Chapter 3 Random Variables and Univariate Probability Distributions

3.1 Mathematical Expectations

- **Definition: Expected Value of $g(X)$.** Suppose X is a r.v. with PMF or PDF $f_X(x)$. Then the expectation of a measurable function $g(X)$ is defined as

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x) dF_X(x) \\ &= \begin{cases} \sum_{x \in \Omega_X} g(x) f_X(x) & \text{DRV}, \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx, & \text{CRV}. \end{cases} \end{aligned}$$

where Ω_X is the support of X.

- **Remarks:**
  - $E[g(X)]$ can be considered as the weighted average of $g(X)$.
  - If $E|g(X)| = \infty$, we say $E[g(X)]$ does not exist.
  - Suppose a is a constant, then $E(a) = a$.

- **Remarks:**
  - The expectation $E(\cdot)$ is a linear operator, namely,

$$E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)].$$

  - $Y = g(X)$ is also a r.v..
  - Let the pdf/pmf of $Y = g(X)$ be $f_Y(y)$, then we can also compute $E[g(X)]$ by

$$\begin{aligned} E[g(X)] &= E(Y) \\ &= \begin{cases} \sum_{y \in \Omega_Y} y f_Y(y), & \text{d.r.v.,} \\ \int_{-\infty}^{\infty} y f_Y(y) dy, & \text{c.r.v..} \end{cases} \end{aligned}$$

3.2 Moments: Mean

- **Definition: Mean.** The mean of a random variable X is defined as

$$\mu_X = E(X) = \begin{cases} \sum_x x f_X(x) & \text{DRV}, \\ \int_{-\infty}^{\infty} x f_X(x) dx, & \text{CRV}. \end{cases}$$

- **Remarks:**
  - The mean μ_X is also called the *expected value* of X, or the *first moment* of X.
  - μ_X is a measure of central tendency for the distribution of X.

3.2 Moments : Mean

- **Example: Cauchy Distribution.** Suppose X follows Cauchy$(0,1)$ distribution, then

$$E|X| = \int_{-\infty}^{\infty} |x| \frac{1}{\pi\left(1+x^2\right)}\, dx = \infty.$$

Mean of Cauchy distribution does not exist.

- **Theorem:** Suppose $E(X^2)$ exists. Then

$$\mu_X = \arg\min_a E(X-a)^2.$$

How to prove?

- **Question:** Does $X = \mu_X$ has the largest probability to occur?

3.2 Moments : Variance

- **Definition: Variance.** The variance of random variable X is defined as

$$\sigma_X^2 = E(X - \mu_X)^2 = \begin{cases} \sum_x (x - \mu_X)^2 f_X(x) & \text{DRV}, \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx, & \text{CRV}. \end{cases}$$

The *standard deviation* of X is given by $\sigma_X = \sqrt{\sigma_X^2}$.

- **Remarks:**
  - σ_X^2 is a measure of the degree of spread of a distribution around its mean.
  - In economics, it is interpreted as a measure of uncertainty or risk. It is often called a measure of *volatility* of X. $\sigma_X^2 = 0$ implies $X \equiv \mu_x$, and there is no uncertainty.

3.2 Moments : Variance

- **Theorem:** $\sigma_X^2 = E(X^2) - \mu_X^2$.

  Remark: σ_X^2 is called the *second central moment*, and $E(X^2)$ is called the *second moment* of X.

- **Theorem:** If $Y = a + bX$, then (i) $\mu_Y = a + b\mu_X$; (ii) $\sigma_Y^2 = b^2\sigma_X^2$.

  Remark: The variance of Y only depends on the *scale* parameter b but not on the *location* parameter a.

- **Theorem: k-th moment and k-th central moment.** The k-th moment of a random variable X is defined as

$$E(X^k) = \begin{cases} \sum_{x \in \omega_X} x^k f_X(x) & \text{DRV}, \\ \int_{-\infty}^{\infty} x^k f_X(x) dx, & \text{CRV}. \end{cases}$$

The k-th central moment of a random variable X is defined as

$$E(X-\mu_X)^k = \begin{cases} \sum_{x\in\Omega_X}(x-\mu_X)^k f_X(x) & \text{DRV}, \\ \int_{-\infty}^{\infty}(x-\mu_X)^k f_X(x)dx, & \text{CRV}. \end{cases}$$

- **Question:** What is the relationship between uncentered moments and centered moments?

- **Example: Petersburg Paradox.** A risk-averse player will not only consider the expected return but also take into account the risk of the game, which is often measured by variance.

