CHAPTER 3 BASIC PROBABILITY

Intersection of Events

Let A and B be two events in the Sample Space S. Their intersection, denoted by $A \cap B$, is the set of all basic outcomes in S that belong to both A and B.

Hence the intersection $A \cap B$ occurs if and only if both A and B occur. We use the term joint probability of A and B to denote the probability of the intersection of A and B.

More generally, given K events, $E_1, E_2 \ldots E_K$, their intersection $E_1 \cap E_2 \cap \ldots \cap E_K$ is the set of all basic outcomes that belong to every E_i $(i = 1, 2 \ldots K)$

Mutually Exclusive

If the events A and B have no common basic outcomes, they are called mutually exclusive and their intersection $A \cap B$, is said to be the empty set, indicating that $A \cap B$ has no members.

Union

Let A and B be two events in the Sample Space, S. Their union, denoted by $A \cup B$, is the set of all basic outcomes in S that belong to at least one of these two events. Hence the union $A \cup B$ occurs if and only if either A or B both occur.

More generally given K events, $E_1, E_2 \ldots E_K$, their union, $E_1 \cup E_2 \cup \ldots \cup E_K$ is the set of all basic outcomes belonging to at least one of these K events.

Collectively Exhaustive

Given the k events, $E_1, E_2 \ldots E_k$, in the sample space S, if $E_1 \cup E_2 \cup \ldots \cup E_k = S$ these k events are said to be collectively exhaustive

Probability and its Postulates

Three definitions of probability:

1. Classical
2. Relative frequency
3. Subjective

Classical Probability

Classical probability is the proportion of times that an event will occur, assuming that all outcomes in a sample space are equally likely to occur.

Dividing the number of outcomes in the sample space that satisfy the event by the total number of outcomes in the sample space determines the probability of an event. The probability of an event A is: $P(A) = \frac{N_A}{N}$

where N_A is the number of outcomes that satisfy the condition of event A and N is the total number of outcomes in the sample space.

Relative Frequency Probability

The relative frequency probability is the limit of the proportion of times that event A occurs in a large number of trials n:

$$P(A) = \frac{n_A}{n}$$

where n_A is the number of A outcomes and n is the total number of trials or outcomes. The probability is the limit as n becomes large or $\rightarrow \infty$

Subjective Probability

Subjective probability expresses an individuals degree of belief about the chance that an event will occur. These subjective probabilities are used in certain management decision procedure

PROBABILITY RULES

Complement Rule:

Let A be an event and $\bar{A}$ its complement. Then the complement rule is as follows:

$$P(\bar{A}) = 1 - P(A)$$

The addition rule of probabilities

Let A and B be two events. Using the addition rule of probability of their union:

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

Conditional Probability

Let A and B be two events. The conditional probability of event A given that event B has occured, is denoted P(A|B) and is found as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided } P(B) > 0$$

The multiplication Rule of Probabilities

Let A and B be two events, Using the multiplication rule of probabilities, the probability of their intersection can be derived from conditional probability as:

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

Statistical Independence

Let A and B be two events. These events are said to be statistically independent, if and only if:

4

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

From the multiplication rule it also follows that:
$P(A|B) = P(A)$ (if $P(B) > 0$)

BIVARIATE PROBABILITIES

Joint and Marginal Probabilities

In the context of bivariate probabilities the intersection probabilities $P(A_i \cap B_j)$ are called joint probabilities.

The probabilities for individual events $P(A_i)$ or $P(B_j)$ are called marginal probabilities.
Can be computed by summing the corresponding row or column.

Independent Events

Let A and B be a pair of events, each broken into mutually exclusive and collectively exhaustive categories denoted by labels. $A_1, A_2 \ldots A_n$ and $B_1, B_2 \ldots B_k$.

If every A_i is statistically independent of every event B_j then A and B are independent events.

