MAT283 Notes

Arjun Vardhan

†

Created: 20th February 2022
Last updated: 5th May 2022

1 Notes

- A **sample space** is the set of all possible outcomes of a random experiment.
- If a sample space contains an at most countable number of elements, it is said to be a discrete sample space.
- An **event** is a subset of a sample space.
- A subset E of sample space S is an event if it belongs to a collection $\mathbb{F}$ of subsets of S which satisfies the following:
  1. $S \in \mathbb{F}$.
  2. If $E \in \mathbb{F}$, then $E^c \in \mathbb{F}$.
  3. If $E_j \in \mathbb{F}$ for $i = 1, 2, 3...$, then $\bigcup_{i=1}^{\infty} E_i \in \mathbb{F}$.

  The collection $\mathbb{F}$ is then called an **event space**.
- Let S be the sample space of a random experiment. A **probability measure** $P : \mathbb{F} \to [0, 1]$ is a set function that assigns real values to events in S such that:
  1. $P(E) \geq 0$ for all $E \in \mathbb{F}$.
  2. $P(S) = 1$.
  3. If $E_1, E_2, ..., E_k, ...$ are mutually disjoint events in S, then $P(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$.
- $P(\phi) = 0$.
- $P(E^c) = 1 - P(E)$.
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.
- If A is an event in a discrete sample space S, then $P(A)$ is the sum of the probabilities of the individual outcomes comprising A.
- If an experiment can result in any one of n equally likely outcomes, and if m of these outcomes together constitute event A, then $P(A) = \frac{m}{n}$.
- The **conditional probability** of an event A, given that an event B has already occurred, is defined as: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, provided $P(B) > 0$.
- Two events A and B are called **independent** if and only if $P(A \cap B) = P(A)P(B)$.
- If two events are independent, then the occurrence or non-occurrence of one does not affect the probability of the other.
- If A and B are independent, then A and B^c are also independent.
- Two **mutually exclusive** (disjoint) events are always dependent.

- Let S be a set and let $\mathbb{P} = \{A_i\}_{i=1}^m$ be a collection of subsets of S. $\mathbb{P}$ is called a partition of S if $S = \bigcup_{i=1}^m A_i$ and if $A_i \cap A_j = \phi$ whenever $i \neq j$.
- **Law of Total Probability:** If the events $\{B_i\}_{i=1}^m$ constitute a partition of the sample space S and if $P(B_i) \neq 0$ for $i = 1, 2, 3..., m$, then for any event A, $P(A) = \sum_{i=1}^m P(B_i)P(A|B_i)$.
- **Baye's Theorem:** If the events $\{B_i\}_{i=1}^m$ constitute a partition of the sample space S and if $P(B_i) \neq 0$ for $i = 1, 2, 3, .., m$, then for any event A such that $P(A) \neq 0$, $P(B_k|A) = \dfrac{P(B_k)P(A|B_k)}{\sum_{i=1}^m P(B_i)P(A|B_i)}$, where $k = 1, 2, 3, ..., m$.
- Consider a random experiment with sample space S. A **random variable** X is a function from S to $\mathbb{R}$ such that for each interval I in $\mathbb{R}$, the set $\{s \in S : X(s) \in I\}$ is an event in S.
- The set $R_X = \{x \in \mathbb{R} : x = X(s), s \in S\}$ is called the space of the random variable X.
- If R_X is at most countable, then X is called a discrete random variable.
- Let X be a discrete random variable. The function $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = P(X = x)$ is called the **probability mass function** of X.
- f can serve as the pmf of a discrete random variable X if and only if $f(x) \geq 0$ for all x within its domain, and if $\sum_x f(x) = 1$.
- If X is a discrete RV, then the function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = P(X \leq x) = \displaystyle\sum_{t \leq x} f(t)$ for $-\infty < x < \infty$, where f is the pmf of X, is called the **cumulative distribution function** of X.
- F can serve as the cdf of discrete RV X if and only if $F(-\infty) = 0$, $F(\infty) = 1$, and if $a < b$, then $F(a) \leq F(b)$ for all $a, b \in \mathbb{R}$.
- If R_X consists of the values $x_1, x_2, ..., x_n$, where $x_1 < x_2 < ... < x_n$, then $f(x_1) = F(x_1)$, and $f(x_i) = F(x_i) - F(x_{i-1})$ for $i = 1, 2, 3, ..., n$.
- An RV X is said to be continuous if and only if there exists a function $f_X : \mathbb{R} \to \mathbb{R}$ such that $f_X(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and $P(a < x < b) = \int_a^b f_X(x)dx$ for any real a, b where $a \leq b$. $f_X(x)$ is called the **probability density function** of X.
- If X is a continuous RV, then $P(a \leq x \leq b) = P(a \leq x < b) = P(a < x \leq b) = P(a < x < b)$.
- If X is a continuous RV, then the function $F : \mathbb{R} \to \mathbb{R}$, defined by $F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$ for $-\infty < x < \infty$, is the cdf of X.
- If F is the cdf and f the pdf of X, then $\dfrac{d}{dx}F(x) = f(x)$.
- Let X be a random variable with space R_X and pdf/pmf f. The nth **moment** about the origin of X, denoted by $E(X^n)$, is defined as $\displaystyle\sum_{x \in R_X} x^n f(x)$ if X is discrete, and $\displaystyle\int_{-\infty}^{\infty} x^n f(x)dx$ if X is continuous, for $n = 1, 2, 3, ...$, provided the sum or integral converge absolutely.
- The **mean** or **expected value** of X, denoted $E(X)$ or μ_X, is defined as $\displaystyle\sum_{x \in R_X} xf(x)$ if X is discrete, and $\displaystyle\int_{-\infty}^{\infty} xf(x)dx$ if X is continuous, for $n = 1, 2, 3, ...$, provided the sum or integral converge absolutely. So the expected value is nothing but the first moment about the origin.
- Let X be an RV and let $Y = g(X)$. If X is discrete with pmf f, then $E(Y) = \displaystyle\sum_x g(x)f(x)$. If X is continuous with pdf f, then $E(Y) = \displaystyle\int_{-\infty}^{\infty} g(x)f(x)dx$.
- Let X be an RV, and let $a, b \in \mathbb{R}$. Then, $E(aX + b) = aE(X) + b$.

