Non-Comprehensive ECE286 Notes

James Li

July 23, 2023

I Counting and Other Shenanigans

Permutations: cares about order; given n items, no. of permutations is n!

- no. of permutations of r out n items:

$$\frac{n!}{(n-r)!}$$

- with repeated items:

$$\binom{n}{n_1, ..., n_m} = \frac{n!}{n_1!...n_m!} \quad ; \quad \sum_{k=1}^{m} n_k = n$$

- Example: permutations of ATLANTIC:

$$\binom{8}{2, 2, 1, 1, 1, 1} = \frac{8!}{2!2!}$$

Partitions: Events that are mutually exclusive; found the same as permutations with repetition:

$$\binom{n}{n_1, ..., n_m} = \frac{n!}{n_1!...n_m!}$$

Combinations: does not care about order; expressed as $\binom{n}{r}$

- **size r combination:** partition with $n_1 = r$, $n_2 =$ n-r:

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

II Probability

- Given that A, B $\subseteq$ S: if A $\cap$ B = , $\mathbb{P}$(A) + $\mathbb{P}$(B) = 1.

Additive Rule: Given that A, B $\subseteq$ S:

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

- for 3 events, given A, B, C $\subseteq$ S, :

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) \\ -\mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$$

Conditional Probability: denoted as $\mathbb{P}$(B|A); probability that event B occurs given A occurs first.

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

Product Rule: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$

Independence: Event A is independent of B is $\mathbb{P}$(A|B) = $\mathbb{P}$(A), $\mathbb{P}$(B|A) = $\mathbb{P}$(B)

- not the same as mutually exclusive – A $\cap$ B = $\emptyset$
- $\mathbb{P}$(A $\cap$ B) = $\mathbb{P}$(A)$\mathbb{P}$(B)

Bayes' Rule: Utilizes the Product Rule and Conditional Probability to derive this equality:

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

III Total Probability

- Suppose A is an event, B_1...B_k is a partition, then:

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Bayes' Rule (retooled): suppose that C_1,..., C_k is a partition:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)} \\ = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\sum_{i=1}^{k} \mathbb{P}(C_i)\mathbb{P}(A|C_i)}$$

- and if B is an element of {C_1,...,C_k}:

$$\mathbb{P}(C_n|A) = \frac{\mathbb{P}(C_n)\mathbb{P}(A|C_n)}{\sum_{i=1}^{k} \mathbb{P}(C_i)\mathbb{P}(A|C_i)}$$

IV Random Variables

Random Variable: Function that maps each element of a sample space to a real number:

- 3 coin flips, RV X is no. of H:
  - $\mathbb{P}$(X = 3) = 1/8
  - $\mathbb{P}$(X = 2) = 3/8
- **discrete** — countable
- **continuous** — in interval of $\mathbb{R}$

IV.I Discrete Probability Distributions

Probability Mass Function (PMF)

- $\sum_x$ f(x) = 1, f(x) = $\mathbb{P}$(X=x), f(x) $\geq$ 0 for X = x

Cumulative Distribution Function (CDF)

- If X has PMF f(x), CDF:

$$F(x) = \sum_{t \leq x} f(t)$$

- F(x) = $\mathbb{P}$(X$\leq$x); i.e) F(1) = f(0) + f(1)

IV.II Continuous Probability Distributions

Probability Mass Function

- $\int_{-\infty}^{\infty} f(x)dx = 1$, $\int_a^b f(x)dx = \mathbb{P}(a < X < b)$

Cumulative Distribution Function

- F(x) = $\int_{-\infty}^{x} f(t)dt$, $\mathbb{P}(a < X \leq b)$ = F(b) - F(a)
- F(∞) = $\int_{-\infty}^{\infty} f(t)dt = 1$

IV.III Joint Distributions

Joint Distribution: a function $f(x, y)$ is a joint PMF of RVs X an Y if:

- $f(x, y) \geq 0$ for all x, y.
- $\mathbb{P}(X = x, Y = y) = f(x, y)$
- **Discrete:** $\sum_x \sum_y f(x, y) = 1$
  - Given A $\subset$ S:

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

- **Continuous:** $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$
  - Given A $\subset$ S:

$$\mathbb{P}(X, Y \in A) = \int_{(x,y) \in A} f(x, y)dxdy$$

IV.III.1 Marginal Distributions

A **marginal distribution** is the distribution of each individual RV given a joint distribution, i.e) g(x) is the marginal distribution of X.

- **Discrete:**

$$g(x) = \sum_y f(x, y) \qquad h(y) = \sum_x f(x, y)$$

- **Continuous:**

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy \qquad h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

IV.III.2 Conditional Distributions

Same idea as conditional probability, but applied to a PDF:

$$f(x|y) = \frac{f(x, y)}{g(y)}$$

where g(y) is the marginal distribution of y.

- **Discrete:**

$$\mathbb{P}(a \leq X \leq b | Y = y) = \sum_{a \leq x \leq b} f(x|y)$$

- **Continuous:**

$$\mathbb{P}(a \leq X \leq b | Y = y) = \int_a^b f(x|y)dx$$

Random Variables X and Y are **independent** if joint distribution $f(x, y)$ and marginal distributions $g(y), h(x)$ can be expressed as:

$$f(x, y) = g(y)h(x)$$

V Expectation, Variance, Covariance

Given RV X has a distribution $f(x)$, the **expected value (or mean), E[X]**:

- **Discrete:**
$$E[X] = \sum_x x f(x)$$
- **Continuous:**
$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Given RV X has a distribution $f(x)$ and g(X) is a function of X, the expected value, E[g(X)]:

- **Discrete:**
$$E[g(X)] = \sum_x g(x) f(x)$$
- **Continuous:**
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Given RVs X and Y have a joint distribution $f(x, y)$ and g(X,Y) is a function of X and Y, the expected value, E[g(X,Y)]:

- **Discrete:**
$$E[g(X, Y)] = \sum_x \sum_y g(x, y) f(x, y)$$
- **Continuous:**
$$E[g(X, Y)] = \int_{-\infty}^{\infty} g(x, y) f(x, y) dx$$

Let X be an RV with distribution f(x) and mean $\mu =$ E[X]. The **variance** of X, σ^2, is:

- **Discrete:**
$$\sigma^2 = E[(x-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$
- **Continuous:**
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$
- σ is known as the **standard deviation**.
- **Useful Formula:**
$$\sigma^2 = E[X^2] - \mu^2 = E[X^2] - E[X]^2$$

Let X and Y be RVs with joint distribution f(x, y) and means μ_X and μ_Y. The **covariance**, σ_{XY}, of X and Y is:

- **Discrete:**
$$\sigma_{XY} = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x, y)$$
- **Continuous:**
$$\sigma_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x, y) dx dy$$
- if X>0, Y>0, $\sigma_{XY} > 0$; if X>0, Y<0, $\sigma_{XY} < 0$.
- **Useful Formula:**
$$\sigma_{XY} = E[XY] - \mu_X \mu_Y$$

Let X and Y be RVs with covariance σ_{XY} and standard deviations σ_X and σ_Y. The **correlation coefficient** of X and Y is:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, -1 \leq \rho_{XY} \leq 1$$

V.I Linear Combinations

Expectation Value:

- The expectation of the RV $aX + Y$ is:
$$E[aX + Y] = aE[X] + E[Y]$$
- Similarly:
$$E[aX + b] = aE[X] + b$$
$$E[g(X, Y) + h(X, Y)] = E[g] + E[h]$$

Variance, Covariance: Suppose X and Y are independent — $f(x, y) = g(x)h(y)$:

$$E[XY] = E[X]E[Y]$$

Recall: covariance

$$\sigma_{XY} = E[XY] - E[X]E[Y]$$

- Independence implies **uncorrelated** ($\sigma_{XY} = 0$)
- Uncorrelated does not imply independence.

Example: Consider the following PDF.

$$f(x, y) = \begin{cases} 1/\pi, \ x^2 + y^2 = 1 \\ 0, \ \text{else} \end{cases}$$

- Check for the function's independence:
$$g(x) = \frac{2}{\pi}\sqrt{1-x^2}$$
$$h(y) = \frac{2}{\pi}\sqrt{1-y^2}$$
- Note that $g(x)h(y) \neq f(x, y)$, implying RVs X and Y are not independent.
- Check for the covariance of the function:
$$E[XY] = E[X] = E[Y] = 0$$
- Thus, the RVs are uncorrelated but not independent.
- **Useful formula:**

$$\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma^2_{XY}$$

VI Distributions

A **uniform distribution** implies that every element in S has the same probability.

- If S = {1,..., n}, $f(k) = \frac{1}{n}$ given k $\in$ S.

A **Bernoulli distribution** is a probability distribution with 2 outcomes, where the probability of achieving a 1 is **p** and a 0 is $\mathbf{q = 1 - p}$.

Binomial Distribution/Bernoulli Process: suppose a binomial event is repeated n times, and RV X is the number of 1's that occurs:

- Notation for binomial distribution: $f(x) = b(x; n, p)$
- Probability of x 1's and n-x 0's in a particular order:

$$p^x(1-p)^{n-x}$$

$$b(x; n, p) = \binom{n}{x} p^x(1-p)^{n-x}$$

Expectation value of a binomial distribution:

$$E[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x(1-p)^{n-x}$$

- Picture Bernoulli as a sum of n trials, and apply linearity:

$$\begin{aligned} E[X] &= E[Y_1] + ... + E[Y_n] \\ &= np \end{aligned}$$

Variance:

$$\begin{aligned} \sigma_X^2 &= \sum_{k=1}^{n} \sigma_{Y_k}^2 \\ &= np(1-p) \end{aligned}$$

A **Multinomial distribution** is the same idea as a binomial distribution but each trial can now have m outcomes rather than 2 outcomes.

- The chance of any outcome i, $\mathbb{P}(E_i)$:

$$\mathbb{P}(E_i) = p_i, \quad \sum_{i=1}^{m} p_i = 1$$

- The multinomial distribution relates the probability of outcome i happening x_i times, $p_i^{x_i}$, with the concept as repeated items in a permutation, or a partition:

$$f(x_1...x_n; p_1...p_m, n) = \binom{n}{x_1, ..., x_m} p_1^{x_1}...p_m^{x_m}$$

A **Hypergeometric distribution** is different in the sense that there is **no replacement** after the object is drawn.

- Given N objects with K successes drawn n times, the chance of having x successes and n-x failures is:

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

- If N can be partitioned into events $a_1...a_k$, the **multivariate hypergeometric distribution** is as follows:

$$f(x_1...x_k; a_1...a_k, N, n) = \frac{\binom{a_1}{x_1}...\binom{a_k}{x_k}}{\binom{N}{n}}$$

- The mean and variance for a hypergeometric distribution is:

$$\mu = \frac{nK}{N}, \quad \sigma^2 = \frac{Kn(N-n)}{N(N-1)}(1 - \frac{K}{N})$$

A **Negative Binomial** is the chance that the k^{th} event occurs on the n^{th} draw, denoted as $b^*(x; k, p)$.

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k(1-p)^{x-k}$$

- How to relate to binomial distribution?[1]

$$b^*(x; k, p) = pb(k-1; x-1, p)$$

A **Geometric Distribution** is a negative binomial distribution with k = 1; when the first success occurs:

$$g(x; p) = b^*(x; 1, p) = p(1-p)^{x-1}$$

- The mean and variance of the geometric distribution are:

$$\mu = \frac{1}{p}, \quad \sigma^2 = \frac{1-p}{p^2}$$

A **Poisson distribution** is the number of times that something happens in one sequence of intervals, *i.e)* number of snow days in a year.

$$p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

- The mean and variance of a Poisson distribution are both λ.
- As such, we can conclude that λ is the average number of occurances per interval, which can be represented with r (rate) and t (length of interval):

$$\lambda = rt = E[X]$$

- How to represent a Poisson distribution with a binomial distribution[2]:

$$p(x; \lambda) = \lim_{n\to\infty, p\to 0} b(x; n, p)$$

A **uniform distribution** for a continuous, uniform RV in [A, B] is:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \le x \le B \\ 0, & else \end{cases}$$

- The mean and variance of a uniform distribution are:

$$\mu = \frac{A+B}{2}, \quad \sigma^2 = \frac{(B-A)^2}{12}$$

A **Gaussian (Normal) distribution**[3] for a normal RV X with mean μ and variance σ^2 is:

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

[1] Check p.44 of Taylor's notes for proof
[2] Check p.48 of Taylor's notes for proof
[3] Is it a PDF? Check p.50.

