Statistical Inference (MA862)

Lecture Slides

Topic 0: Introduction

Website

https://ayonganguly.github.io/ma682.html

Syllabus

- Review of probability theory
- Monte Carlo Simulation
- Point estimation
- Interval estimation
- Testing of hypotheses
- Linear regression
- Basic non-parametric tests.
- Bayesian Analysis
- Markov chain Monte Carlo

Reading Materials

- For Monte Carlo Methods:
  - `https://artowen.su.domains/mc/`
- For Parametric Inference:
  - V. K. Rohatgi and A. K. Md. E. Saleh, An Introduction to Probability and Statistics, Wiley
  - G. Casella and R. L. Berger, Statistical Inference, Duxbury Press
  - B. L. S. Prakasa Rao, A First Course in Probability and Statistics, World Scientific/Cambridge University Press India

Class Times

- In general:
  - Monday 3 pm to 3:55 pm
  - Tuesday 2 pm to 2:55 pm
  - Friday 4 pm to 4:55 pm
- Make-up class (if needed):
  - Thursday 5 pm to 5:55 pm

Exams and Grading Policy

- Will be informed.

Resource Persons

- Instructor:
  - Ayon Ganguly (Email: aganguly@iitg.ac.in)
- Teaching Assistant:
  - Aryan Bhambu (Email: a.bhambu@iitg.ac.in)

Probability

- Let Ω be a non-empty set.
- Let $\mathcal{F}$ be a collection of subsets of Ω such that
  1. $\emptyset \in \mathcal{F}$.
  2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
  3. If $\{A_i\}_{i \geq 1} \subset \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
- A function $P : \mathcal{F} \to [0, \infty)$ is called a probability if
  1. $P(\Omega) = 1$.
  2. If $\{A_i\}_{i \geq 1} \subset \mathcal{F}$ is a sequence of disjoint sets, then

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

- The triplet $(\Omega, \mathcal{F}, P)$ is called probability space.

Conditional Probability

- Let H be an event with $P(H) > 0$. For any arbitrary event A, the conditional probability of A given H is defined by

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

- (Theorem of total probability) Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Then for any event E,

$$P(E) = \sum_i P(E|E_i)P(E_i).$$

- (Bayes rule) Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Then for any event E with $P(E) > 0$,

$$P(E_j|E) = \frac{P(E|E_j)P(E_j)}{\sum_i P(E|E_i)P(E_i)} \quad j = 1,\ 2,\ \ldots.$$

Random Variable

- A function $X : \Omega \to \mathbb{R}$ is called a random variable if for all $x \in \mathbb{R}$,

$$X^{-1}(-\infty, x] = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

- The function $F : \mathbb{R} \to [0, 1]$ defined by $F(x) = P(X \leq x)$ is called the cumulative distribution function (CDF) of X.
- CDF has following properties.
  - F is non-decreasing.
  - F is right continuous.
  - $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
- $F(x-) = F(x) - P(X = x)$.

Discrete Random Variable

- A random variable X is said to have a discrete distribution if there exists an atmost countable set $S \subset \mathbb{R}$ such that $P(X \in S) = 1$.
- Let X be a discrete random variable. The function $f : \mathbb{R} \to [0, 1]$ defined by $f(x) = P(X = x)$ is called probability mass function (PMF).
- For a discrete random variable X,

$$F(x) = \sum_{u \leq x,\, u \in S} f(u).$$

Continuous Random Variable

- A random variable X is said to have a continuous distribution if there exists a non-negative function f on $\mathbb{R}$ such that

$$F(x) = \int_{-\infty}^{x} f(u)du \quad \text{for all} \quad x \in \mathbb{R}.$$

- f is called probability density function (PDF) of X.
- CDF is continuous.
- $P(X = x) = 0$ for all $x \in \mathbb{R}$.

Expectation

- Let X be a discrete random variable with PMF f. The expectation of X is defined by

$$E(X) = \sum_{x \in S} xf(x),$$

provided $\sum_{x \in S} |x| f(x) < \infty$.

