ST2334 Finals Cheatsheet

by: Jiahao

definitions

sample space (S): set of all possible outcomes

- aka *sure event*

sample point: outcome in sample space, $p \in S$

event: subset of sample space, $E \subseteq S$

- no elements: *null event*, $\emptyset$

event operations & relationships

* Applicable to n events

union: $A \cup B = \{x : x \in A \vee x \in B\}$

intersection: $A \cap B = \{x : x \in A \wedge x \in B\}$

complement: $A' = \{x : x \in S \wedge x \notin A\}$

mutually exclusive: $A \cap B = \emptyset$

contained: $A \subset B$

equivalent: $A \subset B \wedge B \subset A \Rightarrow A = B$

others:

- $A \cap A' = \emptyset$
- $A \cap \emptyset = \emptyset$
- $A \cup A' = S$
- $(A')' = A$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup B = A \cup (B \cap A')$
- $A = (A \cap B) \cup (A \cap B')$
- $(A_1 \cup A_2 \cup ... \cup A_n)' = A_1' \cap A_2' \cap ... \cap A_n'$
- $(A_1 \cap A_2 \cap ... \cap A_n)' = A_1' \cup A_2' \cup ... \cup A_n'$

counting methods

multiplication principle: r different experiments performed sequentially, producing $n_1 \times n_2 \times ... \times n_r$ outcomes

addition principle: r different procedures performed sequentially, producing $n_1 + n_2 + ... + n_r$ ways (non-overlapping) to perform an experiment

permutation: selection and arrangement of r objects out of n where order matters (i.e. $\{a, b\} \neq \{b, a\}$)

$$P_r^n = nPr = \frac{n!}{(n-r)!} = n(n-1)(n-2)...(n-(r-1))$$

combination: selection of r objects out of n where order does not matter

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

probability

$P(\cdot)$ is a function on the collection fo events of the sample space S satisfying:

- axiom 1. For any event A, $0 \leq P(A) \leq 1$
- axiom 2. For the sample space, $P(S) = 1$
- axiom 3. For any two mutually exclusive events A and B, $A \cap B = \emptyset$ and $P(A \cup B) = P(A) + P(B)$

properties:

- $P(\emptyset) = 0$
- if $A_1, A_2, ..., A_n$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$
- $P(A') = 1 - P(A)$
- $P(A) = P(A \cap B) + P(A \cap B')$
- **inclusion-exclusion principle:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $A \subset B \Rightarrow P(A) \leq P(B)$

finite sample space with equally likely outcomes: $S = \{a_1, a_2, ..., a_k\}$ and all outcomes are equally likely, so any event occurring is where $A \subset S$

$$P(A) = \frac{|A|}{|S|}$$

probability of repeated event: if the outcome is always the same, then $P(K) = P(A)^n$

conditional probability

for any two events A and B with $P(A) > 0$, the conditional probability of B given that A has occurred is

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

since A has occurred, A becomes the reduced sample space

multiplication rule: $P(A \cap B) = P(A)P(B \mid A)$ if $P(A) \neq 0$

inverse probability formula: $P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$

independence

events are independent ($A \perp B$) iff $P(A \cap B) = P(A)P(B)$

$$P(A) \neq 0 \Rightarrow P(A \mid B) = P(A)$$
$$P(B) \neq 0 \Rightarrow P(B \mid A) = P(B)$$

intuition: A and B if the knowledge of A does not change the probability of B

independence vs mutually exclusive:

- $P(A) > 0 \wedge P(B) > 0, A \perp B \Rightarrow$ not mutually exclusive
- $P(A) > 0 \wedge P(B) >$ $0, A, B$ not mutually exclusive $\Rightarrow A \not\perp B$
- S and $\emptyset$ are independent of any other event
- $A \perp B \Rightarrow A \perp B', A' \perp B, A' \perp B'$

total probability

partition: if $A_1, A_2, ..., A_n$ are mutually exclusive events and $\bigcup_{i=1}^n A_i = S$, then $A_1, A_2, ..., A_n$ is a partition of S (i.e. how to split the sample space up into parts)

law of total probability: given a partition of S, for any even B,

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B \mid A_i)$$

applied to single event A and B:

$$P(B) = P(A)P(B \mid A) + P(A')P(B \mid A')$$

bayes' theorem

give a partition of S, then for any event B and $k = 1, 2, ..., n$,

$$P(A_k \mid B) = \frac{P(A_k)P(B \mid A_k)}{\sum_{i=1}^n P(A_i)P(B \mid A_k)}$$

Applied to single event A and B,

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

random variables

a function X which assigns a real number to every $s \in S$ (mapping of values from sample space to some value representing a property of that value in the sample space)