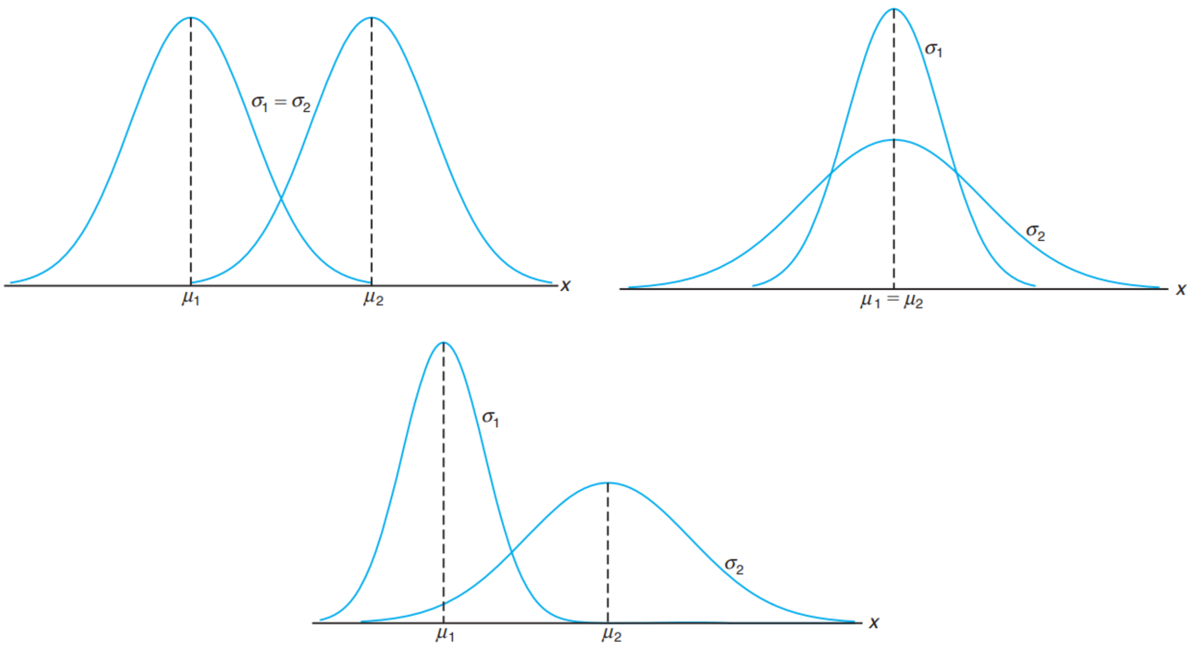


Figure 1: **Top left:** different means, same std dev; **Top right:** same mean, different std dev; **Bottom:** different means and std dev.

A **Standard normal distribution** is a Gaussian distribution where the mean is 0 and the standard deviation is 1:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$

- If X is a standard normal RV:

$$P(A \leq X \leq B) = \frac{1}{\sqrt{2\pi}} \int_{A}^{B} e^{\frac{-t^2}{2}} dt = \Phi(B) - \Phi(A)$$

- If X has PDF $n(x; \mu, \sigma)$, there is no analytical form for the CDF integral and it is inconvenient to compute μ and σ.
- Instead, choose Z $= \frac{X-\mu}{\sigma}$, where Z has PDF $n(z; 0, 1)$:

$$P(X \leq x) = \int_{-\infty}^{x} n(t; \mu, \sigma) dt = P(Z \leq \frac{x-\mu}{\sigma})$$

- Similarly:

$$P(A \leq X \leq B) = \Phi(\frac{B-\mu}{\sigma}) - \Phi(\frac{A-\mu}{\sigma})$$

The **Normal approximation of a binomial PDF** can be represented as:

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

noting that as $n \to \infty$, PDF of Z is $n(x; 0, 1)$.

- Using the property $\mu = np$ and $\sigma = \sqrt{np(1-p)}$:

$$P(X \leq x) = \sum_{k=0}^{x} b(k; n, p) = P(Z \leq \frac{(x+0.5)-\mu}{\sigma})$$

- Note that for the lower z bound (z_1), you subtract 0.5 from the smaller x value, meaning $z_1 = \frac{x_1-0.5-\mu}{\sigma}$, $z_2 = \frac{x_2+0.5-\mu}{\sigma}$.
- If a question asks for the probability of **exactly**, still use normal approximation but with the same x value, meaning $z_1 = \frac{x-0.5-\mu}{\sigma}$, $z_2 = \frac{x+0.5-\mu}{\sigma}$.

VI.I Gamma and Exponential Distributions

The **gamma function**, $\Gamma(\alpha)$ is denoted as:

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x} dx, \ \alpha > 0$$

This gamma function is a generalization of the factorial function, where $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

For a RV X with **gamma distribution** with parameters $\alpha, \beta > 0$:

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{\frac{-x}{\beta}}, & x > 0 \\ 0, & \text{else} \end{cases}$$

with $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$.

Special Cases of the Gamma Distribution include:

- **Chi-squared distribution** (χ^2) with parameter $v \in \mathbb{N}$, where v is the number of degrees of freedom (in experiments, number of independent variables in calculation).

$$f(x; v) = \begin{cases} \frac{1}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})} x^{\frac{v}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{else} \end{cases}$$

  with $\mu = v, \sigma^2 = 2v$.
- **Exponential distribution** with parameter $\beta > 0$:

$$f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{\frac{-x}{\beta}}, & x > 0 \\ 0, & \text{else} \end{cases}$$

  with $\alpha = 1, \mu = \beta, \sigma^2 = \beta^2$.
- **Poisson Process:** Consider the probability that no event occurs in t: $p(0; rt) = e^{-rt}$.
  - Let X be the RV of the time to first event. $\mathbb{P}(X > x) = e^{-rx}$, and $\mathbb{P}(X \leq x) = 1 - e^{-rx}$.
  - $\mathbb{P}(X \leq x)$ is a CDF ($F(x)$) and is the sum (integral) of a PDF ($f(x)$).

$$f(x) = \frac{d}{dx} \mathbb{P}(X \leq x) = re^{-rx}$$

    which is just the exponential distribution with r $= 1/\beta$.