- Let X be an RV with mean μ_X. Its **variance** is defined as $\mathrm{Var}(X) = E((X-\mu_X)^2)$. The positive square root of the variance is called the **standard deviation** of X and denoted σ_X.
- $\mathrm{Var}(X) = E(X^2) - E(X)^2$.
- If $\mathrm{Var}(X)$ exists and $Y = a + bX$, then $\mathrm{Var}(Y) = b^2\mathrm{Var}(X)$.
- **Chebyshev's Inequality:** Let X be an RV with mean μ and standard deviation $\sigma > 0$. Then, $P(|X-\mu| < k\sigma) \geq 1 - \dfrac{1}{k^2}$ for any $k \in \mathbb{R}$, $k > 0$.
- Let X be an RV. A function $M : \mathbb{R} \to \mathbb{R}$ defined by $M(t) = E(e^{tX})$ is called the **moment generating function** of X if this expected value exists for all $t \in (-h, h)$ for some $h > 0$.
- If X is discrete, then $M(t) = \displaystyle\sum_{x \in R_X} e^{tx} f(x)$. If X is continuous, then $M(t) = \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)dx$.
- A discrete RV X is said to have a **Discrete Uniform distribution** if and only if its pmf is of the form $f(x) = \frac{1}{k}$, where $R_X = \{x_1, x_2, ..., x_k\}$ and $x_i \neq x_j$ for $i \neq j$. This distribution represents a random experiment with a finite number of equally likely outcomes.
- A discrete RV X is said to have a **Bernoulli distribution** with parameter p if and only if its pmf is of the form $f(x) = p^x(1-p)^{1-x}$, where $x = 0$ or $x = 1$. If a random experiment has only two possible outcomes, success and failure, with probabilities p and $1-p$ respectively, then the random variable representing the number of successes has a Bernoulli distribution. Such an experiment is referred to as a Bernoulli trial.
- If X is a Bernoulli RV with parameter p, then $E(X) = p$, $\mathrm{Var}(X) = p(1-p)$ and $M_X(t) = (1-p) + pe^t$. All its moments about the origin are equal to p.
- A discrete RV X is said to have a **Binomial distribution** with parameters p and n if and only if its pmf is of the form $f(x) = \binom{n}{x}p^x(1-p)^{n-x}$, where $x = 0, 1, 2, ..., n$. In a random experiment consisting of n Bernoulli trials, this RV represents the total number of successes.
- If X is a Binomial RV, then $E(X) = np$, $\mathrm{Var}(X) = np(1-p)$ and $M_X(t) = ((1-p) + pe^t)^n$.
- A discrete RV X is said to have a **Geometric distribution** with parameter p if and only if its pmf is of the form $f(x) = (1-p)^{x-1}p$, where $x \in \mathbb{N}$. In a random experiment consisting of an infinite sequence of Bernoulli trials, this RV represents the number of the trial on which the first success occurs.
- If X is a Geometric RV, then $E(X) = \dfrac{1}{p}$, $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$ and $M_X(t) = \dfrac{pe^t}{1-(1-p)e^t}$ if $t < \log(1-p)$.
- A discrete RV X is said to have a **Negative Binomial** or **Pascal distribution** with parameters p and r if and only if its pmf is of the form $f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$, where $x \in \mathbb{N}$. In a random experiment consisting of an infinite sequence of Bernoulli trials, this RV represents the number of the trial on which the rth success occurs.
- If X is a Negative Binomial RV, then $E(X) = \dfrac{pr}{1-p}$, $\mathrm{Var}(X) = \dfrac{pr}{(1-p)^2}$ and $M_X(t) = \left(\dfrac{1-p}{1-pe^t}\right)^r$ for $t < -\log p$.
- A discrete RV X is said to have a **Poisson distribution** with parameter $\lambda > 0$ if and only if its pmf is of the form $f(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, where $x \in \mathbb{N}$. It can be used to approximate the Binomial RV when n is very large and p is very small.
- If X is a Poisson RV, then $E(X) = \lambda$, $\mathrm{Var}(X) = \lambda$, and $M_X(t) = e^{\lambda(e^t-1)}$.
- A continuous RV X is said to have a **Uniform distribution** on the interval $[a, b]$ if and only if its pdf is of the form $f(x) = \dfrac{1}{b-a}$, where $a \leq x \leq b$ and $a, b \in \mathbb{R}$.

