ST5201: Basic Statistical Theory
Chapter 1-9: Review

CHOI Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

Thursday 9^{th} November, 2017

Outline

- Announcement
- Some information about final
- Review

Announcement

- Assignment 4 released.
  - Due on 14th of November by 9 pm

Information about final

- Does it cover the things before midterm?
  - Yes. Chapter 1 to 9 are all subject to the final exam.
- Will the tables be provided?
  - Yes.
- How can I review my midterm paper?
  - You can email me (stachoiy@nus.edu.sg) to make an appointment. The last day you can review your paper is 15th of November.

- Sample Space
  - The set that contain all the possible outcomes
  - Can be finite or infinite
- Probability Measure
  - $P(\Omega) = 1$
  - $0 \leq P(A) \leq 1$
  - $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, $\{A_i\}$ are disjoint
- Apply the properties of probability measure to random variables:
  - For discrete r.v.'s, the *summation* of PMF over all the possible values is 1.
  - For cont. r.v.'s, the *integral* of PDF over the support is 1.
  - For discrete r.v.'s, $0 \leq P(X = x_i) \leq 1$ for any x_i
  - For cont. r.v.'s, $f(x) \geq 0$ for any $x \Leftarrow f(x)$ can be larger than 1.
  - $P(X \in A)$ can be calculated by summation (for discrete r.v.) or integral (for cont. r.v.)

Probability

- Sample Spaces With Equally-likely Outcomes
  - Identify the cardinality of the sample space n
  - Identify the cardinality of event of interest m
  - Probability: m/n
- Generally used counting methods:
  - Sampling with replacement: n^r permutations
  - Sampling without replacement: ${}_nP_r$ permutations
  - Sampling without replacement: ${}_nC_r$ combinations
- Conditional Probability
  - $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  - For discrete r.v.'s, $p_{X|Y}(x|y) = \frac{P(X=x, Y=y)}{P(Y=y)}$
  - For cont. r.v.'s, $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$

- Multiplication Law
  - Events: $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
  - Discrete r.v.'s: $p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x)$
  - Cont. r.v.'s: $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$
- Law of Total Probability
  - Events: $P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$, where B_i is a division of Ω
  - Discrete r.v.'s: $p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$
  - Cont. r.v.'s: $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$
- Bayes Rule
  - Events: $P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$, where B_i is a division of Ω
  - r.v.'s: not referred
- Independence
  - Events: $P(A \cap B) = P(A)P(B)$
  - r.v.'s: will be referred later

Random Variables

- Random Variable: a function from *Sample Space* to *Real Numbers*
  - A function of a random variable is also a random variable
  - Example: $X|Y=y$ is a r.v.; $E(X|Y)$ is a r.v.
  - However, $E(X|Y=y)$ is a constant
- Discrete r.v.'s: PMF/CDF/MGF
  - Bernoulli r.v.: 2 outcomes, parameter p, mean p, variance pq
  - Binomial r.v.: $n+1$ outcomes, parameters n and p, mean np, variance npq; can be viewed as summation of n Bernoulli r.v.'s
  - Geometric r.v.: $\Omega = \{1, 2, 3, \cdots\}$, parameter p, mean $1/p$, variance $(1-p)/p^2$; number of trials until first success
  - Negative Binomial r.v.: $\Omega = \{r, r+1, r+2, \cdots\}$, parameter r, p
  - Hypergeometric r.v.
  - Possion r.v.: $\Omega = \{0, 1, 2, 3, \cdots\}$, parameter λ, mean λ, variance λ variance; related to Poisson Process
  - For 2 indept Poisson r.v.'s., $X + Y \sim Pois(\lambda_x + \lambda_y)$

Random Variables

- Continuous r.v.'s
  - Characterization: PDF/CDF/MGF
  - Difference between PDF and PMF: 1. To calculate the probability of an event, we take integral of PDF and summation of PMF; 2. PDF $f(x)$ means $P(X \in (x, x+\Delta)) \approx f(x)\Delta$, and PMF $p(x)$ means $P(X = x) = p(x)$; 3. Hence, $f(x)$ can be larger than 1, but $p(x)$ cannot