- Exponential distributions have a property called **the memoryless property**.

$$P(X \geq s+t \,|\, X \geq s) = P(X \geq t)$$

VII Functions of Random Variables

- Given X has PDF $f(x)$ and $Y = u(X)$, where $u(X)$ is a one-to-one function — each value of X maps one value to Y.
- As such, we can write $X = w(Y) = u^{-1}(Y)$.
- **Discrete PDF:** the probability distribution of RV Y, if we let $g(y)$ be the distribution of Y:

$$g(y) = f(u^{-1}(y))$$

- **Continuous PDF:** the probability distribution of RV Y, if we let $g(y)$ be the distribution of Y:

$$g(y) = f[w(y)]|J|$$

where $|\text{J}| = \frac{d(w(y))}{dy}$ is the Jacobian of the transformation.

Suppose $\text{X}_{1,2}$ are discrete random variables with joint PDF $f(x_1, x_2)$. If we let $\text{Y}_{1,2} = u_{1,2}(X_1, X_2)$, we can find an inverse expression $x_{1,2} = w_{1,2}(y_1, y_2)$.

- **Joint Discrete PDF:** the joint probability distribution of RVs Y_1 and Y_2 is:

$$g(y_1, y_2) = f(w_1(y_1, y_2), w_2(y_1, y_2))$$

- **Joint Continuous PDF:** the joint probability distribution of RVs Y_1 and Y_2 is:

$$g(y_1, y_2) = f(w_1(y_1, y_2), w_2(y_1, y_2))|J|$$

where $|\text{J}|$ is the 2 × 2 determinant:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

If $Y = u(X)$ defines a transformation that is not a one-to-one function between X and Y. If the interval over which X is defined can be partitioned into k mutually disjoint sets such that each of the inverse functions:

$$x_1 = w_1(y), \qquad ..., x_k = w_k(y)$$

Then the probability distribution of Y is:

$$g(y) = \sum_{i=1}^{l} f[w_i(y)]|J_i|$$

If we have distribution Z = X + Y, the distribution of Z, $h(z)$, is:

$$P(Z) = P(X+Y=z) = \sum_w P(X=w)P(Y=z-w)$$

$$\begin{aligned} h(z) &= \int_{-\infty}^{\infty} f(w)g(z-w)dw \\ &= \sum_{w=-\infty}^{\infty} f(w)g(z-w)dw \end{aligned}$$

VII.I Moments and Moment-Generating Functions

The $\mathbf{r}^{th}$ **moment about the origin** of RV X, given $g(X) = X^r, r \in \mathbb{Z}$ is defined as:

$$\mu_r' = E(X^r) = \begin{cases} \sum_x x^r f(x), & \text{discrete case} \\ \int_{-\infty}^{\infty} x^r f(x)dx, & \text{continuous case} \end{cases}$$

with $\mu = \mu_1'$ and $\sigma^2 = \mu_2' - \mu^2$.

The **moment-generating function**, $M_X(t)$, is defined as:

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x), & \text{discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx, & \text{continuous} \end{cases}$$

Moment-generating functions will exist only if the sum or integral above converges. If a moment-generating function of a random variable X does exist, it can be used to generate all the moments of that variable:

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'$$

If X is normal with mean μ and variance σ^2:

$$M_X(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}$$

Linear Combinations of RVs

If RV X has distribution $f(x)$, what is distribution, $g(y)$, given Y = aX:

- **Discrete:** $h(y) = f(y/a)$
- **Continuous:** $h(y) = \frac{1}{|a|} f(y/a)$

Suppose the same X has moment-generating function $M_X(t)$:

$$M_Y(t) = M_X(at)$$

Similarly:

$$M_{aX}(t) = M_X(at)$$

With prior example of Z = X + Y, the moment-generating function for Z is:

$$M_Z(t) = M_X(t)M_Y(t)$$

VIII Sampling

We sample because we can't always take the data of the whole population.

- **Sample data:** $x_1, ..., x_n$, where each data point is a **realization** of a RV, X_i.
- **Sample mean:**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

- **Sample median:** midpoint of sample data set, meaning the data set must be **ordered**.

$$\text{median} = \begin{cases} \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{for n even} \\ x_{\frac{n+1}{2}}, & \text{for n odd} \end{cases}$$

- **Sample mode:** most commonly occurring value

VIII.I Measures of variability

- **Sample variance:**

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

- **Sample std. dev.:** $\sqrt{s^2} = s$
- (n - 1)[4] is referred to as the **degrees of freedom**.

VIII.II Visualization

Histogram

- Uses a relative frequency distribution:

$$\text{relative f} = \frac{\text{frequency}}{\text{total occurrences}}$$

Table 1.7: Relative Frequency Distribution of Battery Life

Class Interval	Class Midpoint	Frequency, f	Relative Frequency
1.5–1.9	1.7	2	0.050
2.0–2.4	2.2	1	0.025
2.5–2.9	2.7	4	0.100
3.0–3.4	3.2	15	0.375
3.5–3.9	3.7	10	0.250
4.0–4.4	4.2	5	0.125
4.5–4.9	4.7	3	0.075

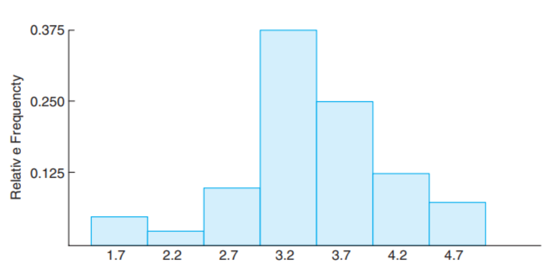


Figure 1.6: Relative frequency histogram.

Box-and-whisker Plot

- The box plot encloses the **interquartile range** — between the 25th and 75th percentile — in a box.
- At the extremes, the whiskers show the extreme observations.