- Let X be a continuous random variable with PDF f. The expectation of X is defined by

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

provided $\int_{-\infty}^{\infty} |x| f(x)dx < \infty$.

Jointly Distributed Random Variables

- A function $\boldsymbol{X} : \Omega \to \mathbb{R}^n$ is called a random vector if for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$\boldsymbol{X}^{-1}(-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_n] \in \mathcal{F}.$$

- For any random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, the joint cumulative distribution function (JCDF) is defined by

$$F_{\boldsymbol{X}}(x_1, \ldots, x_n) = P\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right),$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Discrete Random Vector (DRV)

- A random vector (X, Y) is said to have a discrete distribution if there exists an atmost countable set $S_{X,Y} \subset \mathbb{R}^2$ such that $P((X, Y) = (x, y)) > 0$ for all $(x, y) \in S_{X,Y}$ and $P((X, Y) \in S_{X,Y}) = 1$. $S_{X,Y}$ is called the support of (X, Y).
- Define a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ by

$$f_{X,Y}(x, y) = \begin{cases} P(X = x, Y = y) & \text{if } (x, y) \in S_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

The function $f_{X,Y}$ is called joint probability mass function (JPMF) of the DRV (X, Y).

Expectation of Function of DRV

- Let $(X,\ Y)$ be a DRV with JPMF $f_{X,\,Y}$ and support $S_{X,\,Y}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$. Then the expectation of $h(X,\ Y)$ is defined by

$$E\left(h(X,\ Y)\right) = \sum_{(x,y)\in S_{X,\,Y}} h(x,\ y) f_{X,\,Y}(x,\ y),$$

provided $\displaystyle\sum_{(x,y)\in S_{X,\,Y}} |h(x,\ y)| f_{X,\,Y}(x,\ y) < \infty.$

Continuous Random Vector (CRV)

- A random vector (X, Y) is said to have a continuous distribution if there exists a non-negative integrable function $f_{X, Y} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t, s) ds dt$$

  for all $(x, y) \in \mathbb{R}^2$.
- The function $f_{X, Y}$ is called the joint probability density function (JPDF) of (X, Y).

Expectation of Function of CRV

- Let (X, Y) be a CRV with JPDF $f_{X,Y}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$. Then the expectation of $h(X, Y)$ is defined by

$$E\left(h(X, Y)\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f_{X,Y}(x, y) dx dy,$$

provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| f_{X,Y}(x, y) dx dy < \infty$.

Independent Random Variables

- The random variables $X_1, X_2, \ldots, X_n$ are said to be independent if

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

- For DRV/CRV (X, Y), the condition of independence is equivalent to

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \text{ for all } (x, y) \in \mathbb{R}^2.$$

- If X and Y are independent, then

$$E\left(g(X)h(Y)\right) = E\left(g(X)\right) E\left(h(Y)\right),$$

provided all the expectations exist.

Conditional Distribution for DRV

- Let (X, Y) be a DRV with JPMF $f_{X,Y}(\cdot,\cdot)$. Suppose the marginal PMF of Y is $f_Y(\cdot)$. The conditional PMF of X, given $Y = y$ is defined by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\,,$$

  provided $f_Y(y) > 0$.
- The conditional CDF of X given $Y = y$ is defined by

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{\{u \leq x : (u,y) \in S_{X,Y}\}} f_{X|Y}(u|y)\,.$$

  provided $f_Y(y) > 0$.

Conditional Expectation for DRV

- The conditional expectation of $h(X)$ given $Y = y$ is defined by

$$E(h(X)|Y = y) = \sum_{x:(x,y)\in S_{X,Y}} h(x) f_{X|Y}(x|y)\,,$$

provided it is absolutely summable.

Conditional Distribution for CRV

- Let $f_{X,Y}$ be the JPDF of (X, Y) and let f_Y be the marginal PDF of Y. If $f_Y(y) > 0$, then the conditional PDF of X given $Y = y$ is given by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} .$$

- The conditional expectation of $h(X)$ given $Y = y$ is defined for all values of y such that $f_Y(y) > 0$ and given by

$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y) dx,$$

provided it is absolutely integrable.

