STSM2616
Sample Distribution Theory and Inference
2022

Jocelyne Smith
2020046385

Table of Contents

The importance of missing data

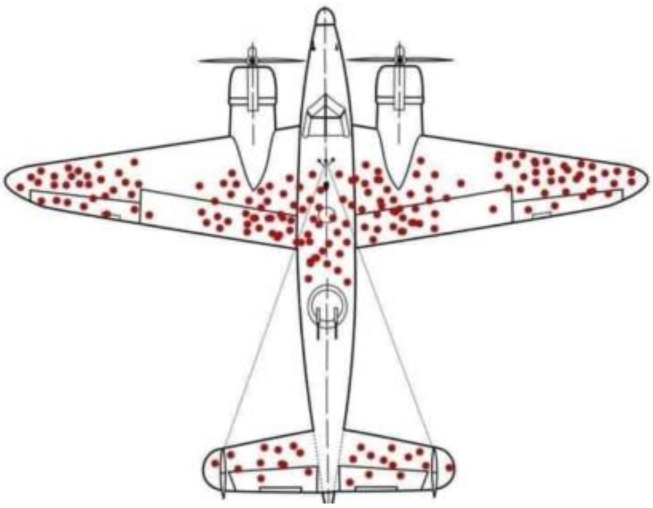
While mindlessly scrolling through social media, I found the following story which highlights what our professor mentioned in our first statistics lecture – the importance of missing data. This story tells of the real-life importance of not only understanding the data we are given, but also understanding the data we are not given.

The Story of Abraham Wald and the Missing Bullet Holes

“During World War II, fighter planes would come back from battle with bullet holes. The Allies found the areas that were most commonly hit by enemy fire. They sought to strengthen the most commonly damaged parts of the planes to predict the number that was shot down.

A mathematician Abraham Wald, pointed out that there was another way to look at the data. Perhaps the reason certain areas of the planes weren't covered in bullet holes was that planes thar were shot in those areas did not return. This insight lead to the armor being re-enforced on the parts of the places where there were no bullet holes.

The story behind the data is arguably more important than the data itself. Or more precisely, the reason behind why we are missing certain pieces of data may be more meaningful than the data we have. Don't only listen to what is being said. Listen more to what is not being said.”



I never really understood how important what our professor mentioned really was until I read this story. Amazing! I couldn’t believe how missing data can sometimes be more meaningful than the data we have.

The Ultimate goal

I believe it is important to always remember the goal when you are working on projects / assignments or studying for tests, for this reason, I am including my goals at the beginning of my portfolio so that I can always remember, "Why am I doing this?".

I have a dream of one day working for SpaceX. The work done at SpaceX has and will continue to change the course of humanity. Just like Tomas Eddison, Einstein and Stephan Hawking, Elon musk will be remembered as someone who changed humanity.

Elon Musk can celebrate that he:

- Donated $5.74 billion to end world hunger.
- Is the CEO of SpaceX:
    - the first private company to deliver a spacecraft to the International Space Station.
    - Which plans to put the first humans on mars in the next 20 years.
    - Is developing a program to actively remove cardon dioxide from the air and use it to power spacecraft.
- Is the CEO of Tesla who has changed the electric car industry and is developing the first self-driving car prototypes.
- Is developing a new underground system of transportation to solve traffic congestion in highly populated areas.
- And many, many more.

I don't have any need to be remembered as a famous like Elon Musk. However, I want to feel that my work has value. The work that I produce is deeply linked to my personal self-worth. I want to produce work I can be proud of and I want to be able to celebrate what I have achieved.

The way I Study

The following must be pointed out for any persons viewing my portfolio for them to understand the structure and organization of my portfolio.

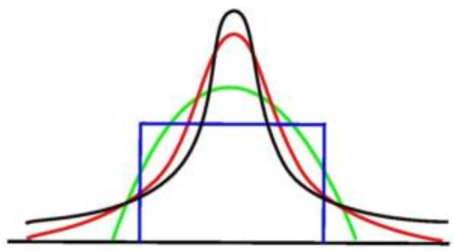
Firstly, a lot of revision and additional notes are included at the beginning of the portfolio as I want to have a place where I store information, I look up often.

Secondly, each Chapter will include:

- A logbook for every subchapter,
- Interview questions and answers,
- A summary,
- A mind-map,
- My personal thoughts and understanding are written in blue,
- Important concepts are highlighted,
- *Extremely important things are written in red, and bold and is preceded but a an Asterix,
- Links to additional information,
- Exercises at the end of every chapter,
- The assignment.

Important Definitions and concepts

Firstly, I think it is important that I revise my statistical in order for me to be able to understand statistical material I am reading and to correctly express during the interview.

Probability Distribution Function and Probability Density Function	What is the difference between a Probability Distribution Function and a Probability Density Function? • Probability distribution function: A function that shows the probabilities for the values that the random variable can take (defined for discrete random variables). • Probability density is defined for the continuous random variables, gives the probability of a certain random variable to assume a certain value.
Average / Mean / $\bar{X}$	"The sum of a collection of numbers divided by the count of the numbers in the collection" – Wikipedia.
Variance σ^2	"Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value." – Wikipedia.
Skewness	"The degree of asymmetry observed in a probability distribution" – Investopedia. Does the median (most) lie to the left-skewed (negative) or the right-skewed (positive) of the mean of the data? The normal distribution has no skewness as it is symmetrical about its mean.
(Excess) Kurtosis	"Kurtosis is a statistical measure that defines how heavily the tails of a distribution differ from the tails of a normal distribution. In other words, kurtosis identifies whether the tails of a given distribution contain extreme values." – Corporate finance institute.  This video explained that kurtosis is how pointy a distribution is in comparison to a normal distribution. A positive kurtosis is very pointy (the black line in the above image has a kurtosis of 1.2), and a negative kurtosis is very flat (the dark blue has a kurtosis of -1.2). A normal distribution has a kurtosis of 0 (green line).
Population *The population mean is the	"A set of similar items or events which is of interest for some question or experiment It can be a group of existing objects, or a hypothetical and potentially infinite group of objects conceived as a generalization from experience." – Wikipedia.

expended value!	Basically, it is all the items or events obtained in a study or potentially, infinite items as an idea / concept.
Sample	"A subset of data selected from a population that used to provide statistical information about the population." - Lumenlearning.com. It is chosen to represent the population and is used to make inference about the population.
Random variable	"A numerical value that describes the outcome of a statistical experiment" - britannica.com. A random variable can be discrete (finite), continuous (infinite) or mixed.
Observation	"A value of something of interest you're measuring or counting during a study or experiment" - statisticshowto.com. "An observation of a random variable is the value that is observed (what actually happened). The random variable itself is the process dictating how the observation comes about." - Wikipedia. An observation is the value that was measured, and a random variable is the process of who's values we are observing for example my eight is 173 cm (observation) when we are comparing female height (random variable).
Probability	"The measure of the likelihood that an event will occur." - opentextbc.ca. How likely is a coin to land on heads? Well obviously, a half.
Event	"An outcome or defined collection of outcomes of a random experiment" - statistics.com. How likely is it that the coin will land on heads? The coin landing on heads is an event. An event is an outcome in a defined collection of outcomes.
Correlation (Measured by means of a correlation co-efficient - formula below)	"A statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate)" - [jmp.com](jmp.com). More simply it looks at the strength of relationship between variables based on their pattern in relation to one another. Variables are positively correlated when they move in the same direction and negatively correlated when they move in opposite directions.
Covariance	"Covariance is a statistical term that refers to a systematic relationship between two random variables in which a change in the other reflects a change in one variable." - implilearn.com. What I understand from the above is that Covariance determines the relationship between the movements of two random variables. (It asks the question does a change in the one variable reflect a change in the other?). The greater the relationship, the more reliant the variables are one another and visa-versa. Positive covariance means there is a direct relationship and negative co-variance means there is an indirect relationship.

Moment Generating Function	"It is an alternative way of writing a random variables probability function" - Wikipedia, that allows us to calculate the functions moments. Moments allow us to make inference about the distribution such as the mean, variance, skewness, kurtosis, etc.
Markov's Inequality (Known mean)	Markov's inequality gives an upper bound for the percentage of a distribution that can be above or equal to a particular positive value. $$P(X \geq a) \leq \frac{E(X)}{a}$$ Furthermore, Markov's inequality says that the chance that a positive random variable is at least k times its mean can be no more than $\frac{1}{k}$.
Chebyshev's Inequality (Known mean and variance)	Chebyshev's theorem is used to find the proportion of observations you would expect to find within a certain number of standard deviations from the mean. $$P(\lvert X-\mu \rvert) > t) \leq \frac{\sigma^2}{t^2}$$ Furthermore, Chebyshev's inequality then states that the probability that an observation will be more than k standard deviations from the mean is at most $\frac{1}{k^2}$.
Law of Large Numbers	The large numbers of theorem states that if the same experiment or study is repeated independently a large number of times, the average of the results of the trials must be close to the expected value. $$P(\lvert \bar{X}_n - \mu \rvert > \varepsilon) \text{ as } n \to \infty$$

Formulae Sheet

The expected value and variance of distributions

The expected value and variance of different functions:

	Expected Value	Variance
$Geometric(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poison(\lambda)$	λ	λ
$Gamma(\alpha,\lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$Normal(\mu,\sigma^2)$	μ	σ^2
$Cauchy$	Does not exist.	Does not exist.
$Uniform(a,b)$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
$Binomial(n,p)$	$n \times p$ Multiplying the number of trials (n) by the probability of successes (p).	$n \times p \times (1 - p)$
$Bernoulli(p)$	p	$p(1-p)$
$Exponential(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Expected value formula

The expected value formula for discrete and continuous variables:

Discrete	Continuous
$E(X) = \sum_i x_i p(x_i)$	$E(X) = \int_{-\infty}^{\infty} xf(x)dx$

Variance formula

The variance of a single random variable:
$Var(x) = E(x^2) - E(x)^2$

The variance between two different random variables:
$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

Basic Rules for the Mean, Variance.

The website states that: *"It is a handy review for someone who has been away from statistics for a while but suddenly finds an article using these Rules."* It was exactly that. I was able to use this resource to review the concepts I've learned last year such as:
The following information is highlighted.

- Multiplying a random variable by a constant value, c, multiplies the expected value by that constant: $E(cX) = cE(X)$.
- Multiplying a random variable by a constant increases the variance by the square of the constant: $VAR(cX) = c^2 VAR(X)$.

The variance of the sum of two or more random variables is equal to the sum of each of their variances only when the random variables are independent.

$$Var(X+Y) = Var(X) + Var(Y).$$

Co-variance formula

If X and Y are dependent.	If X and Y are independent.
$Cov(X, Y) = E(XY) - E(X)E(Y)$	The covariance = 0.

*The co-variance only applies if X and Y are dependent on one another.

Coloration coefficient

$$p = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Moment generating functions

The MGF of different common distributions:

Distribution:	MGF:
$Geometric(p)$	$\frac{pe^t}{1-(1-p)e^t}$ for $t < \ln(1-p)$
$Poison(\lambda)$	$e^{\lambda(e^t-1)}$
$Gamma(\alpha, \lambda)$	$(\frac{\lambda}{\lambda - t})^\alpha$
$Normal(\mu, \sigma^2)$	$e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}$
$Cauchy$	Does not exist.
$Uniform(a, b)$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
$Binomial(n, p)$	$(1 - p + pe^t)^n$
$Bernoulli(p)$	$q + (1-p)e^t$
$Exponential(\lambda)$	$\frac{\lambda}{\lambda - t}$

Frequency distributions

The following information was quoted from professor Verster's notes (1st year statistics).