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

range space: set of real numbers produced by random variable X

$$R_X = \{x \mid x = X(s), s \in S\}$$

notations:

- $\{X = x\} = \{s \in S : X(s) = x\} \subset S$
- $\{X \in A\} = \{s \in S : X(s) \in A\} \subset S$
- $P(X = x) = P(\{s \in S : X(s) = x\})$
- $P(X \in A) = P(\{s \in S : X(s) \in A\})$

describing random variables: (1) range of inputs to outputs, (2) constructing a table/formula

probability distribution

probability distribution: $(x_i, f(x_i))$ where $f(x)$ is the probability function

discrete random variables

number of values in R_X is finite or countable

probability mass function: for a discrete random variable X, the probability (mass) function is:

$$f(x) = \{P(X = x), x \in R_X \\ 0, x \notin R_X$$

properties: $f(x)$ must satisfy the following

- $f(x_i) \geq 0 \forall x_i \in R_X$ (all fractional and ≤ 1)
- $f(x) = 0 \forall x \notin R_X$
- $\sum_{i=1}^\infty f(x_i) = \sum_{x_i \in R_X} f(x_i) = 1$

extension: for any set $B \subset \mathbb{R}$,

$$P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i)$$

continuous random variables

R_X is an interval or collection of intervals

probability density function: quantifies probability that X is in a certain range

properties: $f(x)$ must satisfy the following

- $f(x) \geq 0 \forall x \in R_X$
- $f(x) = 0 \forall x \notin R_X$
- $\int_{-\infty}^\infty f(x)dx = \int_{R_X} f(x)dx = 1$

extension 1: given that $a \leq b$,

$$\begin{aligned} &P(a \leq X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X \leq b) \\ &= P(a < X < b) \\ &= \int_a^b f(x)dx \end{aligned}$$

extension 2:

$$P(X = x) = 0$$

cumulative distribution function

probability distribution over a range (both discrete and continuous)

$$F(x) = P(X \leq x)$$

properties:

- $F(x)$ is always non-decreasing
- ranges of $F(x)$ and $f(x)$ satisfy
  - $0 \leq F(x) \leq 1$
  - for discrete distributions, $0 \leq f(x) \leq 1$
  - for continuous distributions, $f(x) \geq 0$ but *not necessary* that $f(x) \leq 1$

discrete random variables

$$F(x) = \sum_{t \in R_X; t \leq x} f(t) = \sum_{t \in R_X; t \leq x} P(X = t)$$

CDF is a step function and can be represented as such (note that probability is cumulated to reach 1):

$$\begin{aligned} F(x) = \{&0, x < 0 \\ &1/4, 0 \leq x < 1 \\ &3/4, 1 \leq x < 2 \\ &1, 2 \leq x \end{aligned}$$

for any two numbers $a < b$,

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a-)$$

$F(a-)$ is the largest value in R_X that is smaller than a

continuous random variable

$$F(x) = \int_{-\infty}^x f(t)dt$$

$$f(x) = \frac{dF(x)}{dx}$$

for any two numbers $a < b$,

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$$

note that if there are multiple functions per interval and a, b run across multiple intervals, separately integrate each interval with the functions for each interval

expectation

expectation, also known as mean, of random variable is the average value of X after repeating the experiment many times. This value may not be a possible value of X.

discrete random variable:

$$\mu_X = E(X) = \sum_{x_i \in R_X} x_i f(x_i)$$

continuous random variable:

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

properties:

- $E(aX + b) = aE(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- let $g(\cdot)$ be an arbitrary function,

$$E(g(X)) = \sum_{x \in R_X} g(x)f(x)$$

or

$$E(g(X)) = \int_{R_X} g(x)f(x)dx$$

variance

calculates the deviation of X from its mean (expectation)

$$\sigma_X^2 = V(X) = E(X - \mu_X)^2 = E(X^2) - E(X)^2$$

applicable regardless of discrete/continuous random variable.

$$V(X) = \sum_{x \in R_X} (x - \mu_X)^2 f(x)$$

or

$$V(X) = \int_{-\infty}^\infty (x - \mu_X)^2 f(x)dx$$

properties:

- $V(X) \geq 0$ if $P(X = E(X)) = 1$ where X is a constant
- $V(aX + b) = a^2 V(X)$
- standard deviation of X: $\sigma_X = \sqrt{V(X)}$

joint distributions

(X, Y) is a two-dimensional random vector/random variable

range space: $R_{X,Y} = \{(x, y) \mid x = X(s), y = Y(s), s \in S\}$ (effectively looking at all pairs of (x, y); generalizable to n dimensions)

discrete two-dimensional random variable: number of possible values of $(X(s), Y(s))$ is finite or countable (both X and Y are discrete)

continuous two-dimensional random variable: number of possible values of $(X(s), Y(s))$ can assume any value in some region of the Euclidean space $\mathbb{R}^2$ (both X and Y are continuous)

joint probability function

discrete joint probability function

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

properties:

- $f_{X,Y}(x, y) \geq 0 \forall (x, y) \in R_{X,Y}$
- $f_{X,Y}(x, y) = 0 \forall (x, y) \notin R_{X,Y}$
- $\sum_{i=1}^\infty \sum_{j=1}^\infty f_{X,Y}(x_i, y_j) = \sum_{i=1}^\infty \sum_{j=1}^\infty P(X = x_i, Y = y_j) = 1$
- above is the same as $\sum\sum_{(x,y) \in R_{X,Y}} f(x, y) = 1$

let $A \subset R_{X,Y}$,

$$P((X, Y) \in A) = \sum\sum_{(x,y) \in A} f_{X,Y}(x, y)$$

continuous joint probability function

$$\begin{aligned} &P((X, Y) \in D) \\ &= P(a \leq X \leq b, c \leq Y \leq d) \\ &= \int_a^b \int_c^d f_{X,Y}(x, y)dydx \end{aligned}$$

*the order of integration does not matter

properties:

- $f_{X,Y}(x, y) \geq 0 \forall (x, y) \in R_{X,Y}$
- $f_{X,Y}(x, y) = 0 \forall (x, y) \notin R_{X,Y}$
- $\int\int_{(x,y) \in R_{X,Y}} f_{X,Y}(x, y)dxdy = 1$

*focus of this module is ranges where x and y do not depend on each other (not the same as independence)

marginal probability distribution

isolating X or Y from a joint probability distribution (projection of joint distribution to univariate distribution). To find X, use Y, and vice versa

$$P(X = x) = f_X(x) = \sum_y f_{X,Y}(x, y)$$

or

$$f_X(x) = \int_{-\infty}^\infty f_{X,Y}(x, y)dy$$

marginal probability distributions are probability functions.

conditional distribution

distribution of Y given that the random variable X is observed to take the value x

conditional probability function of Y given that $X = x$:

$$f_{Y \mid X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

when given values, it finds $P(Y \mid X = x)$

only defined for x such that $f_X(x) > 0$ (same for y)

$f_{Y \mid X}(y \mid x)$ **is not a probability function of** x**:** the requirements of probability functions do not need to hold

applications: you can also use summation for discrete joint random variables

$$P(Y \leq y \mid X = x) = \int_{-\infty}^y f_{Y \mid X}(y \mid x)dy$$

$$E(Y \mid X = x) = \int_{-\infty}^\infty y f_{Y \mid X}(y \mid x)dy$$

reading discrete joint probability tables

x/y	y_1	y_2	y_3	$f_X(x)$
x_1	a	b	c	a + b + c
x_2	d	e	f	d + e + f
x_3	g	h	i	g + h + i
$f_Y(y)$	a + d + g	b + e + h	c + f + i	1

$E(Y \mid X = x)$ (same steps for $E(X \mid Y = y)$) (using $x = x_1$):

1. sum of probability given $X = x_1 = f_X(x_1)$, $a + b + c = K$
2. divide each value in $X = x_1$ by K, a/K, b/K, c/K
3. multiply each by the corresponding Y value, $\frac{ay_1}{K}$, $\frac{by_2}{K}$, $\frac{cy_3}{K}$
4. sum the values: $E(Y \mid X = x_1) = \frac{ay_1 + by_2 + cy_3}{K}$

$E(X)$ (same steps for $E(Y)$): $x_1 \cdot (a + b + c) + x_2 \cdot (d + e + f) + x_3 \cdot (g + h + i)$

simplified $E(X)$ (same steps for $E(Y)$): $x_1 \cdot f_X(x_1) + x_2 \cdot f_X(x_2) + x_3 \cdot f_X(x_3)$

independent random variables

X does not decide Y and vice versa

X and Y are independent iff for any x and y (all pairs),

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

*must manually check all combinations

product feature: necessary condition for independence: $R_{X,Y}$ needs to be a product space