- **Example: Portfolio Selection.**
  - Assume that the investor likes higher return but lower risk. That is, his/her utility function $U(\mu, \sigma^2)$ is a function of μ and σ^2 such that

$$\begin{aligned} \partial U/\partial\mu \geq 0 &: \text{ The more expected return, the better.} \\ \partial U/\partial\sigma^2 < 0 &: \text{ The smaller risk, the better.} \end{aligned}$$

    An example of $U(\mu, \sigma^2)$ is

$$U(\mu, \sigma^2) = a\mu - \frac{b}{2}\sigma^2,$$

    where $a, b > 0$.
  - Assume that the investor has totally I dollars to be split between the risky asset (z) and the risk-free asset ($I - z$). What is the portfolio that maximizes the utility function ?

3.2 Moments : Portfolio Selection

- – The rate of return on stocks is a random variable Y with mean μ_Y and variance σ_Y^2 .

  – The rate of returns on risk-free asset is constant r, which can be considered as a random variable Z with mean r and variance 0. Usually $r < \mu_Y$.

  – The return of a portfolio is $X = zY + (I - z)Z$. The utility function is

$$U(\mu, \sigma^2) = a[z\mu_Y + (I - z)r] - \frac{b}{2}z^2\sigma_Y^2.$$

  It is maximized when

$$z = \frac{a(\mu_Y - r)}{b\sigma_Y^2}.$$

  – If $b = 0$ (i.e., the investor does not care risk), $U(\mu, \sigma^2)$ is maximized at $z = \infty$, which means the investor will borrow money to invest on the risky asset.

3.2 Moments : Skewness

- **Definition: Skewness.** The third central moment $E[(X-\mu_X)^3]$ is a measure of "skewness" (or asymmetry) of the distribution for X. *Skewness* is defined as

$$S_X = \frac{E[(X-\mu_X)^3]}{\sigma_X^3}.$$

The skewness has been used to measure financial crashes. Negative (positive) skewness indicates a higher (lower) probability of experiencing **large** losses than **large** gains.

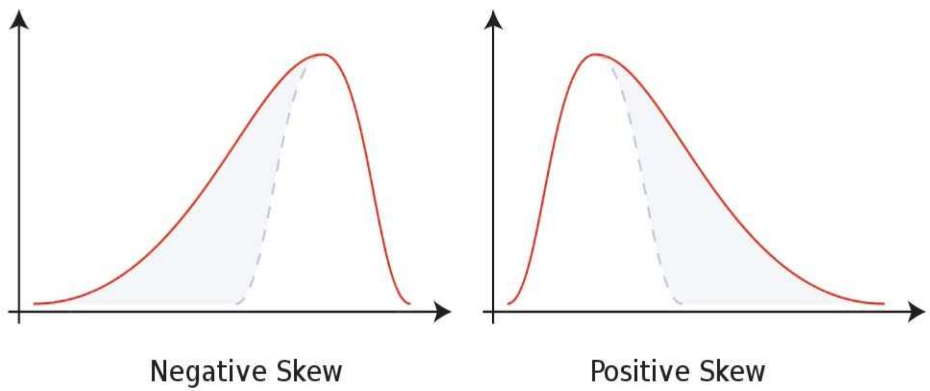

Figure 1: Skewness of Distributions.

3.2 Moments : Kurtosis

- **Definition: Kurtosis.** The fourth central moment $E[(X-\mu_X)^4]$ is a measure of how heavy the tail of a distribution is. *Kurtosis* is defined as

$$K_X = \frac{E[(X-\mu_X)^4]}{\sigma_X^4}.$$

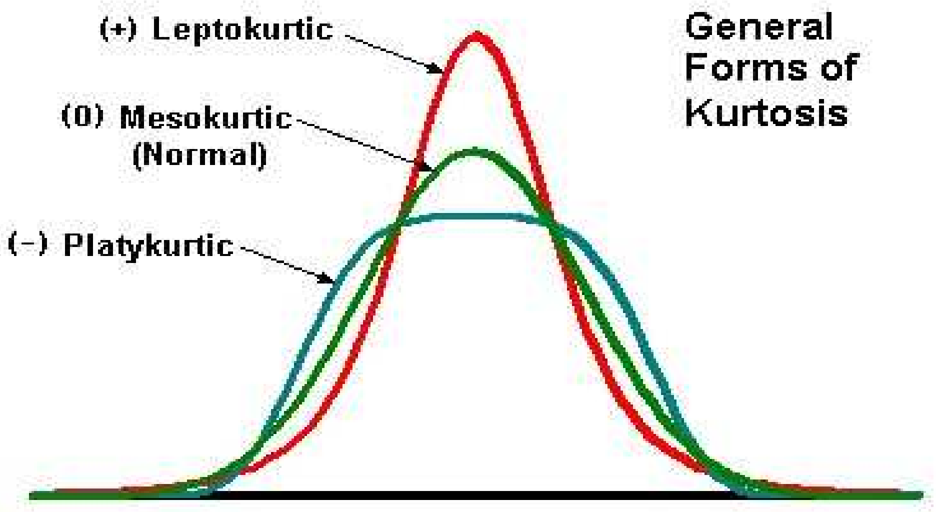

Figure 2: Kurtosis of Distributions.

