Advanced Statistics: Theory and Methods

Dr. Nandini Kannan

email : nandini.kannan@plaksha.edu.in

Chapter 1

Random Variables and Distributions - Review

Let $\mathcal{S}$ denote the sample space for the experiment.

Definition: A **Random Variable** (rv) X is a mapping from the sample space $\mathcal{S}$ to the set of real numbers.

Definition: The cumulative distribution function (cdf) $F(x)$ of the random variable X is

$$F_X(x) = P(X \leq x), \qquad -\infty < x < \infty.$$

The function $F_X(x)$ satisfies the following three conditions:

1. $\lim_{x\to-\infty} F_X(x) = 0; \qquad \lim_{x\to\infty} F_X(x) = 1.$

2. $F_X(x)$ is a nondecreasing function of x; i.e. for all $a, b \in \mathcal{R}$, $a < b$, we

have $F_X(a) \leq F_X(b)$.

3. $F_X(x)$ is right continuous; i.e. $\lim_{x \downarrow x_0} F_X(x) = F_X(x_0)$.

1.1 Discrete Random Variables

Definition: A random variable is said to be **discrete** if it can assume finite or countably many values.

The **probability distribution** or **probability mass function** (pmf) of a discrete random variable is a formula, table or graph that associates a probability with each value of the random variable. The probability distribution is denoted by $p(x)$.

$$p(x) = P(X = x)$$

Properties of the PMF

1. $p(x) \geq 0, \forall x$

2. $\sum_x p(x) = 1$.

The cumulative distribution function of a discrete random variable is a step function with jumps of $p(x)$ at each point x in the **support** of X.

1.2 Continuous Random Variables

A **continuous random variable** is one that can assume values in an interval of the real line.

For continuous random variables, probabilities are computed from a smooth function called the **probability density function (pdf)**, denoted by **f(x)**. The probability that the random variable takes values in an interval is simply the area under f(x) between the end points of the specified interval.

Therefore the probability that the random variable X lies in the interval [a, b] is given by

$$P(a \le X \le b) = \int_a^b f(x)\ dx$$

With this definition of probabilities for a continuous random variable, observe that

$$P(X = k) = \int_k^k f(x)dx = 0.$$

The pdf must satisfy two properties

1. $f(x) \ge 0$ for all x.
2. $\int_{-\infty}^{\infty} f(x)dx = 1.$

Definition: The cumulative distribution function $F(x)$ of a continuous random variable X with pdf $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \qquad -\infty < x < \infty.$$

We have from the previous definition, that

$$f(x) = \frac{dF(x)}{dx} = F'(x).$$

1.3 Transformations

Let X be a random variable with cdf $F_X(x)$. Sometimes, we are interested not in X, but a transformed version of X. Let $Y = g(X)$ be a function of X.

Clearly (!), Y is also a random variable. We can describe the probabilistic behaviour of Y in terms of that of X. We have

$$P[Y \in A] = P[g(X) \in A] \;\text{ for any }\; A.$$

Let $\mathcal{X}$ be the sample space of X. Then

$$g : \mathcal{X} \to \mathcal{Y}$$

where $\mathcal{Y}$ is the sample space of the random variable Y.

We can define the inverse mapping

$$g^{-1} : \mathcal{Y} \to \mathcal{X}$$

where

$$g^{-1}(A) = \{x \in \mathcal{X}; g(x) \in A\}.$$

Therefore

$$\begin{aligned} P[Y \in A] &= P[g(X) \in A] \\ &= P[\{x \in \mathcal{X}; g(x) \in A\}] \\ &= P[X \in g^{-1}(A)]. \end{aligned}$$

This defines the probability distribution of Y. This distribution satisfies Kolmogorov's axioms.

If X is a discrete r.v. with pmf $p_X(.)$, then Y is also discrete. We have

$$\begin{aligned} p_Y(y) &= P(Y = y) = \sum_{x \in g^{-1}(y)} P(X = x) \\ &= \sum_{x \in g^{-1}(y)} p_X(x) \qquad for \;\; y \in \mathcal{Y}. \end{aligned}$$

Here

$$g^{-1}(y) = \{x \in \mathcal{X} : g(x) = y\}.$$

The pmf of Y may be easily obtained by first identifying $g^{-1}(y)$ and then summing the appropriate probabilities.

If X and Y are continuous r.v.'s, it is possible to find simple formulae for the cdf and the the pdf of Y in terms of the cdf and pdf of X.

We have

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) = P(g(X) \le y) \\
&= P[\{x \in \mathcal{X}; g(x) \le y\}] \\
&= \int_{\{x\in\mathcal{X}; g(x)\le y\}} f_X(x)dx.
\end{aligned}
$$

This is called the **method of distribution functions**.

Example: Let X be a random variable with pdf

$$
f_X(x) = \begin{cases} \frac{2x}{\pi^2}, & 0 < x < \pi; \\ 0, & \text{o.w.} \end{cases}
$$

Let $Y = \sin X$. The range of Y is $(0, 1)$.

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) = P[\sin X \le Y] \\
&= P[0 \le X \le x_1] + P[x_2 \le X \le \pi].
\end{aligned}
$$

where $x_1 = \sin^{-1} y, x_2 = \pi - \sin^{-1} y$.