Discrete random variables

Negative Binomial	Hypergeometric
What is the number of trials needed for a given number of successes? $$P(k) = P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r},$$ $$k = 1,2,3,\ldots$$ p = the probability of a success. k = the k^{th} trial. r = the given number of successes.	"When you want to determine the probability of obtaining a certain number of successes without replacement from a specific sample size." - towardsdatascience. $$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$
Poisson	**Bernoulli**
The value x equals the number of times that a certain event occurs in each interval of time. $$p(k) = \mathrm{P}(\mathrm{X} = k)\begin{cases} e^{-\lambda}\frac{\lambda^k}{k!} & k = 0,1, \ldots, n \\ 0 & \end{cases}$$	Bernoulli trail: An experiment where only one of two outcomes are possible, a "success" / 1 and a "failure" / 0. $$P(x) = p(X = x) = p^x(1-p)^{1-x}, x = 0,\ 1$$
Binomial	**Geometric**
Counts how often a particular event (successes) occurs in a fixed number of tries or trials. $$P(x = k)\begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0,1, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$ *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with $p(X_i = 1) = p$. Then $y = X_1 + X_2 + \cdots + X_n$ is a binomial random variable.	The total number of trails up to and including the trail that the experiment has the first success. $$P(k) = P(X = k) = (1-p)^{k-1}p, \quad k = 1,2,3,\ldots$$

Continuous Random Variables

Uniform distribution

	$X \sim U(0,1)$	$X \sim U(a,b)$
Density function:	$f(x)\begin{cases} 1 \text{ if } 0 \le x \le 1 \\ 0 \quad \text{otherwise} \end{cases}$	$f(x)\begin{cases} \frac{1}{b-a} \text{ if } a \le x \le b \\ 0 \quad \text{otherwise} \end{cases}$
Cumulative density function:	$F(x) = x$ for $0 \le x \le 1$	$F(x) = \frac{x-a}{b-a}$ for $a \le x \le b$

Exponential distribution

Density function:	Cumulative density function:
$f(x)\begin{cases}\lambda e^{-\lambda x} & \text{if } x>0\\ 0 & \text{otherwise}\end{cases}$.	$F(x)\begin{cases}0 & \text{if } x\leq 0\\ 1-e^{-lx} & \text{if } x\geq 0\end{cases}$

Gamma distribution

The gamma function is defined as $\Gamma(\alpha)=(\alpha-1)!=\int_0^\infty x^{\alpha-1}\,e^{-x}dx$

Density function of the Gamma random variable: $f(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$

Properties of the gamma density function:

- $f(x)\geq 0$
- $\int_0^\infty f(x)dx=1$
- α is the shape parameter λ and is the scale parameter.

If $\alpha=1$, then $f(x)\ =\lambda e^{-\lambda x}$ (the exponential density function).

***Note that: $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$ and $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$.**

Normal distribution

Density function of the Normal random variable: A continuous random variable X is normally distributed with parameters μ and σ if it has the following density function:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2},-\infty<x<\infty,\quad -\infty<\mu<\infty,\quad \sigma>0$$

- μ is called the mean and it is a measure of location.
- σ is called the standard deviation and it is a measure of spread.

The normal distribution makes use of the following notation $X\sim N(\mu,\sigma^2)$.

*Note: if $\mu=0$ and $\sigma=1$ then we say $f(x)=\frac{1}{\sqrt{2\pi}}\,e^{\frac{1}{2}(x)^2}$ is the standard normal density.

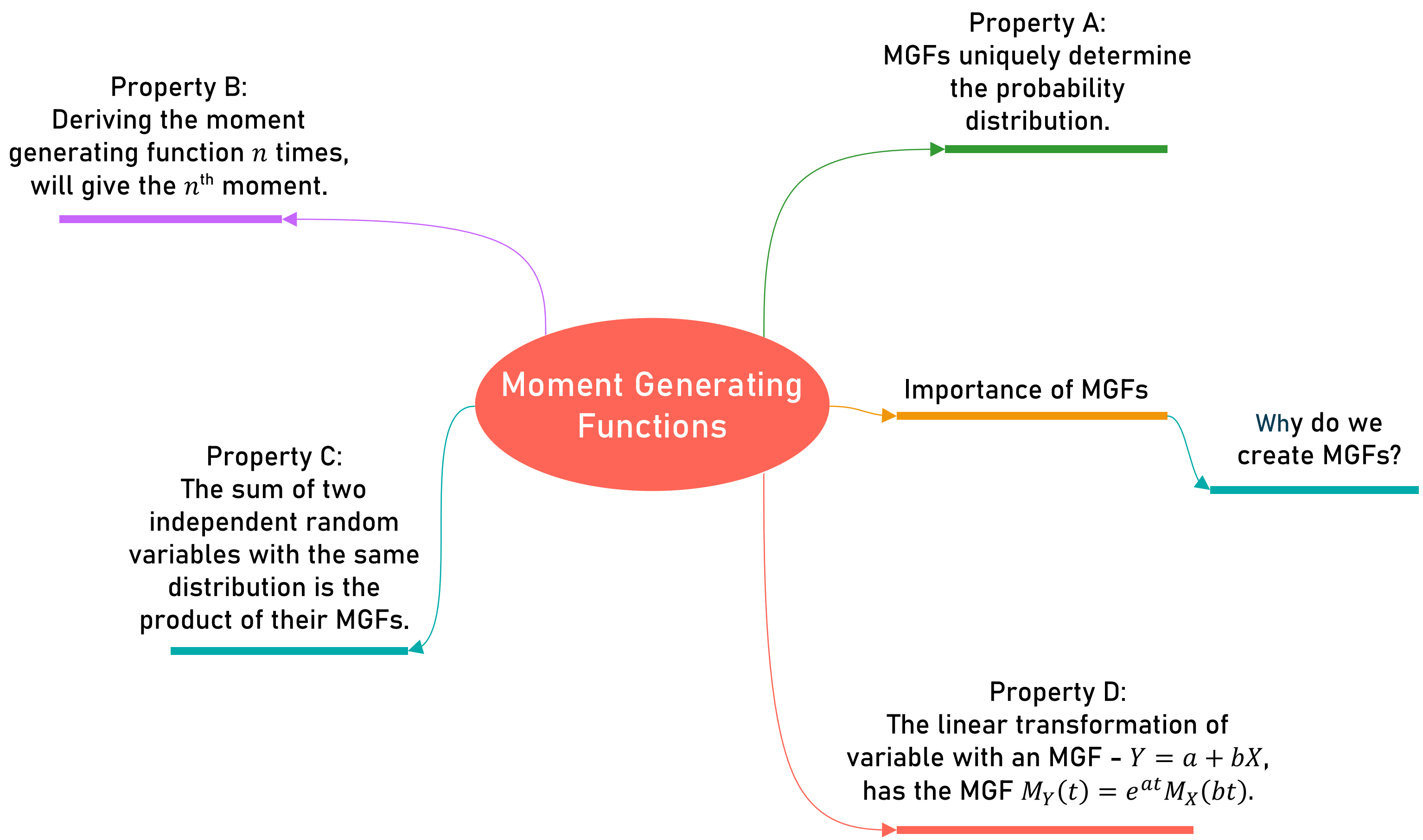
Beta distribution

Density function of the Beta random variable: A continuous random variable X is beta distributed with parameters a and b if it has the following density function:

$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},0\leq x\leq 1,a>0,b>0$$

The beta density is useful for modelling random variables that are restricted to the interval $[0,1]$. Note: The case $a=b=1$ is the uniform distribution.

Moment Generating Functions



Created using Microsoft Visio.

Logbook

The following process was used in the studying of this topic:

1. Firstly, I revisited the work from last year (STSM1624 – Introduction to Probability Theory). As I worked through last year's notes I:
    - Copied the work into this document and read through it thoroughly.
    - Certain concepts I couldn't remember, I highlighted in red. Those topics were googled, and sources were added.
    - Important concepts were revised using the key as stated above under the topic: The way I study.
2. I obtained the written examples from last year and annotated the examples to ensure that I remembered what I learned.
3. I read and completed the interview questions.
4. I included more examples at the end of the chapter.

Interview Questions and Answers

Interview Questions:

- Derive moment-generating-functions (MGFs) known and unknown distributions and understand and defend their usage.

- Why do we create MGFs (or characteristic functions, cumulant generating functions, etc., for that matter)?

Simply it turns a process of integration into differentiation (which is much easier) by:

- The MGF is obtained by integrating the density function multiplied by e^{tx} - exponential family.
- The moments are then derived by differentiating the MGF and substituting zero n times to obtain the n^{th} moment.
- Provides us with an alternative way of expressing a probability distribution in order to make statistical inference from it e.g. mean, variance, skewness and kurtosis.

What is a Moment Generating Function?

"It is an alternative way of writing a random variables probability function" - Wikipedia, that allows us to calculate the functions moments. Moments allow us to make inference about the distribution such as the mean, variance, skewness, kurtosis, etc.

According to risk.net, the shape of any distribution can be described by its various 'moments'. The first four are:

1. "The mean, which indicates the central tendency of a distribution.
2. The second moment is the variance, which indicates the width or deviation.
3. The third moment is the skewness, which indicates any asymmetric 'leaning' to either left or right.
4. The fourth moment is the Kurtosis, which indicates the degree of central 'peakedness' or, equivalently, the 'fatness' of the outer tails."

We can therefor infer the shape of any distribution by calculating its MGF and deriving its moments. These concepts have been discussed in detail under Important Definitions and concepts.

Why is the MGF useful?

*If the MGF can be found, the process of integration (which may be difficult) is replaced by the process of differentiation (which is much easier). The way I understand this, instead of integrating the density function n times to obtain the n^{th} moment, we can simply integrate the function multiplied by e^{tx} once and differentiating n times (while substituting t as 0) to obtain the n^{th} moment.

Source: Introduction to Probability, Statistics and Random Processes.

The moment generating function definition

Discrete	Continuous
$M(t) = \sum_x e^{tx} p(x)$	$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$
Sum over all values of x, the function $p(x)$ multiplied by e^{tx}.	Integrate from the lower bound to the upper bound of x, the function $f(x)$ multiplied by e^{tx}.

Properties of Moment generating functions

The properties of moment generating functions (details follow):

1. If two random variables have the same MGF, they have the same distribution.
2. Deriving the moment generating r times, will give the r^{th} moment.
3. The sum of two independent random variables with the same distribution is the product of their MGFs. (For the sum of gamma distributions: The second parameter needs to be the same.)
4. If Y can be expressed as a linear equation of X (i.e. $Y = a + bX$) and the MGFs of these distributions have the same distribution, then the MGF of Y can be expressed as e to the power of the constant multiplied by the paramater, multiplied by the MGF of X with the parameter of bt i.e. $e^{at}M_X(bt)$.

Property A

If the moment generating function exists for $\boldsymbol{t}$ in an open interval containing zero, it uniquely determines the probability distribution. Thus, if two random variables have the same MGF, they have the same distribution.

Property B

If the moment-generating function exists in an open interval containing zero, then $\boldsymbol{M^r(0) = E(X^r)}$. If the moment generating function can be calculated and it is derived r times (and t is substituted with zero) then we will obtain the r^{th} moment.

Property C

If $\boldsymbol{X}$ has the MGF $\boldsymbol{M_X(t)}$ and $\boldsymbol{Y = a + bX}$, then $\boldsymbol{Y}$ has the MGF $\boldsymbol{M_Y(t) = e^{at}M_X(bt)}$.

Proof:

$$M_{Y(t)} = E(e^{tY}) = E\left(e^{at+btX}\right) = E\left(e^{at}e^{btX}\right) = e^{at}E\left(e^{btX}\right) = e^{at}M_Y(bt)$$

Linear transformations of a random variable with MGF, has an MGF that is as follows:

General Normal Distribution: If Y follows a general normal distribution with parameters μ and σ, the distribution of Y is the same as that of $\mu + \sigma X$, where X follows a standard normal distribution, then:

$$M_Y(t) = e^{\mu t}M_X(\sigma t) = e^{\mu t}e^{\frac{1}{2}\sigma^2 t^2}$$

Property D

If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random variables with MGF's $\boldsymbol{M_X}$ and $\boldsymbol{M_Y}$ and $\boldsymbol{Z = X + Y}$ then $\boldsymbol{M_Z(t) = M_X(t)M_Y(t)}$ on the common interval where both MGF's exist. The MGF of the sum of independent random variables is the product of their MGFs.