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) > 0$$
$$\Rightarrow R_{X,Y} = \{(x, y) \mid x \in R_X; y \in R_Y\} = R_X \times R_Y$$

if $R_{X,Y}$ is not a product space, then X and Y are not independent (visually, it's a rectangular space)

properties:

- if $A \subset \mathbb{R} \wedge B \subset \mathbb{R}$, the events $X \in A$ and $Y \in B$ are independent events in S

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$
$$P(X \leq x; Y \leq y) = P(X \leq x)P(Y \leq y)$$

- for arbitrary functions $g_1(\cdot)$ and $g_2(\cdot)$, $g_1(X) \perp g_2(Y)$
- $f_X(x) > 0 \Rightarrow f_{Y \mid X}(y \mid x) = f_Y(y)$
- $f_Y(y) > 0 \Rightarrow f_{X \mid Y}(x \mid y) = f_X(x)$

checking independence

given a joint probability table (for discrete variables), if there are 0 entries in the table, then $R_{X,Y}$ is not a product space, hence $X \not\perp Y$

more generally, both conditions must hold:

1. $R_{X,Y}$ is positive and is a product space
2. for any $(x, y) \in R_{X,Y}$, $f_{X,Y}(x, y) = C \times g_1(x) \times g_2(y)$ (can be decomposed into parts that all do not depend on each other)
   * $g_1(X)$ and $g_2(Y)$ do not need to be probability functions

joint expectation

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)f_{X,Y}(x, y)$$

or

$$E(g(X, Y)) = \int_{-\infty}^\infty \int_{-\infty}^\infty g(x, y)f_{X,Y}(x, y)dydx$$

covariance

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= \sum_x \sum_y (x - \mu_X)(y - \mu_Y)f_{X,Y}(x, y) \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty (x - \mu_X)(y - \mu_Y)f_{X,Y}(x, y)dxdy \end{aligned}$$

properties:

- $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$
- if $X \perp Y$, $\text{cov}(X, Y) = 0$ (but converse is not true)
- $X \perp Y \Rightarrow E(XY) = E(X)E(Y)$
- $\text{cov}(aX + b, cY + d) = ac \cdot \text{cov}(X, Y)$
  - $\text{cov}(X, Y) = \text{cov}(Y, X)$
  - $\text{cov}(X + b, Y) = cov(X, Y)$
  - $\text{cov}(aX, Y) = a\ \text{cov}(X, Y)$
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab \cdot \text{cov}(X, Y)$
  - $V(aX) = a^2V(X)$
  - $V(X + Y) = V(X) + V(Y) + 2\ \text{cov}(X, Y)$
- $V(a + bX) = b^2V(X)$

joint variance

- $X \perp Y \Rightarrow V(X \pm Y) = V(X) + V(Y)$
- $V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n) + 2\sum_{j>i} \text{cov}(X_i, X_j)$
- $X_1 \perp X_2 \perp \cdots \perp X_n \Rightarrow V(X_1 \pm X_2 \pm \cdots \pm X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$

discrete distributions

discrete uniform distribution

if random variable X assumes values $x_1, x_2, ..., x_k$ with equal probability p, then X follows a discrete uniform distribution

$$f_X(x) = \begin{cases} \frac{1}{k} \text{ if } x = x_1, x_2, ..., x_k \\ 0 \text{ otherwise} \end{cases}$$

- $\mu_X = E(X) = \sum_{i=1}^k x_i f_X(x_i) = \frac{1}{k}\sum_{i=1}^k x_i$
- $\sigma_X^2 = V(X) = \frac{1}{k}\sum_{i=1}^k x_i^2 - \mu_X^2$

bernoulli distribution

bernoulli trial: random experiment with only two possible outcomes: `1` success, `0` failure

bernoulli random variable: let X be the number of successes in a Bernoulli trial

- X only has 2 values ($\{0, 1\}$)
- p: probability of success for a Bernou.li trial

$$f_X(x) = P(X = x) = \begin{cases} p \text{ if } x = 1 \\ 1 - p \text{ if } x = 0 \end{cases} = p^x(1-p)^{1-x}$$

- $X \sim \text{Bernoulli}(p)$ and $q = 1 - p$

$$f_X(1) = p; f_X(0) = q$$

- $\mu_X = E(X) = p$
- $\sigma_X^2 = V(X) = p(1-p) = pq$

bernoulli process: sequence of repeatedly performed independent and identical Bernoulli trials

