BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Measure and Probability Theory

Course Code: 23ST02CC

Unit-III

Probability Theory

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. S. Soundarya
Guest Faculty
Department of Statistics

UNIT - III

PROBABILITY THEORY

Introduction to Probability

The mathematical definition of probability of an event is defined as the ratio of the number of cases in its favour to the total number of cases. Consider S as a sample space and E be an event such that n(S) = n, n(E) = m and each outcome is equally likely. Then

$$p(E) = \frac{n(E)}{n(S)} = \frac{m}{n} = \frac{\text{No. of favourable outcomes of E}}{\text{Total number of possible outcomes}}$$

Probability Terms and Definitions

Sample Point

It denotes one of the possible outcomes. For example, in a deck of cards 4 of hearts is a sample point, similarly, the queen of clubs is a sample point.

Experiment or Trial

A series of trials where the results are always unpredictable. For example, the tossing of a coin, picking a card from a deck of cards, rolling dice, etc.

Random Experiment

An experiment is an operation whose output cannot be predicted with certainty. If in each trail of an experiment conducted under identical conditions, the outcome is not unique, but may be any one of the possible outcomes, then such an experiment is called Random Experiment.

Sample Space

A sample space can be defined as the set of all possible outcomes of an experiment and is denoted by S. The set S = {E_1, E_2, E_3, ... , E_n} is called a sample space of an experiment satisfying the following two conditions:

- Each element of the set S denotes one of the possible outcomes
- The outcome is one and only one element of the set S whenever the experiment is performed.

For example, in a tossing a coin Sample space consists of head and tail S={H,T} and the two coins are tossed then the sample space S ={HH,HT,TH,TT}.

Trail and Events

Any particular performance of a random experiment is called trail and the outcome or combinations of outcomes are termed as event.

Exhaustive Events

The total number of possible outcome of a random experiment is known as the exhaustive events. For example, in a tossing a coin head and tail are the two exhaustive cases.

In drawing two cards from a pack of cards, the exhaustive number of cases is ${}^{52}C_2$, since 2 cards can be drawn out of 52 cards in ${}^{52}C_2$ ways.

Favorable Events

The number of cases favorable to an event in a trail is the number of outcomes which entail the happening of the event.

For example, in throwing of two dice, the number of cases favorable to getting the sum 5 is (2,3),(3,2),(1,4) and (4,1).

Mutually Exclusive Events

Events are said to be mutually exclusive or incompatible if the happening of any one of them precludes the happening of all the others, i.e., if no two or more of them can happen simultaneously in the same trail.

For example, in tossing a coin, both head and tail cannot occur in a single trail.

Equally Likely Events

Outcomes of a trail are said to be equally likely if taking into consideration all the relevant evidences, there is no reason to expect one in preference to the others.

For example, in tossing a coin, getting a head and tail are equally likely events.

Independent Events

Several events are said to be independent if the happening of an event is not affected by the supplementary knowledge concerning the occurrence of any number of the remaining events.

For example, in tossing a unbiased coin, the event of getting a head in the first toss is independent of getting a head in the second, third and subsequent throws.

Dependent Events

If the occurrence of one event influences or affects the probability of another event then the two events are said to be dependent. Now that we are familiar with the definitions and terms of probability let’s head towards the concept of probability, its definition, and other related concepts.

Complimentary event

This type of event denotes the non-happening of events. The complement of an event P is the event, not P (or P’).

Impossible Event

The event that cannot happen is called an impossible event. For example, in tossing a coin it is impossible to get both head and tail at an equal time.

Different approaches to probability

There are three important classes of probabilities:

- Theoretical Probability

- Experimental Probability
- Axiomatic Probability

Theoretical Probability

Theoretical Probability is based on the potential chances of something happening. The type of probability is principally based on the logic behind probability. For example, if a coin is tossed, the theoretical probability of getting a head or a tail will be ½ or o.5.

Experimental Probability

This type of probability is based on the observations of an experiment. The experimental probability can be determined based on the number of potential outcomes by the cumulative/total number of trials. For instance, if a coin is tossed 8 times and the head is recorded 4 times then, the experimental probability for heads is 4/8 or 1/2.

Axiomatic Probability

In axiomatic probability, a set of commands or assumptions are set which fits all types. These axioms are set by Kolmogorov and are called Kolmogorov's three axioms. With the axiomatic method of probability, the chances of existence or non-existence of the events can be quantified.

Complementary Events

The possibility that there will be simply two outcomes that state that an event will occur or will not occur. For example, a person will visit or not visit your residence, getting a government job or not receiving a job, etc. are examples of complementary events. The complement of an event occurring is exactly the reverse of the probability of it not occurring. Some more additional examples are:

- It will rain or not rain tomorrow.
- The trainee will pass the examination or not pass.
- You win the quiz or you won't win.

Probability Density Function

The Probability Density Function (PDF) is the probability function that is outlined for the density of a continuous random variable existing within a certain range of values. The PDF explains the normal distribution and how mean and deviation exist. The standard normal distribution is applied to build a database of statistics, which are frequently used in science to describe the real-valued variables, whose distribution is not identified.

Formulas of Probability

In probability examples, one thing that helps a lot are the formulas and theorem as probability sometimes gets a little confusing, so next will look at the formulas;

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- If A and B are mutually exclusive events i.e $A \cap B = \phi$, then

  $P(A \cup B) = P(A) + P(B)$.
- If A, B, and C are any three events then the formula for addition is given by;

  $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$

- If A, B, and C are any three events such that they are mutually exclusive then;

  $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

Random variables

Let S be a sample space associated with a given random experiment. A real valued function defined on S and taking values in $R(-\infty, \infty)$ is called one dimensional random variable.

A random variable X is a rule which associates uniquely a real number with every elementary event $E_i \in S$, i =1, 2, 3, ... , n i.e, a random variable is a real valued function which maps the sample space on to the real line. Discrete Random Variables and Continuous Random Variables are the two types of a random variable.

Discrete Random Variable

A variable which can assume only a countable number of real values and for which the value which the variable takes depends on chance is called discrete random variable. In other words, a real valued function defined on a discrete sample space is called a discrete random variable. For instance, numbers of members of family, number of students in a class, number of passenger in a bus, tossing a coin and rolling a dice are the example of discrete random variable.

Probability Mass Function

If X is one dimensional discrete random variable taking at most a countable in finite number of values $x_1, x_2, x_3, \ldots$ then it is probabilistic behaviour at each real point described by a function called the probability mass function.

Definition

If X is a discrete random variable with distinct $x_1, x_2, x_3, \ldots x_n, \ldots$ then the function P(x) defined as

$$P_X(x) = \begin{cases} P(X = x_i) & if\ x = x_i \\ 0 & if\ x \neq x_i; i = 1,2,3,\ldots \end{cases}$$

is called the probability mass function of random variable X.