Computing Expectation by Conditioning

$$E(X) = EE(X|Y) = \begin{cases} \displaystyle\sum_y E(X|Y=y)P(Y=y) & \text{for } Y \text{ discrete} \\ \displaystyle\int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy & \text{for } Y \text{ continuous.} \end{cases}$$

Computing Probability by Conditioning

$$P(E) = \begin{cases} \sum_y P(E|Y=y)P(Y=y) & \text{for } Y \text{ discrete} \\ \int_{-\infty}^{\infty} P(E|Y=y)f_Y(y)dy & \text{for } Y \text{ continuous.} \end{cases}$$

Transformation for DRV

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a discrete random vector with JPMF $f_{\boldsymbol{X}}$ and support $S_{\boldsymbol{X}}$. Let $g_i : \mathbb{R}^n \to \mathbb{R}$ for all $i = 1, 2, \ldots, k$. Let $Y_i = g_i(\boldsymbol{X})$ for $i = 1, 2, \ldots, k$. Then $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$ is a discrete random vector with JPMF

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_k) = \begin{cases} \sum_{\boldsymbol{x} \in A_{\boldsymbol{y}}} f_{\boldsymbol{X}}(\boldsymbol{x}) & \text{if } (y_1, \ldots, y_k) \in S_{\boldsymbol{Y}} \\ 0 & \text{otherwise,} \end{cases}$$

where $A_{\boldsymbol{y}} = \{\boldsymbol{x} \in S_{\boldsymbol{X}} : g_i(\boldsymbol{x}) = y_i,\ i = 1, \ldots, k\}$ and $S_{\boldsymbol{Y}} = \{(g_1(\boldsymbol{x}), \ldots, g_k(\boldsymbol{x})) : \boldsymbol{x} \in S_{\boldsymbol{X}}\}$.

Transformation for CRV

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a continuous random vector with JPDF $f_{\boldsymbol{X}}$.

1. Let $y_i = g_i(\boldsymbol{x})$, $i = 1, 2, \ldots, n$ be $\mathbb{R}^n \to \mathbb{R}$ functions such that

$$\boldsymbol{y} = g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_n(\boldsymbol{x}))$$

   is one-to-one. That means that there exists the inverse transformation $x_i = h_i(\boldsymbol{y})$, $i = 1, 2, \ldots, n$ defined on the range of the transformation.

2. Assume that both the mapping and its' inverse are continuous. Assume that partial derivatives $\frac{\partial x_i}{\partial y_j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, exist and are continuous. Assume that the Jacobian of the inverse transformation $J \doteq \det\left(\frac{\partial x_i}{\partial y_j}\right)_{i,j=1,2,\ldots,n} \neq 0$ on the range of the transformation.

Then $\boldsymbol{Y} = (g_1(\boldsymbol{X}), \ldots, g_n(\boldsymbol{X}))$ is a continuous random vector with JPDF $f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(h_1(\boldsymbol{y}), \ldots, h_n(\boldsymbol{y}))|J|$.

Moment Generating Function

- Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector. The MGF of $\boldsymbol{X}$ at $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$ is defined by

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = E\left(\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right)$$

  provided the expectation exists in a neighborhood of origin $\mathbf{0} = (0, 0, \ldots, 0)$.
- Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two n-dimensional random vectors. Let $M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{\boldsymbol{Y}}(\boldsymbol{t})$ for all $\boldsymbol{t}$ in a neighborhood around $\mathbf{0}$, then $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$.

