deviation

$$s = \sqrt{s^2}$$

Quartiles and Percentiles

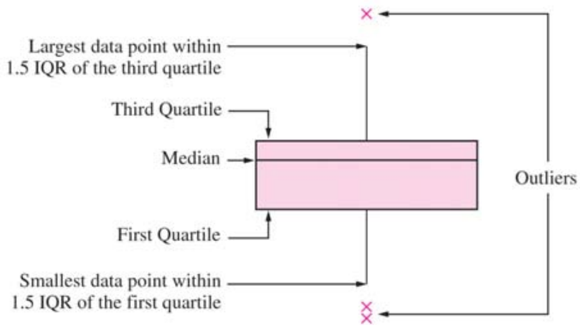
- Quartiles
  - Lower fourth (lower quartile)
  - Upper fourth (upper quartile)
  - Fourth spread (fs)

$$f_s = IQR = \text{upper fourth} - \text{lower fourth}$$

- Percentiles
  - The 25th percentile = Lower quartile (Q1)
  - The 50th percentile = Median (Q2)
  - The 75th percentile = Upper quartile (Q3)

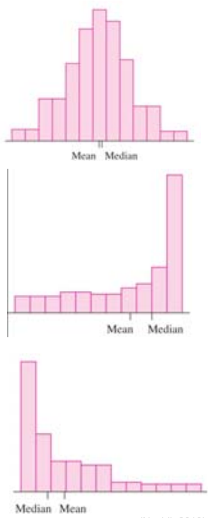
Boxplots

- A boxplot is a graphic that presents the median, the first and third quartiles, and any outliers present in the sample



(Navidi, 2010)

Histogram



(Navidi, 2010)

- A histogram is symmetric if its right half is a mirror image of its left half
  - Mean ≅ Median
- Histograms that are not symmetric are referred to as skewed
  - skewed to the left, or negatively skewed
    - a histogram with a long right-hand tail
    - the mean < the median
  - skewed to the right, or positively skewed
    - A histogram with a long left-hand tail
    - the mean > the median

Intersection, Union, Complement

- Intersection : $A \cap B$
- Union : $A \cup B$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Complement $\overline{A}$

$$P(\overline{A}) = 1 - P(A)$$

- $(A \cup C) \cap (B \cup C) = (A \cap B) \cup C$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

- Venn Diagram

Axioms of Probability

- For any event A, $0 \le P(A) \le 1$
- Axiom 3: $P(S) = 1$
- If A and B are *mutually exclusive* then
  $A \cap B = \phi, \qquad P(A \cup B) = P(A) + P(B)$
- If A_1, A_2, A_3, … , A_k is a finite collection of mutually exclusive events,
  $$P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i)$$
- If A_1 and A_2 are *collectively exhaustive*
  $$A \cup B = S$$

Permutations, Combinations

- Permutations
  - Selection of choices that can't be repeated
  - Chosen without replacement
  - Ordered matters

  $$P_{k,n} = n \times (n-1) \times (n-2) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$
- Combinations
  - Similar to permutations, but order does not matter (dealing cards)
    $$C_{k,n} = \binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$
  - $P_{k,n} = k! \times C_{k,n}$

Conditional Probability

- P(A|B) denotes the probability of event A occurring given event B occurs

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Tree diagram

Multiplication Rule

- The probability of A and B = $P(A \cap B)$
  - $P(A \cap B) = P(A)\,P(B|A)$
  - $P(A \cap B) = P(B)\,P(A|B)$
- Independence
  - A and B are independent if P(A|B) = P(A)
  - This is true if and only if P(B|A) = P(B)
  - If A and B are independent, then $P(A \cap B) = P(A)P(B)$

Law of Total Probability

- Let $A_1, \ldots, A_k$ be mutually exclusive and exhaustive events. Then for any other event B,

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$$
$$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

- Or otherwise restated

$$P(B) = \sum_{i=1}^{k} P(A_i \cap B)$$
$$= \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

Bayes' Theorem

$$P(B\,|\,A) = \frac{P(A\,|\,B)P(B)}{P(A)}$$

$$= \frac{P(A\,|\,B_i)P(B_i)}{P(A\,|\,B_1)P(B_1) + P(A\,|\,B_2)P(B_2) + \ldots P(A\,|\,B_N)P(B_N)}$$

Probability Mass Function

- A Probability Mass Function (pmf), also called *probability distribution*, is a function p(x) that assigns to each possible value x that the random variable X can take, its probability

$$p(x) = P(X = x)$$
$$= P(\text{all } s \in S : X(s) = x)$$

  - $p(x_i) \geq 0$ for each possible value x_i of X
  - $\sum_{\text{all } x_i} p(x_i) = 1$

Cumulative Distribution Function

- The cumulative distribution function (cdf) of a r.v. X is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

  which gives the sum of the probabilities up to that value x

- Any probability distribution must follow the axioms of probability
  - $F(-\infty)=0$; $F(\infty) = 1$
  - $F(x) \geq 0$ and is weakly increasing
  - It is continuous in x

Expected Value

- A random variable has probabilities p associated with outcomes x with set of possible values D and pmf p(x)

$$E[X] = \mu_x = \sum_{x \in D} x \cdot p(x)$$

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

- Properties of expected values

$$E(X_1 + X_2) = E[X_1] + E[X_2]$$

$$E(aX + b) = a \times E[X] + b$$

Variance and Standard Deviation

- Variance

$$V(X) = \sigma^2 = E[(X - \mu)^2] = \sum_{D} (x - \mu)^2 \cdot p(x)$$

$$= E[X^2] - (E[X])^2$$

- Standard deviation

$$\sigma = \sqrt{\sigma^2}$$

- Properties

$$\sigma^2(aX + b) = a^2\sigma^2(X)$$

$$\sigma(aX + b) = |a|\,\sigma(X)$$

Bernouilli Distribution

- Bernouilli Probability Distribution

x	0	1
p(x)	1-p	p

$\mu = E[X] = p$

$\sigma^2 = Var[X] = p(1-p)$

Binomial Distributions Theorem

- X~Bin(n,p) $\binom{n}{x} = \frac{n!}{x!(n-x)!}$
- Cumulative Distribution Function

$$\sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$

- Expected values and variance

$$E[X] = np$$
$$Var[X] = np(1-p)$$

Hypergeometric Distribution

- If *X* is the number of *S*'s in a completely random sample of size *n* drawn from a population consisting of *M S*'s and (*N–M)* *F*'s,

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$E(X) = n\frac{M}{N}$$

$$V(X) = \left(\frac{N-n}{N-1}\right) n \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

Negative Binomial Distribution

- The probability that takes X=x failures to get *r* successes, with probability of success *p* is:

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x, \quad x = 0,1,2, \ldots$$

$$E(X) = \frac{r(1-p)}{p}$$

$$V(X) = \frac{r(1-p)}{p^2}$$