Remarks

The numbers $p(x_i)$; i =1, 2, 3, ... must satisfy the following conditions:

- $P(x_i) \geq 0$
- $\sum_{i=1}^{\infty} p(x_i) = 1$

Continuous Random Variable

A random variable which can assume any value from a specified interval of the form *[a, b]* is known as continuous random variable.

Probability Density Function

If X is a continuous random variable, it will have infinite number of values in any interval however small. The probability that this variable lies in the infinitesimal interval *(x, x+dx)* is expressed as *f(x) dx*, where the function *f(x)* is called probability density function (p.d.f), satisfying the following conditions:

- $f(x) \geq 0$, for all x
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Distribution functions of random variables

Let X be a random variable, the function F defined for all real x by $F(x) = P(X \leq x)$ is called the distribution function (d.f) or cumulative distribution function of the random variable X.

If random variable X is discrete then distribution function is $F(x) = P(X \leq x)$

If X is continuous random variable then distribution function is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\, dx$$

Properties of Distribution Function

- If F is the distribution function of random variable X and if a < b then

$$P(a < X \le b) = F(b) - F(a)$$

- If F is the distribution function of random variable X then

$$\text{(i) } 0 \le F(X) \le 1$$

$$\text{(ii) } F(x) \le F(Y), \text{ if } x < y$$

- If F is the distribution function of random variable X then

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$

$$F(\infty) = \lim_{x \to \infty} F(x) = 1$$

- $$\frac{d}{dx}\left(F(x)\right) = f(x)$$

Example – Probability of Discrete Distribution Function

A random variable X has the following probability function

X	0	1	2	3	4	5	6	7
P(x)	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

(i) Find k,

(ii) Evaluate $P(X < 6)$, $P(X \ge 6)$ and $P(0 < X < 5)$

(iii) Determine the distribution function of *X*

(iv) $P(X \leq a) > 1/2$ find the minimum value of *a*

Procedure

- To find k
- To calculate,

$$P(X < 6) = P(X=0)+ P(X=1)+ P(X=2)+ P(X=3)+ P(X=4)+ P(X=5)$$

$$P(X \geq 6) = 1 - P(X < 6)$$

$$P(0 < X < 5) = P(X=1)+ P(X=2)+ P(X=3)+ P(X=4)$$

- To calculate the distribution function of *X*,

$$F(x) = P(X \leq x)$$

- To calculate minimum value of *a*,

$$P(X \leq a) > 1/2$$

Calculation

(i) To find k

$$\sum_{x=1}^{7} p(x_i) = 1$$

$k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$

$\Rightarrow 10k^2 + 9k - 1 = 0 \Rightarrow (10k - 1)(k + 1) = 0$

$\therefore k = \dfrac{1}{10}$ or $k = -1$(negative)

Hence, $k = \dfrac{1}{10}$

(ii) Evaluate *P(X < 6), P(X ≥ 6)* and *P(0 < X < 5)*

$P(X < 6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$= k + 2k + 2k + 3k + k^2$

$$P(X < 6) = \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100}$$

$P(X \geq 6) = 1 - P(X < 6)$

$$= 1 - \frac{81}{100} = \frac{19}{100}$$

$P(0 < X < 5) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= k + 2k + 2k + 3k = 8k$

$$P(0<X<5) = \frac{8}{10}$$

(iii) Determine the distribution function of *X*

$F(x) = P(X \leq x)$

x	$F(x) = P(X \leq x)$
0	0
1	$k = \frac{1}{10}$
2	$k + 2k = 3k = \frac{3}{10}$
3	$k + 2k + 2k = 5k = \frac{5}{10}$
4	$k + 2k + 2k + 3k = 8k = \frac{8}{10}$
5	$k + 2k + 2k + 3k + k^2 = 8k + k^2 = \frac{8}{10} + \frac{1}{100} = \frac{81}{100}$
6	$k + 2k + 2k + 3k + k^2 + 2k^2 = 8k + 3k^2 = \frac{8}{10} + \frac{3}{100} = \frac{83}{100}$
7	$k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 9k + 10k^2 = \frac{9}{10} + \frac{10}{100} = 1$

(iv) *P(X ≤ a) > 1/2* find the minimum value of *a*

From the distribution function P(X ≤ 4)

$$= \frac{8}{10} = \frac{4}{5} > \frac{1}{2}$$

Therefore a = 4.

Result

- $k = \frac{1}{10}$
- $P(X < 6) = \frac{81}{100}$
- $P(X \geq 6) = \frac{19}{100}$
- $P(0 < X < 5) = \frac{8}{10}$
- $P(X = 0) = 0$, $P(X = 1) = \frac{1}{10}$, $P(X = 2) = \frac{3}{10}$, $P(X = 3) = \frac{5}{10}$, $P(X = 4) = \frac{8}{10}$,

  $P(X = 5) = \frac{81}{100}$, $P(X = 6) = \frac{83}{100}$, $P(X = 7) = 1$
- The minimum value of *a = 4*

Example – Probability of Continuous Distribution Function

Let X be a continuous random variables density function given by

$$f(x) = \begin{cases} kx & ; 0 \leq x < 1 \\ k & ; 1 \leq x < 2 \\ -kx + 3k & ; 2 \leq x < 3 \\ 0 & ; otherwise \end{cases}$$

i. Find the value of k.

ii. Determine the cumulative distribution function.

PROCEDURE:

- To find the value of k is , $\int_{-\infty}^{\infty} f(x)dx = 1$.
- To calculate cumulative distribution function,
  - For any x, such that $-\infty \leq x < 0$
  - For any x, where $0 \leq x < 1$
  - For any x, where $1 \leq x < 2$
  - For any x, where$2 \leq x < 3$
  - For any x, where$x \geq 3$

CALCULATION:

i. Find the value of k.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_0^1 kxdx + \int_1^2 kdx + \int_2^3 -kx + 3k\ dx = 1$$

$$k\left(\frac{x^2}{2}\right)_0^1 + k(x)_1^2 + -k[\left(\frac{x^2}{2}\right)_2^3 + 3(x)_2^3\ ] = 1$$

$$k\left(\frac{1}{2}\right) + \text{k }(1) + \text{k}\left[\frac{3^2}{2} + \frac{2^2}{2} + 3(1)\right] = 1$$

$$\frac{k}{2} + \left[-k\left(\frac{9}{2}\right) + 9k\right] - \left[-k\ \left(\frac{4}{2}\right) + 6k\right]$$

$$\frac{k}{2} + k + \left[(k)\left(\frac{-9}{2} + 9\right) - [(k)(2+6)]\right] = 1$$

$\frac{k}{2} + k + \left[(k)\frac{-9+18}{2} - 4\right]$=1

$\frac{k}{2} + \left[(k)\left[\frac{-9+18-8}{2}\right]\right]$=1

$\frac{k}{2}$+k+$\left[(k)\left(\frac{1}{2}\right)\right] = 1$

$\frac{k}{2} + k + \frac{k}{2}$ = 1

$$\frac{k}{2} + \frac{2k}{2} + \frac{k}{2} = 1$$

$\frac{k+2k+k}{2}$ = 1

$$\frac{4k}{2} = 1$$

2k = 1

K= $\frac{1}{2}$

ii) Determine the cumulative distribution distribution function

For any x, such that -∞ $< x <$0;

