BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Measure and Probability Theory

Course Code: 23ST02CC

Unit-IV

Independence of Sequence

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

Ms. S. Soundarya
Guest Faculty
Department of Statistics

UNIT – IV

INDEPENDENCE OF SEQUENCE

Independence of sequence of events

An event A is said to be independent (or statistically independent) of another event B, if the conditional probability of A given B, i.e., P(AIB) is equal to the unconditional probability of B, i.e., if P(AB)=P(A).

Independence of sequence of Random variables

A sequence of random variables is independent if all of its terms are mutually independent. This means that the value of one term does not affect the probability of the others.

Explanation

- **Independent random variables**

  Two random variables are independent if knowing the value of one does not change the probability of the other.

- **Independent and identically distributed (IID) sequence**

  A sequence of random variables is IID if all of its terms are independent and have the same probability distribution.

Example

- Flipping a fair coin is an example of an IID sequence. Each flip is independent because the coin doesn't remember the result of the previous flip.
- If you repeatedly toss a fair coin, the sequence of random variables $X_1,X_2,X_3,\dots$ is independent because each is the result of a different coin toss.

Conditional probability

The probability of an event that would be affected by another event.

$$P(A|B)=P(A\cap B\ )/P(B)$$

multiplication theorem of probability for independent events.

If A and B are two events with positive probabilities ($P(A) \neq 0$, P(B) #01, then A and B are independent if and only if

$$P(A\cap B)=P(A).\ P\ (B)$$

Proof:

We have:

$$P(A\cap B)=P(A)\ P\ (BIA) = P(B)\ P(AIB);\ P(A)\neq 0,\ P(B)\neq 0$$

If A and B are independent, i.e., A is independent of B and B is independent of A, then, we have

$$P(AIB) = P(A)$$

and

$$P(BIA) = P(B)$$

then we get

$$P(A\cap B)\ P(B)\ P(A\cap B)\ P(A) = P(A) = P(B)\ P(A\cap B)=P(A)\ P\ (B),$$

as required.

$$P(AIB) = P(A)\ P(B|A) = P(B)$$

and implies that A and B are independent events.

Hence, for independent events A and B, the probability that both of these occur simultaneously is the product of their respective probabilities.

Extension of multiplication theorem of probability for n- independent events.

Necessary and sufficient condition for independence of n events $A_1, A_2, ..., A_n$ is that the probability of their simultaneous happening is equal to the product of their respective probabilities, i.e.,

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n)$$

Proof:

If $A_1, A_2, ..., A_n$ are independent events then

$$P(A_2|A_1) = P(A_2),$$

$$P(A_3|A_1 \cap A_2) = P(A_3), ...,$$

$$P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1}) = P(A_n)$$

Hence, we get

$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n)$

Conversely if (3-24) holds, then from (3-22) and (3-24), we get

$$P(A_2) = P(A_2|A_1),$$

$$P(A_3) = P(A_3|A_1 \cap A_2), ...,$$

$$P(A_n) = P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})$$

$\Rightarrow A_1, A_2, ..., A_n$ are independent events

Hence the theorem.

Conditional expectation

Discrete Case

Discrete Case: The conditional expectation or mean value of a continuous function g (X, Y) given that $Y = y_j$, is defined by

$$E [g(X, Y) 1 Y = y_j)] = \sum g (x_i, y_j) P (X = x_i 1 Y = y_j)$$

$$\frac{\sum_{i=1}^{m} p(x_i y_i) P(X = x_i \cap Y = y_j)}{P(Y = y_j)}$$

i.e., E[g (X, Y) |Y = y_jl is nothing but the expectation of the function g (X, y_j) of X w.r. to the conditional distribution of X when y =y_j. In particular, the conditional expectation of a discrete random variable X given Y = y_j is:

$$E\,(XIY = y_j) = \Sigma\, x_i\, P(X= x_i \mid Y = y_j)$$

The conditional variance of X given Y = y_j is likewise given by

$$V(X|Y = y_j) = E\,[\{X- E\,(XIY=y_j)^2 \mid Y = y_j]$$

The conditional expectation of g(X, Y) and the conditional variance of Y given X=x_i may also be defined in an exactly similar manner.