ODDS

The odds in favour of a particular event are given by the ratio of the probability of the event divided by the probability of its complement. The odds in favour of A are:

$$\text{Odds} = \frac{P(A)}{1 - P(A)} = \frac{P(A)}{P(\bar{A})}$$

5. B

Bayes Theorem

Let A_1 and B_1 be two events. Then bayes' theorem states

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B_1|A_1) = \frac{P(A_1|B_1)P(B_1)}{P(A_1)}$$

Bayes' Theorem Alternative Statement

Let $E_1, E_2 \ldots E_k$ be k mutually exclusive and collectively exhaustive events, and let A be some other event. The conditional probability of E_i, given A_1 can be expressed as Bayes' theorem:

$$P(E_i|A_1) = \frac{P(A_1|E_i)P(E_i)}{P(A_1)}$$

$$P(E_i|A_1) = \frac{P(A_1|E_i)P(E_i)}{P(A_1|E_1)P(E_1) + P(A_1|E_2)P(E_2) + \ldots + P(A_1|E_k)P(E_k)}$$

where $P(A_1) = P(A_1 \cap E_1) + P(A_1 \cap E_2) + \ldots + P(A_1 \cap E_k)$

$= P(A_1|E_1)P(E_1) + P(A_1|E_2) + \ldots + P(A_1|E_k)P(E_k)$

1

Probability Distributions DISCRETE

The probability distribution function $P(x)$ of a discrete random variable X represents the probability that X takes on value x as a function of x. That is:

$P(x) = P(X=x)$, for all values of x

$0 \le P(x) \le 1$ and $\sum_x P(x) = 1$

Cumulative Probability distribution

The cumulative probability distribution $F(x_0)$ of a random variable X, represents the probability that X does not exceed the value x_0 as a function of x_0, that is:

$F(x_0) = P(X \le x_0)$

where the function is evaluated at all values of x_0.

Example

X	$P(x)$	$F(x)$
0	0.15	0.15
1	0.30	0.45
2	0.20	0.65
3	0.20	0.85
4	0.10	0.95
5	0.05	1.00

Relationship:

Let X be a random variable with probability distribution $P(x)$ and cumulative probability distribution $F(x_0)$. Then we can show that: $F(x_0) = \sum_{x \le x_0} P(x)$

Where the notation implies that summation is over all possible values of x that are less than or equal to x_0

2.

1. $0 \le F(x_0) \le 1$ for every number x_0; and
2. if x_0 and x_1 are two numbers such that $x_0 < x_1$ then $F(x_0) \le F(x_1)$

Expected values

Expected value $E[x]$ of a discrete random variable X is defined as

$$E[x] = \mu = \sum_x x P(x)$$

where notation indicates that the summation extends over all possible values of x.

The expected value of a random variable is also called the Mean $= \mu$.

Variance and Standard Deviation of a Discrete Random Variable

Let X be a discrete random variable. The expectation of the squared deviation about the mean $(X-\mu)^2$ is called the variance denoted as σ^2 and is given by:

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

The variance of a discrete random variable X can also be expressed as

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

The standard deviation, σ, is the positive square root of the variance

Summary of Properties for linear Functions of a random Variable

Let X be a random variable with mean μ_X and variance σ^2_X and let a and b be any constant fixed numbers. Define the random variable Y as $a+bX$. then the mean and variance of Y are:

$$\mu_Y = E[a+bX] = a + b\mu_X$$

and

$$\sigma^2_Y = Var(a+bX) = b^2\sigma^2_X$$

3

Binomial Distribution $Y \sim B(N, p)$

Begins with Bernoulli model, success or failure
$P(0) = (1-p)$ and $P(1) = p$

Mean is $\mu_x = E[x] = \sum_x x P(x) = 0(1-p) + (1)(p) = p$

Variance is: $\sigma_x^2 = E[(x-\mu_x)^2] = \sum_x (x-\mu_x)^2 P(x)$
$= (0-p)^2(1-p) + (1-p)^2 p = p(1-p)$

- two possible outcomes success + failure
- finite number of trials
- n independent trials
- As trials are independent order is not important and thus 001 100 010 are all the same

$= C_1^3 = \frac{3!}{1!(3-1)!}$

~~Probability of x success in y trials is~~ $\frac{3!}{1!(3-1)!}$

Number of sequences with x successes in n independent trials is:

$C_x^n = \frac{n!}{x!(n-x)!}$ sequences are mutually exclusive

The binomial Distribution

Suppose that a random experiment can result in two possible mutually exclusive and collectively exhaustive outcomes, success and failure, and that p is the probability of success in a single trial. If n independent trials are carried out, the distribution of the number of resulting success x is called the binomial distribution.