- If X is a Uniform RV on $[a, b]$, then $E(X) = \dfrac{b+a}{2}$, $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$, and $M_X(t) = 1$ if $x = 0$ and $M_X(t) = \dfrac{e^{tb} - e^{ta}}{t(b-a)}$ if $x \neq 0$.
- A continuous RV X is said to have an **Exponential distribution** with parameter $\theta > 0$ if and only if its pdf is of the form $f(x) = \dfrac{1}{\theta} e^{-\frac{x}{\theta}}$ if $x > 0$ and $f(x) = 0$ if $x \leq 0$.
- If X is an Exponential RV, then $E(X) = \dfrac{1}{\theta}$ and $\mathrm{Var}(X) = \dfrac{1}{\theta^2}$, and $M_X(t) = \dfrac{\theta}{\theta - t}$ for $t < \theta$.
- A continuous RV X is said to have a **Normal** or **Gaussian distribution** with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if and only if its pdf is of the form $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, where $-\infty < x < \infty$. Here, $f(\mu - x) = f(\mu + x)$. f has a maximum at $x = \mu$.
- If X is a Normal RV, then $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$ and $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.
- A Normal RV X is said to be **Standard Normal** RV if $\mu = 0$ and $\sigma = 1$. Its pdf is given by $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, where $-\infty < x < \infty$.
- If X is a Normal RV with parameters μ and σ, then $Z = \frac{X-\mu}{\sigma}$ is a Standard Normal RV.
- The **gamma function**, denoted $\Gamma(z)$, is defined as $\Gamma(z) = \displaystyle\int_{-\infty}^{\infty} x^{z-1} e^{-x} dx$, where $z \in \mathbb{R}$, $z > 0$.
- $\Gamma(1) = 1$ and $\Gamma(n) = n!$ for all $n \in \mathbb{N}$.
- $\Gamma(z)$ satisfies the functional equation $\Gamma(z) = (z-1)\Gamma(z-1)$ for all $z \in \mathbb{R}$, $z > 1$.
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(-\frac{1}{2}) = -2\sqrt{\pi}$.
- A continuous RV X is said to have a **Gamma distribution** with parameters $\alpha > 0$ and $\theta > 0$ if and only if its pdf is of the form $f(x) = \dfrac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}$.
- If X is a Gamma RV with $\alpha = 1$, then X is an Exponential RV.
- If X is a Gamma RV, then $E(X) = \theta\alpha$, $\mathrm{Var}(X) = \theta^2\alpha$ and $M_X(t) = \left(\dfrac{1}{1-\theta t}\right)^\alpha$, if $t < \frac{1}{\theta}$.
- Let α, β be any two positive real numbers. The **beta function**, denoted $B(\alpha, \beta)$, is defined as $B(\alpha, \beta) = \displaystyle\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}$.
- $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta}$.
- $B(\alpha, \beta) = B(\beta, \alpha)$.
- A continuous RV X is said to have a **Beta distribution** with parameters $\alpha, \beta > 0$ if and only if its pdf is of the form $f(x) = \dfrac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$ if $0 < x < 1$ and $f(x) = 0$ otherwise.
- If X is a Beta RV with $\alpha = \beta = 1$, then X is a Uniform RV.
- If X is a Beta RV, then $E(X) = \dfrac{\alpha}{\alpha + \beta}$, $\mathrm{Var}(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.
- A discrete **bivariate** RV, (X, Y), is an ordered pair of discrete RVs. Its pmf $f : R_X \times R_Y \to \mathbb{R}$, called the **joint pmf** of X and Y, is given by $f(x, y) = P(X = x, Y = y)$.
- Let X, Y be discrete RVs with joint pmf f. The **marginal pmf** of X is defined by $f_X(x) = \sum_{y \in R_Y} f(x, y)$. Similary, $f_Y(y) = \sum_{x \in R_X} f(x, y)$.

