Nonparametric Statistics Notes

Chapters One and Two: Preliminaries

Jesse Crawford

Department of Mathematics
Tarleton State University

Outline

1. Set Theory Review
2. Brief Overview of Measure Theory
3. Probability Theory
4. The Hypergeometric, Binomial, and Normal Distributions
5. Statistical Inference

Sets and Elements

Definition

- A *set* is a collection of mathematical objects, called the *elements* of the set.
- $x \in A$ means that x is an element of the set A.
- $x \notin A$ means that x is not an element of the set A.

Example

- Let $A = \{1, 2, 3\}$
  - $1 \in A$
  - $2 \in A$
  - $3 \in A$
  - $7 \notin A$
- The order that elements are listed in a set doesn't matter.

$$\{4, 8, 9, 15\} = \{8, 15, 4, 9\}$$

Example

Sets can contain infinitely many elements. If the set exhibits an obvious pattern, it may be sufficient to simply list a few elements of the set.

- Let $A = \{5, 10, 15, \ldots\}$
  - $25 \in A$
  - $28 \notin A$

Sets can also be defined with *set builder notation*.

- Let $B = \{x \mid 4 \leq x \leq 9\}$
  - $6 \in B$
  - $10 \notin B$

Here is common notation used for certain important sets of numbers.

- $\mathbb{N} = \{1, 2, 3, \ldots\}$ (Set of Natural Numbers)
- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ (Set of Integers)
- $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$ (Set of Rational Numbers)
- $\mathbb{R}$ (Set of Real Numbers)
- $\mathbb{C}$ (Set of Complex Numbers)

Interval Notation

- $(a, b) = \{x \in \mathbb{R} | a < x < b\}$
- $[a, b] = \{x \in \mathbb{R} | a \leq x \leq b\}$
- $[a, b) = \{x \in \mathbb{R} | a \leq x < b\}$
- $(a, b] = \{x \in \mathbb{R} | a < x \leq b\}$
- $(a, \infty) = \{x \in \mathbb{R} | a < x\}$
- $[a, \infty) = \{x \in \mathbb{R} | a \leq x\}$
- $(-\infty, b) = \{x \in \mathbb{R} | x < b\}$
- $(-\infty, b] = \{x \in \mathbb{R} | x \leq b\}$
- $(-\infty, \infty) = \mathbb{R}$

Unions

Definition

Given two sets A and B, the *union* of A and B is

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

Note: In mathematical logic, “or” means “and/or”, so “$x \in A$ or $x \in B$” means $x \in A$ or $x \in B$ or both.

Example

- If $A = \{1, 2, 3, 4, 5\}$ and $B = \{3, 4, 5, 6, 7\}$, then

$$A \cup B = \{1, 2, 3, 4, 5, 6, 7\}.$$

- If $A = [0, 8)$ and $B = (6, 10]$, then

$$A \cup B = [0, 10].$$

Intersections

Definition

Given two sets A and B, the *intersection* of A and B is

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

Example

- If $A = \{1, 2, 3, 4, 5\}$ and $B = \{3, 4, 5, 6, 7\}$, then
$$A \cap B = \{3, 4, 5\}.$$
- If $A = [0, 8)$ and $B = (6, 10]$, then
$$A \cap B = (6, 8).$$

Universal Sets and Complements

Definition

- In most situations, there is a *universal set* U that contains all elements under consideration.
- The *complement* of a set A is

$$A' = \{x \in U \mid x \notin A\}.$$

Example

- If $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $A = \{1, 2, 3\}$, then

$$A' = \{4, 5, 6, 7, 8, 9, 10\}.$$

- If $U = \mathbb{R}$ and $A = [5, 8)$, then

$$A' = (-\infty, 5) \cup [8, \infty).$$

Set Difference

Definition

Given two sets A and B,

$$A \setminus B = \{x \in A \mid x \notin B\}.$$

Example

- If $A = \{1, 2, 3, 4, 5\}$ and $B = \{3, 4, 5, 6, 7\}$, then

$$A \setminus B = \{1, 2\}.$$

- If $A = [0, 8)$ and $B = (6, 10]$, then

$$A \setminus B = [0, 6].$$

Note that complements are just a special case of set difference, since

$$A' = \{x \in U \mid x \notin A\} = U \setminus A.$$

Subsets and the Empty Set

Definition

A is a subset of B, written $A \subseteq B$, if every element of A is an element of B.