$P(Y=y) = \binom{n}{y} p^y (1-p)^{n-y}$

9

takes the value y, as a function of y, when the value x is fixed for X
This is denoted $P(y|x)$ and is

$$P(y|x) = \frac{P(x,y)}{P(x)}$$

$$P(x|y) = \frac{P(x,y)}{P(y)}$$

Independence
The jointly distributed random variables X and Y are said to be independent if and only if their joint probability distribution is the product of their marginal probability distributions - that is, if

$$P(x,y) = P(x)P(y)$$

for all possible pairs of values of x and y. K random variables are independent if and only if

$$P(X_1, X_2 \ldots X_K) = P(X_1)P(X_2)\ldots P(X_K)$$

Conditional Mean and Variance
Conditional mean is computed using the following:

$$\mu_{y|x} = E[Y|X] = \sum_y (y|x)P(y|x)$$

hold X fixed

Conditional variance is computed using

$$\sigma^2_{y|x} = E[(Y-\mu_{Y|X})^2|X] = \sum ((y-\mu_{y|x})^2|x)P(y|x)$$

Again hold X fixed

COVARIANCE
Let X be a random variable with mean μ_X and let Y be a random variable with mean μ_Y. The expected value $(X-\mu_X)(Y-\mu_Y)$ is called the covariance between X and Y, denoted as $Cov(X,Y)$. For discrete random variables:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)P(x,y)$$

equivalent to $E[XY] - \mu_x\mu_y = \sum_x \sum_y xy\,p(x,y) - \mu_x\mu_y$

Correlation

Let X and Y be jointly distributed random variables. The correlation between X and Y is:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

PROPERTIES: LINEAR SUMS AND DIFFERENCES

$E[X+Y] = \mu_x + \mu_y$

$E[X-Y] = \mu_x - \mu_y$

If $Cov = 0 \Rightarrow Var(X+Y) = \sigma_x^2 + \sigma_y^2$.

If $Cov \neq 0$ $Var(X+Y) = \sigma_x^2 + \sigma_y^2 + 2Cov(X,Y)$

$E[X_1, X_2 .. X_k] = \mu_1 + \mu_2 + \dots + \mu_k$

If cov = 0 for all pairs, $Var(X_1 + X_2 .. + X_k) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2$

$Cov \neq 0$ $Var(X_1 + X_2 + \dots + X_k) = \sum_{i=1}^{k} \sigma_i^2 + 2\sum_{i=1}^{k-1} \sum_{j>i}^{k} Cov(X_i, X_j)$

1 Continuous Distribution

Cumulative Distribution Function

The cumulative distribution function, $F(x)$ for a continuous random variable X expresses the probability that X does not exceed the value x, as a function of x.

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

Probability of a range using a cumulative distribution function

Let X be a continuous random variable with a cumulative distribution function $F(x)$ and let a and b be two possible values of X, with $a < b$. The probability that X lies between a and b is as follows:

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

For continuous random variables it does not matter whether we write $\leq$ or $<$ because the probability that X is precisely equal to b is 0

Probability Density Function

Let X be a continuous random variable, and let x be any number lying in the range of values for the random variables. The probability density function $f(x)$, of the random variable is a function with the following properties:

1. $f(x) \geq 0$ for all values of X
2. The area under the probability density function, $f(x)$ over all values of the random variable X within its range is equal to 1.
3. Suppose that this density function is graphed. Let a and b be two possible values of random variable X, with $a < b$. Then the probability that X lies between a and b is the area under the probability density function between these points

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

2

4 The cumulative distribution function, $F(x_0)$ is the area under the probability density function $f(x)$ up to x_0

$$F(x_0) = \int_{x_m}^{x_0} f(x)\,dx.$$

where x_m is the minimum value of random variable X

- The total area under the curve $f(x)$ is 1
- The area under the curve $f(x)$ to the left of x_0 is $F(x_0)$ where x_0 is any value that the random variable can take

Uniform Distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Mean Variance and Standard Deviation for Continuous R.V.