F(x) = $\int_{-\infty}^{x} f(x)dx = 0$

For any x where 0 $\leq x < 1$

F(x) = $\int_{-\infty}^{0} 0\ dx\ + \int_{0}^{x} kx\ dx$

= k$\int_{0}^{x} x\ dx$

=$\frac{1}{2}\left[\frac{x^2}{2}\right]_0^x$

$= \frac{1}{2}\left[\frac{x^2}{2} - \frac{0}{2}\right] = \frac{x^2}{4}$

F(x) = $\frac{x^2}{4}$

For any x where $1 \leq x \leq 2$`

$$= \int_{-\infty}^{0} 0\, dx + \int_{0}^{1} kxdx + \int_{1}^{x} k\, dx$$

$= k\int_0^1 xdx + k\int_1^x dx$

$= \int_0^1 xdx + \frac{1}{2}\int_1^x dx$

$= \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 + \frac{1}{2}[x]_1^x$

$= \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{2}[x - 1]$

$= \frac{1}{4} + \frac{x}{2} - \frac{1}{2}$

$= \frac{1+2x-2}{4}$

$= \frac{2x-1}{4}$

F(x) $= \frac{2x-1}{4}$

For any x where ,$2 \leq x < 3$

F(x) = $\int_0^1 kxdx + \int_1^2 kdx + \int_2^3 -kx + 3k\, dx = 1$

$=k\int_0^1 xdx + k\int_1^2 dx + k\int_2^x -x+3dx$

$=\frac{1}{2}\int_0^1 xdx + \frac{1}{2}\int_1^2 dx + \frac{1}{2}\int_2^x -x+3dx$

$=\frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 + \frac{1}{2}[x]_1^2 + \frac{1}{2}\left[\frac{-x^2}{2}+3x\right]_2^1$

$=\frac{1}{2}\left[\frac{1}{2}\right] + \frac{1}{2}[2-1] + \frac{1}{2}\left[\left(\frac{-x^2}{2}+3x\right) - \left(\frac{-4}{2}+6\right)\right]$

$=\frac{1}{4} + \frac{1}{2} + \frac{1}{2}\left[\left(\frac{-x^2}{2}+3x\right) - (2+6)\right]$

$=\frac{1}{4} + \frac{1}{2} + \frac{1}{2}\left[\left(\frac{-x^2+6x}{2}\right) - (2+6)\right]$

$=\frac{1}{4} + \frac{1}{2} + \frac{1}{2}\left[\left(\frac{-x^2+6x}{2}\right) + (2-6)\right]$

$=\frac{1+2}{4} + \frac{1}{2}\left[\left(\frac{-x^2+6x}{2}\right) - 4\right]$

$=\frac{3}{4} + \frac{1}{2}\left[\left(\frac{-x^2+6x-8}{4}\right)\right]$

$=\frac{-x^2+6x-8+3}{4}$

$=\frac{-x^2+6x-5}{4}$

F(x) = $\frac{6x-x^2-5}{4}$

For any x , x $\geq$ 3

F(x) = $\int_{-\infty}^0 0dx + \int_0^1 kxdx + \int_1^2 kdx + \int_2^3 -kx+3k\,dx + \int_3^x 0dx$

$=0+k+\left[\frac{x^2}{2}\right]_0^1+k[x]_1^2+\left[-k\frac{x}{2}+3kx\right]_2^3+0$

$=0+k\left(\frac{1}{2}\right)+k(2-1)+\left[\left(-k\left(\frac{9}{2}\right)+9k\right)\left(k\left(\frac{4}{2}\right)+6k\right)\right]$

$=k\left(\frac{1}{2}\right)+k(1)+\left[k\left(\frac{-9}{2}+9\right)-k\left(\frac{-4}{2}\right)+6\right]$

$=k\left(\frac{1}{2}\right)+k(1)+\left[k\left(\frac{-9+18}{2}\right)-k\left(\frac{-4+12}{2}\right)\right]$

$=k\left(\frac{1}{2}\right)+k+k\left[\frac{9}{2}-\frac{8}{2}\right]$

$=\frac{1}{2}\left(\frac{1}{2}\right)+\frac{1}{2}+\frac{1}{2}\left(\frac{1}{2}\right)$

$=\frac{1}{4}+\frac{1}{2}+\frac{1}{4}$

$=\frac{1}{4}+\frac{2}{4}+\frac{1}{4}$

$=\frac{4}{4}$

$= 1$

RESULT:

i. $k=\frac{1}{2}$

ii. Hence the distribution function F(x) is given by

iii. $$f(x) = \begin{cases} 0, for -\infty \le x < 0 \\ \frac{x^2}{4}\ for\ \ 0 \le x < 1 \\ \frac{2x-1}{4} for\ 1 \le x < 2 \\ \frac{x^2+6x-5}{4}\ for\ 2 \le x < 3 \\ 1\ , for\ 3 \le x < \infty \end{cases}$$

Random Vector

Definition:

A random vector is an ordered set of random variables, represented as a column or row vector, where each component is a random variable with its own probability distribution.

Let's denote a random vector as:

X = (X1, X2, ..., Xn)

or

X = [X1, X2, ..., Xn]^T

where:

- X is the random vector

- Xi are the individual random variables (components)

- n is the dimension of the vector

- ^{T} denotes the transpose (column vector)

Properties:

1. Each component Xi has its own probability distribution.

2. The components may be correlated or independent.

3. The dimension of the vector (n) can be any positive integer.

Types of Random Vectors:

1. Discrete Random Vector

2. Continuous Random Vector

3. Mixed Random Vector

1. Discrete Random Vector

A Discrete Random Vector $X = (X_1, X_2, ..., Xn)$ is a random vector where each component Xi takes discrete values.

$$f_{Xi}(xi) = \sum ... \sum f_X(x)\ x_1 ... x_{(i-1)} x_{(i+1)...} x_n$$

Discrete Random Vector Properties:

1. Normalization: $\sum x\ f_X(x) = 1$.

2. Non-negativity: $f_X(x) \geq 0$.

3. Marginal distributions: $fi(xi) = \sum x[-i]\ f_X(x)$.

2. Continuous Random Vector

Definition:

A Continuous Random Vector $X = (X_1, X_2, ..., Xn)$ is a random vector where each component Xi takes continuous values.

$$f_X i(xi) = \int ... \int f_X(x)\ dx_1 ... dx_{(i-1)} dx_{(i+1)...} dx_n$$

Continuous Random Vector Properties:

1. Normalization: $\int ... \int f_X(x)\ dx = 1$.

2. Non-negativity: $f_X(x) \geq 0$.

3. Marginal distributions: $fi(xi) = \int ... \int f_X(x)\ dx[-i]$.

Expectation

The 'average' value of a random phenomenon is also termed as its mathematical expectation or expected value. Once we have constructed the probability distribution for a random variable, to compute a mean or expected value of the random variables, where the weights are probabilities associated with the corresponding values. The mathematical expression for computing the expected value of a discrete random variable X with the probability mass function and computing the expected value of a continuous as random variable X with the probability density function are denoted by E(X).

$$E(X) = \begin{cases} \sum_{i=1}^{n} x_i \, P(X = x_i) & \text{for discrete random variable} \\ \int_{-\infty}^{\infty} x \, f(x) \, dx & \text{for continuous random variable} \end{cases}$$

Properties of Expectation

Property 1 - Addition Theorem of Expectation

Property 1 - Addition Theorem of Expectation

If X and Y are random variables then E(X+Y) = E(X) + E(Y), provided all the expectation exists.