- Let X, Y be discrete RVs with joint pmf f. The joint cdf of X and Y is a function $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $F(x, y) = P(X \leq x, Y \leq y) = \displaystyle\sum_{s \leq x} \sum_{t \leq y} f(s, t)$.
- A bivariate RV (X, Y) is said to be continuous if there exists a function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that $f(x, y) > 0$, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$ and for any subset $A \subseteq \mathbb{R} \times \mathbb{R}$, $P((X, Y) \in A) = \displaystyle\int\int_A f(x, y)dxdy$. f is the **joint pdf** of X and Y.
- Let (X, Y) be a continuous bivariate RV, and let f be its joint pdf. The **marginal pdf** of X is $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$ and similarly for Y, $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$.
- Let (X, Y) be a continuous bivariate RV, and let f be its joint pdf. The joint cdf of X and Y is a function $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $F(x, y) = P(X \leq x, Y \leq y) = \displaystyle\int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)dsdt$. $f(x, y) = \frac{\partial^2 F}{\partial x \partial y}$, whenever this partial derivative exists.
- Let X and Y be any two RVs with joint pdf/pmf f and marginals f_X and f_Y. The **conditional pdf/pmf** g of X given $Y = y$, is defined as $g(x|y) = \dfrac{f(x, y)}{f_Y(y)}$, provided $f_Y(y) > 0$.
- Let X and Y be any two RVs with joint cdf F and marginals F_X and F_Y. X and Y are independent if and only if $F(x, y) = F_X(x)F_Y(y)$ for all $(x, y) \in \mathbb{R}^2$.
- Two discrete RVs X and Y are independent if and only if $P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i)$ for all $(x_i, y_i) \in R_X \times R_Y$.
- Two continuous RVs X and Y are independent if and only if $f(x, y) = f_X(x)f_Y(y)$, for all $(x, y) \in \mathbb{R}^2$.
- The RVs X and Y are said to be **independent and identically distributed (IID)** if and only if they are independent and have the same distribution.
- Let X and Y be RVs with joint pdf/pmf f. The **product moment** of X and Y about the origin, denoted $E(XY)$, is defined as $\displaystyle\sum_{x \in R_X} \sum_{y \in R_Y} xyf(x, y)$ if X, Y are discrete and $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy$ if X, Y are continuous and provided $E(XY) < \infty$.
- The **covariance** between X and Y, denoted by $\mathrm{Cov}(X, Y)$ or σ_{XY}, is defined as $E((X - \mu_X)(Y - \mu_Y))$.
- For arbitrary RVs X and Y, the product moment and covariance may or may not exist. The covariance, unlike variance, can also be negative.
- $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$. Thus, $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
- $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$, where $a, b, c, d \in \mathbb{R}$.
- If X and Y are independent, then $E(XY) = E(X)E(Y)$.
- If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.
- $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$, where $a, b \in \mathbb{R}$.
- $\mathrm{Var}(X + Y + Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\mathrm{Cov}(X, Y) + 2\mathrm{Cov}(Y, Z) + 2\mathrm{Cov}(Z, X)$.
- Let X and Y be two RVs with variances σ_X^2 and σ_Y^2 respectively. The **correlation coefficient** between X and Y, denoted ρ, is defined as $\dfrac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$.
- If X and Y are independent, then the correlation coefficient between them is 0. The converse is not true. If $\rho = 0$, then X and Y are said to be **uncorrelated**.
- Let X be an RV. The **standardization** of X is defined as $X^* = \frac{X - \mu_X}{\sigma_X}$.