Example

- If $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4, 5\}$, then $A \subseteq B$.
- If $A = \{1, 2, 3\}$, and $B = \{1, 2, 8, 9\}$, then $A \not\subseteq B$, because $3 \in A$ but $3 \notin B$.

Definition

The *empty set* is the set $\emptyset = \{\}$ that contains no elements.

- For any set A
  - $A \cup \emptyset = A$
  - $A \cap \emptyset = \emptyset$
  - $\emptyset \subseteq A$
  - $A \subseteq A$
  - $A \subseteq U$, where U is the universal set.

Families of Sets

Definition

- A *family* $\mathcal{F}$ of sets is simply a set whose elements are also sets.
- The union of all sets in the family is

$$\bigcup_{A \in \mathcal{F}} A = \{x \mid x \in A, \text{ for some } A \in \mathcal{F}\}.$$

- The intersection of all sets in the family is

$$\bigcap_{A \in \mathcal{F}} A = \{x \mid x \in A, \text{ for all } A \in \mathcal{F}\}.$$

Example

If $\mathcal{F} = \{[-x, x] \mid x > 0\}$, evaluate the following

$$\bigcup_{A \in \mathcal{F}} A \text{ and } \bigcap_{A \in \mathcal{F}} A.$$

Definition

- If A_i is a set, for every $i \in I$, then
- $$\bigcup_{i \in I} A_i = \{x \mid x \in A_i, \text{ for some } i \in I\}.$$
  $$\bigcap_{i \in I} A_i = \{x \mid x \in A_i, \text{ for all } i \in I\}.$$
- The variable i is called the *index*, and I is the *index set*. When the index set is $I = \{1, 2, 3, \ldots\}$, we can write
  $$\bigcup_{i=1}^{\infty} A_i \text{ and } \bigcap_{i=1}^{\infty} A_i.$$

Example

$$\bigcap_{i=1}^{\infty} (0, 5 + \tfrac{1}{i})$$

The Power Set of a Set

Definition

- Let U be a set.
- The *power set* of U, denoted $\mathcal{P}(U)$, is the set of all subsets of U.

$$\mathcal{P}(U) = \{A \mid A \subseteq U\}.$$

- Note that
  - $A \subseteq U$ if and only if $A \in \mathcal{P}(U)$.
  - $\mathcal{F}$ is a family of subsets of U if and only if $\mathcal{F} \subseteq \mathcal{P}(U)$.

Example

If $U = \{1, 2, 3\}$, then

- $\mathcal{P}(U) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, U\}$.
- $\{1, 3\} \subseteq U$, so $\{1, 3\} \in \mathcal{P}(U)$.
- If $\mathcal{F} = \{\{x\} \mid x \in U\} = \{\{1\}, \{2\}, \{3\}\}$, then $\mathcal{F} \subseteq \mathcal{P}(U)$.

Cartesian Products and Functions

Definition

Given two sets U and V, the *cartesian product* of U and V is the following set of ordered pairs:

$$U \times V = \{(u, v) \mid u \in U \text{ and } v \in V\}.$$

Definition

A *function* f from U to V is a subset of $U \times V$, such that for any $u \in U$, there exists a unique $v \in V$, such that $(u, v) \in f$.

- Functions are often called maps or mappings.
- U and V are the *domain* and *codomain* of f, respectively.
- To indicate that f is a function from U to V, we write $f : U \to V$.
- If $(u, v) \in f$, we write $f(u) = v$.

Images and Inverse Images

Definition

Let $f : U \to V$, and let $A \subseteq U$ and $B \subseteq V$.

- The *image* of A under f is
$$f(A) = \{f(x) \mid x \in A\}.$$
- The *inverse image* of B under f is
$$f^{-1}(B) = \{x \in U \mid f(x) \in B\}.$$
- The *image* or *range* of f is $f(U) = \{f(x) \mid x \in U\}$.