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

$E(Y) = \int_{-\infty}^{\infty} y\, f(y)\, dy$

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

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{XY}(x,y)\, dx\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{XY}(x,y)\, dx\, dy$

$= \int_{-\infty}^{\infty} x\, [\int_{-\infty}^{\infty} f_{XY}(x,y)\, dy]\, dx + \int_{-\infty}^{\infty} y\, [\int_{-\infty}^{\infty} f_{XY}(x,y)\, dx]\, dy$

$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx + \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$

E(X+Y) = E(X) + E(Y)

Property 2 - Multiplication theorem of Expectation

If X and Y are independent random variables, then E(XY) = E(X).E(Y)

$E(XY) = \iint xy\, f_{XY}(x,y)\, dx\, dy$

$= \iint xy\, f_X(x)\, f_Y(y)\, dx\, dy$, X, Y are independent

$= \int x\, f_X(x)\, dx \int y\, f_Y(y)\, dy$

$E(XY) = E(X) . E(Y)$

Property 3

If X is a random variable and 'a' is constant.

(i) $E[a\, \psi (X)] = a\, E[\psi (X)]$ and (ii) $E[\psi (X) + a] = E[\psi (X)] + a$

Where $\psi (X)$ is a function of X, is a r.v and all the expectation are exists.

Proof:

(i) $E\{a\Psi(X)] = \int_{-\infty}^{\infty} a\Psi(x)\, f(x)\, dx = a \int_{-\infty}^{\infty} \Psi(x)\, f(x)\, dx = a\, E[\Psi(x)]$

(ii) $E[\Psi(X) + a] = \int_{-\infty}^{\infty} [\Psi(x) + a]\, f(x)\, dx$

$= \int_{-\infty}^{\infty} \Psi(x)\, f(x)\, dx + a \int_{-\infty}^{\infty} f(x)\, dx$

$= E\{\Psi(X)] + a$ ($\because \int_{-\infty}^{\infty} f(x)\, dx = 1$)

Cor. (i) If $\Psi(X) = X$ then $E(aX) = a\, E(X)$ and $E(X + a) = E(X) + a$

(ii) If $\Psi(X) = 1$ then $E(a) = a$.

Property 4

If X is a random variable and a and b are constants then $E(aX + b) = a\, E(X) + b$ provided all the Expectations exists.

Proof. By definition, we have:

$E(a\, X + b) = \int_{-\infty}^{\infty} (ax + b)\, f(x)\, dx$

$= a \int_{-\infty}^{\infty} x\, f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx$

$= a\, E(X) + b$

Cor. 1. Taking a = 1, b = 0, we get E(X) = E(X)

Cor. 2. If E(aX + b) = 0 then E(X) = -b/a, $a \neq 0$

Remark. If $E(X) = \mu$, then $E(X - \mu) = 0$

i.e. E(X - E(X)) = 0

Property 5

If $X \geq 0$ then $E(X) \geq 0$.

Proof

If x is continuous random variable such that $X \geq 0$ then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xf(x) > 0$$

[If $X \geq 0$ $f(X) = 0$ for $n < 0$] provided the expectation exists.

Property 6

If X and Y are two random variables such that $Y \leq X$, then $E(Y) \leq E(X)$, provided all expectations exists.

Proof

Since $Y \leq X$

We have r.v $Y - X \leq 0 \Rightarrow X - Y \geq 0$.

Hence $E(X-Y) \geq 0$

$E(X) - E(Y) \geq 0$

$E(X) \geq E(Y)$

Therefore, $E(Y) \leq E(X)$.

Variance

The variance of a random variable X is defines as

$$Var(X) = E(X^2) - [E(X)]^2$$

Property

Let X is a random variable then $V(aX+b) = a^2V(X)$ where a and b are constants.

Proof

If Y = aX + b then

$$E[Y] = E(aX + b) = aE[X] + b$$

$$Y - E[Y] = Y - (aE[X] + b)$$

$$= (aX + b) - (aE[X] + b)$$

$$= (aX + b - aE[X] - b)$$

$$= aX - aE[X] + b - b$$

$$= aX - aE[X]$$

$$Y - E(Y) = a(X - E[X])$$

Taking expectation and squaring on both sides we get

$$E[Y - E(Y)]^2 = E[a(X - E(X))]^2$$

$$= a^2 [E[X - E[X]]^2]$$

$$= a^2 [E[X^2 - 2XE[X] + (E[X])^2]$$

$$= a^2 [E[X^2] - 2E[X]\, E[X] + (E[X])^2]$$

$$= a^2 [E[X^2] - 2(E[X])^2 + (E[X])^2]$$

$$= a^2 [E[X^2] - (E[X])^2]$$

$$V(aX + b) = a^2\, V(X)$$

Moments

The rth moment of a variable X· about· any point. x = A. usually denoted by given μ_r' by

$$\mu_r' = \frac{1}{N} \sum_i f_i (x_i - A)^r, \quad \sum_i f_i = N$$

$$= \frac{1}{N} \sum_i f_i d_i^r,$$

where $d_i = x_i - A$.

The rth moment of a variable about the mean $\bar{x}$, usually denoted by μ_r is given by

$$\mu_r = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^r = \frac{1}{N} \sum_i f_i z_i^r$$

where $z_i = x_i - \bar{x}$.

In particular

$$\mu_o = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^o = \frac{1}{N} \sum_i f_i = 1$$

and $\mu_1 = \frac{1}{N} \sum_i f_i (x_i - \bar{x}) = 0$, being the algebraic sum of deviations from the mean. Also

$$\mu_2 = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^2 = \sigma^2$$

These results, viz., $\mu_o = 1$, $\mu_1 = 0$, and $\mu_2 = \sigma^2$, are of fundamental importance and should be committed to memory.

We know that if $d_i = x_i - A$, then

$$\bar{x} = A + \frac{1}{N} \sum_i f_i d_i = A + \mu_1'$$

Relation between moments about mean in terms of moments about any point .