Proof:

$$M_Z(t) = E(e^{tZ}) = E(e^{tX+tY}) = E(e^{tX}e^{tY})$$

From the assumption of independence: $\mathrm{M}_Z(t) = E(e^{tX})E(e^{tY}) = M_Z(t) = M_X(t)M_Y(t)$

The sum of the poison, gamma and normal MGFs

The MGF of the sum, of the MGFs of independent random variables (the same distributions and having an MGF), will follow the same distribution.

Sum of random variable	MGF
$Poisson(\lambda) + Poisson(\mu)$	$e^{(\lambda+\mu)(e^t-1)}$
$Gamma(\alpha_1, \lambda) + Gamma(\alpha_2, \mu)$	$(\frac{\lambda}{\lambda - t})^{\alpha_1+\alpha_2}$
$Normal(\mu, \sigma^2) + Normal(v, \tau^2)$	$e^{(\mu+v)t}e^{\frac{1}{2}(\sigma^2+\tau^2)t^2}$

*Only works for independent random variables!

The sum of the poison distributions' MGF

The sum of the MGF of two independent poison distributions will also have a MGF of a poison distribution. If X is a Poisson with parameter λ and Y is a Poisson with parameter, μ then $X + Y$ is a Poisson with parameter $\lambda + \mu$ is:

$$e^{\lambda(e^t-1)}\, e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}.$$

The sum of the Gamma distributions' MGF

If X follows a gamma distribution with parameters α_1 and λ and Y follows a gamma distribution with parameters α_2 and λ then the MGF of $X + Y$ is:

$$(\frac{\lambda}{\lambda - t})^{\alpha_1}(\frac{\lambda}{\lambda - t})^{\alpha_2} = (\frac{\lambda}{\lambda - t})^{\alpha_1+\alpha_2}.$$

The sum of the Normal distributions' MGF

If X follows a normal distribution with parameters μ, σ^2 and Y follows a normal distribution with parameters, v, τ^2 then the MGF of the sum of the distributions is:

$$e^{\mu t}e^{\frac{1}{2}(\sigma^2+\tau^2)t^2}\, e^{vt}e^{\frac{1}{2}(\sigma^2+\tau^2)t^2} = e^{(\mu+v)t}e^{\frac{1}{2}(\sigma^2+\tau^2)t^2}$$

*Note from last year: The MGF of the product of two independent exponential distributions is the same as the MGF of the gamma distribution.

Examples

Example 1

Find the expected and variance of a Poisson distribution using the moment generating function:

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

Multiply the e^{tx} with λ^x and rewrite as $(e^t\lambda)^x$, (by taking out the common factor x).

$$= e^{-\lambda} e^{\lambda e^t} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} e^{-e^t\lambda}$$

$$= e^{-\lambda} e^{\lambda e^t}$$

$$= e^{\lambda(e^t-1)}$$

Take out the $e^{-\lambda}$ term (it does not need to be part of the summation - it does not include an x variable).

*Remember the distribution of a Poisson: $e^{-\lambda}\frac{\lambda^k}{k!}$. The function can be manipulated to take on a Poisson distribution with parameter: $e^t\lambda$ by including the $e^{-parameter}$ and its inverse. The summation of distribution over all the values of its parameters can take on always equals 1.

Differentiate twice, remember: $e^{ax-b'} = ae^{ax-b}$. Apply the product rule: $\frac{\mathrm{d}}{\mathrm{d}x}[f(x) + g(x)] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$. $\lambda e^{t'} = \lambda e^t$ multiplied by $e^{\lambda(e^t-1)}$ and $e^{\lambda(e^t-1)'} = \lambda e^t e^{\lambda(e^t-1)}$. (As shown in the previous differentiation).

$$M_X'(t) = \lambda e^t e^{\lambda(e^t-1)}$$

$$M_X''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}$$

$$E(X) = M_X'(0) = \lambda$$

$$E(X^2) = M_X''(0) = \lambda^2 + \lambda$$

$$Var(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda$$

And substitute $t = 0$ and use the variance formula: $Var(X) = E(X^2) - E(X)^2$. Simpify.

Example 2

Find the expected and variance of a Gamma distribution using the moment generating function:

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx$$

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

$$= \left(\frac{\lambda^\alpha}{\Gamma(\alpha)}\right)\left(\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha}\right)$$

$$= \left(\frac{\lambda}{\lambda-t}\right)^\alpha = \lambda^\alpha(\lambda-t)^{-\alpha}$$

Multiply the e^{tx} with $e^{-\lambda x}$ and rewrite as $e^{x(t-\lambda)}$, (by taking out the common factor x).

Remove the $\frac{\lambda^\alpha}{\Gamma(\alpha)}$ from the equation (as it has no value that needs to be integrated).

Rewrite as a gamma distribution with parameter $\alpha + 1$. And multiply RHS by $\frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}$. And LHS by the reciprocal. The integration of the new gamma distribution is 1.

$$M_X'(t) = \lambda^\alpha(-\alpha)(\lambda-t)^{-\alpha-1}(-1) = \alpha\lambda^\alpha(\lambda-t)^{-\alpha-1}$$

$$M'_X(0) = \frac{\alpha}{\lambda}$$

$$M_X''(t) = \alpha\lambda^\alpha(-\alpha-1)(\lambda-t)^{-\alpha-2} = \alpha\lambda^\alpha(\alpha+1)(\lambda-t)^{-\alpha-2}$$

$$M_X(0)'' = \frac{\alpha(\alpha+1)\lambda^\alpha}{\lambda^{\alpha+2}} = \frac{\alpha(\alpha+1)}{\lambda^2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

And differentiate again. $Constant \times exponent \times base\ to\ the\ power - 1$. Substitute $t = 0$. Use the variance formula: $Var(X) = E(X^2) - E(X)^2$. Simpify. Substitute $a = 0$ and $t = 0$.

Examples obtained from: studylib.net.

Example 3

Consider a discrete random variable with the following probability distribution:

x	0	2	4	6
$p(x)$	0.4	0.3	0.2	0.1

Construct the moment generating function $\boldsymbol{m(t)}$ for this random variable:

$$m(t) = E(e^{tx}) = e^0(0.4) + e^{2t}(0.3) + e^{4t}(0.2) + e^{6t}(0.1)$$
$$= 0.4 + 0.3e^{2t} + 0.2e^{4t} + 0.1e^{6t}$$

Using the derivative, compute the expected value:

$$m'(t) = 0.6e^{2t} + 0.8e^{4t} + 0.6e^{6t}$$
$$E(X) = m'(0) = 0.6 + 0.8 + 0.6 = 2$$

Compute the expected value using the original definition:

$$E(X) = \sum xp(x) = 0(0.4) + 2(0.3) + 4(0.2) + 6(0.1) = 2$$

Example 4

The Poisson probability distribution with $\boldsymbol{\lambda} = \mathbf{3}$, has moment generating function:

$$\boldsymbol{m(t) = e^{3e^t - 3}}$$

It first and second derivates are given by:

$$\boldsymbol{m'(t) = 3e^t e^{3e^t-3}} \text{ and } \boldsymbol{m''(t) = e^{3e^t-3}(9e^{2t} + 3e^t)}$$

Determine the second moment:

$$E(X) = m''(0) = e^{3e^0-3}\left(9e^{2(0)} + 3e^0\right) = e^0(9+3) = 12$$

Determine the third moment:

I used a computing application Mathaway to compute the derivate.

$$m'''(t) = e^{3e^t-3}(18e^{2t} + 3e^t) + (9e^{2t} + 3e^t)(e^{3e^t-3})(3e^t)$$
$$E(X^3) = m'''(0)e^{3e^t-3}\left(18e^{2(0)} + 3e^{(0)}\right) + \left(9e^{2(0)} + 3e^t\right)\left(e^{3e^{(0)}-3}\right)\left(3e^{(0)}\right)$$
$$= (18+3) + (9+3)(3) = 57$$

Example 5

For a discrete random variable Y with moment generating function given by:

$$\boldsymbol{m(t) = \left(\frac{e^t + 4}{5}\right)^{10}}$$

Is a binomial distribution with $n = 10, p = \frac{1}{5}, q = \frac{4}{5}$. See Formulae sheet.

1. Determine the expected value of Y:
$$E(Y) = np = 2$$

2. Determine the expected value of $\boldsymbol{3Y^2 + 5}$:
$$Var(Y) = np(1-p) = (10)\left(\frac{1}{5}\right)\left(\frac{4}{5}\right) = \frac{28}{5}$$
$$E(3Y^2 + 5) = 3E(Y^2) + 5 = 3\left(\frac{28}{5}\right) + 5 = 21.8$$

3. Determine the variance of $\boldsymbol{4Y + 10}$:
$$Var(4Y + 10) = 4^2 Var(Y) = 16\left(\frac{8}{5}\right) = 25.6$$

Example 6

For a random variable Y with geometric distribution the MGF is given by $\boldsymbol{m(t)} = \frac{\boldsymbol{pe^t}}{\boldsymbol{1-qe^t}}$ where $\boldsymbol{q = 1 - p}$.

1. Demonstrate how the derivative of the MGF may be used to determine the expected value.

$$m(t) = \frac{(1-qe^t)pe^t - pe^{t(-qe^t)}}{(1-qe^t)^2}$$

$$E(Y) = m'(0) = \frac{(1-q) - p(-q)}{(1-q)^2} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

2. Determine the $\boldsymbol{E(Y)}$ and $\boldsymbol{Var(Y)}$ with MGF $\boldsymbol{m(t)} = \frac{\boldsymbol{5e^t}}{\boldsymbol{8-3e^t}}$.

   Y has a geometric distribution with $p = \frac{5}{8}$ and $q = \frac{3}{8}$.

$$m(t) = \frac{5e^t}{8-3e^t} = \frac{\frac{5}{8}e^t}{1-\frac{3}{8}e^t}$$

   $E(Y) = \frac{1}{p} = \frac{8}{5}$ and $Var(Y) = \frac{q}{p^2} = \frac{24}{25}$

3. Determine $\boldsymbol{E(Y)}$ and $\boldsymbol{Var(Y)}$ with MGF $\boldsymbol{m(t) = (0.2 + 0.8e^t)^{25}}$.

   Y has a binomial distribution with $n = 25$ and $p = 0.8$.

   $E(Y) = np = 20$ and $Var(Y) = npq = 4$

4. Determine the $\boldsymbol{E(Y)}$ and $\boldsymbol{Var(Y)}$ with MGF $\boldsymbol{m(t)} = \left(\frac{\boldsymbol{3+2e^{3t}}}{\boldsymbol{5}}\right)^{\boldsymbol{15}}$.

$$m(t) = \left(\frac{3+2e^t}{5}\right)^{15} = (0.6 + 0.4e^t)^{15}$$

   Y has a binomial distribution with $n = 15$ and $p = 0.4$.

   Note:

$$m(3t) = \left(\frac{3+2e^{3t}}{5}\right)^{15}$$

   Therefore $Y = 3X$:

   $E(Y) = E(3X) = 3np = 18$ and $Var(Y) = Var(3X) = 3^2npq = 32.4$

Rice Examples

Example 7: Question 81

Find the moment-generating function of a Bernoulli random variable, and use it to find the mean, variance, and third moment.

$$M(t) = E(e^{tX}) = \left(\sum_{k=0}^{1} e^{tk}\right)(P(X=k)) = e^{0}(1-p) + e^{t}(p)$$
$$= p(e^{t}-1)+1$$

Since X can only take on values 0 and 1 with probabilities $1-p$ and p respectively.

$$E(X) = M'(t) = pe^{t} - p + 1\,' = pe^{t} - 1 + 1 = pe^{t} = pe^{0} = p$$

$$\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2 \\
&= M''(0) - p^2 \\
&= pe^{t\,''} - p^2 \\
&= pe^{t} - p^2 \\
&= pe^{0} - p^2 \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}$$

$$E(X^3) = M'''(0) = pe^{t\,'''} = pe^{t} = pe^{0} = p$$

Example 8: Question 82

Use the result of Problem 81 to find the mgf of a binomial random variable and its mean and variance.

$$M(t) = (p(e^{t}-1)+1)^{n}$$

where n = number of trials

A binomial random variable is the sum of n independent Bernoulli random variables.

$$\begin{aligned}
E(X) = M'(t) &= (p(e^{t}-1)+1)^{n\,'} \\
&= n(p(e^{t}-1)+1)^{n-1}pe^{t} \\
&= n(p(e^{0}-1)+1)^{n-1}pe^{0} \\
&= n(1)^{n-1}p \\
&= n(1)p \\
&= np
\end{aligned}$$

$$E(X^2) = M_X'' = n(p(e^{t}-1)+1)^{n-1}pe^{t\,''} = (pe^{t})(n-1)(n)(p(e^{t}-1)+1)^{n-1}pe^{t}$$
$$+\ n(p(e^{t}-1)+1)^{n-1} + pe^{t}$$

substitute $t=0$

$$= n^2p^2 - np^2 + np$$

Apply chain rule - Derivative of the outside function $\times$ derivative of the inside function.