- If X^* and Y^* are standardizations of the RVs X and Y, then the correlation coefficient between X and Y is equal to the correlation coefficient between X^* and Y^*.
- For any RVs X and Y, $-1 \leq \rho \leq 1$. If $\rho = \pm 1$, then $Y = aX + b$ where $a, b \in \mathbb{R}$, $a \neq 0$.
- Let X and Y be two RVs. A function $M : \mathbb{R}^2 \to \mathbb{R}$ defined by $M(s,t) = E(e^{sX+tY})$, is called the **joint moment generating function** of X and Y if this expected value exists for all s in some interval $(-h, h)$ and for all t in some interval $(-k, k)$.
- $M(s, 0) = E(e^{sX})$ and $M(0, t) = E(e^{tY})$.
- $E(X^k) = \dfrac{\partial^k M(s,t)}{\partial s^k}$, $E(Y^k) = \dfrac{\partial^k M(s,t)}{\partial t^k}$, and $E(XY) = \dfrac{\partial^2 M(s,t)}{\partial s \partial t}$ for $k \in \mathbb{N}$, evaluated at $(0, 0)$.
- If X and Y are independent then $M_{aX+bY}(t) = M_X(at)M_Y(bt)$, where $a, b \in \mathbb{R}$.
- The **conditional mean** or **conditional expected value** of X given $Y = y$ is defined as $\mu_{X|y} = E(X|y) = \sum_{x \in R_X} xg(x|y)$ if X is discrete and $\int_{-\infty}^{\infty} xg(x|y)dx$ if X is continuous.
- $E(Y|x)$ is a function of x. $E_X(E(Y|x)) = E_Y(Y)$.
- Let X and Y be two RVs. If $E(Y|x)$ is a linear function of x, then $E(Y|x) = \mu_Y + \rho\frac{\sigma_X}{\sigma_Y}(x - \mu_X)$, where ρ is the correlation coefficient of X and Y.
- Let X and Y be two RVs and let $h(y|x)$ be the conditional pdf of Y given $X = x$. The **conditional variance** of Y given $X = x$, is defined as $\text{Var}(Y|x) = E(Y^2|x) - (E(Y|x))^2$.
- Let X and Y be two RVs. If $E(Y|x)$ is a linear function of x, then $E(\text{Var}(Y|x)) = (1 - p^2)\text{Var}(Y)$.
- A discrete bivariate RV is said to have a **Bivariate Bernoulli distribution** with parameters p_1, p_2 if and only if its joint pmf is of the form $f(x,y) = \dfrac{1}{x!y!(1-x-y)!}p_1^x p_2^y (1 - p_1 - p_2)^{1-x-y}$ if $x, y \in \{0, 1\}$ and $f(x, y) = 0$ otherwise. Here, $p_1, p_2 > 0$ and $p_1 + p_2 < 1$ and $x + y \leq 1$.
- Let (X, Y) be a Bivariate Bernoulli RV with parameters p_1, p_2. Then, $E(X) = p_1$, $E(Y) = p_2$, $\text{Var}(X) = p_1(1-p_1)$, $\text{Var}(Y) = p_2(1-p_2)$, $\text{Cov}(X,Y) = -p_1p_2$, and $M(s,t) = 1 - p_1 - p_2 + p_1e^2 + p_2e^t$.
- A discrete bivariate RV (X, Y) is said to have a **Bivariate Binomial distribution** with parameters n, p_1, p_2 if and only if its joint pmf is of the form $f(x,y) = \dfrac{n}{x!y!(n-x-y)!}p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$ if $x, y \in \{0, 1, 2, ..., n\}$ and $f(x, y) = 0$ otherwise. Here, $p_1, p_2 > 0$, $p_1 + p_2 < 1$ and $x + y \leq n$.
- Let (X, Y) be a Bivariate Binomial RV. Then, $E(X) = np_1$, $E(Y) = np_2$, $\text{Var}(X) = np_1(1-p_1)$, $\text{Var}(Y) = np_2(1-p_2)$, $\text{Cov}(X,Y) = -np_1p_2$, and $M(s,t) = (1 - p_1 - p_2 + p_1e^2 + p_2e^t)^n$.
- A continuous bivariate RV (X, Y) is said to have a **Bivariate Normal distribution** with parameters $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ if and only if its joint pdf is of the form $f(x,y) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{\frac{1}{2}Q(x,y)}$ if $x, y \in (0, \infty)$ and $f(x, y) = 0$ otherwise.
  Here, $Q(x,y) = \dfrac{1}{1-\rho^2}\left[\left(\dfrac{x-\mu_1}{\sigma_1}\right)^2 - 2\left(\dfrac{x-\mu_1}{\sigma_1}\right)\left(\dfrac{y-\mu_2}{\sigma_2}\right) + \left(\dfrac{y-\mu_2}{\sigma_2}\right)^2\right]$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in (0, 1)$, and $\rho \in (-1, 1)$.
- If $X_1, X_2, ..., X_n$ are continuous RVs, and $Y = u(X_1, X_2, ..., X_n)$, there are three methods to find the cdf and pdf/pmf of Y:
  1. **Distribution Function Method:** The pdf of Y can be found by getting its CDF, $P(Y \leq y) = P(u(X_1, X_2, ..., X_n) \leq y)$ and then differentiating it to obtain the pdf.
  2. **Transformation Method (Univariate Case):** Let f be the pdf of X and let $Y = u(X)$. If u is differentiable and monotonic for all values within R_X such that $f(x) \neq 0$, then we can find the inverse of u, say w, such that $x = w(y)$. Then the pdf of Y is given by $g(y) = f(w(y))|w'(y)|$, provided $u'(x) \neq 0$, and elsewhere, $g(y) = 0$.

