

MATH 203 Fall 2024: Midterm Formula Sheet

- **Numerical summaries:** Suppose $x_1, x_2, \dots, x_n$ be n data values.

► The *sample mean*, $\bar{x}$, is defined as:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

► The *sample median*, m , is the number such that, if the data values were ordered in ascending order from smallest to largest then, m would be in the middle:

- * If n is odd, the sample median m is the number in the $(n+1)/2$ th position
- * If n is even, m is the average of the number in the $(n/2)$ th position and the number in the $(n/2 + 1)$ th position

► The *sample variance*, s^2 , is defined as:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right)$$

- **Empirical rule:** For a perfectly symmetric, mound-shaped distribution, approximately

- 68% of the observations will lie in the range $(\bar{x} - s, \bar{x} + s)$,
- 95% of the observations will lie in the range $(\bar{x} - 2s, \bar{x} + 2s)$,
- 99.7% of the observations will lie in the range $(\bar{x} - 3s, \bar{x} + 3s)$.

- **Chebyshev's rule:** For any distribution and any number $k > 1$: at least

$$\left(1 - \frac{1}{k^2} \right) \times 100\%$$

of observations will fall into the interval $(\bar{x} - ks, \bar{x} + ks)$ regardless of the shape of the sample.

- **Quartiles and quantiles:**

- The first sample (*lower*) quartile is the number, q_L , such that 25% of the observations are $\leq q_L$
- The second sample quartile is the median, m , such that 50% of the observations are $\leq m$
- The third sample (*upper*) quartile is the number, q_U , such that 75% of the observations are $\leq q_U$
- For $0 \leq p \leq 1$, the p th *sample quantile* is the number, q_p , such that $(100 \times p)\%$ of the observations are $\leq q_p$.

$$q_L = q_{0.25} \quad m = q_{0.5} \quad q_U = q_{0.75}$$

► Interquartile range (IQR) is $q_U - q_L$.

- **Boxplot construction:**

- *Median line*
- *Box* (lower to upper quartile) or lower and upper hinge
- *Inner fences:* the upper inner fence is $q_U + 1.5 \times \text{IQR}$, the lower inner fence is $q_L - 1.5 \times \text{IQR}$.
- *Outer fences:* the upper outer fence is $q_U + 3 \times \text{IQR}$, the lower outer fence is $q_L - 3 \times \text{IQR}$.
- *Whiskers:* at largest/smallest data points less extreme than inner fences

- **Set theory:** Suppose S is a set, and A and B subsets of S .
 - ▶ the *intersection* of two sets A and B , $A \cap B$, is the collection of elements that are elements of *both* A and B .
 - ▶ the *union* of two sets A and B , $A \cup B$, is the collection of elements that are either in A , or in B , or in both A and B .
 - ▶ the *complement* of set A , A^c , is the collection of elements of S that are not elements of A .
 - ▶ We have

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

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

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

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

- ▶ *Basic partition:* we have the partition of S via the four sets

$$A \cap B \quad A \cap B^c \quad A^c \cap B \quad A^c \cap B^c$$

- ▶ *de Morgan's Laws:*

$$A^c \cap B^c = (A \cup B)^c \quad A^c \cup B^c = (A \cap B)^c$$

- **Counting principles:**

- ▶ *Multiplication principle:* A sequence of k operations, in which operation i can result in n_i possible outcomes, can result in $n_1 \times n_2 \times \cdots \times n_k$ possible sequences of outcomes.
- ▶ *Permutations:* An ordered arrangement of r distinct objects is called a *permutation*. The number of ways of ordering r distinct objects taken from n without replacement is denoted by

$$P_r^n = \frac{n!}{(n-r)!}$$

- ▶ *Combinations:* The number of *combinations* of n objects taken r at a time (without replacement) is the number of subsets of size r that can be formed. This number is denoted

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

The number of binary sequences of length n containing r 1s is C_r^n .

- **Probability:** Suppose S is a sample space for an experiment, A and B are events.

- ▶ *The rules of probability:*

(I) $P(A) \geq 0$.

(II) $P(S) = 1$.

(III) If $A \cap B = \emptyset$, then

$$P(A \cup B) = P(A) + P(B).$$

► *Consequences of the rules:*

- (i) For any A , $P(A^c) = 1 - P(A)$.
- (ii) $P(\emptyset) = 0$.
- (iii) For any A , $P(A) \leq 1$.
- (iv) For two events A and B , if $A \subseteq B$, then

$$P(A) \leq P(B)$$

► *General Addition Rule:*

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

► *Boole's Inequality:* For a collection of events $A_1, A_2, \dots, A_K$.

$$P\left(\bigcup_{k=1}^K A_k\right) \leq \sum_{k=1}^K P(A_k).$$

- **Conditional probability:** For two events A and B , the conditional probability of A given B is denoted $P(A|B)$, and is defined by

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

We consider this where $P(B) > 0$. It follows that $P(A \cap B) = P(A|B)P(B)$.

A and B are *independent* if $P(A|B) = P(A)$, equivalently if $P(A \cap B) = P(A)P(B)$.

- **Chain rule:** we have that

$$P(A_1 \cap A_2 \cap \dots \cap A_K) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_K|A_1 \cap A_2 \cap A_3 \cap \dots \cap A_{K-1}).$$

- **Probability Theorems:** For two events A and B

Theorem of Total Probability

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Bayes Theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

General versions: Suppose events $B_1, B_2, \dots, B_K$ form a partition of sample space S , and let $A \subseteq S$.

Provided all the conditional probabilities are well-defined, we have

► *Theorem of Total Probability:*

$$P(A) = \sum_{k=1}^K P(A|B_k)P(B_k).$$

► *Bayes Theorem:*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{j=1}^K P(A|B_j)P(B_j)}.$$