Example 9: Question 91

Use the MGF to show that if X follows an exponential distribution, cX $(c > 0)$ does also.

Firstly, we calculate the MGF of X as an exponential function.

$$f_X(x) = \lambda e^{-\lambda x}$$

$$M_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \int_0^\infty \lambda e^{(t-\lambda)x} dx$$

$$= \left[\frac{\lambda}{t-\lambda} e^{(t-\lambda)x} \right]_0^\infty$$

$$= \frac{\lambda}{\lambda - t}$$

We equate $Y = cX$.

$$M_Y(t) = M_X(ct) = \frac{\lambda}{\lambda - ct} = \frac{\frac{\lambda}{c}}{\frac{\lambda}{c} - t}$$

As MGFs uniquely determine the distribution, Y is an exponential random variable with parameter $\frac{\lambda}{c}$.

Example 10: Question 95

Show that if X and Y are independent, their joint moment-generating function factors.

We want to prove that the join MGF of X and Y is the product of the MGF of X and Y.

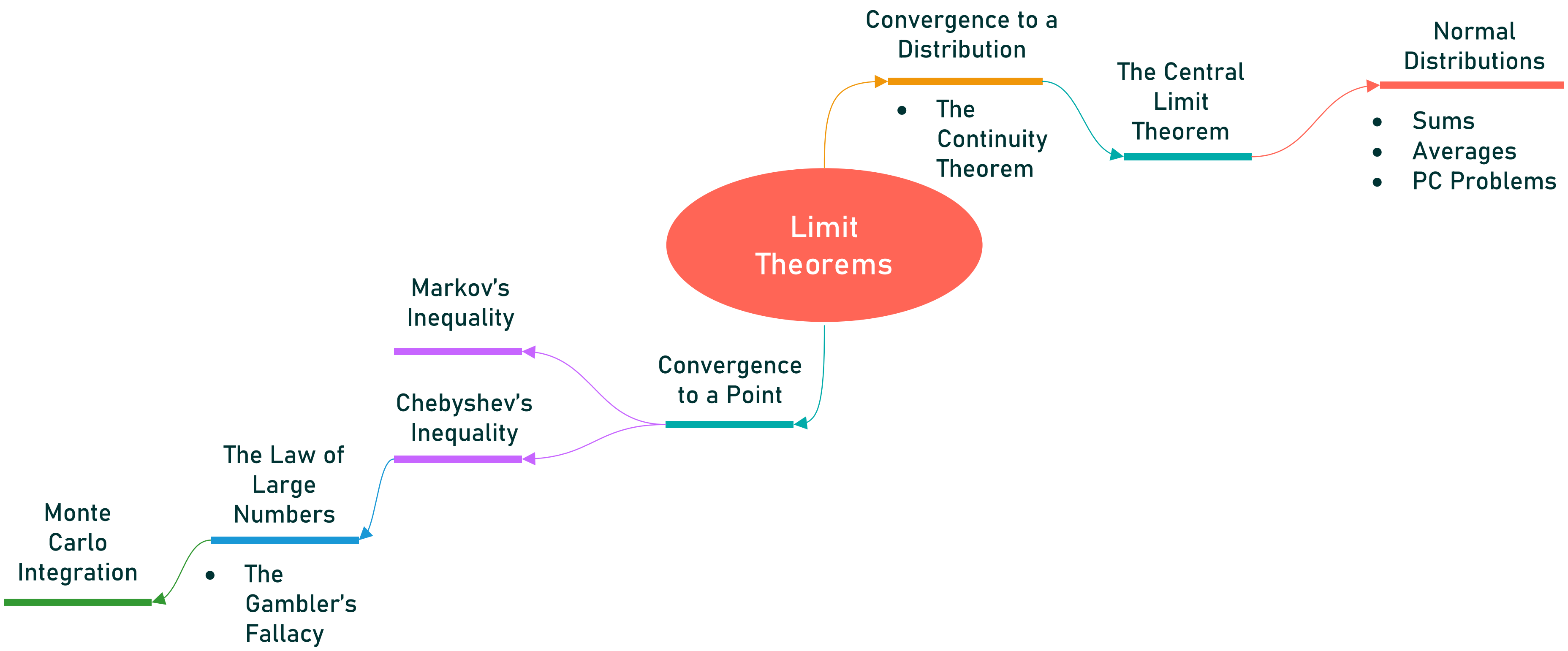
$$M_{X,Y}(s,t) = E(e^{sX+tY}) = E(e^{sX}e^{tY})$$

$$= E(e^{sX})E(e^{tY})$$

$$= M_X(s) M_Y(t)$$

Limit Theorems

How does it all fit together?



Created using Microsoft Visio.

Back to Table of Contents

Logbook

Logbook:

- I researched the history, significance, and interesting information on Markov's inequality, Chebyshev's inequality, and the law of large numbers. The detail of this research is documented below (under the relevant headings).
- I still wasn't extremely confident in the material, so I added examples for Markov and Chebyshev's inequalities.
- I then added more research for the law of large numbers.
- The following process was used for Monte Carlo Integration:
    - Before reading through the pages on Monte Carlo integration in Rice, I read the following article: Mathematical Foundations of Monte Carlo Methods. It provided me with the fundamental theory such as: What is Monte Carlo integration? Why do we make use of Monte Carlo Integration? (The details of these discoveries are documented below).
    - Furthermore, I was able to revise statistical vocabulary and definitions such as Random Variable, Expected value, estimator etc.
    - Afterwards, I watched the following video, Monte Carlo integration which explained the monte Carlo estimate. I learned that according to the law of large numbers that the integrand will converge almost surely to an expected value. I was able to follow through the explanation of the example done in the video and apply the principles in the exercises at the end of the chapter.
    - I read page 179 in Rice and completed the examples on Monte Carlo integration at the end of the chapter. I used quizlet.com to complete the questions. I also used a statistical software R (used in STSM2634) to generate n random variables of the uniform distribution.
- I completed the proof for the central limit theorem and the continuity theorem.
- I read through the learning outcomes and completed the interview questions.
- Lastly, I completed Rice examples for the topic.

Interview Questions

- Understand the difference between convergence to a point (e.g. LLN), and convergence to a distribution (e.g. CLT). What do you understand about the CLT?
  Convergence to a point means to come together to a single value. This applies to the Law of Large numbers which states if the same experiment or study is repeated independently a large number of times, the average of the results of the trials will approach the expected value (the point) as n gets closer to infinity. *The result becomes closer to the expected value as the number of trials is increased.
  Convergence to a distribution means at once variable becomes like another variable as it tends towards finity. This applies to the CLT as the sums of the averages of a large number of variables (from the same distribution) will be a normal distribution.

- Understand how one uses Chebyshev's inequality to prove the law of large numbers and apply the inequality and law to practical problems.
  Chebyshev's inequality that states that the probability of a variable X differing from its mean by some small constant k is less than or equal to the variance of X divided by the square of the constant k.
  We replace X with the sample mean and replace k with epsilon. We express the variance of the sample as a sum of independent identical random variables therefore the variance is n multiplied by the variance. As n tends to infinity, it follows that the right side of the inequality equals 0. - medium.com.

$$\frac{var(\bar{X}_n)}{\varepsilon^2} = \frac{\frac{1}{n^2} var(X_i) \sum_{i=1}^{n} var(X_i)}{\varepsilon^2} = \frac{\frac{n\sigma^2}{n^2}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

- Understand how Monte Carlo integration works and apply in practice.
- Understand and solve problems based on the continuity theorem.
- Understand the implications of the CLT on what we observe in nature and science.
  Lots of things in nature and life are normally distributed since it is averages of underlying processes.

- State and apply the CLT (for random variables, sums, or averages) to solve practical problems.
- Explain how a basic form of the CLT is proved.

We want to show that the MGF of a sequence of independent random variables is equal to the MGF of a standard normal distribution as n tends to infinity.

1. We use the properties of MGFs to calculate the MGF of the sequence of independent random variables.
2. We useTaylor series expansion to evaluate the sequence of MGF calculations.
3. Lastly, we apply the limit that n tends to infinity of the answer of the Taylor series explosion and see that the MGF of the sequence of independent random variables is the MGF of a standard normal variable.
4. We can say that the MGF of a sequence of independent random variables converges in distribution to a standard normal random variable.

Markov's inequality

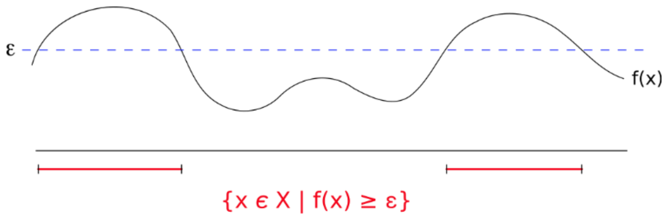
Named after the Russian mathematician Andrey Markov, the inequality appeared earlier in the work of Markov's teacher - Pafnuty Chebyshev. Many sources, refer to Markov's inequality as Chebyshev's first inequality.

What does Markov's inequality state?

Markov's inequality gives an upper bound for the percentage of a distribution that can be above or equal to a particular positive value.

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Basically, saying at most this percentage can be bigger than t.



It tells us where in the function the value the function is bigger than some positive constant value.

"Markov's inequality gives an upper bound for the measure of the set (indicated in red) where $f(x)$ exceeds a given level ε. The bound combines the level ε with the average value of f" – Wikipedia.

Furthermore, Markov's inequality says that the chance that a positive random variable is at least k times its mean can be no more than $\frac{1}{k}$. A good example of this is that if we assume that no income is negative, Markov's inequality states that no more than $\frac{1}{5}$ of the population can have more than 5 times the average income.

Usage

Markov's inequality allows us to define bounds on distributions when we only know the average / expected value.

Why is it important?

Markov's inequality has the following significance:

1. In general, no better bound (using only $E(X)$) is possible.
2. The other major use of Markov's inequality is to prove Chebyshev's inequality.

The following examples were obtained from math.dartmouth.edu.

Example 1

Suppose that the average grade on the upcoming Math $\mathbf{20}$ exam is $\mathbf{70}\%$. Give an upper bound on the proportion of students who score at least $\mathbf{90}\%$.

$P(X \geq 90) \leq \frac{E(X)}{90} = \frac{7}{9} = 0.77778$

This means that at most 77.8% of students can score at least 90%. Since the average is 70% (which is equal to $\frac{7}{9} \times 90$) we can conclude that the other $\frac{2}{9}$ of students must score 0 if the inequality were to apply. The chances of any student scoring zero for a test is very unlikely and therefore we can see that the inequality is not a very good estimate.

Example 2

A coin is weighted so that its probability of landing on heads is $\mathbf{20}\%$. Suppose the coin is flipped $\mathbf{20}$ times. Find a bound for the probability it lands on heads at least $\mathbf{16}$ times.

The actual value is a binomial distribution with $n = 20$ and $p = \frac{1}{5}$. The expected value is $= \frac{1}{5} \times 20 = 4$.

According to Markov's Inequality:

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

Therefore, the probability that the coin lands on heads at least 16 times is greater than or equal to 0.25. This does not feel right intuitively, therefore we compare this estimate with the actual value of the probability is:

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

These values differ significantly, furthermore showing that Markov's inequality is not a very good estimate.

Chebyshev's inequality

What is Chebyshev's inequality?