3.2 Moments : Kurtosis

- **Remarks:**
  - If $X \sim N(\mu, \sigma^2)$, then $E(X-\mu)^4 = 3\sigma^4$. Kurtosis of normal distribution is 3.
  - The *excess kurtosis* of a random variable X is defined as $K_X - 3$.
  - A distribution with positive excess kurtosis is called *leptokurtic*. A leptokurtic distribution has a more acute peak around the mean and fatter tails.
  - A distribution with negative excess kurtosis is called *platykurtic*. A platykurtic distribution has a lower, wider peak around the mean and thinner tails.

- **Remarks:**
  - Relationship between uncentered moments and centered moments. We have

$$E(X-\mu_X)^k = \sum_{i=0}^{k} \begin{pmatrix} k \\ i \end{pmatrix} E(X^i)(-\mu_X)^{k-i},$$

and

$$\begin{aligned} E(X^k) &= E(X-\mu_X+\mu_X)^k \\ &= \sum_{i=0}^{k} \begin{pmatrix} k \\ i \end{pmatrix} E(X-\mu_X)^i \mu_X^{k-i}. \end{aligned}$$

3.3 Quantile

- **Definition: α-Quantile.** Suppose X has a CDF $F_X(x)$, and let $\alpha \in (0,1)$. Then the α-quantile of the distribution $F_X(x)$ is defined as $Q(\alpha)$, which satisfies

$$F_X(Q(\alpha)) = P(X \leq Q(\alpha)) = \alpha.$$

When $F_X(x)$ is strictly increasing, we have

$$Q(a) = F_X^{-1}(\alpha),$$

where $F^{-1}(\alpha)$ is the inverse function of $F_X(x)$.

- **Remarks:**
  - For α-quantile $Q(\alpha)$, we have $\int_{-\infty}^{Q(\alpha)} f_X(x)dx = \alpha$.
  - 0.5-quantile is called the *median*, 0.25-quantile and 0.75-quantile are called *lower* and *upper quartiles*.
  - 0-quantile is the minimum value, 1-quantile is the maximum value.

- **Boxplot:**

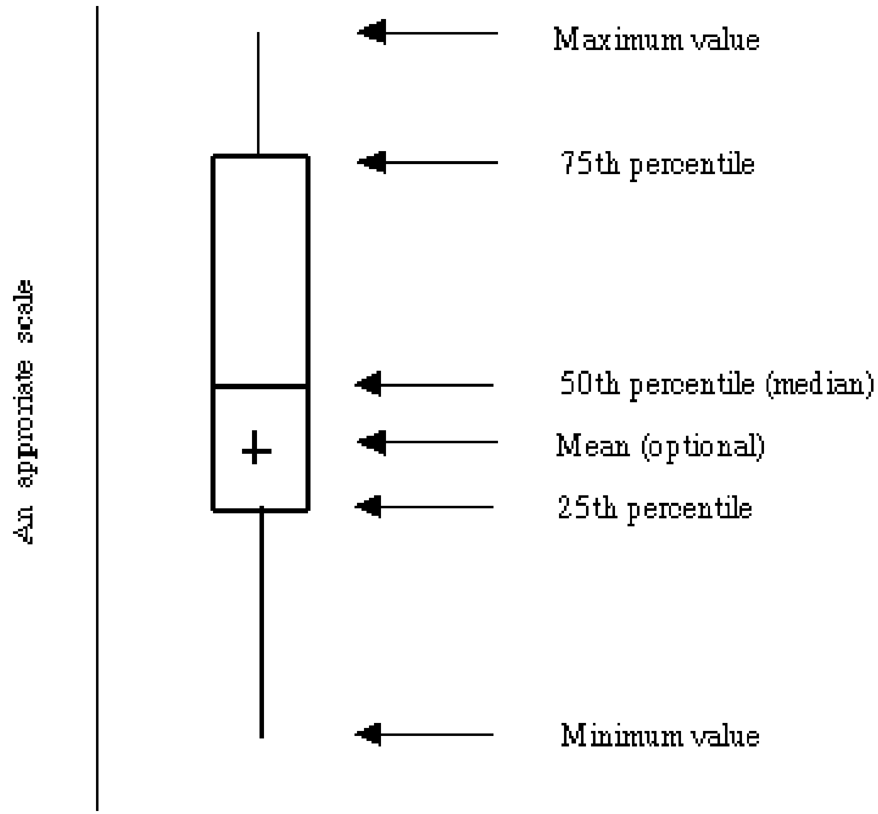

Figure 3: Boxplot.