3. **Transformation Method (Bivariate Case):** Let X, Y have joint pdf f and let $U = Q(X, Y)$ and $V = R(X, Y)$. If $Q(x, y)$ and $R(x, y)$ have single valued inverses, that is, $X = S(U, V)$ and $Y = T(U, V)$, then the joint pdf of U and V is given by the Jacobian, which is defined as $J = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$.
4. **MGF Method:** If X and Y are independent, then the distribution of $X + Y$ can be found by $M_{X+Y}(t) = M_X(t)M_Y(t)$.

- Let there be a random experiment whose outcome is represented by the RV X with pdf/pmf f. Suppose the experiment is repeated n times and that X_k is the RV associated with the kth repetition. Then the collection of RVs $\{X_1, X_2, ..., X_n\}$ is called a **random sample** of size n. $X_1, X_2, ..., X_n$ are independent and identically distributed with common pdf f.

- Given a random sample $\{X_1, X_2, ..., X_n\}$, functions such as the **sample mean** $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and the **sample variance** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, are called **statistics**.

- If $X_1, X_2, ..., X_n$ are mutually independent RVs with respective means $\mu_1, \mu_2, ..., \mu_n$ and respective variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, then the mean and variance of $Y = \sum_{i=1}^{n} a_i X_i$, where $a_i \in \mathbb{R}$, is given by $\mu_Y = \sum_{i=1}^{n} a_i \mu_i$ and $\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$.

- Let $X_1, X_2, ..., X_n$ be observations from a random sample of size n with distribution f. Let $X_{(1)}$ denote the smallest of $\{X_1, X_2, ..., X_n\}$, and similarly let $X_{(2)}$ denote the second smallest of them, and so on. Then the random variables $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are called the **order statistics** of the sample $X_1, X_2, ..., X_n$. In particular, $X_{(r)}$ is called the rth order statistic of $X_1, X_2, ..., X_n$.

- Let $X_1, X_2, ..., X_n$ be a random sample of size n with pdf f. Then the pdf of $X_{(r)}$ is given by $g(x) = \frac{n!}{(r-1)!(n-r)!}(F(x))^{r-1}f(x)(1 - F(x))^{n-r}$, where F is the cdf of f.

- Let $p \in (0, 1)$. A $100p$th **percentile** of the distribution of a random variable X is any real number q satisfying $P(X \leq q) \leq p$ and $P(X > q) \leq 1 - p$.

- The 25th and 75th percentiles of any distribution are called the first and third **quartiles**, respectively. The 50th percentile is called the **median**.

- A **mode** of the distribution of the continuous RV X is the value of x where the pdf of X attains a relative maximum. An RV can have more than one mode.

- Let $X_1, X_2, ..., X_n$ be a random sample. The **sample median** is defined as $M = X_{(\frac{n+1}{2})}$ if n is odd, and $M = \frac{1}{2}(X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})})$ if n is even.