Table 1.8: Nicotine Data for Example 1.5

1.09	1.92	2.31	1.79	2.28	1.74	1.47	1.97
0.85	1.24	1.58	2.03	1.70	2.17	2.55	2.11
1.86	1.90	1.68	1.51	1.64	0.72	1.69	1.85
1.82	1.79	2.46	1.88	2.08	1.67	1.37	1.93
1.40	1.64	2.09	1.75	1.63	2.37	1.75	1.69

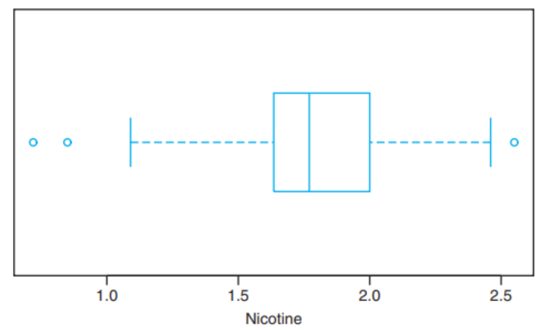


Figure 1.9: Box-and-whisker plot for Example 1.5.

IX Random Sampling

- A **population** is all possible observations; each observation is a realization of a RV.
- A **sample** is a subset of a population.

A **random sample** with n observations in the sample. Observation i is the realization of independent RV X_i where i = 1, ..., n.

$$f(x_1...x_n) = f(x_1)f(x_2)...f(x_n)$$

- A **statistic** is a function of the X_i, such as mean, median, mode.
- A sample is biased if it consistently over/underestimates the statistic of interest.

IX.I Sampling Distribution

The probability distribution of the statistic is called a sampling distribution.

Facts about random sampling:

- If X_1 and X_2 are normal with means $\mu_{1,2}$ and variances $\sigma^2_{1,2}$, the distribution of $X_1 + X_2$ is normal with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.
- If X is normal with μ and σ^2, then X/n has μ/n and σ^2/n^2.
- If $X_1...X_n$ are normal (mean μ, variance σ^2), then $\bar{X}$ has μ and σ^2/n.

[4]See pg. 71-72 on Taylor's notes for proof.

IX.II Central Limit Theorem

Given the sample $X_1 ... X_n$, which are realizations of IID RVs, the sample average is $\bar{X}_n$. Let:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

- The **Central Limit Theorem** states that as n $\to \infty$, the distribution of Z_n converges to the standard normal distribution n(z; 0, 1).
- The normal approximation for $\bar{X}$ is good for $n \geq 30$; should only use approximation for $n < 30$ if population is not too different from a normal distribution.

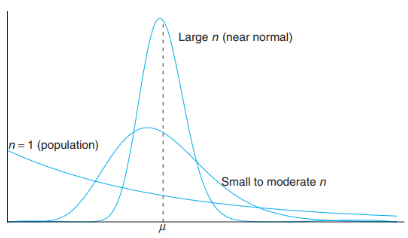


Figure 8.1: Illustration of the Central Limit Theorem (distribution of $\bar{X}$ for $n = 1$, moderate n, and large n).

- Note that the standard deviation of $\bar{X}$ is $\sigma/\sqrt{n}$. Also, as n increases, the variance of the distribution decreases at a rate of $\sqrt{n}$.

Sample Variance Distribution

- Recall:
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
- Let
$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Then χ^2 has a chi-squared distribution with $v = n-1$, where v is known as the **degrees of freedom** (no. of independent pieces of information).

- Suppose a known population mean μ:
$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$
- This is a chi-squared distribution with $v = n$, one less degree of freedom because one DOF is lost when estimating μ.

IX.III t-Distribution

- We use a t-distribution when the variance of the population is not known.
- Consider the statistic:
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$
with
$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$
- Note that if $n \geq 30$, S is close to σ and T follows normal, but if $n < 30$, the **t-distribution is much more accurate.**
- We can also write the statistic as:
$$T = \frac{(\bar{X} - \mu)/\frac{\sigma}{\sqrt{n}}}{\frac{S}{\sigma}} = \frac{Z}{\sqrt{\frac{V}{n-1}}}$$
where Z has the standard normal distribution and V has chi-squared distribution with $n-1$ DOF.
- t-distribution with v DOF is given by:
$$h(t) = \frac{\Gamma[\frac{(v+1)}{2}]}{\Gamma[\frac{v}{2}]\sqrt{\pi v}}(1 + \frac{t^2}{v})^{-\frac{v+1}{2}}$$
- We expect more variability with a t-distribution because we do not know σ exactly; instead we use S as an estimate.

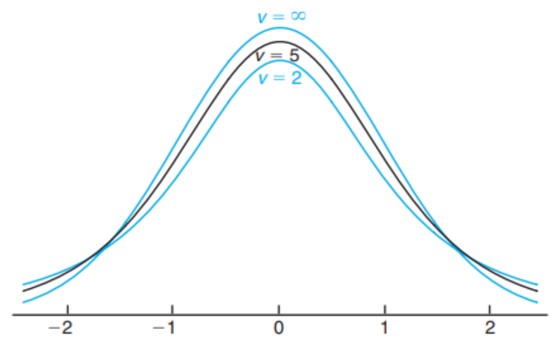


Figure 8.8: The t-distribution curves for $v = 2, 5$, and ∞.

- The t-distribution is used for problems that deal with inference about the population mean (μ) or in problems that involve comparative samples. It requires that $X_1 ... X_n$ be normal.
- It does not relate to Central Limit Theorem.

IX.IV Quantile

- Given a sample $x_1 ... x_n$, the **quantile**, $q(f)$, is value for which a specified fraction f of the data values is less than or equal to q(f).
- **Quantile plot:** $q(f)$ versus f.
$$f_i = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$$
- To sketch, for each data point $i = 1...n$, plot f_i for every x_i (where the sample data is in increasing order).
- **Sample median** — $q(0.5)$; **lower quartile** — $q(0.25)$, 25th percentile; **upper quartile** — $q(0.75)$, 75th percentile.
- Flat regions — data clusters; steep regions — data sparsity.

The quantile function has a close relation with the CDF:

- CDF F(x) is the chance that outcome is less or equal to x — $P(X \leq x)$.
- If F is continuous and strictly increasing, $q = F^{-1}(x)$ — swap axes on graph.