Example

Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then

- $f([0,5]) = [0,25]$
- $f^{-1}([0,25]) = [-5,5]$
- The range of f is $[0,\infty)$.

Injective and Surjective Functions

Definition

Let $f : U \to V$ then

- f is *injective* (also called one-to-one) if for all $u_1, u_2 \in U$,

$$f(u_1) = f(u_2) \text{ implies } u_1 = u_2.$$

- f is *surjective* (also called onto) if $f(U) = V$, that is,

$$\text{for any } v \in V, \text{ there exists } u \in U, \text{ such that } f(u) = v.$$

- A function that is both injective and surjective is called bijective.

Example

Give examples of injective and surjective functions.

Countable and Uncountable Sets

Definition

- A set A is *countably infinite* if there is a bijection between A and the set of natural numbers $\mathbb{N}$.
- A set A is *countable* if it is finite or countably infinite.
- A set is *uncountable* if it is not countable.

The following are equivalent:

- A is countable
- There exists an injective function $f : A \to \mathbb{N}$
- There exists a surjective function $f : \mathbb{N} \to A$.

Example

Which of the following sets are countable?
$\{0, 1, 2, \ldots, 10\}, \mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$

Outline

1. Set Theory Review
2. Brief Overview of Measure Theory
3. Probability Theory
4. The Hypergeometric, Binomial, and Normal Distributions
5. Statistical Inference

σ-algebras

Definition

- Suppose Ω is a set and $\mathcal{F}$ is a family of subsets of Ω.
- Then $\mathcal{F}$ is called a σ-algebra (or σ-field) if the following conditions hold:
  - $\Omega \in \mathcal{F}$
  - $A \in \mathcal{F}$ implies $A' \in \mathcal{F}$
  - $A_1, A_2, \ldots \in \mathcal{F}$ implies $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.
- Elements of $\mathcal{F}$ are called *measurable sets*.
- $(\Omega, \mathcal{F})$ is called a *measurable space*.

Example

Show that a σ-algebra always has these two properties also:

- $\emptyset \in \mathcal{F}$
- $A_1, A_2, \ldots \in \mathcal{F}$ implies $A_1 \cap A_2 \cap \cdots \in \mathcal{F}$.

σ-algebras

Definition

- Suppose Ω is a set and $\mathcal{F}$ is a family of subsets of Ω.
- Then $\mathcal{F}$ is called a σ-algebra (or σ-field) if the following conditions hold:
  - $\Omega \in \mathcal{F}$
  - $A \in \mathcal{F}$ implies $A' \in \mathcal{F}$
  - $A_1, A_2, \ldots \in \mathcal{F}$ implies $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.
- Elements of $\mathcal{F}$ are called *measurable sets*.
- $(\Omega, \mathcal{F})$ is called a *measurable space*.

Example

- Let Ω be any set.
- Prove that $\mathcal{P}(\Omega)$, the power set of Ω, is a σ-algebra.

σ-algebras

Definition

- Suppose Ω is a set and $\mathcal{F}$ is a family of subsets of Ω.
- Then $\mathcal{F}$ is called a σ-algebra (or σ-field) if the following conditions hold:
  - $\Omega \in \mathcal{F}$
  - $A \in \mathcal{F}$ implies $A' \in \mathcal{F}$
  - $A_1, A_2, \ldots \in \mathcal{F}$ implies $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.
- Elements of $\mathcal{F}$ are called *measurable sets*.
- $(\Omega, \mathcal{F})$ is called a *measurable space*.

Example

- Let $\Omega = \mathbb{R}$. There exits a σ-algebra $\mathcal{B}$, called the Borel σ-algebra, such that $(a, b) \in \mathcal{B}$, for all $a, b \in \mathbb{R}$.
- Prove that $[a, b] \in \mathcal{B}$, $\{a\} \in \mathcal{B}$, and $(a, \infty) \in \mathcal{B}$, for all $a, b \in \mathbb{R}$.

Pairwise Disjoint Sets

Definition

If $A \cap B = \emptyset$, then A and B are *disjoint*.