The mean of X μ_X is defined as the expected value of X.

$\mu_X = E[X]$

The variance of X σ_X^2 is defined as the expectation of the squared deviation, $(X-\mu_X)^2$ of the random variable from its mean: $\sigma_X^2 = E[(X-\mu_X)^2]$

or $\sigma_X^2 = E[X^2] - E[X]^2$

For a uniform distribution we have:

$f(x) = \frac{1}{b-a}$ $a \leq x < b$

$\mu_X = E[X] = \frac{a+b}{2}$

$\sigma_X^2 = E[(X-\mu_X)^2] = \frac{(b-a)^2}{12}$

3

Probability Density Function of the Normal Distribution

The probability density function for a normally distributed random variable X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

where μ and σ^2 are any numbers such that $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$ ~~and~~

Properties of the Normal Distribution

Suppose random variable X follows a normal distribution with parameters μ and σ^2. Then:

1. Mean of R.V is μ: $E[X] = \mu$
2. Variance of R.V is σ^2: $Var(x) = E[(x-\mu)^2] = \sigma^2$
3. Shape of probability density function is a symmetric bell shaped curve centered on the mean μ.
4. If we know the mean and variance, we define the normal distribution: $X \sim N(\mu, \sigma^2)$

We use standard normal table to compute cdf values

$$Z = \frac{X-\mu}{\sigma}$$

Let X be a normally distributed random variable with mean μ and variance σ^2. Then random variable $Z = (X-\mu)/\sigma$ has a standard normal distribution of $Z \sim N(0,1)$

It follows that if a and b are any possible values of X with $a < b$, then $P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$

$$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

where Z is the standard normal random variable and F denotes its cumulative distribution function

4

Normal Distribution Approximation for Binomial Distribution

$E[X] = \mu = np$
$Var[X] = \sigma^2 = np(1-p)$

- When $p = 0.5$ and $n = 100$, they have same shaped distribution
- Binomial can be approximated by a normal distribution with the same mean and variance
- Example of central limit theorem
- Normal dist provides a good approximation for the binomial distribution when $np(1-p) > 5$

By using the mean and variance from the binomial distribution, we find that, if the number of trials n is large such that $np(1-p) > 5$, then the distribution of the random variable:

$$Z = \frac{X - E[X]}{\sqrt{Var[X]}} = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard normal distribution

Allows us to find for large N, the probability that the number of successes lies in a given range

$$P(a \le X \le b) = P\left(\frac{a-np}{\sqrt{np(1-p)}} \le \frac{X-np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}}\right)$$

$$= P\left(\frac{a-np}{\sqrt{np(1-p)}} \le Z \le \frac{b-np}{\sqrt{np(1-p)}}\right)$$

Proportion Random variable

Number of successes ÷ trial $\qquad P = \frac{X}{n}$

We can show $\quad \mu = p$ and $\sigma^2 = \frac{p(1-p)}{n}$
& can use this with normal distribution

5.

The Exponential Distribution

Restricted to random variables with positive values and its distribution is not symmetric.

The exponential random variable $T (t > 0)$ has a probability density function $f(t) = \lambda e^{-\lambda t}$ for $t > 0$
where λ is the mean number of independent arrivals per time unit, t is the number of time units until the next arrival.

Then T is said to follow an exponential probability distribution.
Arrivals are independent if an arrival does not affect the probability of waiting time, t, until the next arrival.