- Examples:
  - Uniform r.v.: parameter a and b, mean $(a+b)/2$, variance $(b-a)^2/12$
  - Exponential r.v.: parameter λ, mean $1/\lambda$, variance $1/\lambda^2$
  - Gamma r.v.
  - Beta r.v.
  - Normal r.v.: parameter μ and σ^2, mean μ, variance σ^2 Use Z-table to check probabilities and quantiles

- Functions of a r.v., where $Y = g(X)$
  - Find CDF of Y with $F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$, which is an event about X; then figure out the PDF by derivation
  - Change-of-Variable Technique

Joint Distributions

- Joint dist of 2 discrete r.v.'s
  - Joint PMF: $p(x, y) = P(X = x, Y = y)$
  - Joint Probability for any set C: $P((X, Y) \in C) = \sum_{(x,y)\in C} p(x, y)$
  - Marginal pmf: $p_X(x) = P(X = x) = \sum_y p(x, y)$
- Joint dist. of 2 cont. r.v.'s
  - Joint PDF: integrable function $f(x, y)$; integration over $\mathbb{R}^2$ is 1
  - Joint Probability for any set C: $P((X, Y) \in C) = \int_C \int f(x, y)\, dx\, dy$
  - Marginal pdf: $f_X(x) = \int_y f(x, y)$
  - Generalization to more than 2 r.v.'s
- Difficulty here $P((X, Y) \in C)$:
  - Figure out the region to integrate
  - According to the region, figure out the limits for x and y
  - Integration
  - Example: Let (X, Y) be uniformly distributed over a region C (figure representation), what is $f_{X,Y}$?

Independence, & Conditional Prob

- Conditional Dist.
  - Given $Y = y$, $X|Y = y$ is a new r.v.
  - $X|Y = y$ has its own pdf/pmf, we want to figure that out
  - $p_{X|Y}(x|y) = p_{X,Y}(x, y)/p_Y(y)$, $f_{X|Y}(x|y) = f_{X,Y}(x, y)/f_Y(y)$
- Independence
  - $F(x, y) = F(x)F(y)$ for any x and y, or $f(x, y) = f(x)f(y)/p(x, y) = p(x)p(y)$.
  - Shortcut to show independence: If $f(x, y)$ can be written as the product of a function about x and a function about y, i.e. $f(x, y) = g(x)h(y)$ for any function g and h, then X and Y are indept.
  - When X and Y are independent, $h(X)$ and $g(Y)$ are also independent
  - If X and Y are indept, then $X|Y = y$ is the same with $X \Rightarrow E(X|Y) = E(X)$.

Review: Independence

When X and Y are indept., then $h(X)$ and $g(Y)$ are also indept, as long as $h(\cdot)$ and $g(\cdot)$ are well-defined functions

- $h(x) = x^2$, $h(x) = e^x$, $g(y) = |y|$, $g(y) = 1/(y^2 + 1)$
- It is possible that $h(\cdot)$ and $g(\cdot)$ are not well defined, say, $h(x) = 1/x$ when $X \sim Ber(.5)$, $g(y) = \ln(Y)$ when $Y \sim N(0, 1)$. In this case, take care of the domain of functions and range of r.v.'s
- Another example is the inverse function. For example, when $3X$ and $2Y + 1$ are independent, then X and Y are independent, here $h(x) = x/3$, and $g(y) = (y - 1)/2$.
- However, we cannot say that when X and Y^2 are independent, X and Y are independent, when $X \sim N(0, 1)$, $Y \sim N(0, 1)$. It is impossible to find a function from Y^2 to Y (cannot define the sign of $\sqrt{Y^2}$). On the other hand, in this example, if $Y \sim Binomial(3, 0.4)$, $\sqrt{Y^2}$ will work, and the independence claim stands.