Normal Quantile-Quantile Plot

The normal quantile-quantile plot takes advantage of what is known about the quantiles of the normal distribution. The methodology involves a plot of the empirical quantiles recently discussed against the corresponding quantile of the normal distribution. Now, the expression for a quantile of an N(μ, σ) random variable is very complicated. However, a good approximation is given by:

$$q_{\mu,\sigma}(f) = \mu + \sigma\{4.91[f^{0.14} - (1-f)^{0.14}]\}$$

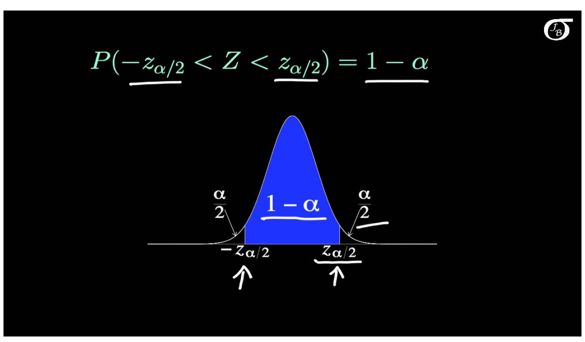
- If the curve of the normal QQ plot is straight, the data is roughly normal.
- Y-intercept is an estimate of population mean (μ), slope is an estimate of standard deviation (σ).

X Estimation

- **Unbiased Estimator** $\longrightarrow \mu_\theta = E[\theta] = \theta$
- Given P($\theta_L < \theta < \theta_U$) = 1 - α, $\theta \in (\theta_{\mathbb{L}}, \theta_{\mathbb{U}})$ is a (1 - α confidence interval.
- Estimating mean: $\bar{x}$ will be an accurate estimate of μ for large n.

  $$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

  for Z = $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.

- To estimate mean using a confidence interval:

  $$P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

- If $\bar{X}$ is used to estimate μ, we can be (1-α)(100%) confident that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.
- If $\bar{X}$ is used to estimate μ, we can be (1-α)(100%) confident that the error will not exceed error **e** when sample size n = $(\frac{z_{\alpha/2}\sigma}{e})^2$.
- **One-Sided Confidence Intervals:** same idea as two-sided, except only 1 error bound: $\theta_U = \bar{x} + z_\alpha\frac{\sigma}{\sqrt{n}}, \theta_L = \bar{x} - z_\alpha\frac{\sigma}{\sqrt{n}}$.
- When the σ is unknown, it is the same idea to estimate the μ, but with a T-distribution instead of a normal distribution.

  $$P(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}) = 1 - \alpha$$

X.I Prediction Intervals

- accounting for the variation of a future observation.

  $$z = \frac{x_0 - \bar{x}}{\sigma\sqrt{1+\frac{1}{n}}}$$

- The prediction interval is as follows:

  $$\bar{x} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}}$$

  where $z_{\alpha/2} = -\Phi^{-1}(\alpha/2)$

- If the x_0 falls out of the range of the prediction interval, it can be considered an **outlier**.

X.II Tolerance Limits

- Third type of confidence interval, concerned about long-range performance
- In a sample with μ and σ, the **tolerance interval** for the middle 95% observations of the population is $\mu \pm 1.96\sigma$.
- Instead, a (1 - γ)100% confidence can be asserted that the given limits contain at least the proportion 1 - α of the measurements.
- Choose interval $x \pm ks$ such that 100(1 - α)% of population is within limit.

X.III Two Samples

- Now we want to estimate $\mu_1 - \mu_2$.
- We know that the sampling distribution is normal with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$.
- As such we define RV Z as:

  $$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- By CLT, the distribution is approximately standard normal — n(x;0,1)

X.III.1 Unknown Variance (Equal)

- unknown variance < 30.

  $$T = \frac{(\bar{x_1} - \bar{x_2}) - \mu_1 - \mu_2}{s_p\sqrt{1/n_1 + 1/n_2}}$$

- Pooled estimate of variance (${S_p}^{\prime 2}$):

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

X.III.2 Unknown Variance (Different)

- If $\sigma_1 \neq \sigma_2$, use statistic:

  $$T' = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

  T' approximately has t-distribution with

  $$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}$$

  degrees of freedom.

X.IV Paired Observations

- Estimation procedure for the difference of 2 means when the samples are not independent and the variances are not necessarily equal.
- Rather, each homogeneous experimental unit receives both population conditions; as a result, each experimental unit has a pair. The two populations are "before" and "after," and the experimental unit is the individual. The setup is as follows:
- Paired samples (X_i, Y_i), i = 1, ..., n; we are interested in their **difference,** D_i**,** $X_i - Y_i$.
- Then, their variance is as follows:

  $$var(D_i) = var(X_i - Y_i) = \sigma_X^2 + \sigma_Y^2 - 2cov(X_i, Y_i)$$

- Helpful variance reduction (take advantage when possible)
- Apply usual CLT/t-dist confidence intervals to sample D_i:

  $$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$$

  where T = $\frac{\bar{D}-\mu_D}{S_d/\sqrt{n}}$ and $t_{\alpha/2}$ is a value of the t-dist with n-1 DOF.

X.V Estimating a Proportion

- Point estimator of proportion p in binomial experiment (Bernoulli process) is given by statistic $\hat{P} = \frac{X}{n}$ — X is no. of successes in n trials.
- By CLT (large n), $\hat{P}$ is approximately distributed normally w mean and variance:

  $$\mu_P = E[\hat{P}] = p; \sigma_{\hat{P}}^2 = \frac{pq}{n}$$

  $$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

  with Z = $\frac{\hat{P}-p}{\sqrt{pq/n}}$.