Definition

- A family of sets A_i, $i \in I$ is called *pairwise disjoint* if

$$A_i \cap A_j = \emptyset, \text{ if } i \neq j.$$

- Also called *mutually exclusive*.

Measures

Definition

- Suppose $(\Omega, \mathcal{F})$ is a measurable space.
- A function $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure* on $(\Omega, \mathcal{F})$ if
  - $\mu(\emptyset) = 0$
  - For any sequence of pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{F}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

- If μ is a measure on $(\Omega, \mathcal{F})$, then $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*.

Example

- Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $A, B \in \mathcal{F}$.
- If $A \subseteq B$, prove $\mu(A) \leq \mu(B)$.
- If $A \subseteq B$, and $\mu(B) < \infty$, prove that $\mu(B \setminus A) = \mu(B) - \mu(A)$.

- A function $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure* on $(\Omega, \mathcal{F})$ if
  - $\mu(\emptyset) = 0$
  - For any sequence of pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{F}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Example

- Let $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. There exists a measure μ on this space called *Lebesgue measure*, such that

$$\mu((a, b)) = b - a, \text{ for all real numbers } a < b.$$

- For all real numbers $a < b$, prove the following:
  - $\mu(\{a\}) = 0$
  - $\mu([a, b]) = b - a$
  - $\mu((0, \infty)) = \infty$
  - $\mu(\mathbb{R}) = \infty$.

Measurable Functions

Definition

- Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces.
- A function $f : \Omega_1 \to \Omega_2$ is *measurable* if

$$f^{-1}(A) \in \mathcal{F}_1, \text{ for any } A \in \mathcal{F}_2.$$

Example

- Let $(\Omega_1, \mathcal{F}_1) = (\mathbb{R}, \mathcal{B})$, and let $(\Omega_2, \mathcal{F}_2) = (\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$.
- Define $f : \mathbb{R} \to \mathbb{Z}$ by $f(x) = \lfloor x \rfloor$, the greatest integer less than or equal to x.
- Prove that f is measurable.

Outline

1. Set Theory Review
2. Brief Overview of Measure Theory
3. Probability Theory
4. The Hypergeometric, Binomial, and Normal Distributions
5. Statistical Inference

Probability Measures

Definition

- A *probability space* is a measure space $(S, \mathcal{F}, P)$, such that $P(S) = 1$.
- S is the *sample space* or *observation space*.
- P is called a *probability measure*.
- A set $A \in \mathcal{F}$ is called an *event*.
- $P(A)$ is the probability of the event A.

- $P(\emptyset) = 0$
- $P(S) = 1$
- For all $A, B \in \mathcal{F}$
    - If $A \subseteq B$, then $P(A) \leq P(B)$.
    - $0 \leq P(A) \leq 1$
    - $P(A') = 1 - P(A)$
    - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probability Measures

Definition

- A *probability space* is a measure space $(S, \mathcal{F}, P)$, such that $P(S) = 1$.
- S is the *sample space* or *observation space*.
- P is called a *probability measure*.
- A set $A \in \mathcal{F}$ is called an *event*.
- $P(A)$ is the probability of the event A.

Example

- $S = \{1, 2, 3, \ldots\}$
- $\mathcal{F} = \mathcal{P}(S)$
- Assume $P(\{s\}) = 2^{-s}$, for all $s \in S$
- Find these probabilities:
  - $P(\{1, 2\})$
  - $P(\{3, 4, 5, \ldots\})$

Discrete Probability Distributions

Definition

- Suppose $(S, \mathcal{P}(S), P)$ is a probability space.
- If S is a countable set, then P is a *discrete* probability measure.
- P is completely determined by the function

$$\begin{aligned} f : S &\to [0,1] \\ f(s) &= P(\{s\}). \end{aligned}$$

- For any $A \subseteq S$,

$$P(A) = \sum_{s \in A} f(s).$$

- f is called the *probability mass function* (p.m.f.) or *probability distribution function* (p.d.f.) of P.
- Note that
  - $0 \leq f(s) \leq 1$, for all $s \in S$
  - $\sum_{s \in S} f(s) = 1$.

