Time Series

Intro to Time Series

- A time series is a sequence of observations in chronological order. Time series data is defined as data gathered on a particular characteristic over a period of time at regular intervals.
- Time series observations are recorded at different time such as daily, weekly, quarterly, monthly, and annually. Sometimes, high frequency data are collected at very short period of time like every second or minute, and so on.
- Some examples of time series data are five years of daily or weekly return of a stock, monthly consumption of coffee over a two-year period, annual data on export of diamond over 25 years, and so on.

Import Prices

- The first thing that we require before analyzing the time series data is to read the data into R and to plot the time series.

- install.packages("BatchGetSymbols")

- first.date <- Sys.Date() - 3 # 3 days
- last.date <- Sys.Date()

- tcs_prices <- BatchGetSymbols::BatchGetSymbols(tickers = stocks, first.date = first.date, last.date = last.date, freq.data = "daily")
- write.csv(tcs_prices,"tcs.csv")

Time Series Components

- There are three types of time series patterns: trend, seasonal, and cyclic. A trend is a long-term increase or decrease in data. A series exhibits seasonality if it is influenced by seasonal factors such as some festive seasons during the year, monsoon, and so on. It could be day of the week, month of the year, or day of the week. Seasonality is always of the fixed period. Cyclic pattern exists when data exhibit rises or fall that are not of the fixed period.
- library(TSA) # Loading library
- data("tempdub") # Load the data
- plot(tempdub)

TSA: Seasonality

- rt=diff(log(tempdub),12)# seasonal differencing for monthly data. For quarterly data, seasonal difference is 4 instead of 12.
- plot(rt)
- adf.test(rt)
- To fit the seasonal model to the **tempdub** dataset, we define the indicator variables that indicate the month to which each of the data points corresponds. The **TSA** package has functionality by which we can extract the month of the dataset.
- We then regress the temperature on the monthly dummy. To avoid dummy variable trap, we will exclude the intercept from the model.

TSA : Seasonality

- month=season(tempdub)# create monthly dummy
- m1=lm(tempdub~month-1)
- summary(m1)
- We can undertake unit root test on residuals to see if the residuals become stationary.
- resid=residuals(m1)
- adf.test(resid)

Stationarity

- A unit root test is a statistical test for the proposition that in an autoregressive statistical model of a time series, the autoregressive parameter is one. It is a test for detecting the presence of stationarity in the series.
- If the variables in the regression model are not stationary, then it can be shown that the standard assumptions for asymptotic analysis will not be valid. In other words, the usual "*t*-ratios" will not follow a *t*-distribution, so we cannot validly undertake hypothesis tests about the regression parameters.
- Stationary time series is one whose mean, variance, and covariance are unchanged by time shift. Nonstationary time series have time varying mean or variance, or both.
- If a time series is nonstationary, we can study its behavior only for a time period under consideration. It is not possible to generalize it to other time periods. Therefore, it is not useful for forecasting purpose.

TSA: Trend

- If a time series has a trend, it is necessary to tackle the trend first. The deterministic trend has to be removed to render the time series stationary.
- plot(ts(sim))
- sim=rnorm(100,mean=0,sd=10) # Normal white noise with mean 0 and Sd=10
- x=5+time(sim)*3+ts(sim)
- x=ts(x)
- plot(x)
- Detrending of deterministic Trend
- model2=lm(x~time(x))
- summary(model2)
- resid2=resid(model2)
- adf.test(resid2) # To check for stationarity of residuals

- The presence of unit root in a time series is tested with the help of Augmented Dickey–Fuller Test (ADF). It tests for a unit root in the univariate representation of time series. For a return series R_t, the ADF test consists of a regression of the first difference of the series against the series lagged k times.
- adf.test(tcs) # ADF test on closing price
- adf.test(rtcs) # ADF test on return series

Forecasting Methods

Simple Moving Average - SMA

Exponential Smoothing - EMA

SMA

$$SMA=(X_1 + X_2 + X_3 + \ldots + X_n)/n$$

```
# Install and load the required packages
install.packages("forecast")
library(forecast)

# Sample time series data
ts_data <- ts(c(10, 15, 20, 25, 30, 35, 40))

# Calculate a 3-period simple moving average (SMA)
sma <- ma(ts_data, order = 3)

# Print the calculated moving average
print(sma)
```