Therefore

$$
\begin{aligned}
F_Y(y) &= \int_0^{x_1} \frac{2x}{\pi^2}dx + \int_{x_2}^{\pi} \frac{2x}{\pi^2}dx \\
&= \frac{x_1^2}{\pi^2} + 1 - \frac{x_2^2}{\pi^2}.
\end{aligned}
$$

■

Definition: The **support** of a random variable or distribution is the set of values for which the pdf (pmf) is nonzero:

$$\mathcal{X} = \{x : f_X(x) > 0\}.$$

Let $\mathcal{Y}$ be the sample space for Y. We have

$$\mathcal{Y} = \{y : y = g(x) \text{ for some } \ x \in \mathcal{X}\}.$$

If the transformation $x \to g(x)$ is monotone, we can obtain simple expressions for $F_Y(y)$.

Theorem 1.3.1. *Let $X \sim F_X(x)$ and $Y = g(X)$.*

(a) If g is an increasing function, then

$$F_Y(y) = F_X[g^{-1}(y)] \textit{ for } \ y \in \mathcal{Y}.$$

(b) If g is a decreasing function, then

$$F_Y(y) = 1 - F_X[g^{-1}(y)] \textit{ for } \ y \in \mathcal{Y}.$$

Proof: (a) If g is an increasing function, it is one-to-one and onto from $\mathcal{X} \to \mathcal{Y}$. In other words, each x goes to only one y, and each y comes from at most one x [one-to one], and for each $y \in \mathcal{Y}$, there is an $x \in \mathcal{X}$ such that $g(x) = y$ (onto). We have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P[\{x \in \mathcal{X}; g(x) \leq y\}] \end{aligned}$$

$$
\begin{aligned}
&= P[\{x \in \mathcal{X}; x \leq g^{-1}(y)\}] \\
&= F_X[g^{-1}(y)].
\end{aligned}
$$

(a) If g is a decreasing function, we have

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\
&= P(X \geq g^{-1}(Y)] \\
&= 1 - F_X[g^{-1}(y)].
\end{aligned}
$$

■

Example: Let X be a random variable with pdf

$$
f_X(x) = \begin{cases} e^{-x}, & x > 0; \\ 0, & \text{o.w.} \end{cases}
$$

Let $Y = \ln X$. Therefore $g^{-1}(y) = e^y$. We have

$$
\frac{d}{dx} g(x) = \frac{d}{dx} \ln x = \frac{1}{x} > 0.
$$

Therefore g is an increasing function. As x ranges from 0 to ∞, y ranges from $-\infty$ to ∞. We have

$$
\begin{aligned}
F_Y(y) &= F_X[g^{-1}(y)] \\
&= F_X(e^y).
\end{aligned}
$$

We have

$$
F_X(x) = \int_0^x e^{-t} dt = 1 - e^{-x}.
$$

Substituting for $F_X(x)$, we have

$$F_Y(y) = 1 - e^{-e^y}.$$

■

Example: Let X be a random variable with pdf

$$f_X(x) = \begin{cases} \alpha x^{\alpha-1}, & 0 < x < 1; \\ 0, & \text{o.w.} \end{cases}$$

Let $Y = -\ln X$. Therefore $g^{-1}(y) = e^{-y}$. We have

$$\frac{d}{dx} g(x) = \frac{d}{dx}(-\ln x) = -\frac{1}{x} < 0.$$

Therefore g is a decreasing function. As x ranges from 0 to 1, y ranges from 0 to ∞. We have

$$F_X(x) = \int_0^x \alpha t^{\alpha-1} dt = x^\alpha.$$

$$\begin{aligned} F_Y(y) &= 1 - F_X[g^{-1}(y)] \\ &= 1 - F_X(e^{-y}) = 1 - e^{-\alpha y}. \end{aligned}$$

■

If the pdf of Y is continuous, it can be obtained by differentiating the cdf.

Theorem 1.3.2. *Let $X \sim f_X(x)$ and $Y = g(X)$, where g is a monotone function. Suppose $f_X(x)$ is continuous on $\mathcal{X}$ and $g^{-1}(y)$ has a continuous* *derivative on $\mathcal{Y}$. Then*

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right|, & y \in \mathcal{Y}; \\ 0, & o.w. \end{cases}$$

Proof: Use the chain rule.

■

Example Let X be a non-negative continuous random variable with pdf $f_X(x)$. Let $Y = X^\alpha, \alpha > 0$. Then

$$\frac{d}{dx} g(x) = \frac{d}{dx} x^\alpha = \alpha x^{\alpha - 1} > 0,$$

which implies $g(.)$ is an increasing function. We have $g^{-1}(y) = y^{\frac{1}{\alpha}}$. Using the theorem, we have

$$f_Y(y) = \begin{cases} f_X\left[y^{\frac{1}{\alpha}}\right] \frac{1}{\alpha} y^{\frac{1}{\alpha}-1}, & y > 0; \\ 0, & \text{o.w.} \end{cases}$$

■

In many applications, the function g may be neither increasing nor decreasing. However, the function may be monotone over certain intervals. It may be possible to divide $\mathcal{X}$ into sets $A_1, \ldots, A_k$ such that $g(.)$ is monotone on each set. We can then modify the previous theorem:

Theorem 1.3.3. *Let $X \sim f_X(x)$ and $Y = g(X)$. Suppose there exists a partition $A_0, A_1, \ldots, A_k$ of $\mathcal{X}$ such that $P[X \in A_0] = 0$ and $f_X(x)$ is con-*

tinuous on each A_i. Suppose there exist functions $g_1(x), \ldots, g_k(x)$ defined on $A_1, A_2, \ldots, A_k$ respectively satisfying

(i) $g(x) = g_i(x), x \in A_i$*;*

(ii) $g_i(x)$ *is monotone on* A_i*;*

(iii) The set $\mathcal{Y} = \{y : y = g_i(x) \quad \text{for some } x \in A_i\}$ *is the same for each* $i = 1, \ldots, k$*; and*

(iv) $g_i^{-1}(y)$ *has a continuous derivative on* $\mathcal{Y}$ *for each* $i = 1, 2, \ldots, k$*.*

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X[g_i^{-1}(y)] \left| \dfrac{d}{dy} g_i^{-1}(y) \right|, & y \in \mathcal{Y}; \\ 0, & o.w. \end{cases}$$

Here A_0 represents the exceptional set.