- If n is large, replace p with $\hat{p} = \frac{x}{n}$ in denominator:

  $$1-\alpha = P\left(\hat{P}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{P}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

- If n is small, solve for p with this fuckery:

  $$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} \pm \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}.$$

- Same with confidence intervals, if we want to be 100(1 - α)% sure that the error does not exceed e, we use sample size:

  $$n = \frac{z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{e^2} \ge \frac{z_{\alpha/2}^2}{4e^2}$$

X.VI Estimating the Variance

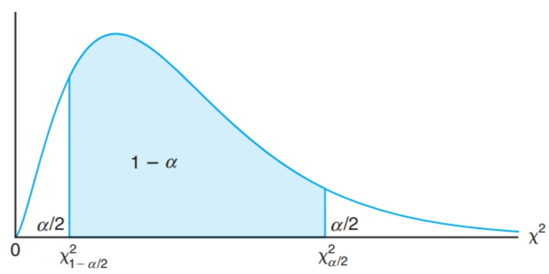
- An interval estimate of σ^2 can be established using the statistic:

  $$W^2 = \frac{(n-1)S^2}{\sigma^2}$$

  which has a chi-squared distribution.

- Chi-squared distribution is not symmetric, cannot apply the same logic to determine a confidence interval.

  $$P(\chi_{1-\alpha/2}^2 \le \frac{(n-1)S^2}{\sigma^2} \le \chi_{\alpha/2}^2) = 1 - \alpha$$



- To determine a confidence interval for σ^2:

  $$P(\frac{(n-1)S^2}{\chi_{\alpha/2}^2} \le \sigma^2 \le \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}) = 1 - \alpha$$

X.VII Maximum Likelihood Estimation

- Method of maximum likelihood is to maximize the likelihood function.
- **Likelihood Function:** RVs $X_1 ... X_n$, parameter θ, JPDF $f(x_1, ..., x_n; \theta)$

  $$\begin{aligned} L(x_1, ..., x_n; \theta) &= f(x_1, ..., x_n; \theta) \\ &= f(x_1; \theta) ... f(x_n; \theta) \\ &= \prod_{i=1}^{n} f(x_i; \theta) \end{aligned}$$

- the philosophy of maximum likelihood estimation evolves from the notion that the reasonable estimator of a parameter based on sample information is that parameter value that produces the largest probability of obtaining the sample.

Example: Poisson Distribution

- Poisson PMF:

  $$f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}$$

- Likelihood Function:

  $$L(x_1, ..., x_n; \mu) = \prod_{i=1}^{n} f(x_i; \mu) = \frac{e^{-n\mu}\mu^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

- Taking the logarithm, we can get rid of the product (log of a product becomes a sum).

  $$\ln L(x_1, ..., x_n; \mu) = -n\mu + \sum_{i=1}^{n} x_i \ln \mu - \ln \prod_{i=1}^{n} x_i!$$

- Maximize likelihood function by taking derivative wrt μ:

  $$\frac{\partial \ln L(x_1, ..., x_n; \mu)}{\partial \mu} = -n + \sum_{i=1}^{n} \frac{x_i}{\mu}$$

- Solve for $\hat{\mu}$ by setting derivative to 0:

  $$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}$$

- Since μ is the mean of the Poisson distribution, the sample average would certainly seem like a reasonable estimator.
- **Important property:**

  $$\prod_{i=1}^{n} e^{x_i} = e^{\sum_{i=1}^{n} x_i}$$

XI Hypotheses

- **Statistic Hypothesis:** assertion or conjecture concerning one or more populations.
- **Role of Probability in Hypothesis Testing:** decision procedure must include an awareness of the probability of a wrong conclusion; rejection of a hypothesis implies that the sample evidence refutes it.

Example:

A sample of 100 revealing 20 defective items is certainly evidence for rejection. Why? If, indeed, p = 0.10, the probability of obtaining 20 or more defectives is approximately 0.002. With the resulting small risk of a wrong conclusion, it would seem safe to reject the hypothesis that p = 0.10.

XI.I Null and Alternative Hypotheses

- **Null Hypothesis:** any hypothesis we wish to test and is denoted by H_0.
- In turn, the rejection of H_0 implies the acceptance of an **alternative hypothesis**, H_1, which represents the question to answer.
- Should arrive at one of two conclusions:
  1. reject H_0 in favor of H_1 bc sufficient evidence
  2. fail to reject H_0 bc of insufficient evidence

XI.II Error

- Rejection of the null hypothesis when it is true is called a **type I error**.
  - **example:** H_0 states $\mu = 68$, probability of type I error is:

    $$\alpha = P(\mu < 67) + P(\mu > 69)$$

  - if n > 30, assume normally distributed, use CLT centered around $\mu = 68$.
- Nonrejection of the null hypothesis when it is false is called a **type II error**.
  - **same example:** it is only necessary to consider the probability of not rejecting H_0 that $\mu = 68$ when the alternative $\mu = 70$ is true.
  - A type II error will result when the sample mean $\bar{x}$ falls between 67 and 69 when H1 is true.

    $$\beta = P(67 \leq \mu \leq 69)$$

    centered around $\mu = 70$.

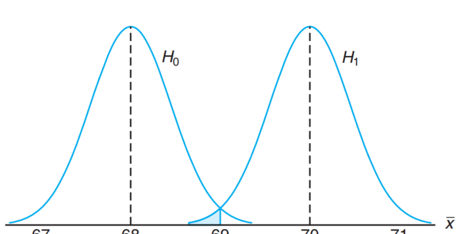


Figure 10.6: Probability of type II error for testing $\mu = 68$ versus $\mu = 70$.

- The probability of committing a type I error, denoted by α, is called **the level of significance**.
- The probability of committing a type II error, denoted by β, is impossible to compute unless we have a specific alternative hypothesis.
- Probability of committing both types of error can be reduced by increasing the sample size.
- The probability of committing a type II error increases rapidly when the true value of μ approaches, but is not equal to, the hypothesized value.
  - **still that example:** if the alternative hypothesis $\mu = 68.5$, we don't mind coming to the conclusion that $\mu = 68$.

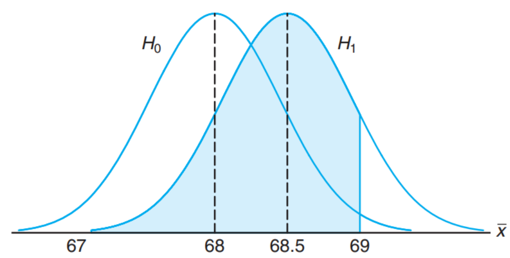


Figure 10.7: Type II error for testing $\mu = 68$ versus $\mu = 68.5$.