EMA

```r
# Install the TTR package (if not installed)
install.packages("TTR")

# Load the TTR package
library(TTR)

# Sample data (replace this with your own data)
data <- c(23, 45, 67, 34, 56, 78, 90, 32, 54, 76, 98)

# Convert the data to a time series object
ts_data <- ts(data)
```

```r
# Print the result
print(ema)
```

EMAt = λ Xt+(1−λ)·EMAt−1

- All Slides taken from Chapter 16 : Introduction to Time Series Analysis , of book , by SAGE Text

Statistical Analysis in Simple Steps Using R

Kiran Pandya • Prashant Joshi • Smruti Bulsari

Course Name :Basic Statistics using GUI-R (RKWard)
Module : Conditional Probability
Week 7 Lecture : 2

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

Conditional Probability

The conditional probability of B, given A, denoted by P(B/A) is defined by

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \text{ if } P(A) > 0.$$

Two events A and B are independent if $P(A \mid B) = P(A)$.
Equivalent conditions are $P(B \mid A) = P(B)$
or $P(A \cap B) = P(A)\,P(B)$

If the events B_1, B_2, … ,B_k constitute a partition of the sample space Ω, where **$P(B_i)$** $\neq 0$ for $i = 1,2, \dots ,k$, then for any event A in Ω,

$$P(A) = \sum_{i} P(B_i \cap A) = \sum_{i} P(B_i)\, \mathbf{P(A/B_i)}$$

Bayes Theorem

If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space Ω, where **$P(B_i)$** $\neq 0$ for $i = 1,2, \ldots, k$, then for any event A in Ω,

$$\mathbf{P(B_r|A) = \frac{P(B_r)\,P(A|B_r)}{\sum_i P(B_i)\,P(A/B_i)}} \quad \textbf{for } r=1,2\ldots,k$$

Belief and Hypothesis

Bayes' Theorem allows us to update our beliefs about the likelihood of an event based on new evidence. Initially, we have a prior probability (A Priori) of an event happening.

As new evidence (B) becomes available, we update our belief to a posterior probability using Bayes' Theorem.

This theorem forms the basis of Bayesian inference, a powerful tool in statistics and machine learning for updating beliefs about hypotheses as new evidence is obtained.

A Priori Probability: - A Priori Probability refers to the initial probability of an event occurring based on logic or prior knowledge, without considering any additional evidence.