Random Variables

Definition

- Let $(S, \mathcal{F}, P)$ be a probability space.
- A *random variable* is a measurable function

$$X : S \to \mathbb{R}.$$

- ($\mathbb{R}$ is equipped with the Borel σ-algebra $\mathcal{B}$.)

Example

$$X : S \to \mathbb{R}$$

$$X(s) = \begin{cases} 0 & \text{if } s = TTT \\ 1 & \text{if } s \in \{HTT, THT, TTH\} \\ 2 & \text{if } s \in \{HHT, HTH, THH\} \\ 3 & \text{if } s = HHH \end{cases}$$

The Distribution of a Random Variable

Notation

If $X : S \to \mathbb{R}$ is a random variable,

$$[X \in A] = \{s \in S \mid X(s) \in A\}$$

Definition

- Suppose $(S, \mathcal{F}, P)$ is a probability space.
- Let $X : S \to \mathbb{R}$ be a random variable.
- Then there is a corresponding probability space $(\mathbb{R}, \mathcal{B}, P_X)$, where

$$P_X(A) = P[X \in A] = P[X^{-1}(A)], \text{ for all } A \in \mathcal{B}.$$

- P_X is the *probability distribution* of X.
- It's often just called the *distribution* of X.

Support of a Random Variable

Definition

If $X : S \to \mathbb{R}$ is a random variable, the *support* of X is the set of all possible values of X:

$$\text{supp}(X) = \{X(s) \mid s \in S\} = X(S).$$

Definition

- A random variable X is *discrete* if $\text{supp}(X)$ is countable.
- If X is discrete, the distribution of X is completely determined by the function

$$f : \mathbb{R} \to [0, 1]$$
$$f(x) = P_X(\{x\}) = P[X = x].$$

- f is called the *probability mass function* (p.m.f.) or *probability distribution function* (p.d.f.) of X.

Discrete Random Variables

Definition

- A random variable X is *discrete* if $\text{supp}(X)$ is countable.
- If X is discrete, the distribution of X is completely determined by the function

$$f : \mathbb{R} \to [0, 1]$$
$$f(x) = P_X(\{x\}) = P[X = x].$$

- f is called the *probability mass function* (p.m.f.) or *probability distribution function* (p.d.f.) of X.
- Note that
  - $0 \leq f(x) \leq 1$, for all $x \in \mathbb{R}$
  - $\sum_{x \in \mathbb{R}} f(x) = 1$.

Expected Value and Variance of a Random Variable

Definition

- Let X be a discrete random variable with p.m.f. f.
- The *expected value* of X is

$$\mu = \mu_X = E(X) = \sum_{x \in \mathbb{R}} x f(x).$$

- Given any function $u : \mathbb{R} \to \mathbb{R}$,

$$E[u(X)] = \sum_{x \in \mathbb{R}} u(x) f(x)$$

- The *variance* of X is

$$\sigma^2 = \sigma_X^2 = \text{Var}(X) = E[(X - \mu_X)^2] = E(X^2) - E(X)^2.$$

- The *standard deviation* of X is $\sigma = \sigma_X = \sqrt{\text{Var}(X)}$.

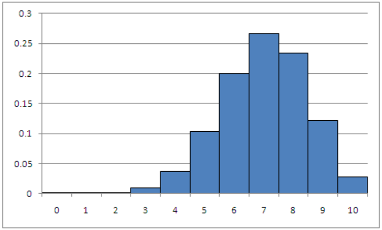


Figure: Binomial distribution with $n = 10$ and $p = 0.7$.

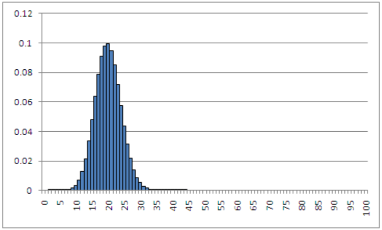


Figure: Binomial distribution with $n = 100$ and $p = 0.2$.

Moment Generating Function of a Random Variable

Definition

- Let X be a random variable.
- Assume there exists $h > 0$, such that

$$M(t) = E[e^{tX}] \text{ converges, for } -h < t < h.$$

- Then M is called the *moment-generating function* (m.g.f.) of X.
- If the above expected value does not exists on some interval $(-h, h)$, then the m.g.f. does not exist.