3.3 Quantile

- **Example:** Customers plan to spend ($thousands)

$$3.8, 1.4, 0.3, 0.6, 2.8, 5.5, 0.9, 1.1.$$

  Find the median of these values.

  – Sorting the values:

$$0.3, 0.6, 0.9, 1.1, 1.4, 2.8, 3.8, 5.5.$$

  – Median is any value between 1.1 and 1.4. Usually, we take $(1.1 + 1.4)/2$.

- **Example: Value at Risk (VaR).** VaR at level α, $V_t(\alpha)$, of a portfolio over a certain time horizon is defined as $P[X_t < -V_t(\alpha)|I_{t-1}] = \alpha$, where X_t is the return on the portfolio over the holding period t, and I_{t-1} is the information available at time $t - 1$.

3.3 Quantile

- **Example:** For symmetric distribution, *e.g.,* normal distribution, mean and median are the same. For skewed distributions, mean and median are different.

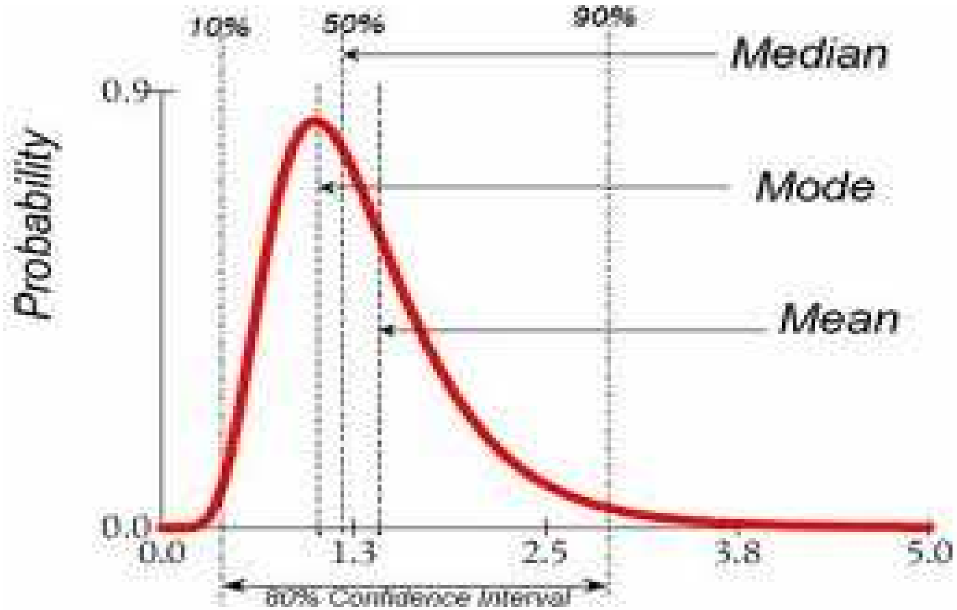

Figure 4: Mean and Median.

- **Difference Between Mean and Median:**
  - Median is the cutoff point that divides the population in half.
  - Mean can be misleading when used to measure the location of highly skewed data. In contrast, median is a more robust measure of the central tendency of a distribution in the sense that it is not much affected by a few *outliers*.
  - **Theorem.** Median $Q(0.5)$ is the optimal solution for minimizing the *mean absolute error*, that is,

$$Q(0.5) = \arg\min_a \{E|X - a|\}.$$

While mean is the optimal solution for minimizing the *mean square error*, that is,

$$E(X) = \arg\min_a \{E(X - a)^2\}.$$

3.4 Moment Generating Function

- **Definition: Moment Generating Function (MGF).** The MGF of a r.v. X is defined as

$$\begin{aligned} M_X(t) &= E[exp(tX)] \\ &= \begin{cases} \sum_{x \in \Omega_X} exp(tx) f_X(x) & \text{DRV}, \\ \int_{-\infty}^{\infty} exp(tx) f_X(x) dx & \text{CRV}. \end{cases} \end{aligned}$$

- **Remarks:** $M_X(t)$ may not exist for some $t \in \mathbb{R}$.