- Change-of-Variable Technique
  - 1 r.v: Let $Y = g(X)$, if $g(\cdot)$ is differentiable and strictly monotonic, then

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

    Note: Find the inverse function of $g(\cdot)$ first, and then take the derivative of this inverse function.
  - 2 r.v.'s: Let $U = g_1(X, Y)$, $V = g_2(X, Y)$, where g_1 and g_2 have cont. partial derivatives, and there exist h_1, h_2 so that $X = h_1(U, V)$, $Y = h_2(U, V)$ for all values X and Y, then

$$f_{U,V}(u, v) = \begin{cases} f_{X,Y}(h_1(u, v), h_2(u, v))|J^{-1}|, & (u, v) \in S^* \\ 0, & \text{otherwise,} \end{cases}$$

    where $J = det \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix}$.

    Note: The result should contain u and v only. Take care of S^*.

Functions of Jointly Dist. r.v.'s

- Convolution
  - Summation of 2 indept. r.v.'s
  - $f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$
  - Later, we introduced MGF to figure out the sum of 2 indept r.v.'s.
- Special case: order statistics
  - For n independent r.v.'s, order them by $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.
  - PDF:

$$f_{X_{(k)}}(x) = \frac{n!}{(n-k)!(k-1)!} f(x) F^{k-1}(x) [1 - F(x)]^{n-k}$$

  - Special case: $f_{X_{(1)}}(x) = n f(x) [1 - F(x)]^{n-1}$
  - Special case: $f_{X_{(n)}}(x) = n f(x) F^{n-1}(x)$

Expectation

- Definition: $E(X) = \sum_x xp(x)$, $E(X) = \int_{-\infty}^{\infty} xf(x)dx$
- Intuition: Long-run average
- Properties
  - $E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$;
  - $E(g(X,Y)) = \int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dx$;
  - For any linear function $h(X_1, X_2, ..., X_n)$, $E[h(X_1, X_2, ..., X_n)] = h(E(X_1), E(X_2), \ldots, E(X_n))$
  - Example:

$$E(aX + bY) = aE(X) + bE(Y),$$

    no matter whether X and Y are indep. or not.

- Markov's Inequality: non-negative r.v.,

$$P(X \geq a) \leq E(X)/a$$

Variance & Standard Deviation

- Motivation: Discribe the “spread” of the r.v.
- Definition: $E[(X-\mu)^2]$, where $\mu = E(X)$; $\mathrm{SD}(X) = +\sqrt{\mathrm{Var}(X)}$
- Properties:
  - $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$
  - If $Y = a + bX$, then $\mathrm{Var}(Y) = b^2\mathrm{Var}(X)$, $\mathrm{SD}(X) = |b|\mathrm{SD}(X)$
  - Only when X and Y are *independent*,

    $$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

    Otherwise $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2Cov(X,Y)$.
- Chebyshev’s Inequality: any r.v.,

$$P(|X-\mu| \geq a) \leq \mathrm{Var}(X)/a^2$$

Covariance & Correlation Coefficient

- Covariance
  - $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$
  - positively correlated/negatively correlated/uncorrelated
  - Remark: Independence $\Rightarrow$ Uncorrelated; Uncorrelated $\not\Rightarrow$ Independence
  - Properties: $\text{Var}(X) = Cov(X, X)$, $Cov(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$.

- Correlation coefficient
  - $Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$
  - Remark: $-1 \leq Corr(X, Y) \leq 1$.
  - Properties: $Corr(a + bX, c + dY) = sign(b)sign(d)Corr(X, Y)$

Conditional Expectation

- Conditional Expectation
  - $E[X|Y=y] = \sum_x x p_{X|Y}(x|y)$ for discrete r.v.;
    $E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$ for cont. r.v.
  - Interpretation: Note that $X|Y=y$ is a new r.v., $E(X|Y=y)$ is the expectation on this r.v.
- Law of Total Expectation
  - $E[X|Y]$ is a function of Y. A function of a r.v. is also a r.v..
  - $$E[E(X|Y)] = E(X)$$
  - Random sum:
    $E(\sum_{i=1}^{N} X_i) = E(E(\sum_{i=1}^{N} X_i|N)) = E(NE(X_1)) = E(N)E(X_1)$.