It can be shown that λ is the same parameter used for the Poisson distribution and that the mean time between occurrences is $1/\lambda$.

The cumulative distribution function is as follows:

$$F(t) = 1 - e^{-\lambda t} \quad \text{for } t > 0$$

Mean of $1/\lambda$ and a variance of $1/\lambda^2$

The probability that the time between arrivals is t_a or less is:

$$P(T \leq t_a) = 1 - e^{-\lambda t_a}$$

The probability that the time between arrivals is between t_b and t_a is:

$$P(t_b \leq T \leq t_a) = (1 - e^{-\lambda t_a}) - (1 - e^{-\lambda t_b})$$
$$= e^{-\lambda t_b} - e^{-\lambda t_a}$$

The random variable T can be used to represent the time until the end of a service time or until the next arrival in a queuing process, beginning at an arbitrary time 0

6

- Model assumptions are the same as those for poisson process
- Poisson distribution provides the probability of X successes or arrivals during a time unit.

In contrast, the exponential distribution provides the probability that a success or arrival will occur during an interval of time t.

Example:

library, exponential, mean service time of 5 mins. What is probability a customer service will take longer than 10 mins?

Let t denote service time in mins. $\lambda = 1/5 = 0.2$ per minute

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

$P(T>10) = 1 - P(T<10)$
$= 1 - F(10)$
$= 1 - (1 - e^{-(0.2)(10)})$
$= e^{-2.0} = 0.1353$

Distribution of Sample Statistics

Sample mean

Let the random variables X_1, X_2, X_n denote a random sample from a population. The sample mean value of these random variables is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The mean of the sampling distribution of the sample mean is the population mean

Standard Normal distribution for the Sample Mean

Whenever the sampling distribution of the sample means is a normal distribution, we can compute a standardised normal random variable Z, that has mean 0 and variance of 1

$$Z = \frac{\bar{X}-\mu}{\sigma_{\bar{x}}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

The sampling distribution of $\bar{x}$ has mean $E[\bar{x}] = \mu$

Has standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ also called standard error.

If the sample size n is not small compared to population size N, the standard error of $\bar{x}$ is as follows:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Example:
Normally distributed with mean 12.2% and standard deviation of 3.6%. Random sample of 9 obtained from this population and the sample mean is computed. What is probability the sample mean will be greater than 14.4%?

2

We know $\mu = 12.2$ $\sigma = 3.6$ $n = 9$

$\bar{x}$ is sample mean, compute standard error:

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{9}} = 1.2$

Then we compute:

$P(\bar{x} > 14.4) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \frac{14.4 - 12.2}{1.2}\right)$

$= P(Z > 1.83) = 0.0336$

Central limit theorem

The mean of a random sample, drawn from a population with any probability distribution, will be approximately normally distributed with mean μ and variance σ^2/n given a large enough sample size

Statement of the central limit theorem

Let $X_1, X_2 \ldots X_n$ be a set of n independent random variables having identical distributions with mean μ and variance σ^2 and $\bar{X}$ as the mean of the random variables. As n becomes large, the central limit theorem states that the distribution of

$$Z = \frac{\bar{X} - \mu_x}{\sigma_{\bar{x}}}$$

approaches the standard normal distribution

The law of large numbers which concludes given a random sample of size n from a population, the sample mean will approach the population mean as the sample size n becomes large regardless of probability distribution

3

Sample Proportion

Let X be the number of successes in a binomial sample of n observations with parameter P. The parameter is the proportion of the population members that have a characteristic of interest. We define the sample proportion as: $\hat{p} = \frac{X}{n}$

X is the sum of a set of n independent Bernoulli random variables, each with probability of success P. As a result $\hat{p}$ is the mean of a set of independent random variables and the results we developed in the previous sections for sample means apply.

The central limit theorem can be used to argue that the probability distribution for $\hat{p}$ can be modeled as a normally distributed random variable.

If random samples are obtained from the population and the success or failure is determined for each observation, then the sample proportion of success approaches P as the sample size increases.