Continuous case

The conditional expectation of g(X, Y) on hypothesis Y = y is given by

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

$$= \int_{-\infty}^{\infty} g(x,y)\frac{f(x,y)}{f_Y(y)}dx$$

In particular, the conditional mean of x given y = y is defined as

$$E\{X/Y=y)\} = \int_{-\infty}^{\infty} x\frac{f(x,y)}{f_Y(y)}dx$$

Similarly,

$$E\{Y/X=x)\} = \int_{-\infty}^{\infty} y\frac{f(x,y)}{f_X(x)}dy$$

The conditional variance of X defined as

$$V(X/Y=y) = E\left[\left(X - E(X/Y=Y)\right)^2 / Y = y\right]$$

$$V(Y/X=x) = E\left[\left(Y - E(Y/X=x)\right)^2 / X = x\right]$$

Theorem - Conditional Expectation of Mean

The expected value of X is equal to the expectation of the conditional expectation of X given that is symbolically,

$$E(X) = E\{E(X/Y)\}$$

$$E\{E(X/Y)\} = E\left\{\sum_i x_i \;\; P(X = x_i / Y = y_j)\right\}$$

$$= E\left\{\sum_i x_i \;\; \frac{P(X = x_i \cap Y = y_j)}{P(y = y_j)}\right\}$$

$$= \sum_j \;\; \left\{\sum_i x_i \;\; \frac{P(X = x_i \cap Y = y_j)}{P(Y = y_j)}\right\} P(Y = y_j)$$

$$= \sum_i x_i \sum_j \;\; P(X = x_i \cap Y = y_j)$$

$$= \sum_j x_i \sum_i \;\; P(X = x_i \cap y = y_j)$$

$$= \sum_j x_i \;\; P(X = x_i) = E(X) \;\; = \mathrm{E(X)}$$

$$\Rightarrow \;\; \mathrm{E\{E\ (X/Y)\} = E(X)}$$

Hence proved.

Theorem - Conditional Expectation of Variance

The variance of X can be regarded as consisting of two parts the expectation of conditional variance and variance of conditional expectation symbolically,

$$\mathrm{Var(X) = E[V(X/Y)] + V\ [E(X/Y)]}$$

$$\mathrm{= E[V(X/Y)] + V\ [E(X/Y)]}$$

$$= E\left\{E(\mathrm{X}^2/\mathrm{Y}) - [E(\mathrm{X}/\mathrm{Y})]^2\right\} + \left[\{E(\mathrm{X}/\mathrm{Y})\}^2\right] - \left[E\{E(\mathrm{X}/\mathrm{Y})\}\right]^2$$

$$= E\left\{E(\mathrm{X}^2/\mathrm{Y})\right\} - E\{E(\mathrm{X}/\mathrm{Y})\}^2 + E\{E(\mathrm{X}/\mathrm{Y})\}^2 - \left[E\{E(\mathrm{X}/\mathrm{Y})\}\right]^2$$

$$= E\left\{E(\mathrm{X}^2/\mathrm{Y})\right\} - \left[E\{E(\mathrm{X}/\mathrm{Y})\}\right]^2$$

$$= E\left\{E(\mathrm{X}^2/\mathrm{Y})\right\} - [E(\mathrm{Y})]^2$$

$$= E\left\{\sum_i x_i^2 P(\mathrm{X} = x_i / \mathrm{Y} = y_j)\right\} - [E(\mathrm{X})]^2$$

$$= E\left\{\sum_i x_i^2 \frac{P(\mathrm{X} = x_i \cap Y = y_j)}{P(\mathrm{Y} = y_j)}\right\} - [E(\mathrm{X})]^2$$

$$= \sum_j \left[\left\{ \sum_i x_i^2 \frac{P(X = x_i \cap Y = y_j)}{P(Y = y_j)} \right\} P(Y = y_j) \right] - [E(X)]^2$$

$$= \sum_j x_i^2 \sum_j P(X = x_i \cap Y = y_j) - [E(X)]^2$$

$$= \sum_j x_i^2 P(X = x_i) - [E(X)]^2$$

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

$$= \text{Var}(X) =$$

$$\Rightarrow \text{Var}(X) = E[V(X/Y)] + V[E(X/Y)]$$

Hence proved.