- The $100p$th sample percentile is defined as $\pi_p = X_{([np])}$ if $p < 0.5$, $\pi_p = M$ if $p = 0.5$, $\pi_p = X_{(n+1-[n(1-p)])}$ if $p > 0.5$. Here $[x]$ denotes the nearest integer to x, M is the sample median and n is the size of the sample.

- The first quartile is also called the lower quartile, and the third quartile is also called the upper quartile. The difference between them is called the **interquartile range**.

- Let $X_1, X_2, ..., X_n$ be a random sample, with distribution f. Then the joint pdf of the sample is given by $f(x_1, x_2, ..., x_n) = \Pi_{i=1}^{n} f(x_i)$.

- Given a random sample $X_1, X_2, ..., X_n$ with pdf $f(x, \theta)$, where θ is a parameter, a statistic is a function T of $x_1, x_2, ..., x_n$ that is independent of the parameter θ.

- The probability distribution of the statistic T is called the **sampling distribution** of T.

- A continuous RV X is said to have **Chi-square distribution** with r **degrees of freedom** if its pdf is of the form $f(x) = \dfrac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}}$ when $0 \leq x < \infty$ and $f(x, r) = 0$ otherwise. Here, $r > 0$. It is denoted $X \sim \chi^2(r)$.
- The Chi-square distribution is equivalent to the Gamma distribution when $\alpha = \frac{r}{2}$ and $\theta = 2$.
- If $r \to \infty$, then the Chi-square distribution tends to the normal distribution.
- If X is a Chi-square RV, then $E(X) = r$ and $\mathrm{Var}(X) = 2r$.
- If $X \sim N(\mu, \sigma^2)$, then $\left(\dfrac{X - \mu}{\sigma}\right)^2 \sim \chi^2(1)$.
- If $X \sim N(\mu, \sigma^2)$ and $X_1, X_2, ..., X_n$ is a random sample from population X, then $\displaystyle\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$.
- If $X \sim N(\mu, \sigma^2)$ and $X_1, X_2, ..., X_n$ is a random sample from population X, then $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, where $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$ is the sample variance.
- A continuous RV X is said to have a **t-distribution** with ν degrees of freedom if its pdf is of the form $f(x) = \dfrac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})(1 + \frac{x^2}{\nu})^{(\frac{\nu+1}{2})}}$, where $x \in \mathbb{R}$, $\nu > 0$. It is denoted $X \sim t(\nu)$.
- If $\nu \to \infty$, then the t-distribution tends to the standard normal distribution.
- If $X \sim t(\nu)$, then $E(X) = 0$ if $\nu \geq 2$. $E(X)$ does not exist if $\nu = 1$. $\mathrm{Var}(X) = \frac{\nu}{\nu-2}$ if $\nu \geq 3$. $\mathrm{Var}(X)$ does not exist if $\nu = 1$ or $\nu = 2$.
- If $Z \sim N(0, 1)$ and $V \sim \chi^2(\nu)$, and if Z and V are independent, then $W = \dfrac{Z}{\sqrt{\frac{V}{\nu}}} \sim t(\nu)$.
- If $X \sim N(\mu, \sigma^2)$ and $X_1, X_2, ..., X_n$ is a random sample from population X, then $\dfrac{\overline{X_n} - \mu}{\frac{S_n}{\sqrt{n}}} \sim t(n-1)$.
- If $X_1, X_2, ..., X_n$ are mutually independent random variables such that $X_i \sim N(\mu_i, \sigma_i^2)$, then the random variable $Y = \displaystyle\sum_{i=1}^{n} a_i X_i$ is normal RV with mean $\mu_Y = \displaystyle\sum_{i=1}^{n} a_i \mu_i$ and variance $\sigma_Y^2 = \displaystyle\sum_{i=1}^{n} a_i^2 \sigma_i^2$.
- A continuous RV X is said to have an **F-distribution** with ν_1 and ν_2 degrees of freedom if its pdf is of the form $f(x) = \dfrac{\Gamma(\frac{\nu_1+\nu_2}{2})(\frac{\nu_1}{\nu_2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})(1 + \frac{\nu_1}{\nu_2}x)^{(\frac{\nu_1+\nu_2}{2})}}$ if $0 \leq x < \infty$ and $f(x) = 0$ otherwise, where $\nu_1, \nu_2 > 0$. It is denoted $X \sim F(\nu_1, \nu_2)$.
- F-distribution tends to the normal distribution when ν_1 and ν_2 become very large.
- If $X \sim F(\nu_1, \nu_2)$, then $E(X) = \frac{\nu_2}{\nu_2 - 2}$ if $\nu_2 \geq 3$ and $E(X)$ does not exist if $\nu_2 = 1, 2$. $\mathrm{Var}(X) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ if $\nu_2 \geq 5$ and $\mathrm{Var}(X)$ does not exists if $\nu_2 = 1, 2, 3, 4$.
- If $X \sim F(\nu_1, \nu_2)$, then $\frac{1}{X} \sim F(\nu_2, \nu_1)$.
- If $U \sim \chi^2(\nu_1)$ and $V \sim \chi^2(\nu_2)$, and if U and V are independent, then $\dfrac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}} \sim F(\nu_1, \nu_2)$.
- Let $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$ be random samples of size n and m, where $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. Then the statistic $\dfrac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n-1, m-1)$. Here, S_1^2 and S_2^2 are the sample variances of the first and second sample respectively.