- **Theorem.** Suppose the MGF $M_X(t)$ exists for t in some small neighborhood of 0. Then $M_X(0) = 1$.

3.4 Moment Generating Function

- **Theorem.** If $M_X(t)$ exists for t in some neighborhood of 0, then for $k = 1, 2, \cdots,$

$$M_X^{(k)}(0) = E(X^k).$$

- **Proof:** For any given integer $k > 0$ and all $t \in (-\epsilon, \epsilon)$, we have

$$\begin{aligned} M_X^{(k)}(t) &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} dF_X(x) \\ &= \int_{-\infty}^{\infty} \frac{d^k}{dt^k}(e^{tx}) dF_X(x) \\ &= \int_{-\infty}^{\infty} x^k e^{tx} dF_X(x). \end{aligned}$$

Setting $t = 0$, we have

$$M_X^{(k)}(0) = \int_{-\infty}^{\infty} x^k dF_X(x) = E(X^k).$$

3.4 Moment Generating Function

- **Remark.**
  - Every moment of X can be computed by differentiating $M_X(t)$ at the origin, provided $M_X(t)$ exists for $t \in (-\epsilon, \epsilon)$.
  - $M_X^{(1)}(0) = \mu_X$, $M_X^{(2)}(0) = E(X^2) = \sigma_X^2 + \mu_X^2$.

- **Theorem.** Suppose $Y = a + bX$, where a and b are two constants, and the MGF $M_X(t)$ of X exists for t in a small neighborhood of 0. Then the MGF

$$M_Y(t) = e^{at} M_X(bt)$$

  for all t in a small neighborhood of 0.

- **Example:** Let $Y = \frac{X-\mu}{\sigma}$, then $E(Y) = 0$, $var(Y) = 1$, and

$$M_Y(t) = exp\{-\frac{\mu}{\sigma}t\} M_X(\frac{t}{\sigma}).$$

- **Example: Bernoulli Distribution.** For the Bernoulli(p) distribution

$$f_X(x) = \begin{cases} p & x = 1, \\ 1-p & x = 0, \end{cases}$$

where $0 < p < 1$. We have

$$\begin{aligned} E(X) &= p, \\ var(X) &= p(1-p), \\ M_X(t) &= pe^t + 1 - p, \ -\infty < t < \infty. \end{aligned}$$

3.4 Moments and MGF of DRV

- **Example: Binomial Distribution.** For the Binomial distribution $B(n, p)$,

$$f_X(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x},\ x = 0, 1, \cdots, n.$$

Find

(1) μ_X by direct formula. ($E(X) = np$).

(2) σ_X^2 by direct formula. ($\sigma_X^2 = np(1-p)$).

(3) $M_X(t)$. ($M_X(t) = (pe^t + 1 - p)^n$.)

(4) Use $M_X(t)$ to find μ_X and σ_X^2.

- **Hints:**
  - Binomial r. v. can be viewed as sum of Bernoulli r.v., $X = \sum_{i=1}^n X_i$, $X \sim B(n, p)$, $X_i \sim$ Bernoulli(p).
  - Binomial expansion $(x+y)^n = \sum_{i=0}^n C_n^i x^i y^{n-i}$.

- **Example: Poisson Distribution** The pmf of Poisson distribution is as follows

$$f_X(X=x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \cdots .$$

Then

(1) $E(X) = \lambda$.

(2) $E(X^2) = \lambda^2 + \lambda$, $\sigma_X^2 = \lambda$.

(3) $M_X(t) = exp\{e^t\lambda - \lambda\}$.

(4) When $n \to \infty$, but $np \to \lambda$,we can use a Poisson(λ) distribution to approximate a binomial distribution,

$$\begin{aligned} M_B(t) &= (pe^t + 1 - p)^n \\ &= [1 + \frac{np(e^t-1)}{n}]^n \\ &\to e^{\lambda(e^t-1)} = M_P(t). \end{aligned}$$

- **Example: Waiting Time.** Consider a telephone operator who, on the average, handles five calls every 3 minutes. What is the probability that there will be no calls in the next minute? at least two calls?

- **Solution:**

  – Five calls every 3 minutes : $\lambda = E(X) = 5/3$.

  – $P(\text{“no call in the next minute”}) = P(X = 0) = 0.189$.

  – $P(\text{“at least two calls in the next minute”}) = P(X \geq 2) = 1 - 0.189 - \frac{5}{3}e^{-5/3} = 0.496$.