The difference between the expected number of sample successes - the sample size multiplied by P - and the number of successes in the sample might actually increase.

Sampling Distribution of the Sample Proportion

Let $\hat{p}$ be the sample proportion of successes in a random sample from a population with proportion of successes P, then:

Binomial $E[X] = nP$ $var[X] = nP(1-P)$

1. Sampling Distribution of $\hat{p}$ has a mean of P:

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{1}{n}E[X] = P$$

4

2 Sampling Distribution of $\hat{p}$ has standard deviation:
$\sigma^2_{\hat{p}} = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{p(1-p)}{n}$ $sd = \sqrt{\frac{p(1-p)}{n}}$

3 and if the sample size is large, the random variable.

$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$

is approximately distributed as a standard normal. The approximation is good if. $np(1-p) > 5$.

EXAMPLE

A random sample of 270 homes taken from large population of older home to estimate proportion of homes with unsafe wiring. In fact, 20% of homes have unsafe wiring, what is the probability that the sample proportion will be between 16% and 24%.

Here we have $p = 0.2$ $n = 270$

We compute standard deviation of sample proportion $\hat{p}$ as follows:

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.2(1-0.2)}{270}} = 0.024$

The required probability is: $P(0.16 < \hat{p} < 0.24)$

$= P\left(\frac{0.16-p}{\sigma_{\hat{p}}} < \frac{\hat{p}-p}{\sigma_{\hat{p}}} < \frac{0.24-p}{\sigma_{\hat{p}}}\right)$

$= P\left(\frac{0.16-0.2}{0.024} < z < \frac{0.24-0.2}{0.024}\right)$

$P(-1.67 < z < 1.67)$

$= 0.9050$.

Called at 90.5% acceptance interval

5.

Sample variance

Let $X_1, X_2 \ldots X_n$ be a random sample of observations from a population. The quantity

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is called the sample variance and its square root, S is called the sample standard deviation

Given a specific random sample, we could compute the sample variance, and the sample variance would be different for each random sample because of differences in sample observation

We use $(n-1)$ as divisor. In a random sample of n observations we have n different independent values or degrees of freedom. But after we know the computed sample mean, there are only n-1 different values that can be uniquely defined $E[S^2] = \sigma^2$

If we can assume that the underlying population distribution is normal, then it can be shown that the sample variance and the population variance are related through probability distribution known as the chi-square distribution

Chi-Square Distribution of Sample and Population Variances

Given a random sample of n observations from a normally distributed population whose population variance is σ^2 and whose resulting sample variance is S^2, it can be shown that

6

$\chi^2_{(n-1)} = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}$

has a distribution known as the chi-square (χ^2) distribution with n-1 degrees of freedom

Used in statistics because it provides a link between the sample and the population variances

The chi-square distribution with n-1 degrees of freedom is the distribution of the sum of squares of n-1 independent standard normal random variables

The assumption of an underlying normal distribution is more important for determining probabilities of sample variances than it is for determining probabilities of sample means

Sampling Distribution of the Sample Variance

Let s^2 denote the sample variance for a random sample of n observations from a population with a variance σ^2.

1 The sampling distribution of s^2 has mean σ^2.

$E[s^2] = \sigma^2$

2 The variance of the sampling distribution of s^2 depends on the underlying population distribution. If that distribution is normal then:

$Var(s^2) = \frac{2\sigma^4}{n-1}$

3 If the population distribution is normal, then $\chi^2_{(n-1)} = \frac{(n-1)s^2}{\sigma^2}$ is distributed as the chi-square distribution with n-1 degrees of freedom ($\chi^2_{(n-1)}$).

Central Limit Theorem from Linear Sum of random variable.