If X is a random variable, and its m.g.f. exists, then

$$E(X^r) = M^{(r)}(0), \text{ for any } r = 1, 2, \ldots$$

Cumulative Distribution Function of a Random Variable

Definition

- The *cumulative distribution function* (c.d.f.) of the random variable X is

$$F : \mathbb{R} \to [0, 1]$$
$$F(x) = P[X \leq x].$$

- Also called the distribution function.

- Let X be a discrete random variable.
- The *distribution* or *probability distribution* of X is the probability measure

$$P_X : \mathcal{B} \to [0,1]$$
$$P_X(A) = P[X \in A]$$

- The *probability mass function* (p.m.f.) or *probability distribution function* (p.d.f.) of X is

$$f : \mathbb{R} \to [0,1]$$
$$f(x) = P[X = x]$$

- The *cumulative distribution function* (c.d.f.) or *distribution function* (d.f.) of the random variable X is

$$F : \mathbb{R} \to [0,1]$$
$$F(x) = P[X \leq x].$$

Identically Distributed Random Variables

Definition

- Let X and Y be two random variables.
- Then X and Y have the same distribution if

$$P_X = P_Y$$

- We say that X and Y are *identically distributed*.

Given two discrete random variables X and Y, the following are equivalent:

- X and Y are identically distributed, $P_X = P_Y$
- X and Y have the same p.m.f., $f_X = f_Y$
- X and Y have the same c.d.f., $F_X = F_Y$
- X and Y have the same m.g.f., $M_X = M_Y$

Continuous Random Variables

Definition

- Let X be a random variable, and suppose there is a function

$$f : \mathbb{R} \to [0, \infty)$$

  such that, for any real numbers $a < b$,

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

- Then the distribution of X is *continuous*, and f is the *probability density function* (p.d.f.).
- For any $A \in \mathcal{B}$,

$$P(X \in A) = \int_A f(x)\, dx.$$

  - $0 \leq f(x)$, for all $x \in \mathbb{R}$
  - $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

Discrete vs. Continuous Random Variables

- Discrete
- f is the probability mass function (p.m.f.)
- $\sum_{x \in \mathbb{R}} f(x) = 1$
- $P(X \in A) = \sum_{x \in A} f(x)$
- $E[u(X)] = \sum_{x \in \mathbb{R}} u(x) f(x)$

- Continuous
- f is the probability density function (p.d.f.)
- $\int_{\mathbb{R}} f(x)\, dx = 1$
- $P(X \in A) = \int_A f(x)\, dx$
- $E[u(X)] = \int_{\mathbb{R}} u(x) f(x)\, dx$

Given two continuous random variables X and Y, the following are equivalent:

- X and Y are identically distributed, $P_X = P_Y$
- The p.d.f.'s of X and Y are equal *almost everywhere*,

$$\mu(\{x \mid f_X(x) \neq f_Y(x)\}) = 0.$$

- X and Y have the same c.d.f., $F_X = F_Y$
- X and Y have the same m.g.f., $M_X = M_Y$

Outline

1. Set Theory Review
2. Brief Overview of Measure Theory
3. Probability Theory
4. The Hypergeometric, Binomial, and Normal Distributions
5. Statistical Inference

Combinations

- Consider a set of n objects.
- The number of different subsets of size r is

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

Example

How many five card poker hands are there?

Example

How many distinct permutations are there of the letters

$$AAAAAAFFF$$

The Hypergeometric Distribution

Example

- A car dealership has 20 cars: 12 Fords and 8 Chevrolets.
- 5 cars are selected at random without replacement.
- What's the probability of selecting exactly 3 Fords?