Proof: Use the chain rule.

■

Example Let $X \sim f_X(x)$ and $Y = |X|$. We have $g(x) = |x|$. Let $A_1 = (-\infty, 0)$, $A_2 = (0, \infty)$. On A_1, we have

$$g_1^{-1}(y) = -y,$$

and on A_2, we have

$$g_2^{-1}(y) = y.$$

We have

$$
\begin{aligned}
f_Y(y) &= f_X(-y)|-1| + f_X(y)|1| \\
&= f_X(-y) + f_X(y).
\end{aligned}
$$

If

$$
f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathcal{R},
$$

then

$$
f_Y(y) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-y^2/2}, & y > 0; \\ 0, & \text{o.w.} \end{cases}
$$

$\blacksquare$

Example Let $X \sim f_X(x)$ and $Y = X^2$. We have $g(x) = x^2$. We know that $g(x)$ is monotone on $A_1 = (-\infty, 0)$ and, $A_2 = (0, \infty)$. Let $A_0 = \{0\}$. On A_1 we have

$$
g_1^{-1}(y) = -\sqrt{y},
$$

and on A_2, we have

$$
g_2^{-1}(y) = \sqrt{y}.
$$

We have

$$
\begin{aligned}
f_Y(y) &= f_X(-\sqrt{y}) \left| \frac{-1}{2\sqrt{y}} \right| + f_X(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| \\
&= \frac{1}{2\sqrt{y}} [f_X(-\sqrt{y}) + f_X(\sqrt{y})].
\end{aligned}
$$

If

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathcal{R},$$

then

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} [2e^{-y/2}] \\ &= \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y/2}, \qquad y > 0. \end{aligned}$$

■

1.4 Moments of Random Variables

Definition Let X be a discrete random variable with probability mass function $p(x)$. The **Expected value** or **mean** of X is given by

$$\mu = E(X) = \sum_x x \;\; p(x)$$

provided that $\sum_x |x| p(x) < \infty$. If the sum diverges,the expectation is undefined.

Definition Let X be a continuous random variable with probability density function $f(x)$. The **Expected value** or **mean** of X is given by

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

provided that $\int |x| f(x) \; dx < \infty$. If the integral diverges,the expectation is undefined.

Theorem 1.4.1. *Let X be a random variable. The expected value of the random variable $g(X)$ is*

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x) \; p(x)$$

if X is discrete, and

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x) \; f(x) \, dx.$$

if X is continuous.

Definition: Let X be a discrete random variable with probability mass function $p(x)$ and mean μ. The **Variance** σ^2 of X is

$$\sigma^2 = E(X-\mu)^2 = \sum_x (x-\mu)^2 p(x).$$

Definition: Let X be a continuous random variable with probability density function $f(x)$ and mean μ. The **Variance** σ^2 of X is

$$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \; f(x) \, dx.$$

The positive square root of the variance is called the **standard deviation**.

Theorem 1.4.2. *The variance of a random variable X is*

$$\sigma^2 \;=\; E(X^2) - \mu^2.$$

Definition: The $k-th$ *moment* of a random variable X is

$$\mu_k^{'} = E(X^k).$$

The $k-th$ *central moment* is

$$\mu_k = E[(X - \mu)]^k,$$

where $\mu = \mu_1^{'} = E(X)$.

The first moment of a random variable is its' mean, while the first central moment is 0. The second central moment is the variance of X. The third central moment μ^3 measures the symmetry of the distribution of X about its' mean. The dimension-free **measure of skewness** is given by

$$\nu_1 = \frac{\mu^3}{\sigma^3}.$$

The index is zero when the distribution is symmetric about the mean. Negative values are associated with distributions **skewed to the left**, whereas ν_1 tends to be positive when the distribution of X is **skewed to the right**.

The fourth central moment μ^4 provides an indication of the "peakedness" or "kurtosis" of a distribution. A dimension-free **measure of kurtosis** is

$$\nu_2 = \frac{\mu^4}{\sigma^4}.$$

The peakedness of a distribution is compared to the Gaussian distribution which has $\nu_2 = 3$. A distribution is "less peaked" (platykurtic) if $\nu_2 < 3$ and "more peaked" (leptokurtic) if $\nu_2 > 3$.

Definition: Let X be a random variable with cdf $F_X(.)$. The **moment generating function** (mgf) of the random variable X, denoted by $M_X(t)$ is defined as

$$M_X(t) = E(e^{tX}), \tag{1.1}$$

provided the expectation exists in some neighbourhood of 0. We have

$$m_X(t) = \begin{cases} \sum_x e^{tx} p(x), & \text{if } X \text{ is discrete;} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Example Let

$$f_X(x) = \frac{1}{2} e^{-x/2}, \qquad x > 0.$$

We have

$$\begin{aligned} M_X(t) &= \frac{1}{2} \int_0^\infty e^{tx} e^{-x/2} dx \\ &= \frac{1}{2} \int_0^\infty e^{\left(t - \frac{1}{2}\right)x} dx \\ &= \frac{1}{1-2t} \qquad if \quad t < \frac{1}{2}. \end{aligned}$$

If $t > \frac{1}{2}$, the integral is infinite.