If $X_1, X_2 \dots X_n$ represent the result of individual random events, or observed random variable.

$$X = X_1 + X_2 + \dots X_n$$

$E[X] = n\mu$ $\quad$ $Var[X] = n\sigma^2$

The central limit theorem states that the resulting sum, X is normally distributed and can be used to compute a random variable, z, with a mean of 0 and variance of 1

$$z = \frac{X - E[X]}{\sqrt{Var[X]}} = \frac{X - n\mu}{\sqrt{n\sigma^2}}$$

In addition if X is divided by n to obtain a mean of $\bar{X}$, then a corresponding z with a mean of 0 and a variance of 1 can also be computed:

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4

Mean and Variance of a Binomial Probability Distribution

Let X be the number of successes in n independent trials each with probability of success P. Then X follows a binomial distribution with mean

$\mu = E(X) = nP$

and variance

$\sigma^2 = E[(X-\mu_x)^2] = nP(1-P)$

Derivation of mean and variance

- Consider n independent trials of Bernoulli, probability of success = P.
- Let $X_i = 1$ if the i^{th} trial results in success and 0 otherwise
- The random variables $X_1, X_2 \dots X_n$ are therefore n independent Bernoulli variables, each with probability of success P.
- The total number of success of X is as follows:

$X = X_1 + X_2 + \dots + X_n$

- Thus binomial random variables can be expressed as the sum of independent Bernoulli random variables
- Mean and variance for Bernoulli random variables used in binomial mean ver.
- We know $E[X_i] = P$ and $\sigma_{X_i}^2 = P(1-P)$ for all $i = 1, 2 \dots n$.
- Then for binomial we have:

$E[X] = E[X_1 + X_2 + \dots X_n] = E[X_1] + E[X_2] + E[X_n] = np$

- Since the Bernoulli random variables are independent, the covariance between any pair of them is zero and...

$\sigma_X^2 = \sigma^2(X_1 + X_2 + X_n)$

$\sigma_X^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 \dots + \sigma_{X_n}^2$

$\sigma_X^2 = nP(1-P)$

- n independent trials
- Probability P of success and 1-P of failure
- Only two outcomes
- Each trial independent
- Same probability for each trial

5.

Example:

Real estate agent has 5 contacts, probability of making a sale is 0.4

- Find probability at most 1 sale
- Probability between 2 and 4 sales inclusive

$P(\text{at most } 1) = P(X \leq 1) = P(X=0) + P(X=1) = 0.078 + 0.259 = 0.337$

$P(0 \text{ sales}) = \binom{5}{0}(0.4)^0(0.6)^5 = 0.078$

$P(1 \text{ sale}) = \binom{5}{1}(0.4)^1(0.6)^4 = 0.259$

$P(2 \leq X \leq 4) = P(2) + P(3) + P(4) = 0.346 + 0.230 + 0.077 = 0.653$

$P(2) = \binom{5}{2}(0.4)^2(0.6)^3 = 0.346$

$P(3) = \binom{5}{3}(0.4)^3(0.6)^2 = 0.230$

$P(4) = \binom{5}{4}(0.4)^4(0.6)^1 = 0.077$

POISSON DISTRIBUTION

Characterised as the number of occurrences or success of a certain event in a given continuous interval such as time

Assumptions

Assume that an interval is divided into a very large number of equal subintervals so that the probability of the occurrence of an event in any subinterval is small.

1. The probability of the occurrence of an event is constant for all subintervals
2. There can be no more than one occurrence in each subinterval.
3. Occurrences are independent.

Can create from binomial limit as $p \to 0$ and $n \to \infty$, $\lambda = np$

Poisson Distribution Function:

The random variable X is said to follow the poisson distribution if it has the probability distribution:

6.

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2$$

Where:

$P(x)$ is probability of x success over a given time given λ.

λ expected number of success per unit time $\lambda > 0$

Mean and variance are:

$\mu_x = E[x] = \lambda \quad \sigma^2 = E[(x-\mu)^2] = \lambda.$

The sum of Poisson random variables is also a Poisson random variable. Thus the sum of K Poisson random variables, each with mean λ, is a poisson random variable with mean $K\lambda$.

Example

Computer centre, computer system experienced 3 component failures during past 100 days. Find:

- Probability of no failures in a given day
- Probability of 1 or more?
- Probability of atleast 2 in a 3-day period.