$\mu = 1/N\ \Sigma\ f_i(x_i - \bar{x}) = 1/N\ \Sigma\ f_i(x_i - A + A - \bar{x})$

$= 1/N\ \Sigma\ f_i(d_i + A - \bar{x})$, where $d_i = x_i - A$

Using (3.17), we get

$\mu_r = 1/N\ \Sigma\ f_i(d_i - \mu)^r$

$= 1/N\ \Sigma\ f_i\ [\ d_i^r - C_1^r\ d_i^{r-1}\ \mu^1 + C_2^r\ d_i^{r-2}\ \mu^2 - ... + (-1)^r\ \mu^r\ ]$

$= \mu_r' - C_1^r \mu_{r-1}'\ \mu^1 + C_2^r\ \mu_{r-2}'\ \mu^2 - ... + (-1)^r\ \mu^r$

In particular, on putting r = 2, 3 and 4 we get

$\mu_2 = \mu_2' - \mu^2$

$\mu_3 = \mu_3' - 3\mu_2' \mu + 2\mu^3$

$\mu_4 = \mu_4' - 4\mu_3' \mu + 6\mu_2' \mu^2 - 3\mu^4$

Conversely,

$\mu_r' = 1/N \Sigma f_i(x_i - A)^r = 1/N \Sigma f_i(x_i - \bar{x} + \bar{x} - A)^r$

$= 1/N \Sigma f_i(x_i + \mu)^r$

where $x_i - x = z_i$ and $x = A + \mu_1'$

Thus $\mu_r = 1/N \Sigma f_i (z_i + {}^rC_1 z_i^{r-1} \mu_1' + {}^rC_2 z_i^{r-2} \mu_1'^2 + ... + \mu_1'^r)$

$= \mu_r' + {}^rC_1 \mu_{r-1}' \mu_1' + {}^rC_2 \mu_{r-2}' \mu_1'^2 + ... + \mu_1'^r$

In particular, putting r = 2, 3 and 4 and noting that $\mu_1 = 0$, we get

$\mu_2 = \mu_2' + \mu_1'^2$

$\mu_3 = \mu_3' + 3\mu_2' \mu_1' + \mu_1'^3$

$\mu_4 = \mu_4' + 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 + \mu_1'^4$

These formulae enable us to find the moments about any point, once the mean and the moments about mean are known.

Effect of Change of Origin and Scale on Moments:

Let $u = \frac{x - A}{h}$, so that $x = A + hu$, $\bar{x} = A + h\bar{u}$ and $x - \bar{x} = h(u - \bar{u})$

Thus, rth moment of x about any point $x = A$ is given by

$$\mu_r' = \frac{1}{N} \sum_i f_i (x_i - A)^r = \frac{1}{N} \sum_i f_i (hu_i)^r = h^r \cdot \frac{1}{N} \cdot \sum_i f_i u_i^r$$

And the rth moment of x about mean is

$$\mu_r = \frac{1}{N} \sum_i f_i (x_i - \bar{x})^r = \frac{1}{N} \sum_i f_i [h(u_i - \bar{u})]^r$$

$$= h^r \frac{1}{N} \sum_i f_i (u_i - \bar{u})^r$$

Thus the rth moment of the variable x about mean is h^r times the rth moment of the variable u about its mean.

Sheppard's Corrections for Moments.

In case of grouped frequency distribution, while .calculating moments we assume that the frequencies are concentrated at the middle point of the crass intervals. If the distribution is symmetrical or slightly symmetrical and the class intervals are not greater than one-twentieth of the range, this assumption is very nearly true. But since the assumption is not in general true, some error, called the 'grouping error', creeps into the calculation of the moments. W.F. Sheppard proved that if

(i) the frequency distribution is continuous, and

(ii) the frequency tapers off to zero in both directions,

the effect due to grouping at the mid-point of the intervals can be corrected by the following fonnulae, known as Sheppard's corrections:

$$\mu_2 \text{ (corrected)} = \mu_2 - \frac{h^2}{12}$$

$$\mu_3 \text{ (corrected)} = \mu_3$$

$$\mu_4 \text{ (corrected)} = \mu_4 - \frac{1}{2} h^2 \mu_2 + \frac{7}{240} h^4$$

where h is the width of the class interval.

Charlier's Checks: The following identities

$$\sum f(x+1) = \sum fx + N; \sum f(x+1)^2 = \sum fx^2 + 2\sum fx + N$$
$$\sum f(x+1)^3 = \sum fx^3 + 3\sum fx^2 + 3\sum fx + N$$
$$\sum f(x+1)^4 = \sum fx^4 + 4\sum fx^3 + 6\sum fx^2 + 4\sum fx + N,$$

are often used in checking the accuracy in the calculation of first four moments and are known as Charlier's Checks.

Pearson's β and γ Coefficients

Karl Pearson defined the following four coefficients, based upon the first four moments about mean:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \ \gamma_1 = +\sqrt{\beta_1} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}, \ \gamma_2 = \beta_2 - 3$$

$$\alpha_1 = \frac{\mu_1}{\sigma} = 0, \ \alpha_2 = \frac{\mu_2}{\sigma^2} = 1, \ \alpha_3 = \frac{\mu^3}{\sigma^3} = \sqrt{\beta_1} = \gamma_1; \ \alpha_4 = \frac{\mu_4}{\sigma^4} = \beta_2$$

Factorial Moments:

Factorial moment of order r about the origin of the frequency distribution X_i |f;, (i = 1,2, ... n), is defined' as

$$\mu'_r = 1/N \ \Sigma \ f_i \ x_i^r$$

where $x^r = x(x-1)(x-2)...(x-r+1)$ and $N = \Sigma f_i$

Thus the factorial moment of order r about any point x = a is given by

$$\mu_r^a = 1/N \ \Sigma \ f_i \ (x_i-a)^r$$

where $x\text{-}a^r = (x-a)(x-a-1)...(x-a-r+1)$

we have

$\mu'_1 = 1/N \Sigma f_i x_i = \mu'_1$ (about origin) = Mean(x)

$\mu'_2 = 1/N \Sigma f_i x_i^2 = 1/N \Sigma f_i x_i(x_i-1)$

$= 1/N \Sigma f_i x_i^2 - 1/N \Sigma f_i x_i = \mu'_2 - \mu'_1$

$\mu'_3 = 1/N \Sigma f_i x_i(x_i - 1)(x_i - 2)$

$= 1/N \Sigma f_i x_i^3 - 3/N \Sigma f_i x_i^2 + 2/N \Sigma f_i x_i$

$= \mu'_3 - 3\mu'_2 + 2\mu'_1$

$\mu'_4 = 1/N \Sigma f_i x_i(x_i - 1)(x_i - 2)(x_i - 3)$

$= 1/N \Sigma f_i x_i^4 - 6/N \Sigma f_i x_i^3 + 11/N \Sigma f_i x_i^2 - 6/N \Sigma f_i x_i$

$= \mu'_4 - 6\mu'_3 + 11\mu'_2 - 6\mu'_1$

Conversely, we will get

$\mu_1' = \mu(1), \mu_2' = \mu(2)' + \mu(1),$

$\mu_3' = \mu(3)' + 3\mu(2)' + \mu(1)$, and

$\mu_4' = \mu(4)' + 6\mu(3)' + 7\mu(2)' + \mu(1).$

Absolute Moments:

For the frequency distribution $X|f_i$ i = I. 2 •... II. the 11h absolute moment of the variable about the origin is given by

$$\frac{1}{N}\sum_{i=1}^{n} f_i \,|\, x_i^r \,|, N = \Sigma f_i$$

where $|\, x_i^r \,|$ represents the absolute or modulus value of x_i^r .