Hypergeometric Distribution

- N_1 = number of objects of type 1
- N_2 = number of objects of type 2
- Random sample without replacement.
- n = sample size.
- X = number of objects in sample of type 1

$$P(X = x) = f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N_1+N_2}{n}}, \text{ for } x \leq n, x \leq N_1, n - x \leq N_2.$$

Independent Events

Definition

Two events A and B are *statistically independent* if

$$P(A \cap B) = P(A)P(B).$$

Example

- $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
- All outcomes are equilikely.
- $A = \{HHH, HHT, HTH, HTT\}$ (First coin is heads)
- $B = \{HHH, HHT, THH, THT\}$ (Second coin is heads)

Independence of Multiple Events

Definition

A family of events $\{A_i \mid i \in I\}$, is statistically independent if

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}),$$

for any subfamily of events $\{A_{i_1}, \ldots, A_{i_k}\}$.

Independent Random Variables

Definition

Two random variables X and Y are *statistically independent* if

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B), \text{ for all } A, B \in \mathcal{B}$$

The Binomial Distribution

Example

- A football player kicks 10 field goals.
- Chance of making each field goal is 80%.
- The field goals are statistically independent.
- Find the probability of making exactly 7 of the field goals.

Binomial Distribution

- Sequence of n trials
- Each trial has only two possible outcomes, “success” and “failure”
- $p =$ probability of “success” on a single trial
- Trials are statistically independent
- $X =$ number of “successes” that actually occur

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

Example

Find the probability of making between 6 and 9 field goals inclusive?

The Normal Distribution

Definition

- Suppose $\mu \in \mathbb{R}$, and $\sigma^2 > 0$.
- The *normal distribution* with mean μ and variance σ^2, is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \ -\infty < x < \infty.$$

Notation: $N(\mu, \sigma^2)$

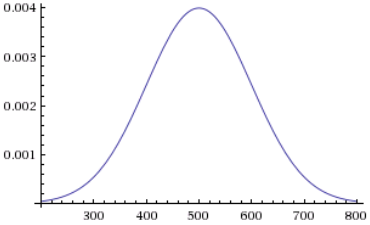


Figure: Normal distribution with $\mu = 500$ and $\sigma = 100$.

Standard Normal Distribution

Definition

The distribution $N(\mu = 0, \sigma^2 = 1)$ is the *standard normal distribution*.

Standardizing a Normal Random Variable

If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

The Central Limit Theorem

Theorem (5.6-1)

- *Suppose $X_1, X_2, \ldots$ is a sequence of IID random variables,*
- *from a distribution with finite mean μ*
- *and finite positive variance σ^2.*
- *Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, for $n = 1, 2, \ldots$*
- *Then, as $n \to \infty$,*

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \Rightarrow N(0, 1).$$

Informal Statement of CLT

Informal CLT

- Suppose $X_1, \ldots, X_n$ is a random sample
- from a distribution with finite mean μ
- and finite positive variance σ^2.
- Then, if n is sufficiently large,

$$\overline{X} \approx N(\mu, \sigma^2/n), \text{ and}$$

$$\sum_{i=1}^{n} X_i \approx N(n\mu, n\sigma^2).$$

- Conventionally, values of $n \geq 30$ are usually considered sufficiently large, although this text applies the approximation for lower values of n, such as $n \geq 20$.

Normal Approximation to the Binomial Distribution

If $np \geq 5$ and $n(1-p) \geq 5$, then

$$b(n,p) \approx N(\mu = np, \sigma^2 = np(1-p)).$$

Outline

1 Set Theory Review

2 Brief Overview of Measure Theory

3 Probability Theory

4 The Hypergeometric, Binomial, and Normal Distributions

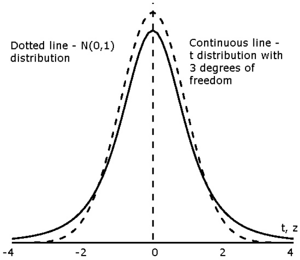
5 Statistical Inference

Definition

- Let r be a positive integer.
- The t-distribution with r degrees of freedom is given by

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, \quad -\infty < t < \infty.$$

- As $r \to \infty$, $t(r) \Rightarrow N(0,1)$.