- generates sequence of independent and identically distributed Bernoulli random variables $X_1, X_2, ..., X_n$

binomial distribution

number of successes in n trials of a Bernoulli process; denoted by X

- $X \sim \text{Bin}(n, p)$ where n is the number of trials and p is the probability of success

$$P(X = x) = \binom{n}{x}p^x(1-p)^{n-x}$$

- $E(X) = np$, $V(X) = np(1-p)$
- normal distribution may approximate binomial distribution

negative binomial distribution

number of independent and identically distributed Bernoulli trials needed until the kth success occurs

- $X \sim \text{NB}(k, p)$

$$f_X(x) = P(X = x) = \binom{x-1}{k-1}p^k(1-p)^{x-k}$$

- equivalent to probability of x trial success and $k - 1$ success with $x - 1$ trials
  - $\binom{x-1}{k-1}p^{k-1}(1-p)^{x-k} \times p$
- $E(X) = \frac{k}{p}$, $V(X) = \frac{(1-p)k}{p^2}$

geometric distribution

number of independent and identically distributed Bernoulli trials needed until first success occurs

- $X \sim \text{Geom}(p)$

$$f_X(x) = P(X = x) = (1-p)^{x-1}p$$

- $E(X) = \frac{1}{p}$, $V(X) = \frac{1-p}{p^2}$
- special case of negative binomial distribution

poisson distribution

number of events occurring in a fixed period of time or fixed region

- $X \sim \text{Poisson}(\lambda)$ where $\lambda > 0$ is the expected number of occurrences during the given period/region

$$f_X(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

- k is the number of occurrences of such event
- $E(X) = \lambda$, $V(X) = \lambda$

$$P(X \leqslant k) = e^{-\lambda}\left(\lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + ... + \frac{\lambda^k}{k!}\right)$$

poisson process: continuous time process with rate α; count the number of occurrences within some interval of time

- properties:
  - expected number of occurrences in interval of length T is αT
  - no simultaneous occurrences
  - number of occurrences in disjoint time intervals are independent
- follows $\text{Poisson}(\alpha T)$ distribution

approximation to binomial: let $X \sim \text{Bin}(n, p)$; $n \to \infty, p \to 0$ such that $\lambda = np$ is a constant

- then $X \sim \text{Poisson}(np)$

$$\lim_{p\to 0;\, n\to\infty} P(X=x) = \frac{e^{-np}(np)^x}{x!}$$

- recommended values: n big, p small
  - $n \geqslant 20 \wedge p \leqslant 0.05$, OR
  - $n \geqslant 100 \wedge np \leqslant 10$

continuous distributions

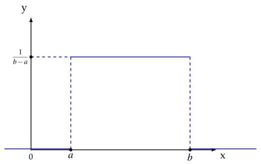
continuous random variable X follows the following distributions f they have the following probability density function

continuous uniform distribution

over interval (a, b)

$$f_X(x) = \begin{cases} \frac{1}{b-a} \text{ if } a \leqslant x \leqslant b \\ 0 \text{ otherwise} \end{cases}$$

- $X \sim U(a, b)$
- $E(X) = \frac{a+b}{2}$, $V(X) = \frac{(b-a)^2}{12}$



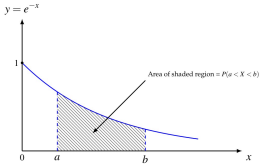
$$F_X(x) = \begin{cases} 0 \text{ if } x < a \\ \frac{x-a}{b-a} \text{ if } a \leqslant x \leqslant b \\ 1 \text{ if } x > b \end{cases}$$

exponential distribution

parameter $\lambda > 0$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} \text{ if } x \geqslant 0 \\ 0 \text{ if } x < 0 \end{cases}$$

- $X \sim \text{Exp}(\lambda)$
- $E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$
- can be written with μ if $\mu = 1/\lambda$



$$\begin{aligned} F_X(x) &= P(X \leqslant x) \\ &= \int_0^x \lambda e^{-\lambda t}\, \mathrm{d}t \\ &= \left[-e^{-\lambda t}\right]_0^x \\ &= 1 - e^{-\lambda x} \end{aligned}$$

- if $x \geqslant 0$, else $F_X(x) = 0$

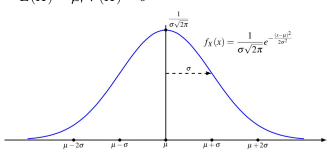
memory-less property: if $X \sim \text{Bin}(\lambda)$, then $P(X > s + t \mid X > s) = P(X > t)$

normal distribution

parameter μ and σ^2

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{(-x-\mu)^2/2\sigma^2}$$

- $X \sim N(\mu, \sigma^2)$
- $E(X) = \mu$, $V(X) = \sigma^2$



properties:

- same σ_2 but different $\mu \to$ same shape, different center
- as σ_2 increases $\to$ curve flattens

approximation to binomial: let $X \sim \text{Bin}(n, p)$; $n \to \infty$

- then $X \sim\approx N(0, 1)$

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1-p)}}$$

- recommended values: $np > 5$ and $n(1-p) > 5$
- continuity correction: adjusting ranges by $\pm 1/2$
  - $a \leqslant X \Rightarrow -1/2$
  - $a < X \Rightarrow +1/2$
  - $X \leqslant b \Rightarrow +1/2$
  - $X < b \Rightarrow -1/2$

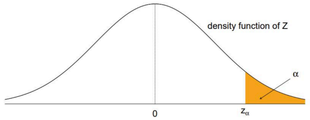
standard normal

given $X \sim N(\mu, \sigma^2)$, let $Z = \frac{X-\mu}{\sigma}$, $Z \sim N(0, 1)$

- $\varphi(\cdot) = f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$: pdf
- $\Phi(\cdot) = \int_{-\infty}^z \varphi(t)\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2}\,\mathrm{d}t$: cumulative function
  - $P(x_1 < X < x_2) \Rightarrow P(z_1 < Z < z_2) \Rightarrow \Phi(\frac{x_2-\mu}{\sigma}) - \Phi(\frac{x_1-\mu}{\sigma})$

properties:

1. $P(Z \geqslant 0) = P(Z \leqslant 0) = \Phi(0) = 0.5$
2. $\forall z, \Phi(z) = P(Z \leqslant z) = P(Z \geqslant -z) = 1 - \Phi(-z)$
3. $Z \sim N(0, 1) \Rightarrow -Z \sim N(0, 1)$
4. $Z \sim N(0, 1) \Rightarrow \sigma Z + \mu \sim N(\mu, \sigma^2)$
5. $P(Z > z_\alpha) = \alpha$



quantile: upper (αth) quantile where $0 \leqslant \alpha \leqslant 1$ is x_α that satisfies $P(X \geqslant x_\alpha) = \alpha$

- common z_α values:
  - $z_{0.05} = 1.645$
  - $z_{0.01} = 2.326$
- symmetrical about 0, so $P(Z \geqslant z_\alpha) = P(Z \leqslant -z_\alpha) = \alpha$

population

population: all possible outcomes/observaations of a survey/experiment; size is N

- population mean: μ_X
- population variance: σ_X^2

sample: subset of population; size is n

finite population: finite number of elements

infinite population: infinitely large number of elements

random sampling

sample of n members taken from a given population: $\binom{N}{n}$

- every member has the same probability of being selected
- yields sample that resembles the population; reducing chance that sample is seriously biased

sampling infinite population: let X be a random variable with pdf $f_X(x)$

- let $X_1, X_2, \ldots, X_n$ be independent random variables with the same distribution as X
- $(X_1, X_2, \ldots, X_n)$ is a random sample of size n
- $f_{X_1, X_2, \ldots, X_n} = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n)$
- sampling from a finite population with replacement $\Rightarrow$ sampling from an infinite population

sampling distribution: probability distribution of a statistic

statistics

function of n observations in sample; statistics are samples

sample mean:

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

- $\mu_{\overline{X}} = E\left(\overline{X}\right) = \mu_X$
  - in the "long run"
- $\sigma_{\overline{X}}^2 = V\left(\overline{X}\right) = \sigma_X^2/n$

sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2$$

standard error: spread of sampling distribution (standard deviation): SE or $\sigma_{\overline{X}}$

- how much $\overline{X}$ tends to vary from sample to sample of size n

law of large numbers: if $X_1, X_2, \ldots, X_n$ are independent random variables with the same mean μ and variance σ_2, then for any $\varepsilon \in \mathbb{R}$,

$$n \to \infty \Rightarrow P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) \to 0$$

- i.e. as sample size increases, the probability that sample mean differs from population mean goes to 0

central limit theorem: if $\overline{X}$ is a mean of random sample size n from population with mean μ and finite variance σ^2, then

$$n \to \infty \Rightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to Z \sim N(0, 1)$$

$$\Rightarrow \overline{X} \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