Chebyshev's theorem gives an estimate of the percentage of observations you would expect to find within a certain number of standard deviations from the mean. *Slightly stronger than Markov's but still not an extremely accurate estimate.

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

Chebyshev's inequality then states that the probability that an observation will be more than k standard deviations from the mean is at most $\frac{1}{k^2}$.

k	Min % within $\boldsymbol{k}$ standard deviations of the mean
2	75%
3	87.5%

Its practical usage is similar to the 68–95–99.7 rule, which applies only to normal distributions. empirical rule, is a shorthand used to remember the percentage of values that lie within an interval estimate in a normal distribution: 68%, 95%, and 99.7% of the values lie within one, two, and three standard deviations of the mean, respectively. - Wikipedia.

Usage

Chebyshev's inequality enables me to derive bounds on probabilities when both the mean and the variance, of the probability distribution are known.

"While in principle Chebyshev's inequality asks about distance from the mean in either direction, it can still be used to give a bound on how often a random variable can take large values, and will usually give much better bounds than Markov's inequality" - math.dartmouth.edu.

Importance

Chebyshev's inequality was used to prove his version of the law of large numbers.

Example 3

Suppose we have sampled the weights of dogs in the local animal shelter. The sample has a mean of 9 kg with a standard deviation of 2 kg.

With the use of Chebyshev's inequality, we know that at least 75% of the dogs that we sampled have weights that are two standard deviations from the mean.

Two times the standard deviation gives us: $2 \times 2 = 4$.
This tells us that 75% of the dogs have weight from 5 to 13 kg.

Subtract and add twice the standard deviation from the mean of 9.

The following examples were obtained from math.dartmouth.edu.

Example 4

Suppose a fair coin is flipped 100 times. Find a bound on the probability that the number of times the coin lands on heads is at least 60 or at most 40.

Let X be the number of times the coin lands on heads. We know X has a binomial distribution with expected value $n \times p = 100 \times \frac{1}{2} = 50$ and variance $n \times p \times (1 - p) = 100 \cdot 0.5 \cdot (1 - 0.5) = 25$.

By Chebyshev, we have:
$P(X < 40 \cup X > 60) = P(|X - \mu| \geq 10) \leq \frac{25}{10^2} = \frac{1}{4}$.

The average 50 must be subtracted/added by what number to what to obtain 40 and 60 respectively?

The actual value of the probability is:

$$P(X < 40 \cup X > 60) = \binom{100}{40} 0.5^{40} \times 0.5^{60} + \binom{100}{60} 0.5^{60} \times 0.5^{40}$$
$$\approx 0.056888$$

The actual probability is close to 5%.

Example 5

Let's revisit Example $\mathbf{3}$ in which we toss a weighted coin with probability of landing heads $\mathbf{20}\%$. Doing this $\mathbf{20}$ times, Markov's inequality gives a bound of $\frac{\mathbf{1}}{\mathbf{4}}$ on the probability that at least $\mathbf{16}$ flips result in heads.

Using Chebyshev's inequality:

$$\begin{aligned} P(X \geq 16) &= P(0 \leq X \leq 16) \\ &= P(-8 \leq X \leq 16) \\ &= P(|X - 4| \geq 12) \quad \text{(Since X can't be negative)} \\ &\leq \frac{Var(X)}{12^2} \\ &= \frac{20 \times 0.2 \times 0.8}{144} \\ &= \frac{3.2}{144} \\ &= \frac{1}{45} \end{aligned}$$

We manipulate the equation as follows: $-8 - 4 = -12$ and $16 - 4 = -12$

This is a much better bound than given by Markov's inequality, but still far from the actual probability.

The law of Large Numbers

In the 16th century, mathematician Gerolama Cardano recognized the Law of Large Numbers but never proved it. In 1713, Swiss mathematician Jakob Bernoulli proved this theorem in his book, Ars Conjectandi. He also discovered the fundamental mathematical constant e. - Investopedia.

What is the law of large numbers?

In a sequence of random variables, Z_n id such that $P(|Z_n - \alpha| > \varepsilon) \to \infty$ as $n \to \infty$ (the probability that the average minus a scalar approaches zero as n approaches infinity) for any $\varepsilon > 0$ where α is some scalar then Z_n is said to converge in probability to α.

$$\lim_{n\to\infty} P(|Z_n - \alpha| > \varepsilon) = 0$$

If we take a sample of n observations and we average all the observations. Then, the sample mean (the average of the observations) will approach the expected value (the actual value) as the sample becomes larger and larger. – https://study.com.

We use the notation $Z_n \overset{p}{\to} \alpha$ to denote convergence in probability.

What I understand from this is that the large numbers of theorem states that if the same experiment or study is repeated independently a large number of times, the average of the results of the trials will approach the expected value as n gets closer to infinity. *The result becomes closer to the expected value as the number of trials is increased.

If there were a content to guess the number of sweets in a jar. If you take a single guess it would be far from the actual value. However, as you take more and more guesses, and you take the average of all those guesses you will eventually obtain a value that is extremely close to the actual number of sweets in the jar.

The Gambler's fallacy

The gambler's fallacy (also known as the Monte Carlo fallacy), is a false belief that a certain random event is less likely or more likely to happen based on the outcome of a previous event or series of events.

Example 6: Proof of the law of large numbers

Let $X_1, X_2, \ldots, X_i$ be as sequence of independent random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

$$\text{Let } \bar{X} = \frac{1}{n}\sum_{n=1}^{n} x_i \text{ Then for any } \varepsilon > 0,$$

$$P(|Z_n - \alpha| > \varepsilon) \to 0 \text{ as } n \to \infty$$

$\lim\limits_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$

OR

$\bar{X}_n \overset{p}{\to} \mu$

$\bar{X}_n$ converges in probability to μ.

We first find the $E(\overline{X_n})$ and $Var(\overline{X_n})$:

Calculating the Expected value and Variance of $\bar{X}$ gives:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

The expectation does not apply to the constant or the summation but the variable that is summed over.

$$= \frac{1}{n}\sum_{i=1}^{n} \mu$$

The Expected value is simply the mean.

$$= \frac{n\mu}{n}$$

$$= \mu$$

The summation of n variables μ is $n\mu$ (the variables are independent and identically distributed).

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$$

The variance of a constant is the constant squared.

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2$$

The variance is sigma squared.

$$= \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

The summation of n variables σ^2 is $N\sigma^2$ (variables are independent and identically distributed).

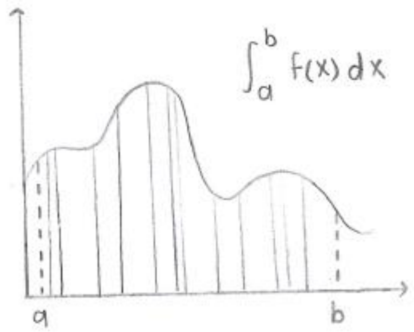
From Chebyshev's inequality: $P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$ $\quad P(|X - \mu X| > \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}$

Thus $P(|\bar{X}_n - \mu_{\bar{X}_n}| > \varepsilon) \leq \frac{\sigma_{\bar{X}_n}^2}{\varepsilon^2}$ $\quad P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$ as $n \to \infty$

Monte Carlo Integration

What is Monte Carlo Integration?

Theory: Monte Carlo integration is an technique used to find an approximation of the average of a random variable (e.g. height of the adult population) by taking random samples from the population and calculating its average.



More specifically to estimate the integral of a function by averaging the random samples of the function's value.

Why do make use of Monte Carlo Integration?

Many things in life are too hard to evaluate when it involves very large numbers. Let's say for example I want to calculate the average height of the adult population of a country, it would mean I have to measure the height of each person making up that population, summing up the numbers and dividing them by total number of people measured. It would be very time consuming and very difficult.

The Monte Carlo Estimate

The approximation of the average value of the random variable X, (the height of the adult population of a given country), is equal to the sum (the Σ sign) of the height of N adults randomly chosen from that population (the samples / x_n), divided by the number N (the sample size).

Represents the average of the sum of the samples taken from the population of size N.

$$\hat{I}(f) = \frac{1}{N}\sum_{n=1}^{N} x_n$$

Represents the $Approximation(E(X))$ where $E(X)$ is the $Approximation\big(Average(X)\big)$ and X represents the random variable we are interested in.

The law of large numbers applies to Monte Carlo Integration: The difference between the approximation and the actual result, gets smaller as the sample size increases.

Convergence in distribution

Let $X_1, X_2, \ldots$ be a sequence of random variables with cumulative distribution functions $F_1, F_2, \ldots$ and let X be a random variable with distribution function F. We say that X_n converges in distribution to X if

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

As the number of random variables in a sequence with cumulative distribution functions increase it strives closer and closer to a random variable with distribution function.

If this is the case, then $X \approx X_n$ if n is large.

We sometimes use the notation $X_n \xrightarrow{d} X$ to denote convergence in distribution.

Continuity theorem

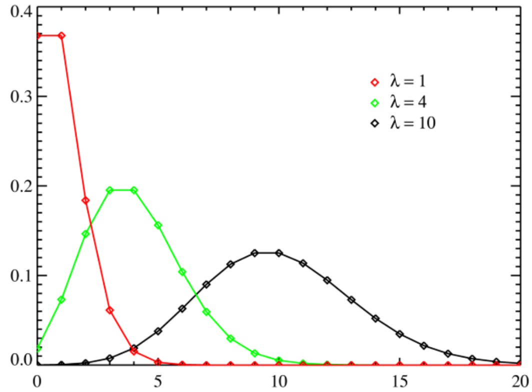
"Used to prove the central limit theorem and it is one of the major theorems concerning characteristic functions." – Rice.

Let F_n be a sequence of cumulative distribution functions with the corresponding moment-generating function M_n. Let F be a cumulative distribution function with the moment-generating function M. If $M_n(t) \to M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity points of F.

$X_n \xrightarrow{d} X$ if and only if $\lim_{n\to\infty} F_n(x) = F(x)$
if and only if $\lim_{n\to\infty} M_n(t) = M(t)$

Approximation of the Poisson distribution

A standardized Poisson (probability of a given number of events happening in a fixed interval of time) variable converges in distribution to a standard normal variable as λ approaches infinity.



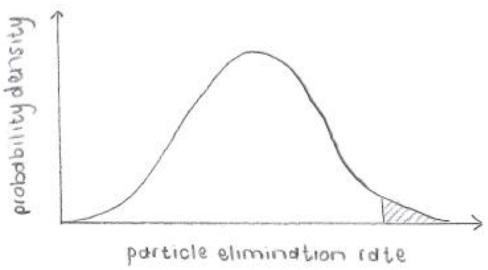
Standardization – The Z-score and P-value

1. First, you calculate a z-score for the sample mean value. To standardize a distribution (obtain a distribution with mean zero and standard deviation one) we make use of the z-score: $\frac{x-\mu}{\sigma}$. The z-score tells you how many standard deviations away the value you are interested in is away from the mean.
2. Then, you find the p-value for your z-score using a z-table. Every z-score has an associated p-value that tells you the probability of all values below or above that z-score occurring. This is the area under the curve left or right of that z-score.

Example 7 – Rice Example B

Suppose that the number of insurance claims, N, filed in a year is Poisson distributed with $\mathbf{E(N) = 10,000}$. Use the normal approximation to the Poisson to approximate $\mathbf{P(N > 10,200)}$.

$$P(X > 950) = 1 - P(X < 950)$$
$$= 1 - P\left(\frac{X-900}{\sqrt{900}} < \frac{950-900}{\sqrt{900}}\right)$$
$$= 1 - \phi\left(\frac{5}{3}\right)$$
$$= 1 - 0.95154$$
$$= 0.04846$$



I used the following source from StackExchange to solve this problem. I learned how to solve problem. I learned how to solve problems using the continuity theorem by standardizing the variable and using the z-table to calculate the probability.

The Central limit theorem

What is the Central Limit Theorem?

The CLT states that: Averages are normally distributed with parameters μ and $\frac{\sigma^2}{n}$ as n tends to infinity.

$$\overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \; as\, n \; \rightarrow \infty$$

What I understand from the CLT is that the sums of the averages of a large number of variables (from the same distribution) will be a normal distribution. Similarly, if you have a random variable that follows a normal distribution, we may assume that random variable is an average or sum.