Modes of Convergence

- Almost sure convergence
- Convergence in probability
- Convergence in r-th mean
- Convergence in distribution

Almost Sure Convergence

- Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\Omega,\ \mathcal{F},\ P)$. We say that X_n converges almost surely or with probability (w.p.) 1 to a random variable X if

$$P\left(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}\right) = 1.$$

Convergence in Probability

- Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\Omega, \mathcal{F}, P)$. We say that X_n converges in probability to a random variable X if for any $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \to 0 \quad \text{as } n \to \infty.$$

Convergence in r-th Mean

- Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\Omega, \mathcal{F}, P)$. For $r = 1, 2, 3, \ldots$, we say that X_n converges in r^{th} mean to a random variable X if

$$E|X_n - X|^r \to 0 \quad \text{as } n \to \infty.$$

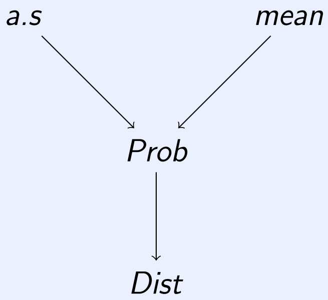
Convergence in Distribution

- Let $\{X_n\}$ be a sequence of RVs and X be a RV. Let $F_n(\cdot)$ and $F(\cdot)$ denote the CDF of X_n and X, respectively. We say that X_n converges in distribution to a random variable X if

$$F_n(x) \to F(x) \quad \text{as } n \to \infty$$

for all x where F is continuous.

Relationship among Modes



Strong Law of Large Numbers

- Let $\{X_n\}$ be a sequence of i.i.d. RVs with finite mean μ. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\{\overline{X}_n\}$ converges to μ almost surely.

Central Limit Theorem

- Let $\{X_n\}$ be a sequence of i.i.d. RVs with mean μ and variance $\sigma^2 < \infty$. Then, $\dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ converges to a standard normal random variable in distribution, *i.e.*, as $n \to \infty$,

$$P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq a\right) \to \Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Sampling Distribution Based on Normal

- A CRV X is said to have a Normal distribution or Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if the PDF of X is given by

$$f(x) = \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\} \quad \text{for all } x \in \mathbb{R}. \right.$$

- A CRV X is said to have a Gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\lambda > 0$ if the PDF of X is given by

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- For any positive integer n, a gamma distribution with $\alpha = \frac{n}{2}$ and $\lambda = \frac{1}{2}$ is also known as χ^2-distribution with n degrees of freedom.

Sampling Distribution Based on Normal

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(0,1)$ random variables. Then

$$\sum_{i=1}^{n} X_i^2 \sim \chi_n^2.$$

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Then $\overline{X}$ and S^2 are independently distributed and

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Sampling Distribution Based on Normal

- A CRV X is said to have a Student's t-distribution (or simply, t-distribution) with n degrees of freedom if the PDF of X is given by

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad \text{for } t \in \mathbb{R}.$$

- We will use the notation $X \sim t_n$ to denote that the RV X has a t-distribution with n degrees of freedom.

Sampling Distribution Based on Normal

- Let $X \sim N(0,1)$ and $Y \sim \chi^2_n$ be two independent RVs. Then the RV $T = \dfrac{X}{\sqrt{Y/n}} \sim t_n$.
- Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Then

$$\frac{\sqrt{n}\left(\overline{X} - \mu\right)}{S} \sim t_{n-1}.$$

Sampling Distribution Based on Normal

- A CRV X is said to have a F-distribution with n and m degrees of freedom if the PDF of X is given by

$$f(x) = \frac{1}{B\left(\frac{n}{2},\ \frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} \left(1 + \frac{n}{m}x\right)^{-\frac{n+m}{2}} \quad \text{for } x > 0.$$

- We will use the notation $X \sim F_{n,m}$ to denote that the RV X has a F-distribution with n and m degrees of freedom.

Sampling Distribution Based on Normal

- Let $X \sim \chi^2_n$ and $Y \sim \chi^2_m$ are two independent RVs. Then

$$F = \frac{X/n}{Y/m} = \frac{mX}{nY} \sim F_{n,\,m}.$$

- Let $X_1,\, X_2,\, \ldots,\, X_n \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1,\, Y_2,\, \ldots,\, Y_m \overset{i.i.d.}{\sim} N(\mu_2,\, \sigma_2^2)$. Also, assume that X_i's and Y_j's are independent. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$, and $S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$. Then

$$\frac{\sigma_2^2 S_X^2}{\sigma_1^2 S_Y^2} \sim F_{n-1,m-1}.$$