- i.e. for large n, sums and means of random samples drawn from a population follow the normal distribution closely
  - if sample comes from normal population, then $\overline{X}$ is normally distributed as well
- rule of thumb:
  - symmetric population: $n = 15 - 20$
  - moderately skewed: $n = 30 - 50$
  - extremely skewed: $n = 1000$
- convergence in distribution: for any x

$$\lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leqslant x\right) = \Phi(x)$$

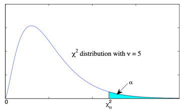
sampling distributions

chi-squared (χ^2) distribution

let $Z_1, Z_2, \ldots, Z_n$ be n independent and identically distributed standard normal random variables

- a random variable with the same distribution as $Z_1^2 + Z_2^2 + \ldots + Z_n^2$ is a χ^2 random variable with n degree of freedom
- $\chi^2(n)$

$$P(Y > \chi^2(n;\alpha)) = \alpha.$$



properties:

1. $Y \sim \chi^2(n) \Rightarrow E(Y) = n; V(Y) = 2n$
2. for large n, $\chi^2(n) \approx N(n, 2n)$
3. if Y_1 and Y_2 are independent χ^2 random variable with m, n degrees of freedom, then $Y_1 + Y_2$ is $\chi^2(m+n)$
4. has a long right tail

distribution of $\frac{(n-1)S^2}{\sigma^2}$: where $X_i \sim N(\mu, \sigma^2)$ has $\chi^2(n-1)$

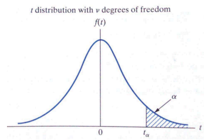
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{\sigma^2}$$

t-distribution

suppose $Z \sim N(0, 1)$ and $U \sim \chi^2(n)$ and $Z \perp U$:

$$T = \frac{Z}{\sqrt{U/n}} \sim t(n)$$

$$P(T > t_{n;\alpha}) = \alpha.$$



properties:

1. $n \to \infty \Rightarrow t(n) \to N(0, 1)$ (when $n \geqslant 30$, replace with $N(0, 1)$)
2. $T \sim t(n) \Rightarrow E(T) = 0; V(T) = n/(n-2), n > 2$
3. graph: symmetric about vertical axis and ressembles grapah of standard normal (but flatter)
4. used if σ (stdev) is unknown and sample size is small
   - otherwise, if sample size is large, use z but continue using S instead of σ

relation to normal distribution: if $X_1, \ldots, X_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

F-distribution

suppose $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are independent

$$F = \frac{U/m}{V/n} \sim F(m, n)$$

properties:

1. $X \sim F(m, n) \Rightarrow E(X) = \frac{n}{n-2}, n > 2; V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$
2. $F \sim F(n, m) \Rightarrow \frac{1}{F} \sim F(m, n)$
3. $F(m, n; \alpha) \Rightarrow P(F > F(m, n; \alpha)) = \alpha$
4. $F(m, n; 1-\alpha) = 1/F(n, m; \alpha)$

estimator

rule about how to calculate an estimate based on information in the sample

unbiased estimator: let $\hat{\Theta}$ be the estimator of θ

$$E\left(\hat{\Theta}\right) = \theta$$

- $E(S^2) = \sigma^2$

maximum error of estimate: replace σ with S if necessary and $z_{\alpha/2}$ with $t_{n-1;\alpha/2}$ if variance not known or sample size too small

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

reversing the formula:

$$n \geqslant \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

confidence intervals

interval estimator: rule for calculating, from a sample, an interval (a, b) with some level of certainty that the parameter of interest lies in

- quantified as degree of confidence/confidence level $(1-\alpha)$
- (a, b) is the $(1-\alpha)$ confidence interval

$$P(a < \mu < b) = 1 - \alpha$$

- written as

$$\overline{X} \pm E$$

interpretation: given some sample statistic, the population parameter is either contained within (or not) the confidence interval

- when repeated over many samples, about $100(1-\alpha)\%$ of the confidence intervals will contain the population parameter

2 populations

independent samples

usually focused on $\mu_1 - \mu_2 = \delta$

known and unequal variance

populations are normal OR both samples are large

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0, 1)$$

confidence interval:

$$\left(\overline{X} - \overline{Y}\right) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

large and unknown variance

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx N(0, 1)$$

confidence interval:

$$\left(\overline{X} - \overline{Y}\right) \pm z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

small, equal, unknown variance

equal variance assumption: if $1/2 \leqslant S_1/S_2 \leqslant 2$

pooled estimator: estimates σ^2; follows $t_{n_1+n_2-2}$ distribution

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \approx N(0, 1)$$

confidence interval:

$$\left(\overline{X} - \overline{Y}\right) \pm t_{n_1+n_2-2;\alpha/2} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

large and equal variance

$$\left(\overline{X} - \overline{Y}\right) \pm z_{\alpha/2} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

dependent samples/paired data

each pair is independent from each other; let $D_i = X_i - Y_i$ and $\mu_D = \mu_1 - \mu_2$