- Moment Generating Function
  - $M_X(t) = E(e^{tX})$: a function of t, not r.v.
  - Characterization of a r.v. (similar as CDF/PDF/PMF)
  - Calculate moments:

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)|_{t=0}$$

  - For indept. r.v.'s $X_1, X_2, ..., X_t$

$$M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

  - Linear Transformation: $M_{aX+b}(t) = e^{bt} M_X(at)$.
  - With MGF, find out the summation of two indept Poisson r.v. is still Poisson r.v., the summation of two indept. normal r.v. is still normal r.v.
  - MGF for common distributions

MGF: examples

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \leq p \leq 1$	$p(1-p)^{x-1}$ $x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \leq p \leq 1$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

MGF: examples-cont'd

	Probability density function , $f(x)$	Moment generating function, $M(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters $(s, \lambda), \lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

Properties of Normal Distribution

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

- $E(X) = \mu$, $\text{Var}(X) = \sigma^2$, $\text{SD}(X) = \sigma$
- $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$
- If $Y = a + bX$, then
  - $Y \sim N(a + b\mu, b^2\sigma^2)$
  - $M_Y(t) = e^{(a+b\mu)t + b^2\sigma^2 t^2/2}$.

If X and Y are independent, $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, then

- $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Suppose (X, Y) is a bivariate normal vector with parameters μ_X, μ_Y, σ_X, σ_Y, ρ, then

- $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$
- $Corr(X, Y) = \rho$, $Cov(X, Y) = \rho\sigma_X\sigma_Y$
- $X|Y = y \sim N(\mu_X + \rho\sigma_X(y - \mu_Y)/\sigma_Y, (1 - \rho^2)\sigma_X^2)$
  $Y|X = x \sim N(\mu_Y + \rho\sigma_Y(x - \mu_X)/\sigma_X, (1 - \rho^2)\sigma_Y^2)$
- $E(X|Y) \sim N(\mu_X, \rho^2\sigma_X^2)$, $E(Y|X) \sim N(\mu_Y, \rho^2\sigma_Y^2)$

- Three types of convergence: a sequence of r.v.'s
  - Convergence in distribution: point-wise convergence of CDF $F_{X_n}(x) \to F_X(x)$, for any x where $F_X(x)$ is cont.
  - Convergence in probability: relevant with sample space $P(|X_n - X| \leq \epsilon) \to 0$, as $n \to \infty$
  - Almost sure convergence: relevant with sample space; point-wise convergence for r.v.'s X_n ("points" means "outcomes") $P(X_n(\omega) \to X(\omega)) = 1$
  - Property: a.s. convergence $\Rightarrow$ Convergence in Prob $\Rightarrow$ Convergence in Dist.

Properties of convergence:

- If $X_n \xrightarrow{P} \mu$ and $g(\cdot)$ is a continuous function, then $g(X_n) \xrightarrow{P} g(\mu)$
- If $X_n \xrightarrow{d} \mu$ and $g(\cdot)$ is a continuous function, then $g(X_n) \xrightarrow{d} g(\mu)$

- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$
- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n Y_n \xrightarrow{P} XY$

- Slutsky's Theorem: If $X_n \xrightarrow{d} X$ in distribution and $Y_n \xrightarrow{P} a$, a is a constant, then
  - $Y_n X_n \xrightarrow{d} aX$
  - $Y_n + X_n \xrightarrow{d} X + a$

- Law of Large Number (LLN)
  - Conditions: independent, share μ and σ
  - Results: $\bar{X}_n \overset{a.s./P}{\Longrightarrow} \mu$
  - Application: Monte Carlo method for integration
- Central Limit Theorem (CLT)
  - Conditions: i.i.d, μ, σ, and mgf exists in a neighborhood of 0
  - Results: $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \xrightarrow{d} Z \Leftrightarrow \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} Z$, $Z \sim N(0,1)$
  - Applications in many fields, especially for unknown distribution.
  - Normal approximations of Poisson distribution

- Sample mean: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
  - With LLN, $\bar{X} \to E(X)$; with CLT, the limiting dist. of sample mean is clear
- Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - $E(S^2) = \text{Var}(X)$; When the variance of S^2 converges to 0 (which usually holds), $S^2 \to \text{Var}(X)$
- Sample standard deviation: $S = \sqrt{S^2}$
  - If $S^2 \to \text{Var}(X)$, then $S \to SD(X)$; S is biased
  - $\sqrt{n}\frac{\bar{X}-\mu}{S} \to N(0,1)$
- Delta Method
  - If $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$, and the function g has nonzero derivative at θ, then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2\sigma^2)$$