- **Example: Uniform Distribution.** A CRV X follows a uniform distribution on [a, b] if its PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

(1) $E(X^k) = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}$.

(2) $E(X) = \frac{a+b}{2}$.

(3) $\sigma_X^2 = \frac{1}{12}(b-a)^2$.

(4) $M_X(t) = \frac{1}{(b-a)t}[e^{tb} - e^{ta}]$.

- **Example: Beta Distribution.** A CRV with Beta(α, β) distribution has the PDF

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 \le x \le 1,$$

where $\alpha > 0$, $\beta > 0$. Then

(1) $E(X) = \frac{\alpha}{\alpha+\beta}$.

(2) $E(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$, $var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

(3) $M_X(t) = 1 + \sum_{j=1}^{n} \left(\prod_{i=0}^{j-1} \frac{\alpha+i}{\alpha+\beta+i} \right) \frac{t^j}{j!}$.

- **Example: Normal Distribution.** A normally distributed random variable, $X \sim N(\mu, \sigma^2)$, has the pdf

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

where $-\infty < \mu < \infty$, $\sigma > 0$.

(1) $E(X) = \mu$.

(2) $\mathrm{Var}(X) = \sigma^2$.

(3) $M_X(t) = exp\{\frac{\sigma^2 t^2}{2} + \mu t\}$, $-\infty < t < \infty$.

(3) $E[(X-\mu)^{2k}] = \frac{1}{\sqrt{\pi}}\Gamma(k+\frac{1}{2})2^k\sigma^{2k}$.

- **Remark.** The normal distribution was discovered in 1733 by Abraham De Moiver, and then it was used to predict the location of astronomical bodies by Gauss. It is the most important distribution in probability theory.

- **Stein's Lemma:** A normally distributed random variable, $X \sim N(\mu, \sigma^2)$, has the PDF
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\},$$
where $-\infty < \mu < \infty$, $\sigma > 0$. Let $g(\cdot)$ be a differentiable function satisfying $E|g'(X)| < \infty$, then
$$E\left[g(X)(X-\mu)\right] = \sigma^2 E\left[g'(X)\right].$$

- **Corollary:** Let $g(X) = (X-\mu)^{k-1}$, then
$$E\left[(X-\mu)^k\right] = \begin{cases} \sigma^{2n} \prod_{i=1}^{n}(2i-1) & k = 2n, \\ 0 & k = 2n-1. \end{cases}$$

3.4 Moments and MGF of CRV

- **Example: Log-normal Distribution.** X follows a log-normal(μ, σ^2) distribution if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}\, exp\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\} & 0 < x < \infty, \\ 0 & \text{otherwise}, \end{cases}$$

  - Let $Y = \log(X)$, then $Y \sim N(\mu, \sigma^2)$.
  -

$$\begin{aligned} E(X^k) &= E(e^{kY}) = M_Y(k) \\ &= exp\{\frac{\sigma^2 k^2}{2} + \mu k\}. \end{aligned}$$

  - $M_X(t)$ does not exist for $t > 0$.

- **Example: Gamma Distribution** A CRV X follows a Gamma(α,β) ($\alpha, \beta > 0$) distribution if

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & 0 < x < \infty, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$. The gamma function Γ satisfies $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

(1) $E(X) = \alpha\beta$.

(2) $E(X^2) = \alpha(\alpha+1)\beta^2$.

(3) $\sigma_X^2 = E(X^2) - E^2(X) = \alpha\beta^2$.

(4) $M_X(t) = (1-\beta t)^{-\alpha}$ for $t < 1/\beta$.

3.4 Moments and MGF of CRV

- **Example: Chi-square Distribution.** A CRV X follows $\chi^2(k)$ with degrees of freedom k if

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2} & 0 < x < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

$\chi^2(k)$=Gamma($\frac{k}{2}$,2).

(1) $E(X) = k$.

(2) $var(X) = 2k$.

(3) $E(X^l)) = \frac{2^l \Gamma(l+\frac{k}{2})}{\Gamma(\frac{k}{2})}$.

(4) $M_X(t) = (1-2t)^{-k/2}$ for $t < 1/2$.

- **Example: Exponential Distribution.** A CRV X follows an Exponential(β) ($\beta > 0$) distribution if

$$f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & 0 < x < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Exponential(β)=Gamma(1,β).

(1) $E(X) = \beta$.

(2) $var(X) = \beta^2$.

(3) $M_X(t) = \frac{1}{1-\beta t}$, $t < \beta^{-1}$.