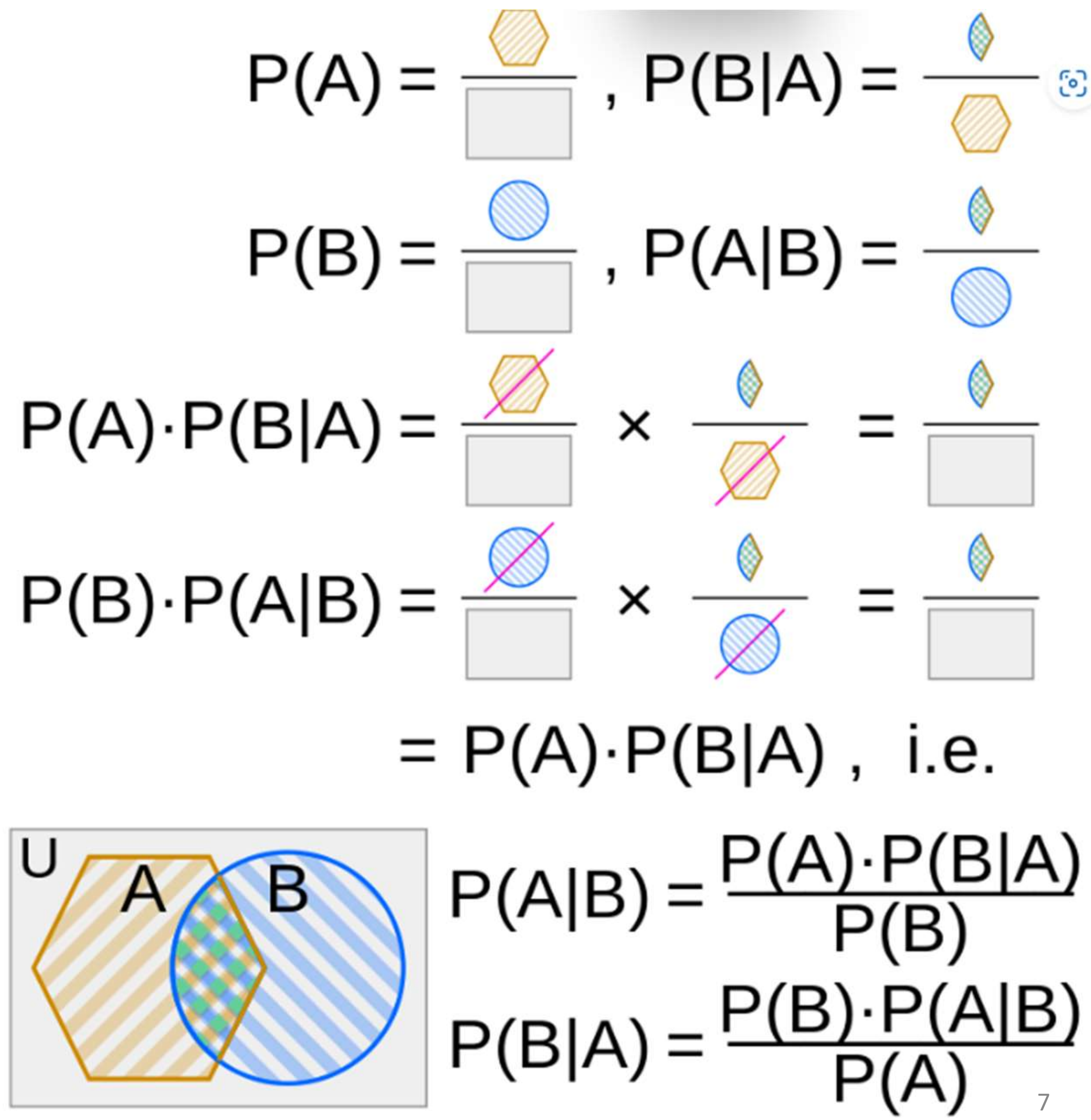
Posterior Probability: - Posterior Probability refers to the probability of an event occurring after considering new evidence or information.

Bayes

One of the many applications of Bayes' theorem is Bayesian inference, a particular approach to statistical inference. When applied, the probabilities involved in the theorem may have different probability interpretations. With Bayesian probability interpretation, the theorem expresses how a degree of belief, expressed as a probability, should rationally change to account for the availability of related evidence. Bayesian inference is fundamental to Bayesian statistics.

Bayes' rule and computing conditional probabilities provide a solution method for a number of popular puzzles, such as the Three Prisoners problem, the Monty Hall problem, the Two Child problem and the Two Envelopes problem.

Bayes' theorem - Wikipedia



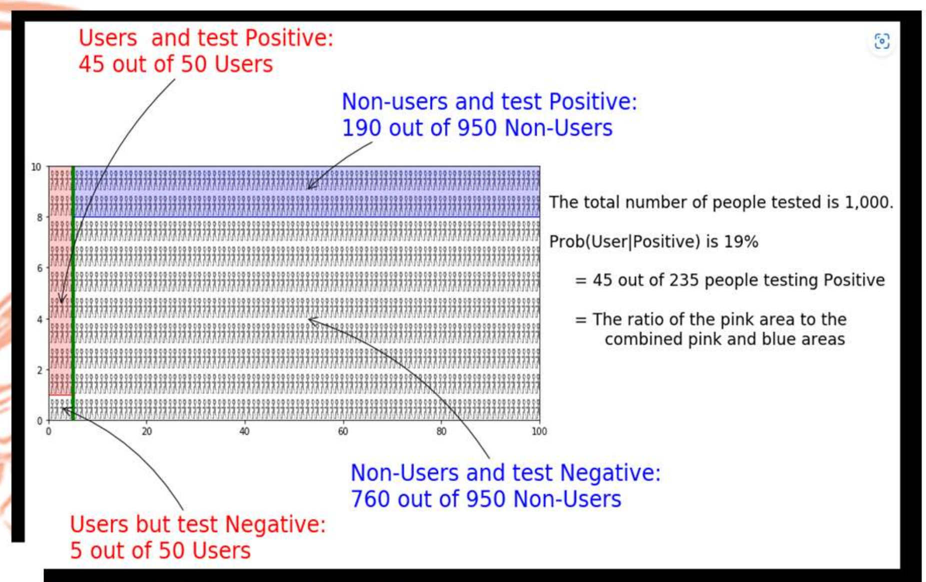
Example

Suppose, a particular test for whether someone has been using cannabis is 90% sensitive, meaning the true positive rate (TPR) = 0.90. Therefore, it leads to 90% *true* positive results (correct identification of drug use) for cannabis users.