  - Generalization of CLT and asymptotic normality of MLE

χ^2-distribution and t-distribution

- χ^2-distribution
  - If $X_1, \cdots, X_n \overset{i.i.d.}{\sim} N(0,1)$, then $\sum_{i=1}^n X_i^2 \sim \chi_n^2$
  - Expectation: n; Variance: $2n$; mgf: $(1-2t)^{-n/2}$, $t < 1/2$
  - If $X \sim \chi_n^2$, indept. with $Y \sim \chi_m^2$, $X + Y \sim \chi_{n+m}^2$
  - If the data is normal distributed, $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$
- t-distribution
  - If $X \sim N(0,1)$, indept. with $Y \sim \chi_n^2$, then $\frac{X}{\sqrt{Y/n}} \sim t_n$
  - Expectation: 0 when $n > 1$; Variance: $n/(n-2)$ when $n > 2$; mgf does not exist
  - If the data is normal distributed, $\sqrt{n}\frac{\bar{X}-\mu}{S} \sim t_{n-1}$
  - When the df $n \geq 30$, t_n is very close to standard normal distribution

- Use χ^2-table and t-table: Find the df first, and then identify the quantile
- Use χ^2 distribution and t-distribution to construct confidence interval when data is normal distributed

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

$$\sigma^2 : \left(\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}, \quad \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)} \right)$$

Parameter Estimation

- Parametric model: Estimate the parameter and the whole distribution is known
  - Remark: for any value in the parameter space, there is a corresponding PDF/PMF
  - For example: $f(x) = \begin{cases} c, & 0 < x, y < 1 \\ 0, & \text{otherwise} \end{cases}$ is not parametric distribution, as c can be found as a *fixed* value.
    $f(x) = \begin{cases} 4c, & 0 < x < 1/2 \\ 4(1-c)/3, & 1/2 < x < 1 \\ 0, & \text{otherwise} \end{cases}$ is a parametric distribution with parameter space $(0, 1)$
- Estimators
  - An estimator is a function of $X_1, \cdots, X_n$, which is also a r.v.

Estimators

- Method of Moments:
  - For K unknown parameters, calculate K lower order moments
  - Express the parameters with these moments
  - Substitute the moments with sample moments to have the estimator
  - Consistent
- Maximum Likelihood Estimates
  - Likelihood function: $L_n(\theta) = f(X_1, X_2, \cdots, X_n|\theta)$
  - MLE: maximizer of $L_n(\theta)$
  - Method 1: figure out the maximizer of $L_n(\theta)$ directly
  - Method 2: Find log-likelihood $l_n(\theta) = \ln L_n(\theta)$, calculate the derivative of $l_n(\theta)$ and solve the equation $l_n'(\theta) = 0$. Verify the solution satisfies that $l_n(\theta)$ achieves the maximum
  - Consistent
  - Asymptotic Normality: Fisher information $I(\theta) = -E(l''(\theta))$ under the smoothness condition,

$$\sqrt{nI(\theta)}(\hat{\theta} - \theta) \to N(0, 1)$$

- Assessment of Estimators
  - Sampling distribution: distribution of $\hat{\theta}_n$
  - Consistency: $\hat{\theta} \to \theta$ in probability. Remark: MLE and MM are both consistent
  - Unbias: $E(\hat{\theta}_n) = \theta$, for any n
  - Variance: $\mathrm{Var}(\hat{\theta}_n)$ is small
  - Mean Squared Error: $\mathrm{MSE}(\hat{\theta}_n) = E(\hat{\theta}_n - \theta)^2 = \mathrm{Bias}^2 + \mathrm{Var}$
  - Remark: MSE converges to 0 $\Rightarrow \hat{\theta}_n$ is consistent
- Cramer-Rao Lower Bound
  - For any unbiased estimator $\hat{\theta}_n$,

$$\mathrm{Var}(\hat{\theta}_n) \geq 1/(nI(\theta)), \qquad \text{any } n.$$

  - Efficiency: $(nI(\theta))^{-1}/\mathrm{Var}(\hat{\theta}_n)$
  - If $\mathrm{Var}(\hat{\theta})_n = 1/(nI(\theta))$ ($\hat{\theta}_n$ has efficiency 1), then $\hat{\theta}_n$ is *efficient*.