- treat $D_1, D_2, \ldots, D_n$ as random sample with mean μ_D and variance σ_D^2

$$T = \frac{\overline{D} - \mu_D}{S_D/\sqrt{n}};$$

$$\overline{D} = \frac{\sum_{i=1}^n D_i}{n}$$

$$S_D^2 = \frac{\sum_{i=1}^n \left(D_i - \overline{D}\right)^2}{n-1}$$

- $n < 30$ and population normally distributed: $T \sim t_{n-1}$
- $n \geqslant 30$: $T \sim N(0, 1)$

confidence interval:

$$\overline{D} \pm t_{n-1;\alpha/2} \frac{S_D}{\sqrt{n}}$$

- replace $t_{n-1;\alpha/2}$ with $z_{\alpha/2}$ if $n \geqslant 30$

hypothesis testing

1. set competing hypotheses: null and alternative
2. set level of significance
3. identify tet statistic, distribution, and rejection criteria
4. compute the observed test statistic value
5. conclusion

null vs alternative hypothesis

looking for ways to reject that there is no change and show that there is significance

null hypothesis (H_0): assumed truth (i.e. no change)

alternative hypothesis (H_1): contrasting hypothesis; what we want to prove

outcome: reject or fail to reject hypothesis

one-sided hypothesis test: $H_1 : \mu > \overline{x}$ or $H_1 : \mu < \overline{x}$ (right or left leaning hypothesis test)

two-sided hypothesis test: $H_1 : \mu \neq \overline{x}$ (both sides)

level of significance

conclusions:

- reject H_0 and conclude H_1
- do not reject H_0 and conclude H_0

	Do not reject H_0	Reject H_0
H_0 is true	Correct decision	Type 1 error
H_0 is false	Type 2 error	Correct decision

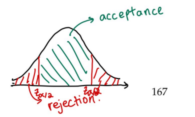
level of significance α: probability of making a type 1 error

- more serious so control this

power of test $1-\beta$: probability of making type 2 error

test statistic, distribution, and rejection region

quantifies how unlikely it is to observe the sample assuming the null hypothesis is true



calculation & conclusion

check if sample statistic falls within rejection region

- if so, sample is improbable, so reject H_0
- else, failed to reject H_0

hypothesis testing with mean

population distribution is normal or n is sufficient large

use $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ as test statistic and rejection region is $P\left(|Z| > z_{\alpha/2}\right) = \alpha$

- if two-tailed test: $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$
- if one-tailed test: $Z < -z_{\alpha/2}$ (for left tail) or $Z > z_{\alpha/2}$ (for right tail)

unknown variance: use t-distribution instead with $n-1$ degree of freedom with S replacing σ; use standard normal distribution iff $n \geqslant 30$

- swap Z check with t check

p-value

probability of obtaining a test statistic at least as extreme than the observed sample value, given that H_0 is true; i.e. observed level of significance

- two-sided test: p-value $= P(|Z| > |z|) = 2P(Z > |z|) = 2P(Z < -|z|)$
- left-tailed test: p-value $= P(Z < -|z|)$
- right-tailed test: p-value $= P(Z > |z|)$

rejection: if p-value $< \alpha$, reject H_0, else do not reject

relation with confidence intervals

confidence intervals can be used to perform two-sided tests so by constructing a confidence interval given some confidence $100(1-\alpha)\%$, then if $\overline{X} \notin$ CI, then we can reject H_0

hypothesis testing with independent samples

$$H_0 : \mu_1 - \mu_2 = \delta_0$$

where δ_0 is usually 0

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

if σ^2 is known, and population is normally distributed or n_1, n_2 are sufficiently large

otherwise, if σ^2 is unknown but samples are sufficiently large

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0, 1)$$

using the same rejection regions as when dealing with single samples

if σ^2 is unknown but equal, distribution are normal and samples are small, then use t-distribution and pooled estimator instead

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \delta_0}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

hypothesis testing with dependent data/paired data

$$T = \frac{\overline{D} - \mu_{D_0}}{S_D/\sqrt{n}} \sim t_{n-1}$$

if sample size $\geqslant 30$, $T \sim N(0, 1)$