Characteristic functions

The characteristic function of a random variable X is is denoted by $\varphi_X(t)$ and is defined as,

$$\varphi_X(t) = E(e^{itx}), \quad t \in R$$

Properties Characteristic functions

Property 1. For all real t, we have

(i) $\varphi(0) = \int_{-\infty}^{\infty} dF(x) = 1$

(ii) $|\varphi(t)| \leq 1 = \varphi(0)$

Property 2. $\varphi(t)$ is continuous everywhere, i.e., $\varphi(t)$ is continuous function of t in $(-\infty, \infty)$. Rather $\varphi(t)$ is uniformly continuous in 't'.

Proof.

For $h \neq 0$,

$\varphi(x + h) - \varphi_{(x)}(t) = |\int e^{i(t+h)x} - e^{itx}\, dF(x)$

$\leq \int | e^{i(t+h)x} - e^{itx} |\, dF(x) = \int | e^{ihx} - 1 |\, dF(x)$

Property 3. $\varphi_X(-t)$ and $\varphi_X(t)$ are conjugate functions, i.e., $\varphi_X(-t) = \varphi_X(t)$, ... where a is the complex conjugate of a.

Proof.

$\varphi_X(t) = E\,(e^{itX}) = E\,[\cos tX + i \sin tX]$

$\varphi_X(-t) = E\,[\cos (-t)X + i \sin (-t)X] = E\,(\cos tX - i \sin tX) = E\,(e^{-itX}) = \varphi_X(-t)$

Inversion formula

The characteristic function $\varphi_X(t)$ is a complex - Valued function that encodes information about the distribution of a random Variables x- The inversion formula provides a way to extract this information and retrieve the original PDF or PMF.

Inversion formula

$$P(X=x) = 1/2\,\pi \int [\, e^{itx}\, \varphi_X(t)]\, dt$$

1. Normal distribution:

$$F(x) = 1/2\,\pi \int [\, e^{itx}\, e^{\,(i\mu t - \sigma 2t2/2)}]\, dt$$

Modes of convergence

Modes of convergence of sequences of random variables: almost sure convergence, convergence in probability and convergence in distribution. It turns out there is a chain of strict implications for modes of convergence:

almost sure ⇒ in probability ⇒ in distribution.

We also give some prominent examples in the Strong Law of Large Numbers, the Weak Law of Large Numbers, the Central Limit Theorem, the Borel-Cantelli lemma, and the Poisson Limit Theorem.

Almost sure convergence

Consider a sequence of random variables X_1, X_2, X_3, . . . that is defined on an underlying sample space S. For simplicity, let us assume that S is a finite set, so we can write

$$S=\{s_1, s_2, . . ,s_k\}$$

Remember that each X_n is a function from S to the set of real numbers. Thus, we may write

$$X_n(s_i)=x_{ni}, \qquad \text{for } i = 1, 2, . . . ,k.$$

After this random experiment is performed, one of the s_i's will be the outcome of the experiment, and the values of the X_n's are known. If s_j is the outcome of the experiment, we observe the following sequence:

$$x_{1j}, x_{2j}, x_{3j}, . . .$$

The sequence $X_n(s)$ converged when s = H and did not converge when s = T. In general, if the probability that the sequence $X_n(s)$ converges to $X(s)$ is equal to 1, we say that X_n converges to X almost surely and write

$$X_n \xrightarrow{a.s} X$$

Definition

A sequence of random variable $X_1, X_2, X_3, . . .$ convergence almost surely to a random variable X, shown by $X_n \xrightarrow{a.s} X$, if

$$P\left(\left\{s \in S : \lim_{n\to\infty} X_n(s) - X(s)\right\}\right) = 1$$

This is denoted $X_n \xrightarrow{a.s} X$. An equivalent definition is if

$$P\left(\left\{\lim_{n\to\infty} X_n(s) \neq X(s)\right\}\right) = 0 \cdot$$

Convergence in Probability

Convergence in probability is stronger than convergence in distribution. In particular, for a sequence $X_1, X_2, X_3, . . .$ to converge to a random variable $\overline{X}$, we must have that