The test is also 80% specific, meaning true negative rate (TNR) = 0.80. Therefore, the test correctly identifies 80% of non-use for non-users, but also generates 20% false positives, or false positive rate (FPR) = 0.20, for non-users.

Assuming 0.05 prevalence, meaning 5% of people use cannabis, what is the probability that a random person who tests positive is really a cannabis user

$$
\begin{aligned}
P(\text{User}|\text{Positive}) &= \frac{P(\text{Positive}|\text{User})P(\text{User})}{P(\text{Positive})} \\
&= \frac{P(\text{Positive}|\text{User})P(\text{User})}{P(\text{Positive}|\text{User})P(\text{User}) + P(\text{Positive}|\text{Non-user})P(\text{Non-user})} \\
&= \frac{0.90 \times 0.05}{0.90 \times 0.05 + 0.20 \times 0.95} = \frac{0.045}{0.045 + 0.19} \approx 19\%
\end{aligned}
$$



If 1,000 people were tested:

- 950 are non-users and 190 of them give false positive (0.20 × 950)
- 50 of them are users and 45 of them give true positive (0.90 × 50)

The 1,000 people thus yields 235 positive tests, of which only 45 are genuine drug users, about 19%.

Course Name :Basic Statistics using GUI-R (RKWard)
Module : Additional Concepts
Week 7 Lecture : 3

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

Expectation

For a linear function of the r.v X given by h(X) = aX+b, the expectation is

E(aX+b) = a E(X) + b

$$E(x) = \sum_{i=1}^{r} x_i . P(X = x_i)$$

Bivariate Random Variables

- In certain occasions, it is more meaningful to study two variables simultaneously rather than studying them in an isolated manner.
- (e.g.) Advertising expenditure and Sales, Height and weight, Sales revenue of two shops of a company

Joint Probability Mass Function

If two random variables are jointly involved in an experiment, their outcomes may be generated by bivariate probability function.

If X and Y are discrete random variables, their joint probability mass function is denoted by

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

A fair coin is tossed three times

X = number of heads in three tossings

Y = difference in absolute value between number of heads and number of tails

Find the joint probability mass function

The sample space for three coin tosses is {(HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), (TTT)}, where H represents heads and T represents tails.

Now, let's calculate the joint probabilities:

1. When X = 0 (no heads):
    1. (TTT): Y = |0 - 3| = 3
    2. P(X=0, Y=3) = P(TTT) = (1/2)^3 = 1/8
2. When X = 1 (one head):
    1. (HTT), (THT), (TTH): Y = |1 - 2| = 1
    2. P(X=1, Y=1) = P(HTT) + P(THT) + P(TTH) = 3*(1/2)^3 = 3/8
3. When X = 2 (two heads):
    1. (HHT), (HTH), (THH): Y = |2 - 1| = 1
    2. P(X=2, Y=1) = P(HHT) + P(HTH) + P(THH) = 3*(1/2)^3 = 3/8
4. When X = 3 (three heads):
    1. (HHH): Y = |3 - 0| = 3
    2. P(X=3, Y=3) = P(HHH) = (1/2)^3 = 1/8

- The distribution of X is P (X = x) = Summation for y , P (x,y) where the sum is over all possible y in the range of Y.
- Similarly P (Y = y) = Summation for X , P (x,y)

Conditional Probability Distribution

For each possible value x of X, as y varies over the range of Y the probabilities P (Y = y / X = x) define a probability distribution over the range of Y. The probability distribution, which may depend on the given value x of X, is called the conditional distribution of Y given X = x.

$$P[Y = y | X = x] = \frac{P[X = x, Y = y]}{P[X = x]}$$

$$E(X+Y) = E(X) + E(Y)$$
$$E(X-Y) = E(X) - E(Y)$$

If X and Y are independent random variables,

$$V(X+Y) = V(X) + V(Y)$$
$$V(X-Y) = V(X) + V(Y)$$

- The <u>covariance</u> of two random variables X and Y, denoted by Cov(X,Y) is the expected value of [X-E(X)] times [Y-E(Y)]. Mathematically,

$$\begin{aligned}Cov(X,Y) &= E[X-E(X)]\,[Y-E(Y)]\\ &= E(XY) - E(X)\,E(Y)\end{aligned}$$

Probability Mass Function

Let $P(x_i)$ be the probability that the random variable X assumes the value x_i [i.e., $P(X=x_i)$]. A probability mass function is a function that assigns probabilities to the values of a discrete random variable such that the following two conditions are satisfied.