What is the Significance of the CLT?

Lots of things in nature and life are normally distributed since it is averages of underlying processes. And we can use this to our advantage.

The central limit theorem tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach normality as the sample size n increases. *It does not matter what the distribution of $\boldsymbol{X}$ is. We only care about the sample average.

The mean of one sample from a population and the population mean can differ. The CLT allows us to make a very good estimate of population mean by using a normal graph of the sample mean. Thus, as the sample size increases the sampling error will decrease.

Source: simplypsychology.org.

Using the CLT to calculate probability

*I should verify that $\boldsymbol{Z_n}$ has a mean of $\mathbf{0}$ and a variance of $\mathbf{1}$.

Standardizing Z_n:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

*Sometimes we make $\boldsymbol{n = 1}$ and still use this approximation, as a single variable may be the sum of other variables e.g. having $\boldsymbol{X = X_1 \sim Bin(m, p)}$ and taking $\boldsymbol{Z_1 \approx Z}$ because we see $\boldsymbol{X}$ as:

$$\sum_{i=1}^{m} \boldsymbol{Y_i}$$

where $\boldsymbol{Y_i \sim Bernoulli(p)}$.

The Standardized variable converges in distribution to a random variable Z that converges in distribution to a standard normal distribution.

$$Z_n \xrightarrow{d} Z \sim N(0,1)$$

Therefore we can use the equations to calculate probability using the CLT.

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \phi(x) \quad \text{for all } -\infty < x < \infty$$

$$\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \le x\right) = \Phi(x)$$

$$\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right) = \Phi(x)$$

In the case that we know the value of n, if I standardizes X I can assume that it converges in distribution to a standard normal (mean zero and variance 1).

$$\frac{S_n - nE(X)}{\sqrt{nVar(X)}} \overset{d}{\approx} N(0,1)$$

Approximation of the Uniform, Exponential, and binomial distributions

Like the examples before, the standard uniform and standard exponential distributions converge to a normal distribution as the value of their parameters increase.

Example 8: Example F

Since a binomial random variable is the sum of independent Bernoulli random variables, its distribution can be approximated by a normal distribution. According to the following source:

- The approximation is best when the binomial distribution is symmetric i.e. p = $\frac{1}{2}$.
- Furthermore, the rule of thumb is that the approximation is reasonable when $np > 5$ and $n(1 - p) > 5$.

Suppose that a coin is tossed 100 times and lands heads up 60 times. Should we be surprised and doubt that the coin is fair?

100 trials with probability of success = $\frac{1}{2}$ (heads or tails)

Expected Value(X) = np = 50

Variance(X) = np(1-p) = 50($\frac{1}{2}$) = 25

The number of heads is fairly small in this calculation, we should calculate the probability of its more extreme case i.e

$$P(X > 60) = \frac{X-50}{\sqrt{5}} > \frac{60-50}{\sqrt{5}}$$

$$= 1 - \phi(2)$$

$$= 1 - 0.97725$$

$$= 0.02275$$

which is very small, so the fairness of the coin is called questionable.

In this section of work, I learned how to standardize a distribution using a z-score. I also learned how to use a z-table to look up probability of all values below or above that z-score occurring. Using (and the examples) I am now able to solve problems using the Central Limit theorem.

Example 9: Proof of the CLT

The proof in Rice was a bit difficult to understand. For this reason, I used the Dr. Morné Sjölander's slides and complied the following proof:

Let $X_1, X_2, \dots, X_i$ be as sequence of independent random variables with mean μ and variance σ^2 and the common distribution F and moment generating function M defined.

MGFs Property C:
$Y = a + bX$ implies $M_Y(t) = e^{at}M_X(bt)$.
We have $a = 0$ and $b = \frac{1}{\sqrt{n}\sigma}$
$Z_n = a + bS_n = 0 + \frac{1}{\sqrt{n}\sigma}S_n$
$M_{Z_n}(t) = M_{S_n}\left(\frac{t}{\sqrt{n}\sigma}\right)$

$$\text{Let} \quad S_n = \sum_{n=1}^{n} x_i$$

We will prove the theorem for the case where $\mu = 0$.

$$\text{And} \quad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n(0)}{\sigma\sqrt{n}} = \frac{S_n}{\sigma\sqrt{n}}$$

$$M_{S_n}(t) = [M(t)]^n$$

$$M_{Z_n}(t) = M_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

We must show that the MGF of Z_n tends to the MGF of the standard normal.

$M(s)$ has a Taylor series expansion about zero.

$$M(s) = 0 + sM'(0) + \frac{1}{2}s^2M''(0) + \varepsilon_s$$

The details of how Taylor series expansion works is detailed in Example 6 on the next page.

where $\frac{\varepsilon_s}{s^2} \to 0$ as $s \to 0$.

where ε_s goes "faster" to 0 than s^2 as $s \to 0$ (as ε_s has terms with s^3, s^4 etc.).

$$M'(0) = E(X) = 0$$

$$M''(0) = E(X^2) = Var(X) + E(X)^2 = \sigma^2 + 0^2 = \sigma^2$$

$$n \to \infty$$

$$\frac{t}{\sigma\sqrt{n}} \to 0$$

And

$$M(s) = 1 + 0 + \frac{1}{2}\sigma^2 s^2 + \varepsilon_s$$

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \varepsilon_n$$

Where $\frac{\varepsilon_n}{\frac{t^2}{n\sigma^2}} \to 0$ as $n \to \infty$

ε_n goes "faster" to 0 than $\frac{t^2}{n\sigma^2}$ as $\frac{t}{\sqrt{n}\sigma} \to 0$ i.e. as $n \to \infty$ ε_n goes "faster" to 0 than $\frac{t^2}{n\sigma^2} \times \frac{\sigma^2}{2}$ i.e. $\frac{t^2}{2n}$ as $n \to \infty$ thus ε_n can be left out if $n \to \infty$

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \varepsilon_n\right)^n$$

It can be shown that if $a_n \to a$ then

$$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = e^a$$

$a_n = \frac{t^2}{2}$ so $a = \frac{t^2}{2}$ (as $a_n \to \frac{t^2}{2}$)

Because ε_n goes "faster" to 0 than $\frac{t^2}{2}$ we can ignore it

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left(1 + \frac{\frac{t^2}{2}}{n} + \varepsilon_n\right) = e^{\frac{t^2}{2}}$$

$M_{Z_n}(t) \to e^{\frac{t^2}{2}}$ where $e^{\frac{t^2}{2}}$ is the MGF of the standard normal distribution.

i.e. $M_{Z_n}(t) \to M_Z(t)$ / Z_n converges in distribution to Z

Back to Table of Contents

Rice Examples

Example 10: Question 3

Suppose that the number of insurance claims, N, filed in a year is Poisson distributed with $E(N) = 10,000$. Use the normal approximation to the Poisson to approximate $P(N > 10,200)$.

$$P(N > 10\,200) = 1 - P(N < 10\,200)$$

$$= 1 - P\left(\frac{N - 10\,000}{\sqrt{10000}} < \frac{10\,200 - 10000}{\sqrt{10\,000}}\right)$$

$$= 1 - \phi(2)$$

$$= 1 - 0.97725$$

$$= 0\,2275$$

Example 11: Question 6

Using moment-generating functions, show that as $\alpha \to \infty$ the gamma distribution with parameters α and λ, properly standardized, tends to the standard normal distribution.

Given $X_1, X_2, \ldots, X_n$ are gamma random variables with paramaters α and λ such that

$$\lim_{n \to \infty} \alpha_n = \infty$$

$$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X_n)}}$$

the MGF of a gamma is $M_{X_n}(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_n}$

The $E(X_n) = \frac{\alpha_n}{\lambda}$ and $Var(X_n) = \frac{\alpha_n}{\lambda^2}$

$$M_{Z_n}(t) = \left(e^{-\frac{\frac{\alpha_n}{\lambda}}{\frac{\sqrt{\alpha_n}}{\lambda}}}\right)\left(M_{X_n}\left(\frac{t}{\frac{\sqrt{\alpha_n}}{\lambda}}\right)\right)$$

$$= e^{-\sqrt{\alpha_n}\,t}\left(\frac{\lambda}{\lambda - \frac{\lambda t}{\sqrt{\alpha_n}}}\right)$$

Working with the log of this expression

$$\ln(M_{Z_n}(t)) = -\sqrt{\alpha_n}\,t + \alpha_n\left(\frac{\sqrt{\alpha_n}}{\sqrt{\alpha_n} - t}\right) = -\sqrt{\alpha_n}\,t + \alpha_n\left(1 + \frac{t}{\sqrt{\alpha_n} - t}\right)$$

"A Taylor series is a clever way to approximate any function as a polynomial with an infinite number of terms. Each term of the Taylor polynomial comes from the function's derivatives at a single point." - www.khanacademy.org.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 \ldots$$

In the more compact sigma notation, this can be written as:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Where $f^n(a)$ denotes the n^{th} derivative of f evaluated at the point a. (The derivative of order zero of f is defined to be f itself and $(x - a)^0$ and $0!$ are both defined to be 1.)

To find the limit of the expression when we let $n \to \infty$, we use power series function $\ln(1+x)$

substituting $x = \frac{t}{\sqrt{\alpha_n} - t}$ and expanding the formula for $K=1$ and $K=2$

$$\ln M_{z_n}(t) = -\sqrt{\alpha_n}\, t + \alpha_n \sum_{k=1}^{\infty} (-1)^{k-1} \left(\frac{\left(\frac{t}{\sqrt{\alpha_n} - t}\right)^k}{k} \right)$$

$$= \underbrace{-\sqrt{\alpha_n}\, t + \alpha_n \frac{t}{\sqrt{\alpha_n} - t}}_{① =} - \underbrace{\left(\frac{\alpha_n}{2}\right)\left(\frac{t^2}{(\sqrt{\alpha_n} - t)^2}\right)}_{② =} + \underbrace{\alpha_n \sum_{k=3}^{\infty} (-1)^{k-1} \left(\frac{\frac{t}{\sqrt{\alpha_n} - t}}{k} \right)}_{③}$$

① =

$$\frac{-\sqrt{\alpha_n}(t)(\sqrt{\alpha_n} - t) + \alpha_n(t)}{\sqrt{\alpha_n} - t}$$

$$= t^2 \frac{\sqrt{\alpha_n}}{\sqrt{\alpha_n} - t} \to t^2(1) = t^2$$

when $\alpha_n \to \infty$

② =

$$\lim_{\alpha_n \to \infty} \left(\frac{\alpha_n}{2}\right)\left(\frac{t^2}{(\sqrt{\alpha_n} - t)^2}\right)$$

$$= \lim_{\alpha_n \to \infty} \frac{t^2}{2} \left(\frac{(\sqrt{\alpha_n})^2}{(\sqrt{\alpha_n} - t)^2} \right)$$

$$= \frac{t^2}{2}(1)$$

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

③

the limit is zero when $\alpha_n \to \infty$ because the power of α_n is always 1 in the numerator, where its always larger than 1 in the denominator (at least $3/2$ increasing by $1/2$ each time)

Therefore

$$\lim_{\alpha_n \to \infty} \ln M_{z_n}(t) = t^2 - \frac{t^2}{2} + 0 = \frac{t^2}{2}$$

$$\lim_{\alpha_n \to \infty} M_{z_n}(t) = e^{\frac{t^2}{2}}$$

which is the MGF of a standard normal variable

Example 12: Question 10

A six-sided die is rolled 100 times. Using the normal approximation, find the probability that the face showing a six turns up between 15 and 20 times. Find the probability that the sum of the face values of the 100 trials is less than 300.