■

Theorem 1.4.3. *If the mgf $M_X(t)$ of X exists in a neighbourhood of 0, the derivatives of all orders exist at $t = 0$ and may be obtained by differentiating under the integral (or summation), i.e.*

$$M_X^n(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = E(X^n).$$

<u>Proof:</u> We have

$$\begin{aligned}
\frac{d}{dt}M_X(t) &= = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)dx \\
&= \int_{-\infty}^{\infty} \frac{d}{dt}(e^{tx}) f(x)dx \\
&= E(Xe^{tX}).
\end{aligned}$$

$$\frac{d}{dt}\, M_X(t)|_{t=0} = E(X).$$

■

<u>Remark:</u> Since

$$\begin{aligned}
M_X(t) &= E(e^{tX}) = E\left(1 + tX + \frac{t^2}{2!}X^2 + \dots\right) \\
&= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \dots
\end{aligned}$$

$E(X^n)$ is the coefficient of $t^k/k!$ in the above expansion.

Chapter 2

Special Families of Distributions

2.1 Discrete Population Models

2.1.1 Bernoulli and Binomial Distribution

Definition: Consider an experiment that can result in one of two outcomes. We classify these outcomes as *Success* and *Failure*. The probability of Success is denoted by p. Such a trial is called a **Bernoulli trial**.

Examples

1. Toss a fair coin: Heads and Tails
2. Testing a blood sample for Absence or Presence of a particular disease
3. Testing items in a factory: Defective or Nondefective

Definition: For any Bernoulli trial, we define the random variable X as follows: if the trial results in a Success, $X = 1$; otherwise $X = 0$. This is called the Bernoulli random variable and its' pmf is given by:

$$p(x) = \begin{cases} 1-p, & \text{if } x = 0; \\ p, & \text{if } x = 1; \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x = 0, 1; \\ 0, & \text{otherwise.} \end{cases}$$

We have $\mu = E(X) = p$ and $\sigma^2 = Var(X) = p(1-p)$. Here p is a **parameter** of the distribution.

Binomial Experiment: A binomial experiment is an experiment that has the following properties:

1. The experiment consists of n identical trials.
2. Each trial can result in one of two possible outcomes. These outcomes will be classified as Success S, and Failure F.
3. The probability of success on a single trial is equal to p and remains constant from trial to trial. The probability of failure is then $1 - p$ which is denoted by q.
4. The trials are independent.

We are interested in X, the number of successes in the n trials. The random variable X can take values $0, 1, \ldots, n$. The random variable X is called a **Binomial random variable**.

We usually write

$$X \sim Bin(n, p).$$

We have

$$p(k) = P(X = k) = \binom{\text{n}}{\text{k}} p^k q^{n-k}, \qquad k = 0, 1, \ldots, n$$

where

$$\binom{\text{n}}{\text{k}} = \frac{n!}{k!(n-k)!}$$

Theorem 2.1.1. *If $Y \sim Bin(n,p)$, then the mgf of Y is*

$$M_Y(t) = (pe^t + q)^n. \tag{2.1}$$

Proof: We have

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k} \\ &= (pe^t + q)^n. \end{aligned}$$

The last equality follows from the Binomial theorem.

■

Result:

The **mean** of the Binomial random variable $X \sim Bin(n, p)$ is given by

$$\mu = E(X) = n\ p$$

The **variance** of the binomial random variable is

$$\sigma^2 = n\ p\ (1 - p)$$

2.1.2 Hypergeometric Distribution

The **Hypergeometric Experiment** has the following properties:

1. A random sample of n items is selected **without replacement** from N items. M of the N items are of Type I (Successes) and $N - M$ are of Type II (failure).

Let X denote the number of successes in the sample. X is said to be a hypergeometric r.v.

The probability of X taking the value k is given by

$$h(k : N, n, M) = \frac{\begin{pmatrix} \text{M} \\ \text{k} \end{pmatrix} \begin{pmatrix} \text{N-M} \\ \text{n-k} \end{pmatrix}}{\begin{pmatrix} \text{N} \\ \text{n} \end{pmatrix}},$$

where

$$\max\{0, n - (N - M)\} \leq k \leq \min\{n, M\}$$

<u>Result</u>

The **mean** of the Hypergeometric random variable X is given by

$$\mu = E(X) = \frac{n\ M}{N}$$

The **variance** is

$$\sigma^2 = \frac{N-n}{N-1}\, n \left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right).$$

The term $\frac{N-n}{N-1}$ is called the **finite population factor**.

Relation between the Hypergeometric and the Binomial

Consider the hypergeometric distribution. Let $p = M/N$ be the proportion of Type I items in the population. Suppose N is large with n and p remaining fixed, i.e.

$$N \to \infty, \qquad M \to \infty, \qquad \frac{M}{N} \to p.$$

Then the hypergeometric probabilities converge to the binomial probabilities. For large samples, there is practically no difference between sampling with and without replacement.