1. $0 \leq P(x_i) \leq 1$, $i=1,2,\ldots,r$
2. $\sum_{i=1}^{r} P(x_i) = 1$

Probability Density Function

The function f(x) is a probability density function for the continuous random variable X, defined over the set of real numbers, R, if

f(x) $\geq$ 0 for all $x \in R$

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

$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

Marginals

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

Independence

$f(x,y) = f_X(x)\, f_Y(y)$, for all x and y

Expectation

The expected value of the random variable $g(X,Y)$

is $$E[g(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy$$

Course Name :Basic Statistics using GUI-R (RKWard)

Module : Distribution

Week 7 Lecture : 4

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

Hypergeometric Distribution

- The hypergeometric distribution is a discrete probability distribution that describes the probability of a specified number of successes (drawn from a finite population without replacement) in a fixed number of draws.

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

- N is the population size.
- K is the number of successes in the population.
- n is the number of draws.
- k is the number of successes in the draws.

The notation $\binom{a}{b}$ represents a binomial coefficient, which is the number of ways to choose b items from a set of a items.

Mean (μ):

$$\mu = \frac{n \cdot K}{N}$$

Variance (σ^2):

$$\sigma^2 = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$$

- Consider an urn with N balls, K of which are white and N-K are red. Suppose we draw a sample of ‘n’ balls at random without replacement from the urn, then the probability of getting k white balls out of ‘n’ is

- When the population is finite and the sampling is done without replacement, we obtain Hypergeometric Distribution.

- The hypergeometric distribution differs from the binomial distribution in that it's concerned with sampling without replacement, whereas the binomial distribution deals with sampling with replacement.

```
# Function to calculate the probability of getting k white balls out of n
get_probability <- function(N, K, n, k) {
  choose(K, k) * choose(N - K, n - k) / choose(N, n)
}

# Parameters
N <- 10  # Total number of balls in the urn
K <- 6   # Number of white balls in the urn
n <- 5   # Number of balls drawn in the sample

# Calculate probabilities for all possible numbers of white balls
possible_k <- 0:n
probabilities <- sapply(possible_k, function(k) get_probability(N, K, n, k))

# Plot the distribution
barplot(probabilities, names.arg = possible_k, xlab = "Number of White Balls",       ylab = "Probability", main = "Distribution of White Balls in Sample")
```

When N is large and n is relatively small compared to N, there is not much difference between sampling with replacement and sampling without replacement and the formula for the binomial distribution with the parameters n and $p = M/N$ may be used to approximate hypergeometric probabilities. ($n < 5\% (N)$)

As $N \to \infty$, Hypergeometric $\to$ Binomial with parameters n and $p = M/N$.

Negative Binomial

- Describes the number of successes in a sequence of independent and identically distributed Bernoulli trials (binary outcomes with a fixed probability of success) before a specified number of failures (or non-successes) occur. In other words, it gives the probability of observing a certain number of succ

The probability mass function (PMF) of the negative binomial distribution is given by:

$$P(X = r) = \binom{r+k-1}{r} \cdot p^r \cdot (1-p)^k$$

- X is the number of trials until the k-th failure occurs.
- p is the probability of success on each trial.
- r is the number of successes.
- k is the number of failures.

Mean (μ):

$$\mu = \frac{k \cdot (1-p)}{p}$$

Variance (σ^2):

$$\sigma^2 = \frac{k \cdot (1-p)}{p^2}$$

- The negative binomial distribution is useful in situations where we're interested in the number of trials required to achieve a certain number of successes, rather than just the number of successes in a fixed number of trials.