The rth absolute moment of the variable about the mean $\bar{x}$ is given by

$$\frac{1}{N}\sum_{i=1}^{n} f_i \,|\, x_i - \bar{x} \,|^r$$

Example:

The first four moment of a distribution about the values 4 of the variables are -1.5, 17,-30 and 108.

Find the moments about mean β_1 and β_2. Find also the moments.

i) The origin and

ii) The point x=2

PROCEDUR:

$\mu_2=\mu'_2-\mu'^2_1$

$\mu_3=\mu'_3-3\mu'_2\mu'_1+2(\mu'^3_1)$

$\mu_4=\mu'_4-4\mu'_3\mu'_1+6\mu'_2\mu'^2_1-3\mu'^4_1$

$\beta_1=\frac{\mu_3^2}{\mu_2^3}$ $\beta_2=\frac{\mu_4}{\mu_2^2}$

Mean=A+μ'_1

$\mu'_2=\mu_2+\mu'^2_1$

$\mu'_3=\mu_3+3\mu_3\mu'_1+\mu'^3_1$

$\mu'_4=\mu_4+4\mu_3\mu'_1+6\mu_2\mu'^2_1+\mu'^4_1$

CALCULATION:

In the usual notations, we are given A=4 and

μ'_1=-1.5 , μ'_2=17, μ'_3=-30 and μ'_4=108

Moments about mean; μ_1=0

$\mu_2=\mu'_2-\mu'^2_1$

$=17-(-1.5)^2$

$=17-2.25$

μ_2 =14.75

μ_3=μ'_3-$3\mu'_2\mu'_1$+2(μ'^3_1)

=-30-3(17)(-1.5)+2(-1.5)^3

=-30+76.5-6.75

μ_3=39.75

μ_4=μ'_4-4$\mu'_3\mu'_1$+6$\mu'_2\mu'^2_1$-3μ'^4_1

=108-4(-30)(-1.5)+6(17)(-1.5)^2-3(-1.5)^4

=108-108+229.5-15.1875

μ_4 =142.3125

Hence,

$\beta_1=\frac{\mu_3^2}{\mu_2^3}=\frac{(39.75)^2}{(14.75)^3}$

β_1=0.4924

$\beta_2=\frac{\mu_4}{\mu_2^2}=\frac{(142.3125)}{(14.75)^2}$

β_2=0.6541

Also, Mean=A+μ'_1

=4+(-1.5)

=2.5

Moments about origin, we have

Mean=2.5, μ_2 =14.75,μ_3=39.75 andμ_1=142.31 (approx)

We know Mean=A+μ'_1, whereμ'_1 is the first moment about the point x=A

Taking, A=0 we get the first moment about origin as, μ'_1=mean=2.5

We get,

$\mu'_2=\mu_2+\mu'^2_1$

$=14.75+(2.5)^2$

$=14.75+6.25$

$\mu'_2=21$

$\mu'_3=\mu_3+3\mu_3\mu'_1+\mu'^3_1$

$=39.75+3(14.75)(2.5)+(2.5)^3$

$=39.75+110.625+15.625$

$\mu'_3=166$

$\mu'_4=\mu_4+4\mu_3\mu'_1+6\mu_2\mu'^2_1+\mu'^4_1$

$=142.3125$

$=142.3125+4(39.75)(2.5)+6(14.75)(2.5)^2+(2.5)^4$

$=142.3125+397.5+553.125+39.0625$

$=1132$

Moments about the point,

X=2 we have , mean=A+μ'_1

Taking A=2, the first moment about the point x=2 is,

μ'_1=mean-2

$=2.5-2$

$\mu'_1=0.5$

Hence,

$\mu'_2=\mu_2+\mu'^2_1$

$=14.75+0.25$

$\mu'_2=15$

$\mu'_3=\mu_3+3\mu_3\mu'_1+\mu'^3_1$

$=39.75+3(14.75)(0.5)+(0.5)^3$

$=39.75+22.125+0.125$

$\mu'_3=62$

$\mu'_4=\mu_4+4\mu_3\mu'_1+6\mu_2\mu'^2_1+\mu'^4_1$

$=142.3125+4(39.75)(0.5)+6(14.75)(0.5)^2+(0.5)^4$

$=142.3125+79.5+22.125+0.0625$

$\mu'_4=244$

RESULT:

i) The origin is,

$\mu_2=14.75$

$\mu_3=39.75$

$\mu_4=142.3125$

$\beta_1=0.4924$

$\beta_2=0.6541$

ii) The point x=2 is,

$\mu'_1=0.5$

μ'_2=15

μ'_3=62

μ'_4=244

Calculate the first four moments of the following distribution about the mean and hence β_1 and β_2.

X	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

PROCEDURE:

i)Moments about the points:

$\mu_1' = \frac{1}{N}\sum fd$

$\mu_2' = \frac{1}{N}\sum fd^2$

$\mu_3' = \frac{1}{N}\sum fd^3$

$\mu_4' = \frac{1}{N}\sum fd^4$

ii)Moments about the mean:

$\mu_2 = \mu_2' - (\mu_1')^2$

$\mu_3 = \mu_3' - 3\,\mu_2'\,\mu_3' + 2\,(\mu_1')^3$

$\mu_4 = \mu_4' - 4\,\mu_3\,\mu_1' + 6\,\mu_2'\,(\mu_1')^2 - 3(\mu_1')^4$

$\beta_1 = \frac{\mu_3 2}{\mu_2 3}$

$\beta_2 = \frac{\mu_4}{\mu_2 2}$

CALCULATION:

x	f	d=x-4	fd	fd^2	fd^3	fd^4
0	1	-4	-4	16	-64	256
1	8	-3	-24	72	-216	648
2	28	-2	-56	112	-224	448
3	56	-1	-56	56	-56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256
Total	256	0	0	512	0	2816