We have

$$P(X = k|N, M, n) = \frac{\begin{pmatrix} M \\ k \end{pmatrix}\begin{pmatrix} N-M \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

$$= \begin{pmatrix} n \\ k \end{pmatrix} \frac{M!(N-M)!(N-n)!}{(M-k)!(N-M-n+k)!N!}$$

$$= \binom{n}{k} \frac{M(M-1)\dots(M-k+1)(N-M)\dots(N-M-n+k+1)}{N^k \left[1.\left(1-\frac{1}{N}\right)\dots\left(1-\frac{(n-1)}{N}\right)\right]}$$

$$\to \binom{n}{k} p^k (1-p)^{n-k}.$$

If n is small relative to N, the probability of success does not change much from draw to draw. So the binomial distribution may be used as an approximation if $n/N \leq 0.05$.

The binomial distribution can arise when a simple random sample is selected **with replacement**, and for each unit selected it is determined whether or not the unit possesses a specified property.

The hypergeometric arises when the samples are drawn **without replacement**.

2.1.3 Geometric Distributions

Consider an experiment with the following properties:

1. The experiment consists of a sequence of identical trials.
2. Each trial can result in one of two possible outcomes. These outcomes will be classified as Success S, and Failure F.
3. The probability of success on a single trial is equal to p and remains constant from trial to trial. The probability of failure is then $1 - p$ which is denoted by q.
4. The trials are independent. The experiment terminates as soon as the *first* success occurs.

We are interested in X, the number of trials required to obtain the *first* success. The random variable X can take values $1, 2, \ldots,$. The random variable X is called a **Geometric random variable**.

The probability of X taking the value x is given by

$$P(X = x) = pq^{x-1}, \qquad x = 1, 2, \ldots.$$

The cdf for the geometric distribution is given by

$$F(x) = 1 - q^x, \qquad x = 1, 2, \ldots.$$

Also, $F(x) = 0, x < 1$, and $F(x) = F(m)$ when $m \leq x < m + 1, m = 1, 2, \ldots$.

Result

The **mean** of the geometric random variable is given by

$$\mu = E(X) = \frac{1}{p}$$

The **variance** is

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

In the general formula the term

$$p^k q^{x-k}$$

is the probability of obtaining a success on the $x-$th trial preceded by $k-1$ successes and $x-k$ failures in some specified order. The term

$$\begin{pmatrix} \text{x-1} \\ \text{k-1} \end{pmatrix}$$

counts the number of possible strings.

2.1.4 Poisson Distribution and the Poisson Process

The distribution of count data is often modeled using the **Poisson distribution**.

Let N_t be the total number of occurrences in an interval of length t, i.e. in $(0, t)$.

The Poisson distribution is derived from the **Poisson process** which possesses the following properties, called the **Poisson postulates**:

1. $N_0 = 0$.

2. The probability of exactly one occurrence during a very small interval $[t, t + h]$ of length h, is proportional to the length of the interval, i.e.

$$P(N_h = 1) = \lambda h + o(h).$$

3. The probability of more than one occurrence in the interval $[t, t + h]$ is $o(h)$.

4. The number of occurrences in non-overlapping intervals are independent.

Remark: Here $o(h)$ is a quantity of smaller order of magnitude than h, i.e.

$$\frac{o(h)}{h} \to 0 \quad \text{as} \quad h \to 0.$$

Result: Under these postulates, we have

$$P(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \qquad k = 0, 1, \ldots. \tag{2.2}$$

Proof: We have

$$\begin{aligned}
P_n(t+h) &= P(N_{t+h} = n) \\
&= P(N_t = n, N_{t+h} - N_t = 0) + P(N_t = n-1, N_{t+h} - N_t = 1) \\
&= \sum\nolimits_{k=2}^{n} P(N_t = n-k, N_{t+h} - N_t = k) \\
&= P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + o(h) \qquad \text{using P1, P3} \\
&= (1 - \lambda h)P_n(t) + \lambda h P_{n-1}(t) + o(h) \qquad \text{using P2.}
\end{aligned}$$

Therefore

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}.$$

Letting $h \to 0$, we have

$$P_n^{'}(t) = -\lambda P_n(t) + \lambda P_{n-1}(t). \tag{2.3}$$

For $n = 0$, we get

$$P_0(t+h) = P_0(t)[1 - \lambda h] + o(h).$$

Rearranging the terms as before, and letting $h \to 0$, we have

$$P_0^{'}(t) = -\lambda P_0(t)$$

$$\Rightarrow P_0(t) = Ke^{-\lambda t}.$$

Since $P_0(0) = 1$, we have

$$P_0(t) = e^{-\lambda t}.$$

Using the differential equation in (2.3) and the initial conditions, we get

$$\begin{aligned} P_1^{'}(t) &= -\lambda P_1(t) + \lambda P_0(t) \\ &= -\lambda P_1(t) + \lambda e^{-\lambda t}. \end{aligned}$$

This implies

$$P_1(t) = \lambda t e^{-\lambda t}.$$

Proceeding in a similar fashion, we can derive the general case:

$$\begin{aligned} & P_n^{'}(t) + \lambda P_n(t) = \lambda P_{n-1}(t) \\ & \Rightarrow e^{\lambda t}[P_n^{'}(t) + \lambda P_n(t)] = \lambda e^{\lambda t} P_{n-1}(t) \\ & \Rightarrow \frac{d}{dt}[e^{\lambda t} P_n(t)] = \lambda e^{\lambda t} P_{n-1}(t) \\ = \;& \lambda e^{\lambda t} \left[\frac{e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} \right] \\ & \Rightarrow P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \qquad n = 0, 1, \ldots. \end{aligned}$$

■

Definition: A random variable is said to have a Poisson distribution $P(\lambda)$ if

$$p_X(x|\lambda) = P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, \ldots; \lambda > 0. \tag{2.4}$$

The probabilities are clearly non-negative. Further

$$
\begin{aligned}
\sum_{k=0}^{\infty} p_X(x|\lambda) &= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^x}{x!} \\
&= e^{-\lambda}e^{\lambda} = 1,
\end{aligned}
$$

since the infinite sum is the Taylor expansion for e^{λ}.