Confidence Interval

- Confidence Interval
  - CI is a *random* interval (L, U)
  - $100(1-\alpha)\%$ CI for θ means that

$$P(\theta \in (L, U)) \geq 1 - \alpha,$$

    and $1-\alpha$ is called *confidence level*
  - A general set-up for (approximate) $100(1-\alpha)\%$ CI:

$$(\hat{\theta} - z_{\alpha/2}\frac{1}{\sqrt{nI(\theta)}}, \hat{\theta} + z_{\alpha/2}\frac{1}{\sqrt{nI(\theta)}}),$$

    where $\hat{\theta}$ is MLE, $z_{\alpha/2}$ is $\Phi^{-1}(1-\alpha/2)$, n is the sample size, $I(\theta)$ is Fisher information
  - Meaning: if we construct this interval for N times, about $(1-\alpha)N$ of these intervals contain θ

- **Elements in a hypothesis test**:
  - Null and alternative hypotheses, H_0 and H_a
  - Test statistic and testing criteria
  - Significance level
  - p value and interpretation

- We are expecting making some error, since no one knows the truth

$$\text{type I error} = \text{erroneous rejection of } H_0 \text{ while } H_0 \text{ is true.}$$
$$\text{type II error} = \text{erroneous retention of } H_0 \text{ while } H_1 \text{ is true.}$$

- Significance level $0 < \alpha < 1$: the probability of committing a type I error.

LRT for two simple hypotheses

Suppose that $H_0 : \theta = \theta_0$ and $H_a : \theta = \theta_1$ are simple hypotheses with $\theta_0, \theta_1 \in \Theta$. The likelihood-ratio test statistic is defined by

$$R = \frac{\text{lik}(\theta_0)}{\text{lik}(\theta_1)}$$

which is a function of the sample $x_1, \cdots, x_n$. A LRT with significance level $0 < \alpha < 1$ is a test that has a rejection region of the form $\{R \leq c\}$ where $c \geq 0$ is chosen so that $P(R \leq c | H_0) = \alpha$

Generalized Likelihood Ratio Test

The LR test statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_a : \theta \in \Theta_1 \equiv \Theta_0^c$ (i.e., $\Theta_0 \cup \Theta_1 = \Theta$) is defined

$$\Lambda = \frac{\max_{\theta \in \Theta_0} \text{lik}(\theta)}{\max_{\theta \in \Theta} \text{lik}(\theta)} = \frac{\max_{\theta \in \Theta_0} \text{lik}(\theta)}{\text{lik}(\hat{\theta})} \leq 1$$

which is a function of the sample $x_1, \cdots, x_n$, where $\hat{\theta}$ is the mle of θ. A generalized LRT with significance level $0 < \alpha < 1$ is a test that has a rejection region of the form $\{\Lambda \leq \lambda_0\}$ where $0 \leq \lambda_0 \leq 1$ is chosen so that $P(\Lambda \leq \lambda_0 | H_0) = \alpha$

The two-sided Z Test

The LRT with significance level α for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ for $N(\mu, \sigma^2)$ with σ known is a test that has a rejection region given by

$$\left\{ |\bar{X} - \mu_0| \geq Z(\alpha/2) \frac{\sigma}{\sqrt{n}} \right\}$$

The two-sided t Test

The LRT with significance level α for testing $H_0 : \mu = \mu_0$ against $H_a : \mu \neq \mu_0$ for $N(\mu, \sigma^2)$ with unknown σ is a test that has a rejection region (red in the below graph) given by

$$\left\{ |\bar{X} - \mu_0| \geq t_{n-1}(\alpha/2) \frac{S}{\sqrt{n}} \right\}$$

where $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X - \bar{X})^2$