- **power of a test:** defined as 1 - β

XI.II.1 One or Two-Tailed Tests

- **One-sided:** $H_0 : \theta = \theta_0; H_1 : \theta <$ or $> \theta_0$.
- **Two-sided:** $H_0 : \theta = \theta_0; H_1 : \theta \neq \theta_0$.

XI.III P-Values

In testing hypotheses in which the test statistic is discrete, the critical region may be chosen arbitrarily and its size determined. If α is too large, it can be reduced by making an adjustment in the critical value.

- **P-value:** lowest level (of significance) at which the observed value of the test statistic is significant.

XI.III.1 P-Value Approach

1. State null and alternative hypotheses.
2. Choose an appropriate test statistic.
3. Compute the P-value based on the computed value of the test statistic.
4. Use judgment based on the P-value and knowledge of the scientific system.

XI.IV Single Mean

XI.IV.1 Variance Known (Z-Distribution)

- Simply put, non-rejection region for H_0 is:

  $$z \in [-z_{\alpha/2}, z_{\alpha/2}]$$

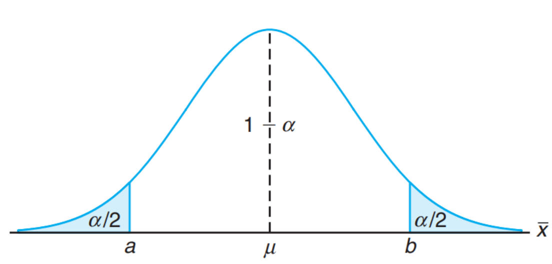


Figure 10.9: Critical region for the alternative hypothesis $\mu \neq \mu_0$.

- **P-Value Calculation:**

$$\begin{aligned} P &= P(\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| > |z|) \\ &= P(|Z| > |z|) \\ &= 2P(Z > |z|) \end{aligned}$$

XI.IV.2 Variance Unknown (T-Distribution)

- Same idea as above but use t-distribution instead of CLT.

XI.V Two Samples: Two Means

- Writing the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \mu_1 - \mu_2 = d_0 \\ H_1 &: \mu_1 - \mu_2 \neq d_0 \end{aligned}$$

- Use statistic z:

$$z = \frac{(\bar{x}_1 - \bar{x}_2 - d_0)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

- Non-rejection region is still $z \in [-z_{\alpha/2}, z_{\alpha/2}]$ for two-tailed tests.
- For one-tailed test, use upper or lower limit of z depending on the direction of the less/greater than sign in the alternative hypothesis.

XI.V.1 Unknown, Equal Variances

- Use pooled t-test, recall S_p^2:

$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}} \\ s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \end{aligned}$$

- Non-rejection range of H_0 is:

$$t \in [-t_{\alpha/2,n_1+n_2-2}, t_{\alpha/2,n_1+n_2-2}]$$

XI.V.2 Unknown, Non-equal Variances

- Use T': just look at Section 10.3.2 I'm too lazy to rewrite the formulas.
- Non-rejection range of H_0 is:

$$t' \in [-t_{\alpha/2,v}, t_{\alpha/2,v}]$$

XI.V.3 Paired Observations

Just look down I don't even get this, just know that $\bar{d}$ is the same idea as section 10.4.

XI.VI Choice of Sample Size

Table 10.3: Tests Concerning Means

H_0	Value of Test Statistic	H_1	Critical Region
$\mu = \mu_0$	$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$; σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$; $v = n-1$, σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \frac{(\bar{x}_1-\bar{x}_2)-d_0}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$; σ_1 and σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \frac{(\bar{x}_1-\bar{x}_2)-d_0}{s_p\sqrt{1/n_1+1/n_2}}$; $v = n_1+n_2-2$, $\sigma_1 = \sigma_2$ but unknown, $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \frac{(\bar{x}_1-\bar{x}_2)-d_0}{\sqrt{s_1^2/n_1+s_2^2/n_2}}$; $v = \frac{(s_1^2/n_1+s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$, $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ or $t' > t_{\alpha/2}$
$\mu_D = d_0$ paired observations	$t = \frac{\bar{d}-d_0}{s_d/\sqrt{n}}$; $v = n-1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

XI.VI.1 One Sample

- For a specific alternative hypothesis, say $\mu = \mu_0 + \delta$, the power of the test is:

$$\begin{aligned} 1-\beta &= P(\bar{X} > a \text{ when } \mu = \mu_0 + \delta) \\ \beta &= P(\bar{X} < a \text{ when } \mu = \mu_0 + \delta) \\ \beta &= P\Big[\frac{\bar{X}-(\mu_0+\delta)}{\sigma/\sqrt{n}} < \frac{a-(\mu_0+\delta)}{\sigma/\sqrt{n}}\Big] \end{aligned}$$

- Take this as a normal distribution with $\mu = \mu_0 + \delta$:

$$\beta = P(Z < z_\alpha - \frac{\delta}{\sigma/\sqrt{n}})$$
$$-z_\beta = z_\alpha - \frac{\delta\sqrt{n}}{\sigma}$$

- We can then conclude that we choose the sample size to be for a **one-sided test** (works both ways):

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{\delta^2}$$

- For a **two-sided test**:

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{\delta^2}$$

XI.VI.2 Two Samples

- Same idea as one sample: new Z is defined as this:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (d_0 + \delta)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

- We can conclude that for **two-tailed test:**

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2}$$

- **One-tailed:**

$$n = \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\delta^2}$$

XI.VII Variances: One and Two Samples

- Writing the null and alternative hypotheses:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2$$

- Use chi-squared statistic to base our decision to reject H_0 or not:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

- Non-rejection region (α, **two-tailed**) of H_0:

$$\chi^2 \in [\chi^2_{1-\alpha/2}, \chi^2_{\alpha/2}]$$

- Non-rejection region (α, **one-tailed**, $\sigma^2 < \sigma_0^2$) of H_0:

$$\chi^2 \geq \chi^2_{1-\alpha}$$

- Non-rejection region (α, **one-tailed**, $\sigma^2 > \sigma_0^2$) of H_0:

$$\chi^2 \leq \chi^2_\alpha$$