$P(|X_n - X| \geq \epsilon)$ goes to 0 as $n \to \infty$, for any $\epsilon > 0$. To say that X_n converges in probability to X, we write

$$X_n \xrightarrow{p} X$$

Definition of convergence in probability

A sequence of random variable $X_1, X_2, X_3, \ldots$ converges in probability to a random variable X, shown by $X_n \xrightarrow{p} X$, if

$$\lim_{n \to \infty} P(|X_n - X| \geq \varepsilon) = 0 \quad \forall \varepsilon < 0$$

Convergence in distribution

Convergence in distribution is in some sense the weakest type of convergence. All it says is that the CDF of $\overline{X}_n$'s converges to the CDF of X as n goes to infinity. It does not require any dependence between the X_n's and X. We saw this type of convergence before when we discussed the central limit theorem. To say that X_n converges in distribution to X, we write

$$X_n \xrightarrow{d} X$$

Definition of convergence in distribution

A sequence of random variable $X_1, X_2, X_3, \ldots$ converges in distribution to a random variable X, shown by $X_n \xrightarrow{d} X$, if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

for all *x* at which $F_X(x)$ is continuous.

convergence in moments

A sequence of random variables Yn is called asymptotically uniformly integrable if

$$\lim_{M\to\infty} \limsup_{n\to\infty} E[|Yn|\ 1\{|Yn| > M\}] = 0.$$

Uniform integrability is the missing link between convergence in distribution and convergence of moments.

Helly-Bray theorem

If the sequence of distribution functions (F_n (x)) converges to the distribution function F(x) at all the points of continuity of the latter and g(x) is bounded continuous function over the line R^1 $(-\infty,\infty)$, then

$$\lim \int g(x)\, dF, (x) = \int g(x)\, dF\,(x) \ldots\ldots(1)$$

Proof:

cos tx and sin tx are continuous and bounded functions of x for all t and hence from (1) we get ,

$$\lim \int \cos tx\, dF_n\,(x) = \int \sin tx\, dF\,(x)$$

$$\lim \int \sin tx\, dF_n\,(x) = \int \sin tx\, dF\,(x)$$

$$\lim \int (\cos tx + i \sin tx)\, dF_n(x) = \int (\cos tx + i \sin tx)\, dF(x)$$

$$\lim \int e^{itx}\, dF_n(x) = \int e^{itx}\, dF(x)$$

Convolution of distributions

Probability Generating Function for the Sum of Independent Variables (Convolutions).

If X and Y are non-negative independent, integral valued discrete random variables with respective probability generating functions,

$$P(s) = \sum p_k\, s^k,\ p_k = P(X=k) \quad\text{and}\quad R(s) = \sum r_k\, s^k,\ r_k = P\,(Y = k),\ k=0$$

it is possible to deduce probability generating function for the variable Z = X + Y, which is a also clearly integral valued, in terms of P (s) and Q (s).

Let w denote P (Z = k). The event Z = k is the union of the following mutually exclusive events,

$$(X=0 \cap Y=k), (X = 1 \cap Y=k-1), (X=2 \cap Y=k-2),..., (X=k \cap Y = 0) \text{ and}$$

since the variables X and Y are independent, each joint probability is the product of the appropriate individual probabilities. Therefore, the probability wx = P (Z =k) given by: is ω_k

$$= p_0 r_k + p_1 r_{k-1} + p_2 r_{k-2} + ... + p_k r_o, \text{ for all integral } k \geq 0.$$

The new sequence of probabilities (ω_k)defined in terms of the sequences (p_k) and (r_k) is called the convolution of these sequences and is denoted by

$$\{\omega_k\} = \{p_k\} * \{r_k\}.$$