$X \sim$ Binomial (Counts the number of successes in a fixed amount of trials)

number of trials = 100

probability of success = $1/6$

Expected Value = $np = 100 \cdot \frac{1}{6}$

Variance = $np(1-p) = 100 \cdot \frac{1}{6} \cdot \frac{5}{6}$

$$P(15 \leq X \leq 20) = \frac{15-\frac{100}{6}}{\sqrt{\frac{125}{9}}} \leq \frac{X-\frac{100}{6}}{\sqrt{\frac{125}{9}}} \leq \frac{20-\frac{100}{6}}{\sqrt{\frac{125}{9}}}$$

$$= -0.45 \leq \frac{X-\frac{100}{6}}{\sqrt{\frac{125}{9}}} \leq 0.89$$

$$\approx \phi(0.89) - \phi(-0.45)$$

$$= \phi(0.89) - 1 + \phi(0.45)$$

$$= 0.81327 - 1 + 0.67364$$

$$= 0.48691$$

Let $X_1 \dots X_{100}$ be independent uniform random variables on $\{1,2,3,4,5,6\}$

$E(X) = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

$Var(X) = E(X^2) - E(X)^2 = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) - 3.5^2$

$= 2.917$

$$P\left(\sum_{i=1}^{100} X_i < 300\right) = P\left(\frac{\sum_{i=1}^{100} X_i - (3.5)(100)}{\sqrt{(100)(2.917)}}\right) < \frac{300-(3.5)(100)}{\sqrt{(100)(2.917)}})$$

$$= \phi(-2.93)$$

$$= 1 - \phi(2.93)$$

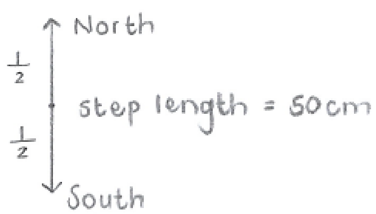
$$= 1 - 0.99831$$

$$= 0.00169$$

Example 13: Question 13

A drunkard executes a “random walk” in the following way: Each minute he takes a step north or south, with probability $\frac{1}{2}$ each, and his successive step directions are independent. His step length is $50\ cm$. Use the central limit theorem to approximate the probability distribution of his location after $1\ h$. Where is he most likely to be?

$S_{60} = \sum_{i=1}^{60} X_i$ where X_i is iid and uniform on $\{-50, 50\}$



CLT: $\dfrac{S_n - nE(X)}{\sqrt{n\,Var(X)}} \overset{D}{\approx} N(0,1)$

$E(X) = -50(\frac{1}{2}) + 50(\frac{1}{2}) = 0$

(probability X values X_i can take on)

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= -50^2(\tfrac{1}{2}) + 50^2(\tfrac{1}{2}) - 0^2 \\ &= 2500 \end{aligned}$$

$$\frac{S_{60} - (60)(0)}{\sqrt{(60)(2500)}} = \frac{S_{60}}{\sqrt{150\ 000}} \overset{D}{\approx} N(0,1)$$

$$S_{60} \overset{D}{\approx} N(0, 150\ 000)$$

The expected value of a normal is its mean, therefor we can assume that the drunkard is most likely where he started.

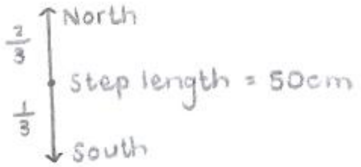
According to [Wikipedia](), X is normally distributed with mean μ and variance σ^2, then a linear transform $aX + b$ (for some real numbers a and b) is also normally distributed:

$$aX + b \sim N(a\mu + b, a^2\sigma^2).$$

Example 14: Question 14

Answer Problem 13 under the assumption that the drunkard has some idea of where he wants to go so that he steps north with probability $\frac{2}{3}$ and south with probability $\frac{1}{3}$.

Similar to above,



$$E(X) = (\tfrac{1}{3})(-50) + (\tfrac{2}{3})(50) = \frac{50}{3}$$

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= (\tfrac{1}{3})(-50^2) + (\tfrac{2}{3})(50^2) + (\tfrac{50}{3})^2 \\ &= \frac{20\,000}{9} \end{aligned}$$

$$\frac{S_{60} - (60)(\frac{50}{3})}{\sqrt{(60)(\frac{20\,000}{9})}} = \frac{S_{60} - 1000}{\sqrt{\frac{400\,000}{9}}} \overset{D}{\approx} N(0,1)$$

$$S_{60} \overset{D}{\approx} N(1000, \frac{400\,000}{3})$$

The drunkard is expected to be 10 m to the North

Example 15: Question 15

Suppose that you bet $\$5$ on each of a sequence of 50 independent fair games. Use the central limit theorem to approximate the probability that you will lose more than $\$75$.

$S_{50} = \sum_{i=1}^{50} X_i$, where X_i is uniform iid random variables on $\{-5, 5\}$

$$E(X) = (-5)(\tfrac{1}{2}) + (5)(\tfrac{1}{2}) = 0$$

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= (-5^2)(\tfrac{1}{2}) + (5^2)(\tfrac{1}{2}) - 0 \\ &= 25 \end{aligned}$$

$$\frac{S_{50} - 50(0)}{\sqrt{50(25)}} = \frac{S_{50}}{\sqrt{1250}} \overset{D}{\approx} N(0,1)$$

$$P(S_{50} < -75) = P\left(\frac{S_{50}}{\sqrt{1250}} < \frac{-75}{\sqrt{1250}}\right)$$

lose = negative

$$\begin{aligned} &\approx \phi(-2.12) \\ &= 1 - \phi(2.12) \\ &= 1 - 0.98300 \\ &= 0.017 \text{ aproximate probability} \end{aligned}$$

Example 16: Assignment 1

Brief and Mark allocation:

The most difficult/interesting problem I encountered in Chapter 5:

Find a problem based on continuity and/or limit theorems and type it out in Word (using Word's Equation Editor) or in LaTeX. Solve the problem (or retype the solution).

- The problem is interesting or practical, and well explained. [2, 2]
- Solution is well structured, easy to understand. [2, 2]
- Problem and solution are provided and are complete. [6]
- Write a small explanation of why your problem is interesting. [4]
- Explain what you learned by typing up these problems and solutions (critical cross-field outcomes are also relevant). [2]

[20]

The reason I chose this problem is that it requires me to use skills learned in STSM2634 (probability distributions in R), MATM1624 (integrating trigonometric functions) as well as theory learned in STSM2616 (Monte Carlo Integration, the Law of Large Numbers) and additional research (plotting distributions in R, Fresnel C integration etc.) to solve.

Furthermore, it practically illustrated the law of large numbers while simultaneously allowing me to practice my Monte-Carlo integration skills. I found this problem quite difficult at first. I understood the content, but I did not fully know to approach the question. After googling "Monte Carlo Integration in R" I had a better idea of how to approach this question. I had to revisit my notes from calculus last year to do the integration.

What I learned from this problem was how to:

- Integrate skills learned in different modules:
  - Using R software to make relevant statistical analysis,
  - Combining topics learned in chapter 5 such as Monte Carlo Estimation and the law of large numbers.
- Research and combine information form multiples sources into a single solution,
- Understand what additional work will allow me to fully understand a certain concept,
- Revisit work from last year and apply it in new context i.e. I had to revisit my notes from calculus last year to do the integration.

Rice - Question 19

a. Use the Monte Carlo method with $\boldsymbol{n = 100}$ and $\boldsymbol{n = 1000}$ to estimate $\int_0^1 \boldsymbol{cos(2\pi x)\, dx}$. Compare the estimates to the exact answer.
b. Use Monte Carlo to evaluate $\int_0^1 \boldsymbol{cos(2\pi x^2)\, dx}$. Can you find the exact answer?

To answer question 19, I used a statistical program R to obtain the samples and do the calculations.

Question 19 A Part A

Firstly, I calculated the actual value of the integral:

$$\int_0^1 cos(2\pi x)\, dx = \frac{1}{2\pi x}[\sin(2\pi x)]_0^1$$
$$= \frac{1}{2\pi}(0 - 0) = 0$$

Then, I found the estimate:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_n) = \frac{1}{100}\sum_{i=1}^{100} \cos(2\pi x_n)$$

Then, using R, I:

1. Generated 100 random uniform variables using:
   `x = runif(100, 0, 1)`

2. And used the values generated previously substituting them into the following equation:
   `y = sum(cos(2*pi*x))/100`

The answer I obtained was: `-0.03961502.`

Question 19 A Part B

I repeated the above process by calculating the estimate:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_n) = \frac{1}{1000}\sum_{i=1}^{1000} \cos(2\pi x_n)$$

Using R to:

1. Generated 1000 random uniform variables:
   `x = runif(1 000, 0, 1)`

2. And used the values generated previously substituting them into the following equation:
   `y = sum(cos(2*pi*x))/1000`

The answer I obtained was: `0.05248159.`

The following tables compares the answers of both evaluations. The results of this question is different from what we expected. We expected the sample of a thousand to the closed to the actual value of the integral. For this reason we conclude that the sample we used was simply too small in order to show a significance change in the approximation. In the next question we use a sample of a million.

n	Our Approximation	Actual Value
100	-0.03961502	0
1000	0.05248159	0

Question 19 B

I am unable to find the integral myself, thus I used the following code:

```
f = c(cos(2*pi*x*x))
fresnelC(f)
```

And found the actual answer: $\int_0^1 f(x) = 0.2441.$

Thereafter, I found the estimate:

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} f(x_n) = \frac{1}{1\ 000\ 000}\sum_{i=1}^{1\ 000\ 000} \cos\left(2\pi x_n^2\right)$$

And similar to above, using R, I:

1. Generated 1000000, random uniform variables using:
   `x = runif(1 000 000, 0, 1)`

3. And used the values generated previously substituting them into the following equation:
   `y = sum(cos(2*pi*x*x))/1 000 000`

The answer I obtained was: `0.2442175.`

The table shows that the approximation is remarkably close to the actual value (the fourth decimal is wrong). Thus, we can conclude that the larger the sample, the more accurate our approximation will be to the actual value.

n	Our Approximation	Actual Value
1 000 000	0.2442175	0.2441

Question 19 - Additional

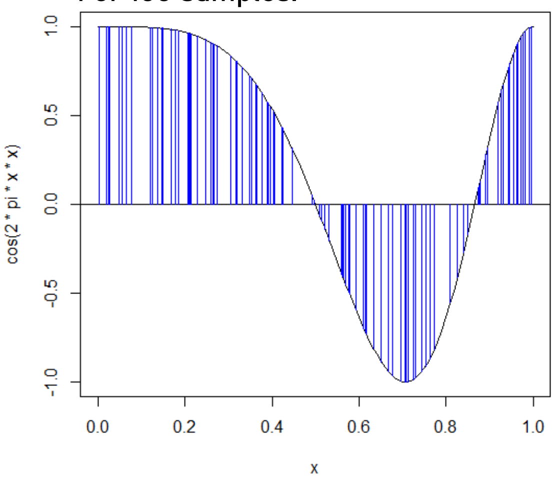
Although not necessary, I included the following to better my understanding of the concepts practiced in this question:

I used the following code to draw the distribution of $f(x)$. The blue lines represent a sample drawn from the uniform distribution. I wanted to be able to visualize the area under the graph that is calculated for each question.

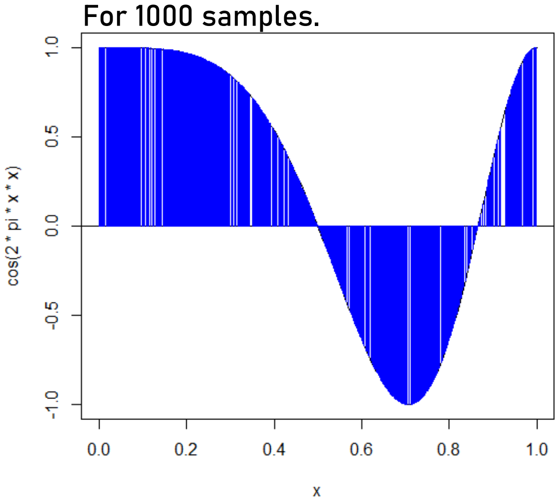
```
curve(expr = cos(2*pi*x), from = 0, to = 1)
abline(h = 0, col="black")


segments(x0 = x,
         x1 = x,
         y0 = cos(2*pi*x),
         y1 = 0,
         col = "blue")
```

Where x represents the sample drawn from the uniform distribution.