Theorem 2.1.2. *If $X \sim P(\lambda)$, then the mgf of X is*

$$
M_X(t) = e^{-\lambda}e^{\lambda e^t}. \tag{2.5}
$$

<u>Proof:</u> We have

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=0}^{\infty} \frac{e^{-\lambda}(\lambda e^t)^x}{x!} \\
&= e^{-\lambda}\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
&= e^{-\lambda}e^{\lambda e^t}.
\end{aligned}
$$

The last equality follows from the Taylor expansion.

■

Result

The **mean** of the Poisson random variable is given by

$$\mu = E(X) = \lambda t$$

The **variance** is

$$\sigma^2 = \mu$$

2.1.5 Poisson as a limiting form of the Binomial

Theorem: Let $X \sim Bin(n, p)$. Then

$$P(X = x) \to \frac{e^{-\mu}\mu^x}{x!},$$

as $n \to \infty, p \to 0$ with $np \to \mu$.

2.2 Continuous Population Models

2.2.1 Uniform or Rectangular Distribution

Definition: A random variable X is said to have a uniform distribution on the interval (a,b) if its pdf is given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a}, & a < x < b; \\ 0, & \text{otherwise.} \end{cases}$$

We write $X \sim U(a,b)$

Theorem 2.2.1. *Let $X \sim U(a,b)$. Then*

$$\mu = \frac{a+b}{2} \tag{2.6}$$

and

$$\sigma^2 = \frac{(b-a)^2}{12}. \tag{2.7}$$

■

Theorem 2.2.2. *Let $X \sim U(a,b)$. Then*

$$M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}, \qquad t \neq 0. \tag{2.8}$$

■

2.2.2 Gamma Distribution

The Gamma distribution has been widely used in the reliability and survival literature to model the lifetime of mechanical or biological systems.

<u>Definition:</u> The random variable X is said to have a Gamma Distribution if its pdf is given by

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

where $\alpha, \beta > 0$. The parameter α is called the shape parameter (influences peakedness), and β is called the scale parameter (influences spread). We write $X \sim G(\alpha, \beta)$. Here $\Gamma(\alpha)$ is the **gamma function** defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Integrating the function by parts, we get

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1).$$

When $\alpha = n$, a positive integer, repeated applications of the above result yield the following:

$$\Gamma(n) = (n-1)!$$

Theorem 2.2.3. *If $X \sim G(\alpha, \beta)$, we have*

$$M_X(t) = \frac{1}{(1-\beta t)^\alpha} \qquad t < \frac{1}{\beta}. \tag{2.9}$$

■

Theorem 2.2.4. *If $X \sim G(\alpha, \beta)$, we have*

$$\mu = \alpha\beta \qquad\qquad \sigma^2 = \alpha\beta^2. \tag{2.10}$$

■

Remarks:

1. When $\alpha = 1$, we get the exponential distribution with pdf

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

   where $\beta > 0$. We write $X \sim E(\beta)$.

2. If $\alpha = p/2$, where p is an integer and $\beta = 2$, then X is said to have the ***chi squared distribution with p degrees of freedom***. We write $X \sim \chi^2(p)$.

The Lack of Memory Property of the Exponential

The exponential distribution has the lack of memory property. In other words, regardless of the age of the product, there is no wearing out and the product is "as good as new". Mathematically, for $s > t \geq 0$,

$$P(X > s|X > t) = P(X > s - t). \tag{2.11}$$

The exponential distribution is the only distribution with the lack of memory property. This is called a characterization of the distribution.

Relationship to the Poisson Process

If events follow a Poisson process with rate λ (the average number of events per unit time), and if T represents the waiting time from any starting point until the occurrence of the next event, then $T \sim E(1/\lambda)$. We have

$$\begin{aligned} P(T > t) &= P(\text{no occurrences in interval of length } t) \\ &= 1 - e^{-\lambda t}. \end{aligned}$$

Let T_1 represent the waiting time until the first event, T_2 the waiting time between the occurrence of the first and second event etc. Then $T_1 + \ldots T_r$ has a Gamma distribution with parameter $\alpha = r$ and $\beta = 1/\lambda$, i.e. the waiting time until r events have occurred is distributed as a Gamma random variable.

Theorem 2.2.5. *If $X \sim E(\beta)$, then $Y = X^{\frac{1}{\gamma}}$ has a Weibull distribution with parameters γ and β, with pdf given by*

$$f_Y(y|\gamma, \beta) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta}, \qquad y > 0; \gamma, \beta > 0. \tag{2.12}$$

■

Applications of the Gamma and Exponential Distributions

Example: The response time X at a certain on-line computer terminal (the elapsed time between the end of a user's inquiry and the beginning of the system's response to that inquiry) has an exponential distribution with average response time equal to 5 seconds.

(a) Find the probability that the response time is at most 10 seconds.

(b) Find the probability that the response time exceeds 15 seconds.