Test Statistics for the Normal Distribution

Proposition

- Consider a random sample $X_1, \ldots, X_n$ from a $N(\mu, \sigma^2)$ population.
- $X_1, \ldots, X_n$ are IID, and $X_i \sim N(\mu, \sigma^2)$, for all i.
- $$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$
- $$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Critical Values for Normal and t-distributions

Definition

- Let $Z \sim N(0,1)$ and $T \sim t(r)$.
- Let $\alpha \in (0,1)$.
- We define z_α and $t_\alpha(r)$ as follows:

$$P[Z > z_\alpha] = \alpha.$$

$$P[T > t_\alpha(r)] = \alpha.$$

α	$z_{\alpha/2}$	$t_{\alpha/2}(30)$
0.10	1.645	1.697
0.05	1.96	2.042
0.01	2.575	2.750

A Hypothesis Testing Example

Example

- Assume that packages of M&M's have a $N(\mu, \sigma^2)$ distribution.
- A sample of 31 packages of M&M's had sample mean $\overline{X} = 235.1$ grams and sample standard deviation $S = 5.7$ grams.
- Perform the following hypothesis test at the $\alpha = 0.05$ significance level

$$H_0 : \mu = 232.5 \text{ vs. } H_1 : \mu \neq 232.5.$$

Hypothesis Test: Mean of a Normal Distribution

- Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be a random sample.
- For the testing problem

$$\mathrm{H}_0 : \mu = \mu_0 \text{ vs. } \mathrm{H}_1 : \mu \neq \mu_0,$$

- the **test statistic** is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$

- The **null distribution** of T is

$$T \sim t(n-1).$$

- The **decision rule** is

$$\begin{aligned} &\text{Reject } \mathrm{H}_0\text{, if } |T| \geq t_{\alpha/2}(n-1) \\ &\text{Do not reject } \mathrm{H}_0\text{, otherwise.} \end{aligned}$$

- The **critical region** is $C = \{t \in \mathbb{R} \mid |t| \geq t_{\alpha/2}(n-1)\}$.

Hypothesis Testing Terminology

Definition

- A **statistic** is any variable computed based on a sample of data.
- A **test statistic** is a statistic used to perform a hypothesis test.
- The **null distribution** of a test statistic is its distribution under the assumption that H_0 is true.
- The **critical region** or **rejection region** of a test is the set of all values of the test statistic that result in rejecting H_0.

Two Types of Errors

- Type I error: Rejecting H_0 when it is true.
- Type II error: Not rejecting H_0 when it is false.

Given the Null Hypothesis Is

Your Decision Based On a Random Sample		True	False
	Reject	Type I Error	Correct Decision
	Do Not Reject	Correct Decision	Type II Error

Two Types of Errors in Decision Making

- The **significance level** of a test is

$$\alpha = \max\{P(\text{type I error}) \mid H_0 \text{ is true}\}.$$

Given the Null Hypothesis Is

Your Decision Based On a Random Sample		True	False
	Reject	Type I Error	Correct Decision
	Do Not Reject	Correct Decision	Type II Error

Two Types of Errors in Decision Making

- The **significance level** of a test is

$$
\begin{aligned}
\alpha &= \max\{P(\text{type I error}) \mid \text{H}_0 \text{ is true}\} \\
&= \max\{P(\text{rejecting H}_0) \mid \text{H}_0 \text{ is true}\} \\
&= \max\{P(T \in C) \mid \text{H}_0 \text{ is true}\}
\end{aligned}
$$

p-values

Definition

- The *p*-**value** is the smallest significance level at which the null hypothesis would be rejected for a given observation.
- Also called the **observed significance level.**
- It is the probability of all values more extreme than T under the null distribution.
- The smaller the *p*-value is, the stronger the evidence is **against** H_0.

Confidence Intervals

Definition

A $1-\alpha$ **confidence interval** for a parameter θ is a random interval $[L, U]$, such that

$$P(L \le \theta \le U) = 1-\alpha.$$

- Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be a random sample.
- A $1-\alpha$ confidence interval for μ is

$$\left[\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

- Also written as

$$\overline{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$$

Hypothesis Testing Conclusions

- Rejecting H_0 means there is strong evidence that H_0 is false.
- Not rejecting H_0 merely means there is a lack of strong evidence against H_0.
  There is not strong evidence in favor of anything, including H_0.