- Let a population be described by random variable X with pdf $f(x;\theta)$. A random sample is a portion of the population and has the same distribution as the population. The data obtained after sampling, i.e, $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$ is called the **sample data**. A **statistical inference** is a statement about the population based on the sample data.
- The three types of statistical inferences are point estimation, hypothesis testing and prediction.
- In **point estimation**, we attempt to find the parameter θ of the distribution function $f(x;\theta)$ from the sample information. The form of the distribution is assumed to be known and only the unknown parameter is estimated.
- Let X be a population with pdf $f(x;\theta)$. The set of all admissible values of θ is called a **parameter space** and is denoted by Ω. $\Omega = \{\theta \in \mathbb{R}^m : f(x;\theta) \text{ is a pdf}\}$, for some $m \in \mathbb{N}$.
- Any statistic that can be used to guess θ is called an **estimator** of θ. The numerical value of this statistic is called an **estimate** of θ. The estimator is denoted by $\widehat{\theta}$.
- There are several methods for finding an estimator of θ:
  1. **Moment Method:** Let $X_1, X_2, ..., X_n$ be a random sample from population X with pdf $f(x;\theta_1,\theta_2,...,\theta_m)$. Let $E(X^k) = \int_{-\infty}^{\infty} f(x;\theta_1,\theta_2,...,\theta_m)dx$ be the kth **population moment** about 0. Let $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ be the kth **sample moment** about 0. The estimator for $\theta_1, \theta_2, ..., \theta_m$ is found by equating the first m population moments to the first m sample moments, i.e, $M_1 = E(X)$, $M_2 = E(X^2)$, and so on.
  2. **Maximum Likelihood Method:** Let $X_1, X_2, ..., X_n$ be a random sample from population X with pdf $f(x;\theta)$. The **likelihood function** of the sample is given by $L(\theta) = \Pi_{i=1}^{n} f(x_i;\theta)$, for $\theta \in \Omega$. The value of θ that maximises $L(\theta)$ is called the **maximum likelihood estimator** of θ and denoted $\widehat{\theta}$.
- Let $\widehat{\theta}$ be a maximum likelihood estimator of θ and let $g(\theta)$ be a function of θ. Then the maximum likelihood estimator of $g(\theta)$ is given by $g(\widehat{\theta})$.
- An estimator of θ is said to be an **unbiased estimator** if and only if $E(\widehat{\theta}) = \theta$. If $\widehat{\theta}$ is not unbiased, it is called a biased estimator.
- Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be two unbiased estimators of θ. $\widehat{\theta}_1$ is said to more efficient than $\widehat{\theta}_2$ if $\text{Var}(\widehat{\theta}_1) < \text{Var}(\widehat{\theta}_2)$. The ratio $\eta(\widehat{\theta}_1, \widehat{\theta}_2) = \dfrac{\text{Var}(\widehat{\theta}_1)}{\text{Var}(\widehat{\theta}_2)}$ is called the **relative efficiency** of $\widehat{\theta}_1$ with respect to $\widehat{\theta}_2$.
- An unbiased estimator $\widehat{\theta}$ of θ is said to be a uniform minimum variance unbiased estimator of θ if and only if $\text{Var}(\widehat{\theta}) < \text{Var}(\widehat{T})$ for any unbiased estimator $\widehat{T}$ of θ.
- If $\widehat{\theta}$ is unbiased then $\text{Var}(\widehat{\theta}) = E((\widehat{\theta} - \theta)^2)$.
- Alternatively, $\widehat{\theta}$ is a uniform minimum variance unbiased estimator of θ if it minimises the variance $E((\widehat{\theta} - \theta)^2)$.
-