Example: Assume that arrival times at a drive-through window follow a Poisson process with an average rate $\lambda = 0.2$ arrivals per minute. Find the probability that the third customer arrives within 20 minutes of opening.

2.2.3 Normal or Gaussian Distribution

The most well known and widely used continuous random variable is the **Normal** random variable. The form of the distribution was discovered early in the history of probability as an approximation to binomial probabilities by Abraham de Moivre. Laplace and Gauss proposed the distribution as a "law of errors" to describe the variability of measurement errors in the physical sciences.

Definition: The random variable X is said to have a normal distribution if its probability density function is given by

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty, \tag{2.13}$$

where μ and σ^2 are the parameters of the distribution. We can show that

$$E(X) = \mu \qquad Var(X) = \sigma^2.$$

We write

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

Theorem 2.2.6. *Suppose $X \sim N(\mu, \sigma^2)$. Then*

$$Z = \frac{(X - \mu)}{\sigma} \sim N(0, 1)$$

the **standard normal distribution**.

Proof: We have

$$Z = \frac{X-\mu}{\sigma} \Rightarrow X = \sigma Z + \mu.$$

This implies $dx = \sigma dz$, and

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} |\sigma| = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \tag{2.14}$$

■

To show the function in (2.14) is a valid pdf, we need to show

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1.$$

We have

$$\begin{aligned}
\left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz\right]^2 &= \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz\right)\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt\right) \\
&= \frac{4}{2\pi} \int_0^{\infty} \int_0^{\infty} e^{-(z^2+t^2)/2} dz dt.
\end{aligned}$$

We can make a transformation to polar coordinates. Let

$$z = r\cos\theta \qquad t = r\sin\theta.$$

Then

$$dzdt = rd\theta dr \qquad t^2 + z^2 = r^2.$$

The double integral becomes

$$\begin{aligned}
\frac{4}{2\pi} \int_0^{\infty} \int_0^{\pi/2} e^{-r^2/2} r d\theta dr &= \frac{4}{2\pi} \frac{\pi}{2} \int_0^{\infty} r e^{-r^2/2} dr \\
&= \int_0^{\infty} r e^{-r^2/2} dr = -e^{-r^2/2}\Big|_0^{\infty} = 1.
\end{aligned}$$

$\blacksquare$

Remark: We can use this integral to show that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{2.15}$$

We have shown that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1 \Rightarrow \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{1}{2}.$$

Let

$$t = z^2/2 \Rightarrow z^2 = 2t \qquad \Rightarrow dt = zdz.$$

We have

$$\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2t}} e^{-t} dt = \frac{1}{2}$$
$$\Rightarrow \Gamma\left(\frac{1}{2}\right) = \int_{0}^{\infty} t^{-1/2} e^{-t} dt = \frac{2\sqrt{\pi}}{2} = \sqrt{\pi}.$$

$\blacksquare$

Remarks:

1. The mode (point on the horizontal axis where the curve is maximum) occurs at $x = \mu$. [Differentiate the pdf wrt x and set the derivative equal to 0.]

2. The graph of this function is a symmetric bell shaped curve, with the point of symmetry being μ.

3. The points of inflection occur at $x = \mu \pm \sigma$. The inflection points are where the curve changes from concave to convex.

4. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.

Theorem 2.2.7. *If $Z \sim N(0,1)$, we have*

$$M_Z(t) = e^{t^2/2}. \tag{2.16}$$

Proof: We have

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz} e^{-z^2/2} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}[z^2 - 2zt]\right] dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}[z^2 - 2zt + t^2]\right] \exp\left[t^2/2\right] dz \\
&= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z-t)^2\right] dz \\
&= e^{t^2/2},
\end{aligned}$$

since the last integral is the $N(t,1)$ pdf.

■

Theorem 2.2.8. *If $X \sim N(\mu, \sigma^2)$, we have*

$$M_X(t) = e^{\mu t} e^{\frac{t^2\sigma^2}{2}}. \tag{2.17}$$

Proof: We have

$$\begin{aligned} M_X(t) &= E(e^{tX}) = E[e^{t(\sigma Z+\mu)}] \\ &= E[e^{\sigma t Z}]e^{\mu t} \\ &= \exp[\mu t]\exp[(\sigma t)^2/2]. \end{aligned}$$

■

Theorem 2.2.9. *If $X \sim N(\mu, \sigma^2)$, we have*

$$E(X) = \mu \qquad\qquad Var(X) = \sigma^2. \tag{2.18}$$

■

Definition: If $X \sim N(\mu, \sigma^2)$, the cumulative distribution function is given by

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \tag{2.19}$$

Theorem 2.2.10. *Let $Z \sim N(0,1)$. Then $Y = Z^2$ has the chi-squared distribution with 1 degree of freedom [$G(\alpha = 1/2, \beta = 2)$].*

Proof: The pdf of Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty.$$

Let $Y = Z^2$. This defines a transformation between the value of Z and Y that is not one-to-one.

The inverse solutions of $y = z^2$ are $z = \pm\sqrt{y}$. Let $g_1^{-1}(y) = -\sqrt{y}$ and $g_2^{-1}(y) = \sqrt{y}$.

The Jacobian of the transformation is given by

$$J_1 = \frac{d}{dy}(-\sqrt{y}) = \frac{-1}{2\sqrt{y}}; \qquad J_2 = \frac{d}{dy}(\sqrt{y}) = \frac{1}{2\sqrt{y}}.$$

The pdf of Y is given by

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} \left| \frac{-1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-y/2} \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{\sqrt{2\pi}}\, y^{1/2-1} e^{-y/2}$$

for $y > 0$. This is the pdf of a chi-squared random variable with 1 degree of freedom.