- **Example: Double Exponential Distribution.** A CRV X follows a double exponential distribution if

$$f_X(x) = \frac{1}{2\beta}\exp(-\frac{|x-\alpha|}{\beta}),\ -\infty < x < \infty.$$

(1) $E(X) = \alpha$.

(2) $var(X) = 2\beta^2$.

(3) $M_X(t) = \frac{e^{\alpha t}}{1-\beta^2 t^2},\ |t| < \beta^{-1}$.

- **Definition: Identical Distribution.** Let X and Y be two r.v. with cdf's $F_X(x)$ and $F_Y(y)$, respectively. If two cdf's are the same, i.e. $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$, then we say that X and Y are *identically distributed.*

- **Remarks:**

  – If X and Y are identically distributed, then for any function $g(\cdot)$,

$$E[g(X)] = E[g(Y)].$$

  – Identity of the distributions of X and Y does not imply $X = Y$.

  – **Example:** Suppose a fair coin is tossed n times, and let X be the number of heads obtained, Y be the number of tails obtained. Then $F_X = F_Y$, but $X + Y = n$.

3.4 Moment Generating Function

- **Theorem: Uniqueness of MGF.** Suppose two r.v. X and Y with $M_X(t)$ and $M_Y(t)$ existing in some neighborhood of 0 denoted as $N_\epsilon(0)$. Then X and Y have the same $M_X(t)$ and $M_Y(t)$ for all t in $N_\epsilon(0)$, if and only if $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$.

- Given some $M_X(t)$ in neighborhood of 0, suppose we can find some distribution $F_X(x)$, then $F_X(x)$ must be the only distribution that generates $M_X(t)$.

- **Example:** A DRV X has $M_X(t) = 1/2 + 1/4e^{-t} + 1/4e^t$. Then its PMF is

$$f_X(x) = \begin{cases} 1/4 & x = 1, \\ 1/2 & x = 0, \\ 1/4 & x = -1. \end{cases}$$

- **Example:** Suppose a random variable X has mean 0, variance 2, and MGF

$$M_X(t) = a(1 + be^{-2t} + e^{-t} + e^t + ce^{2t}).$$

- **Solution:** From $M(0) = 1$, $M'(0) = 0$, $M''(0) = 2$, we can find $a = 1/5$, $b = c = 1$.

- **Example:** If a DRV X has $M_X(t) = \frac{1-r}{1-re^t}$. What is the distribution of X?

- **Solution:**

$$\begin{aligned} M_X(t) &= (1-r)\sum_{x=0}^{\infty}(re^t)^x \\ &= \sum_{x=0}^{\infty}(1-r)r^x e^{xt}. \end{aligned}$$

Therefore, $f_X(x) = (1-r)r^x$ for $x = 0, 1, 2, \cdots$.

3.4 Moment Generating Function

- **Remarks:**
  - If the MGF $M_X(t)$ exists in a neighborhood of 0, it uniquely characterizes a distribution function.
  - The set of moments $E(X^k)$, $k = 1, 2, \cdots$, does not uniquely characterizes a distribution function.

- **Example:** Consider two distributions

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi x}} exp\{-(\log x)^2/2\}, \\ f_2(x) &= f_1(x)[1 + sin(2\pi \log x)], \end{aligned}$$

for $x > 0$. $E(X_1^k) = E(X_2^k)$ for $k = 1, 2, \cdots$.

3.4 Moment Generating Function

- **Definition: Converge in Distribution (Weak Convergence).** Let $\{X_n\}$ be a sequence of r.v.s with CDF's $\{F_n(x)\}$. Let X be a r.v. with CDF F_X. If $F_n(x) \to F_X(x)$ as $n \to \infty$ for all x's where $F_X(x)$ is **continuous**, we say X_n *converges in distribution* to X, denoted by

$$X_n \xrightarrow{d} X.$$

- **Theorem : Convergence of MGF.** Suppose $\{X_n, n = 1, 2, \cdots\}$ is a sequence of random variables, each with MGF $M_n(t)$ and CDF $F_n(x)$. Furthermore, suppose that

$$\lim_{n\to\infty} M_n(t) = M_X(t)$$

for all t in a neighborhood of 0, and $M_X(t)$ is a MGF of some random variable X with CDF F_X. Then

$$X_n \xrightarrow{d} X.$$

3.4 Moment Generating Function

- **Theorem: Monotone Convergence.** If $0 \le g_n(x) \le g_{n+1}(x)$ for $n \ge 1$, then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g_n(x)dF(x) = \int_{-\infty}^{\infty} \lim_{n\to\infty} g_n(x)dF(x).$$