  Suppose a basketball player has a free throw success rate of 70%. We're interested in knowing how attempts and probability to make 5 successful free throws are related.
-

```r
# Parameters
p <- 0.70  # Probability of success (making a free throw)
r <- 5     # Number of successes desired

# Number of attempts
attempts <- 1:20  # Adjust the range as needed

# Compute PMF for each number of attempts
pmf <- dnbinom(attempts - r, size = r, prob = p)

# Plot the distribution
plot(attempts, pmf, type = "h", lwd = 2, col = "blue",
   xlab = "Number of Attempts", ylab = "Probability",
   main = "Negative Binomial Distribution")
```

Course Name :Basic Statistics using GUI-R (RKWard)

Module : Poisson Distribution

Week 7 Lecture : 5

Harsh Pradhan, Assistant Professor,
Institute of Management Studies, BHU
https://bhu.ac.in/Site/FacultyProfile/1_5?FA000562

Poisson Distribution

The probability mass function (PMF) of a Poisson distribution expresses the probability of observing a certain number of events k in a fixed interval of time or space, given the average rate of occurrence λ. The PMF of the Poisson distribution is given by the formula:

$$P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$$

Where:

- $P(X = k)$ is the probability of observing k events,
- e is the base of the natural logarithm (approximately equal to 2.71828),
- λ is the average rate of occurrence (also known as the rate parameter),
- k is the number of events observed,
- $k!$ denotes the factorial of k, which is the product of all positive integers up to k (e.g., $5! = 5 \times 4 \times 3 \times 2 \times 1$).

The Poisson distribution is often used to model the number of occurrences of rare events in a fixed interval of time or space, assuming that the events happen independently of each other, with a constant average rate of occurrence.

```
# Set the rate parameter (lambda)
lambda <- 2

# Generate random Poisson values
random_values <- rpois(10, lambda)
print(random_values)

# Calculate probabilities for specific values
prob_0 <- dpois(0, lambda)
prob_5 <- dpois(5, lambda)
```

- Events happen independently in time or space with, on average, λ events per unit time or space.
- Radioactive decay

  λ=2 particles per minute
- Lightening strikes

  λ=0.01 strikes per acre
- Radioactive decay
- X=# of particles/hour
- λ =2 particles/min * 60min/hour=120 particles/hr

$$P(x = 125) = \frac{120^{125} e^{-120}}{125!}, \text{ x=0, 1, 2, ...}$$

Poisson Approximation for the Binomial Distribution

- For Binomial Distribution with large n, calculating the mass function is pretty nasty
- Good news:

when $n \to \infty, \quad \pi \to 0, \quad n\pi \to$ a constant λ

Binomial$(n, \pi) \to$ Poisson(λ), i.e.:

$$\frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x} \to \frac{e^{-\lambda}\lambda^x}{x!}$$

- So for those nasty "large" Binomials ($n \geq 100$) and for small π (usually ≤ 0.01), we can use a Poisson with $\lambda = n\pi$ (≤ 20) to approximate it!

Suppose 1 in 5000 light bulbs are defective. Let X denote the number of defective light bulbs in a group of size 10000. What is the chance that at least 3 of them is defective?

Geometric Distribution

The geometric distribution $G(p)$ describes the probability distribution of the number of trials X needed to get the first success.

$P(X=k)=(1-p)^{(k-1)}\cdot p$

where:

k is the number of trials until the first success.

p is the probability of success on each trial.

Mean and Variance:

- The mean (expected value) of the geometric distribution is $1/p$, and the variance is (1-p)/(p*p).

Effect Size

pwr.t.test(d = 0.5, # Effect size sig.level = 0.05, # Significance level

power = 0.8, # Desired power type = "one.sample", # Type of t-test alternative = "two.sided") # Two-sided test

This function will return the minimum sample size required to detect the specified effect size with 80% power at a 5% significance level for a one-sample, two-sided t-test.

- It's important to note that the effect size is not a measure of statistical significance but rather a measure of practical significance or the magnitude of the effect. A larger effect size means that the difference or effect is more substantial and easier to detect with a given sample size and statistical power.

- Small effect size: d = 0.2
- Medium effect size: d = 0.5
- Large effect size: d = 0.8

```
# calculate Cohen's d for two independent samples:
lsr::cohensD(my.csv.data[["CSE_1"]],my.csv.data[["CSE_2"]])
# test  of means
lsr::cohensD(my.csv.data[["CSE_1"]],mu=3.9)
```

Package/Library is like Language

Documentation like Dictionary