■

Normal Approximation to the Binomial

Let $X \sim B(n, p)$. Then

$$Z = \frac{X - np}{\sqrt{npq}}$$

is approximately distributed as a standard normal random variable as $n \to \infty$.

Recall

$$M_X(t) = (pe^t + q)^n.$$

Let

$$Z = \frac{X - np}{\sqrt{npq}}.$$

Then

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) = \exp\left[\frac{-npt}{\sqrt{npq}}\right] E\left(\exp\left[\frac{tX}{\sqrt{npq}}\right]\right) \\
&= \exp\left[\frac{-npt}{\sqrt{npq}}\right]\left[p\exp\left(\frac{t}{\sqrt{npq}}\right) + q\right]^n \\
&= \left[p\exp\left(\frac{qt}{\sqrt{npq}}\right) + q\exp\left(\frac{-pt}{\sqrt{npq}}\right)\right]^n \\
&= \left[1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \\
&\to e^{t^2/2} \qquad \text{as } n \to \infty.
\end{aligned}$$

■

2.2.4 Beta Distribution

Definition: The random variable X is said to have a Beta Distribution if its pdf is given by

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 < x < 1 \tag{2.20}$$

where $\alpha, \beta > 0$. We write $X \sim Beta(\alpha,\beta)$. Here $B(\alpha,\beta)$ is the **beta function** defined by

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{2.21}$$

Remarks:

1. Since the range of X is the unit interval, the Beta distribution is used to model proportions.

2. If $\alpha = \beta = 1$, we get the Uniform distribution.

Theorem 2.2.11. *If $X \sim Beta(\alpha, \beta)$, we have*

$$E(X^n) = \frac{B(n+\alpha, \beta)}{B(\alpha, \beta)} \qquad n > -\alpha. \tag{2.22}$$

<u>**Proof:**</u> We have

$$\begin{aligned} E(X^n) &= \frac{1}{B(\alpha,\beta)} \int_0^1 x^n x^{\alpha-1}(1-x)^{\beta-1} dx \\ &= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1} dx. \end{aligned}$$

The integral is the kernel of a $B(n+\alpha, \beta)$ random variable provided

$$n + \alpha > 0 \Rightarrow n > -\alpha.$$

Therefore

$$E(X^n) = \frac{B(n+\alpha, \beta)}{B(\alpha, \beta)}.$$

■

Theorem 2.2.12. *If $X \sim Beta(\alpha, \beta)$, we have*

$$E(X) = \frac{\alpha}{\alpha+\beta} \qquad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{2.23}$$

■

2.2.5 Cauchy Distribution

Definition: The random variable X is said to have a Cauchy Distribution if its pdf is given by

$$f(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2} \qquad -\infty < x < \infty \tag{2.24}$$

where $\theta \in \mathcal{R}$. We write $X \sim C(\theta)$.

We have

$$\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{1}{1+(x-\theta)^2} = \frac{1}{\pi}\arctan(x-\theta)|_{-\infty}^{\infty} = 1.$$

Remarks:

1. The Cauchy is a symmetric bell shaped distribution.

2. It has thicker tails than the normal.

Theorem 2.2.13. *Let $X \sim C(\theta)$. The median is given by θ.*

Proof: The median of an absolutely continuous distribution is defined as the value x for which

$$P(X \leq x) = 0.5.$$

For the Cauchy distribution, we have

$$F_X(x) \;=\; \int_{-\infty}^{x} \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$$

$$
\begin{aligned}
&= \frac{1}{\pi}\arctan(t-\theta)\Big|_{-\infty}^{x} \\
&= \frac{1}{2}+\frac{1}{\pi}\arctan(x-\theta). \qquad (2.25)
\end{aligned}
$$

Setting $F_X(x) = 0.5$, we have

$$x = \theta.$$

■

2.2.6 Lognormal Distribution

Definition: If X is a random variable that is positive, and $Y = \log X$ is normally distributed, then X is said to have a Lognormal distribution with pdf given by

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}, \qquad x>0, \qquad (2.26)$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$.

Remarks:

1. To show that the pdf integrates to 1, use the transformation $Y = \log X$.

2. We have

   $$E(X) = e^{\mu+(\sigma^2/2)},$$

   and

   $$Var(X) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}.$$

These expressions can be obtained by using the relationship to the normal distribution.

3. The lognormal is a right skewed distribution that resembles the Gamma.

2.2.7 Laplace or Double Exponential Distribution

<u>Definition:</u> The random variable X is said to have the double exponential distribution if its pdf is given by

$$f(x|\mu,\sigma) = \frac{1}{2\sigma}e^{-\frac{|x-\mu|}{\sigma}} \qquad x \in \mathcal{R}, \tag{2.27}$$

where $\mu \in \mathcal{R}$ and $\sigma > 0$.

<u>Remarks:</u>

1. The double exponential is a symmetric distribution with heavier tails than the normal. It is not bell shaped.

2. We have

$$E(X) = \mu \qquad Var(X) = 2\sigma^2.$$

3. It is not differentiable at $x = \mu$.

2.2.8 Chi-Squared Distribution

The pdf of the **chi-squared** random variable with ν degrees of freedom is

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

where ν is a positive integer.

Clearly this is a special case of the Gamma.