i)Moments about the points x=4 are,

$\mu_1' = \frac{1}{N}\sum fd = 0$

$\mu_2' = \frac{1}{N}\sum fd^2 = \frac{512}{256} = 2$

$\mu_3' = \frac{1}{N}\sum fd^3 = 0$

$\mu_4' = \frac{1}{N}\sum fd^4 = \frac{2816}{256} = 11$

ii) Moments about the mean are

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2$

$= 2 - (0)^2$

μ_2=2

μ_3= μ_3'-3 μ_2' μ_3'+2 $(\mu_1')^3$

=0 -3(2)(0)+2(0)^3

μ_3 =0

μ_4 = μ_4'-4 μ_3 μ_1'+6 μ_2' $(\mu_1')^2$-3$(\mu_1')^4$

=11-4(0)(0)+6(2 (0)^2-3(0)^4

μ_4 =11

β_1= $\frac{\mu_{3^2}}{\mu_{2^3}}$

=$\frac{0^2}{2^3}$

β_1 =0

β_2 =$\frac{\mu_4}{\mu_{2^2}}$

=$\frac{11}{2^2}$

=$\frac{11}{4}$

β_2=2.75

Basic Inequality

In some cases there are multiple answers to a problem or the situation requires something that is not exactly equal to another value. When a mathematical sentence involves something other than an equal sign, an inequality is formed.

An algebraic inequality is a mathematical sentence connecting an expression to a value, a variable, or another expression with an inequality sign.

Listed below are the most common inequality signs:

- $>$ "greater than"
- $\geq$ "greater than or equal to"
- $\leq$ "less than or equal to"
- $<$ "less than"
- $\neq$ "not equal to"

Markov Inequality

Statement:

Let $X \geq 0$ be a non-negative random variable (discrete or continuous), and let $a > 0$.

Then:

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

Proof:

Let X be a finite random variable

& $a > 0$ be a real number, then

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

Let X be a discrete random variable with P.M.F f(x)

F(x) = Σ x P(x)

= Σ x P(x) + Σ x P(x)

≥ Σ x P(x)

≥ Σ a P(x)

= a Σ P(x)

= a P(x≥a)

Thus, P(x≥a) ≤ E(x) / a

Hence Proved.

Example:

A coin is weighted so that its probability of landing on heads is 20%, independently of other flips. Suppose the coin is flipped 20 times. Use Markov's inequality to bound the probability it lands on heads at least 16 times.

Solution: We actually do know this distribution; the number of heads is X ∼ Bin(n = 20, p = 0.2). Thus, E [X] = np = 20 · 0.2 = 4. By Markov's inequality:

$$\mathbb{P}(X \geq 16) \leq \frac{\mathbb{E}[X]}{16} = \frac{4}{16} = \frac{1}{4}$$

Let's compare this to the actual probability that this happens:

$$\mathbb{P}(X \geq 16) = \sum_{k=16}^{20} \binom{20}{k} 0.2^k \cdot 0.8^{20-k} \approx 1.38 \cdot 10^{-8}$$

This is not a good bound, since we only assume to know the expected value. Again, we knew the exact distribution, but chose not to use any of that information (the variance, the PMF, etc.).

Chebyshev's Inequality

If X be a random variable with mean (μ) and variance (σ^2), then for any positive number K,

1) $P[|X-\mu| \geq K\sigma] \leq 1/K^2$

2) $P[|X-\mu| < K\sigma] \geq 1 - 1/K^2$

Proof:

Case (i):

If x is a Continuous random variable w.r.to ,

$\sigma^2 = \sigma_x^2 = E[(x - E(x))^2]$; $E(X) = \mu$

$\sigma^2 = E[(x - \mu)^2]$

$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$

$= \int_{-\infty}^{\mu} (x - \mu)^2 f(x)\, dx + \int_{\mu - K\sigma}^{\mu + K\sigma} (x - \mu)^2 f(x)\, dx + \int_{\mu + K\sigma}^{\infty} (x - \mu)^2 f(x)\, dx$

$\sigma^2 \geq \int_{-\infty}^{\mu - K\sigma} (x - \mu)^2 f(x)\, dx + \int_{\mu + K\sigma}^{\infty} (x - \mu)^2 f(x)\, dx$

I - integral

$-\infty \leq x \leq \mu - k\sigma$

$x \leq \mu - k\sigma$

$x - \mu \leq -k\sigma$

Square on both sides $\Rightarrow (x - \mu)^2 \geq k^2\sigma^2$

II - integral

$\mu + k\sigma \leq x \leq \infty$

$\sigma^2 \geq k^2\sigma^2 \left[\int_{-\infty}^{\mu - k\sigma} f(x)dx + \int_{\mu + k\sigma}^{\infty} f(x)dx \right]$

$\sigma^2 \geq k^2\sigma^2$ [$P(x \leq \mu - k\sigma) + P(x \geq \mu + k\sigma)$]

$\sigma^2 \geq k^2\sigma^2$ [$P(x - \mu \leq -k\sigma) + P(x - \mu \geq k\sigma)$]

$\sigma^2 \geq k^2\sigma^2$ [$P(|x - \mu| \geq k\sigma)$]

$1/k^2 \geq P(|x - \mu| \geq k\sigma) \rightarrow$ ①

multiply on both sides,

$P(|x - \mu| \geq k\sigma) \leq 1/k^2$

add ① on both sides,

$1 + P(|x - \mu| \geq k\sigma) \leq 1 + 1/k^2$

$P(|x - \mu| \leq k\sigma) \geq 1 - 1/k^2 \rightarrow$ ②

Hence proved.

Problem:

Lets revisit the example in Markov's inequality section earlier in which we toss a weighted coin independently with probability of landing heads p = 02. Upper bound the probability it lands on heads at least 16 times out of 20 ips using Chebyshev's inequality

Solution: Because X Bin(n = 20, p = 02)

$$\mathbb{E}[X] = np = 20 \cdot 0.2 = 4$$

and:

$$\mathsf{Var}(X) = np(1-p) = 20 \cdot 0.2 \cdot (1 - 0.2) = 3.2$$

Note that since Chebyshev's asks about the difference in either direction of the RV from its mean, we must weaken our statement first to include the probability X ≤ - 8. The reason we chose 8 is because

Chebyshev's inequality is symmetric about the mean (difference of 12; 4 ± 12 gives the interval [-8, 16]):

$$
\begin{aligned}
\mathbb{P}(X \geq 16) &\leq \mathbb{P}(X \geq 16 \cup X \leq -8) && \text{[adding another event can only increase probability]} \\
&= \mathbb{P}(|X-4| \geq 12) && \text{[def of abs value]} \\
&= \mathbb{P}(|X-\mathbb{E}[X]| \geq 12) && [\mathbb{E}[X]=4] \\
&\leq \frac{\mathsf{Var}(X)}{12^2} && \text{[Chebyshev's inequality]} \\
&= \frac{3.2}{12^2} = \frac{1}{45}
\end{aligned}
$$

This is a much better bound than given by Markov's inequality, but still far from the actual probability. This is because Chebyshev's inequality only takes the mean and variance into account. There is so much more information about a RV than just these two quantities!