- **Theorem: Dominated Convergence.** If $|g_n(x)| \le \bar{g}(x)$ for all $n \ge 1$, $\int_{-\infty}^{\infty} \bar{g}(x)dF(x) < \infty$, and $\lim_{n\to\infty} g_n(x) = g(x)$ except for $x \in N$, where N is a set with probability zero, then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g_n(x)dF(x) = \int_{-\infty}^{\infty} g(x)dF(x).$$

- **Theorem.** Let $F_X(x)$ and $F_Y(y)$ be two CDF's both of which have bounded support. Then $F_X(z) = F_Y(z)$ for all $z \in (-\infty, \infty)$ if and only if $E(X^k) = E(Y^k)$ for al integers $k = 1, 2, ...$

- **Example: Poisson Approximation.** The MGF of the binomial distribution $B(n, p)$ is

$$M_B(t) = (pe^t + 1 - p)^n,$$

and the MGF of the Poisson distribution $P(\lambda)$ is

$$M_P(t) = e^{\lambda(e^t - 1)}.$$

- Because

$$M_B(t) = (1 + \frac{np(e^t - 1)}{n})^n.$$

When $n \to \infty$ and $np \to \lambda$, we can use the Poisson(λ) distribution to approximate the binomial distribution.

- **Example:** A typesetter, on the average, makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two error in five pages?

- **Solution 1:**

  – Use binomial distribution: Assume that making error in typing a word follows a Bernoulli distribution with $p = 1/500$. Then the number of errors in 1,500 words X follows a binomial distribution $B(1500, p)$.

$$\begin{aligned} P(\text{"no more than two errors"}) &= P(X \leq 2) \\ &= \sum_{x=0}^{2} \binom{1500}{x} \left(\frac{1}{500}\right)^{x} \left(\frac{499}{500}\right)^{1500-x} \\ &= 0.4230. \end{aligned}$$

- **Solution 2:**
  - Use Poisson approximation: Let $\lambda = np = 3$, then

$$
\begin{aligned}
P(X \leq 2) &= \sum_{x=0}^{2} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} + e^{-\lambda}\frac{\lambda}{1!} + e^{-\lambda}\frac{\lambda^2}{2!} \\
&= 0.4232.
\end{aligned}
$$

3.5 Characteristic Function

- **Definition: Characteristic Function.** The *characteristic function* of a r.v. X with cdf $F_X(x)$ is defined as

$$\begin{aligned} \varphi_X(t) &= E(e^{itX}) \\ &= \int_{-\infty}^{\infty} e^{itx} dF_X(x), \end{aligned}$$

where $i = \sqrt{-1}$ is the imaginary unit, and

$$e^{itx} = cos(tx) + isin(tx).$$

- **Remarks:**

  - $\varphi_X(t) = M_X(it)$ if mgf $M_X(\cdot)$ exists.
  - $\varphi_X(0) = 1$.
  - Characteristic function always exists, because

$$|\varphi_X(t)| \leq \int_{-\infty}^{\infty} |e^{itx}| dF_X(x) = 1.$$

  - The characteristic function is the Fourier transform of the distribution function $F_X(x)$. For continuous r.v., pdf $f_X(x)$ can be recovered from the characteristic function by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t) dt.$$

3.5 Characteristic Function

- **Theorem:** Suppose the k-th moment of X exists. Then

$$\varphi_X^{(k)}(0) = i^k E(X^k).$$

- **Theorem: Uniqueness of CF.** Suppose two r.v. X and Y with $\varphi_X(t)$ and $\varphi_Y(t)$. Then X and Y are identically distributed if and only if $\varphi_X(t) = \varphi_Y(t)$ for **all** $t \in \mathbb{R}$.

- **Remarks:** It is important to check all t on the real line $\mathbb{R}$. But for mgfs, it is only necessary to check t in a neighborhood of zero.

- **Theorem : Convergence in CF.** Let $\{X_n\}$ be a sequence of r.v.'s with distribution functions $F_n(x)$ and characteristic functions $\{\varphi_n(t)\}$. Let X be a random variable with distribution function $F_X(x)$ and characteristic function $\varphi_X(t)$. Let $n \to \infty$.
  - If $F_n(x) \to F_X(x)$ for all continuous points of $F_X(x)$, then $\varphi_n(t) \to \varphi_X(t)$ for every $t \in \mathbb{R}$.
  - Further, if for every $t \in \mathbb{R}$, $\varphi_n(t) \to \varphi_X(t)$ and $\varphi_X(t)$ is continuous at $t = 0$, then $X_n \stackrel{d}{\to} X$.

- **Remarks:** Convergence in distribution is equivalent to convergence of characteristic functions.