We know failure per day $= 3/100 = 0.03 = \lambda$.

1. $P(\text{no failures}) = P(X=0 \mid \lambda = 0.03) = \frac{e^{-0.03}\lambda^0}{0!} = 0.970446$

2. Probability at least one failure:

$P(X \geq 1) = 1 - P(X=0) = 1 - 0.970446 = 0.029554$

3. $P(\text{at least 2 failures in 3 days}) = P(X \geq 2 \mid \lambda = 0.09) \quad \lambda = 3(0.03) = 0.09$

$P(X \geq 2 \mid \lambda = 0.09) = 1 - P(X \leq 1) = 1 - [P(X=0) + P(X=1)]$

$1 - (0.913931 + 0.082254)$

$= 0.003815.$

2

Poisson Approximation to Binomial Distribution

- p approaching 0 and n large
- Generally when $\lambda = np \leq 7$ we can approximate

Let X be the number of successes resulting from n independent trials, each with probability of success P. The distribution of the number of successes, X, is binomial with mean nP.

If the number of trials n is large and nP is only of moderate size ($nP \leq 7$), this distribution can be approximated by the Poisson distribution with $\lambda = nP$. The probability distribution of the approximating distribution is then:

$$P(x) = \frac{e^{-np}(np)^x}{x!} \text{ for } x = 0, 1, 2 \ldots$$

If small sample - binomial.

Geometric

- Ask girl to dance, X is number I need to ask to find a partner
- If first girl accepts, $X=1$.
- If 1st declines and next accepts, $X=2$ and so on
- When $X=n$, means I've failed on first $n-1$ tries and succeeded on n^{th} try
- Probability of failing on first try is $(1-p)$, on first two is $(1-p)(1-p)$
- Probability of failing on first $n-1$ tries is $(1-p)^{n-1}$
- Probability of succeeding on n^{th} try is P.

Thus $P(X=n) = (1-p)^{n-1} p$. Geometric Distribution

when $P(x=n+1)$ is a multiple of $P(X=n)$

- Probability it will take more than n tries is:

$$P(X>n) = (1-p)^n$$

Expected value - infinite sum. multiply X times $P(x)$ for $X=1, 2, 3 \ldots$ get

$$\Sigma = p + 2p(1-p) + 3p(1-p)^2 + \ldots + np(1-p)^{n-1}$$

8.

Multiply both sides by $(1-p)$ and we have:

$$(1-p)\Sigma = p(1-p) + 2p(1-p)^2 + 3p(1-p)^3 + \dots np(1-p)^n$$

Subtract Σ from $\Sigma 1$ to get:

$$\Sigma - (1-p)\Sigma = p[1 + (1-p) + (1-p)^2 \dots (1-p)^n] = p(1/p)$$

$$p\Sigma = p(1/p)$$

$$\Sigma = 1/p$$

Variance is $\frac{1-p}{p}$ $k = 1, 2, 3 \dots$

JOINT PROBABILITY DISTRIBUTION

Let X and Y be a pair of discrete random variables. Their joint probability distribution expresses the probability that simultaneously X takes a specific value x and Y takes value y, as a function of x and y

$$P(x,y) = P(X = x \cap Y = y)$$

Derivation of the Marginal Probability Distribution

Let X and Y be a pair of jointly distributed random variables. In this context the probability distribution of the random variable X is called its marginal probability distribution and is obtained by summing the joint probabilities over all possible values - that is

$$P(x) = \Sigma P(x,y)$$

Similarly, marginal probability distribution of random variable Y is:

$$P(y) = \sum_x P(x,y)$$

Properties of Joint Probability Distribution of Discrete Random Variables

1. $0 \leq P(x,y) \leq 1$ for any pair of x and y
2. $\Sigma P(x,y) = 1$

Conditional Probability Distribution

X and Y pair of jointly distributed discrete random variables. The conditional probability distribution of random variable Y, given that the random variable X takes on value x, expresses the probability that Y