Holder's Inequality

Let a, b > 0 and p, q > 1 satisfy

$$p^{-1} + q^{-1} = 1.$$

Then

$$p^{-1} a^{p} + q^{-1} b^{q} \geq ab$$

with equality if and only if $a^{p} = b^{q}$.

Proof: Fix b > 0. Let

$$g(a; b) = p^{-1} a^{p} + q^{-1} b^{q} - ab.$$

We require that $g(a; b) \geq 0$ for all a. Differentiating wrt a for fixed b yields $g^{(1)}(a; b) = a^{p-1} - b$, so that g(a; b) is minimized (the second derivative is strictly positive at all a) when $a^{p-1} = b$, and at this value of a, the function takes the value

$$p^{-1} a^{p} + q^{-1} (a^{p-1})^{q} - a(a^{p-1}) = p^{-1} a^{p} + q^{-1} a^{p} - a^{p} = 0$$

as, by equation ($p^{-1} + q^{-1} = 1$), $1/p + 1/q = 1 \Longrightarrow (p-1)q = p$. As the second derivative is strictly positive at all a, the minimum is attained at the unique value of a where $a^{p-1} = b$, where, raising both sides to power q yields $a^{p} = b^{q}$.

Theorem:

Suppose that X and Y are two random variables, and p, q > 1 satisfy ($p^{-1} + q^{-1} = 1$). Then

$$|E_{X,Y}[XY]| \leq EX,Y[|XY|] \leq \{EX[|X|^{p}]\}^{1/p} \{E_{fY}[|Y|^{q}]\}^{1/q}$$

Proof: (Absolutely continuous case: discrete case similar) For the first inequality,

$$EX,Y[|XY|] = \iint |xy| f_{X,Y}(x, y)\, dx\, dy \geq \int\int xy f_{X,Y}(x, y)\, dx\, dy = E_{X,Y}[XY]$$

And

$$E_{X,Y}[XY] = \int\int xy f_{X,Y}(x, y)\, dx\, dy \geq \iint -|xy| f_{X,Y}(x, y)\, dx\, dy = -E_{X,Y}[|XY|]$$

So

$$-E_{X,Y}[|XY|] \leq E_{X,Y}[XY] \leq E_{X,Y}[|XY|] \qquad \therefore\ |E_{X,Y}[XY]| \leq E_{X,Y}[|XY|].$$

For the second inequality, set

$$a = \frac{|X|}{\{\mathsf{E}_X[|X|^p]\}^{1/p}} \qquad b = \frac{|Y|}{\{\mathsf{E}_{f_Y}[|Y|^q]\}^{1/q}}.$$

and taking expectations yields, on the left hand side,

$$p^{-1}\frac{\mathsf{E}_X[|X|^p]}{\mathsf{E}_X[|X|^p]} + q^{-1}\frac{\mathsf{E}_{f_Y}[|Y|^q]}{\mathsf{E}_{f_Y}[|Y|^q]} = p^{-1} + q^{-1} = 1$$

and on the right hand side

$$\frac{\mathsf{E}_{X,Y}[|XY|]}{\{\mathsf{E}_X[|X|^p]\}^{1/p}\{\mathsf{E}_{f_Y}[|Y|^q]\}^{1/q}}$$

and the result follows.

Note: here we have equality if and only if

$$P_{X,Y}[|X|^{p} = c|Y|^{q}] = 1$$

for some non zero constant c.

Minkowski's Inequality

Statement

If $x, y \in l^p$ for $1 \leq p < \infty$ and $1/p + 1/q = 1$ and $x + y \in l^p$, then,

$$(\Sigma_{i=1}^{n} |x_i + y_i|^p)^{1/p} \leq (\Sigma_{i=1}^{n} |x_i|^p)^{1/p} + (\Sigma_{i=1}^{n} |y_i|^p)^{1/p}$$

Proof:

If $p = 1$, then, inequality follows from triangular inequality of real numbers.

$$\sum|x_i + y_i|^p = \sum |x_i + y_i|^{(p-1)} |x_i + y_i|$$

$$\leq \sum|x_i + y_i|^{(p-1)} (|x_i| + |y_i|)$$

$$\leq \sum |x_i + y_i|^{(p-1)} |x_i| + \sum |x_i + y_i|^{(p-1)} |y_i|$$

By using Holder's inequality on R.H.S of Inequality (under certain Circumstances)

$$\sum |x_i + y_i|^p \leq (\sum |x_i|^p)^{1/p} (\sum |x_i + y_i|^q)^{1/q}$$

$$\leq (\sum |x_i|^p)^{1/p} (\sum |x_i|^q + \sum |y_i|^q)^{1/q}$$

$$\leq (\sum |x_i|^p)^{1/p} (\sum |x_i|^q)^{1/q} + (\sum |x_i|^p)^{1/p} (\sum |y_i|^q)^{1/q}$$

Dividing both sides by, $(\sum |x_i + y_i|^p)^{1/q}$

$$(\sum |x_i + y_i|^p)^{1 - 1/q} \leq (\sum |x_i|^p)^{1/p} + (\sum |y_i|^q)^{1/q}$$

$$(1 - 1/q) = 1/p$$

$$1 - 1/q = 1/p$$

$$p + q = pq$$

$$p = (p-1)q$$

$$(\sum|x_i + y_i|^p)^{1/P} \leq (\sum |x_i|^p)^{1/p} + (\sum |y_i|^p)^{1/p}$$

Hence proved.

Jensen's Inequality

Suppose that X is a random variable with expectation μ, and function g is convex and finite. Then

$$E_X [g(X)] \geq g(E_X [X])$$

with equality if and only if, for every line a + bx that is a tangent to g at μ

$$P_X[g(X) = a + bX] = 1.$$

that is, g(x) is linear.

Proof:

Let l(x) = a + bx be the equation of the tangent at x = μ. Then, for each x, g(x) ≥ a + bx as in the figure. Thus

$$E_X[g(X)] \geq E_X[a + bX] = a + bE_X[X] = l(\mu) = g(\mu) = g(E_X[X])$$

as required. Also, if g(x) is linear, then equality follows by properties of expectations. Suppose that

$$E_X [g(X)] = g(E_X [X]) = g(\mu)$$

but g(x) is convex, but not linear. Let l(x) = a + bx be the tangent to g at μ. Then by convexity

$$g(x) - l(x) > l(x)\ dF_X(x) > 0$$

and hence

$$E_X[g(X)] > E_X[l(X)].$$

But l(x) is linear, so $E_X[l(X)] = a + bE_X[X] = g(\mu)$, yielding the contradiction

$$E_X[g(X)] > g(E_X[X]).$$

If g(x) is concave, then

$$E_X [g(X)] \leq g(E_X [X])$$

$g(x) = x^2$ is convex, thus

$$E_X [X^2] \geq \{E_X [X]\}^2$$

g(x) = log x is concave, thus

$$E_X [\log X] \leq \log \{E_X [X]\}$$

Hence proved.