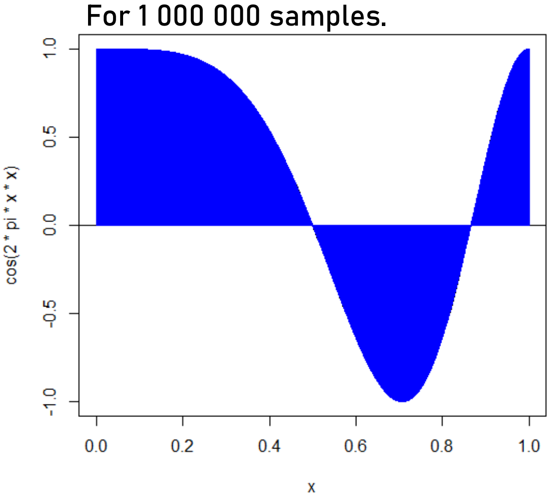
For 100 samples.



For 1000 samples.

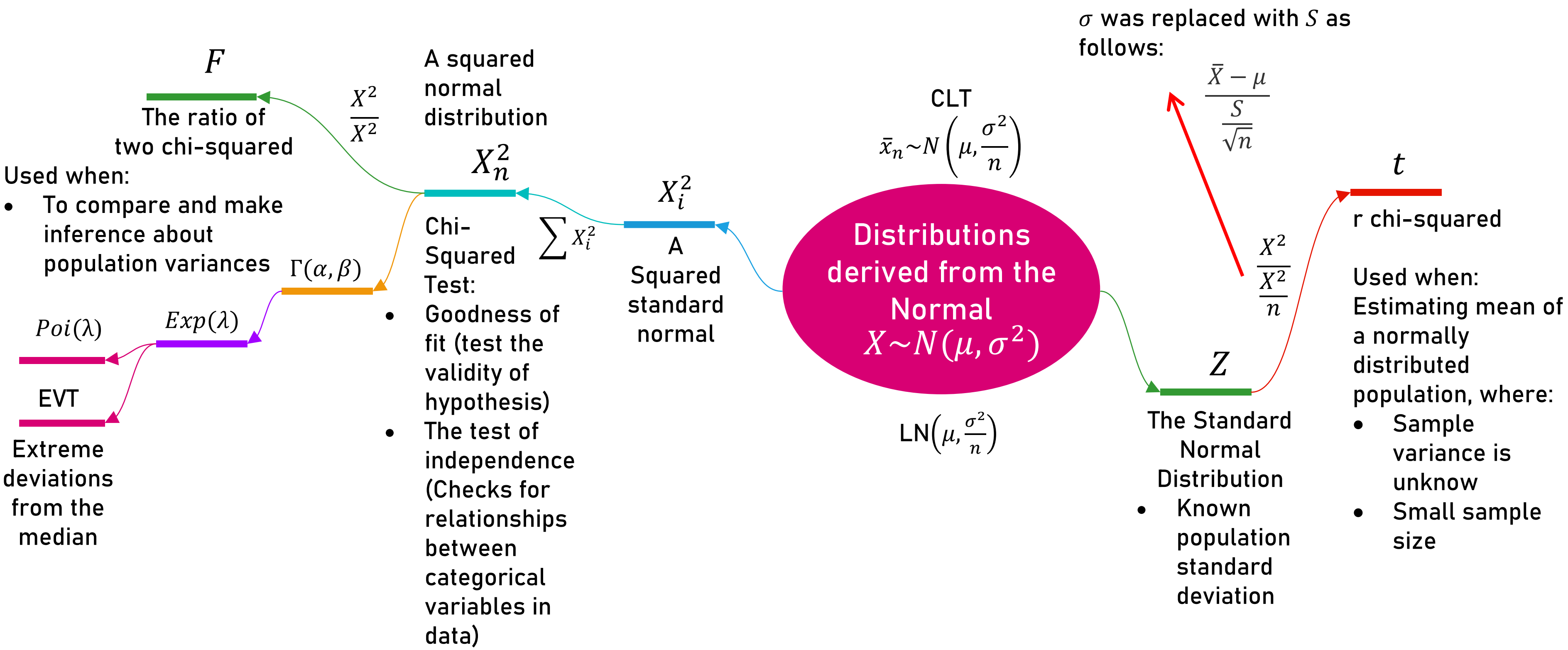


For 1 000 000 samples.



integration, the more accurate the calculation is (as the area underneath the curve gets more filled). This clearly shows the Law of Large numbers in practice.

Distributions Derived from the Normal Distribution



Learning Outcomes and Interview Questions

Construct distributions from the Normal distribution:

- Understand and explain how the Normal, t, χ^2, F distributions are related.
- Research where these distributions might be used in future.

Logbook

The following process was used in the completion of this topic:

- I firstly read through the slides of Dr. Morné Sjölander. I copied relevant work and tried to explain my own understanding from his work.
- I secondly read through Rice and copied the relevant work.
- Thereafter, I googled the usage of the different distributions.
- I then completed my mind-map of the section.
- Lastly, I completed the examples at the end of Rice and did my assignment.

The chi-square distribution

Definition

The chi-square distribution is defined as follows:

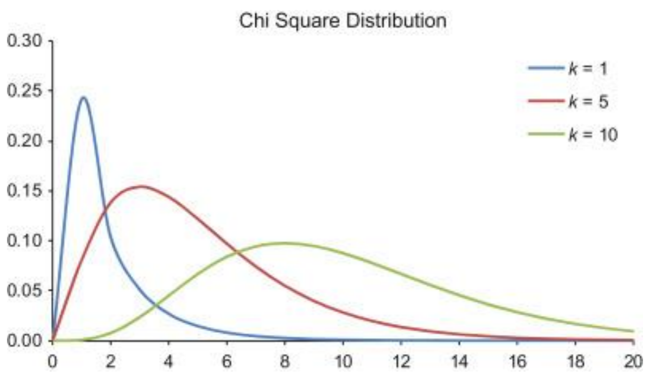
If $\boldsymbol{Z}$ is a standard normal random variable, the distribution of $\boldsymbol{U} = \boldsymbol{Z}^2$ is called the chi-square distribution with 1 degree of freedom.

- It is denoted as $U \sim X_1^2$ which is a special case of the gamma distribution with parameters $\frac{1}{2}$ and $\frac{1}{2}$.
- The expected value /$E(U) = 1$ and the variance / $Var(U) = 2$.

If $\boldsymbol{U_1}, \boldsymbol{U_2}, \dots, \boldsymbol{U_n}$ are independent chi-square random variables with 1 degree of freedom, then the distribution of If $\boldsymbol{V} = \boldsymbol{U_1} + \boldsymbol{U_2} + \cdots + \boldsymbol{U_n}$ is called the chi-square distribution with n degrees of freedom and is denoted by $\boldsymbol{U} \sim \boldsymbol{X_n^2}$.

- It is denoted as $V \sim \chi_n^2$ which is a special case of the gamma distribution with parameters $\frac{n}{2}$ and $\frac{1}{2}$.
- The expected value /$E(V) = n$ and the variance / $Var(V) = 2n$.

Obtained from: Dr. Morné Sjölander's slides.



Obtained from: sciencedirect.com.

Important characteristics

The chi-square distribution with 1 degree of freedom is denoted χ^2 . It is useful to note that if $X \sim N(\mu, \sigma^2)$, then $\frac{(X-\mu)}{\sigma} \sim N(0, 1)$, and therefore $\left[\frac{X-\mu}{\sigma}\right]^2 \sim \chi^2$.

From this I understand that if X follows a normal distribution with parameters μ and σ^2, then the standardized X distribution will follow a normal. And the square of the standardized X distribution follows a chi-squared distribution.

Usage and Purpose – the Chi-Square test

"Chi-square is a statistical test used to examine the differences between categorical variables from a random sample in order to judge goodness of fit between expected and observed results." – Investopedia.

It applies to categorical variables and is especially useful when those variables are nominal (where order doesn't matter, like marital status or gender).

Since chi-square applies to categorical variables, it is most used by researchers who are studying survey response data. This type of research can range from demography to consumer and marketing research to political science and economics.

There are two types of chi-square tests:

- Goodness of fit test - used to test the validity of a hypothesis made about the population based on a random sample. The sample must be mutually exclusive, drawn from a large enough sample and drawn from independent variables.

  The formula for a chi-square test is:

  $$X_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

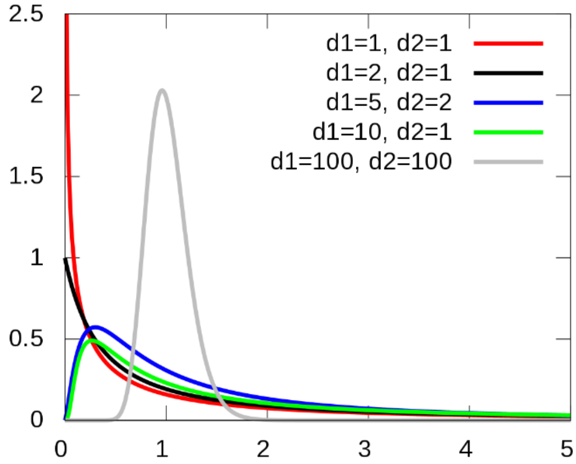
  Where:
  c = degrees of freedom,
  O = observed values,
  E = expected values.

- The test of independence that asks if there is a relationship between two categorical variables in data.

The F distribution

Let U and V be independent chi-squared random variables with m and n degrees of freedom, respectively. The distribution of $W = \frac{\frac{U}{m}}{\frac{V}{n}}$ is called the F distribution with m and n degrees of freedom is denoted by $F_{m,n}$.

The Expected value of W / $E(W) = \frac{n}{n-2}$ if $n > 2$.



The F distribution is positively skewed (the mean is greater than the median) and only defined for positive values.

Uses

The main use of F-distribution is testing the hypothesis about the equality of two population variances or the equality of two or more population means.

The main use of F-distribution is to test whether two independent samples have been drawn for the normal populations with the same variance. For instance, college administrators would prefer two college professors grading exams to have the same variation in their grading. For this, the F-test can be used, and after examining the p-value, inference can be drawn on the variation.

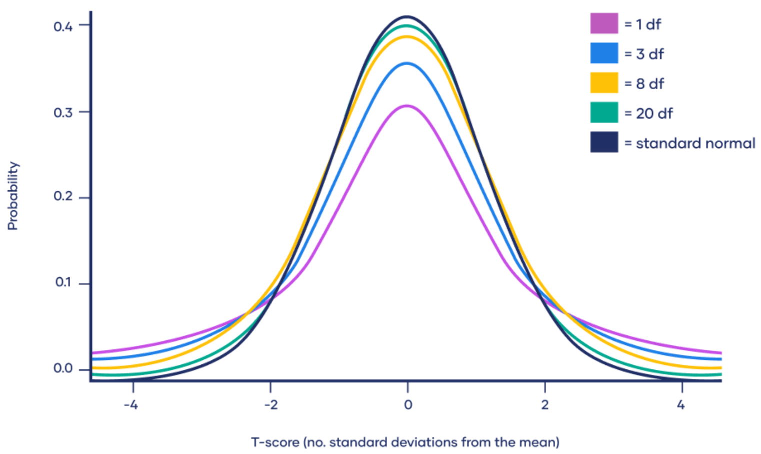
Makes use for ANOVA – A test that allows us to consider the parameters of several populations at once, without getting into some of the problems that confront us by conducting hypothesis tests on two parameters at a time.

The t distribution

The t-distribution is used when data are approximately normally distributed, which means the data follow a bell shape, but the population variance is unknown. The variance in a t-distribution is estimated based on the degrees of freedom of the data set $(n - 1)$.

The t distribution and the normal distribution:

- Both are symmetric.
- Both have a mean of zero.
- The normal distribution assumes that the population standard deviation is known. The t-distribution does not make this assumption.
- The t-distribution is most useful for small sample sizes, when the population standard deviation is not known, or both.
- As the sample size increases, the t-distribution becomes more similar to a normal distribution.



The shape of the t-distribution depends on the degrees of freedom. The curves with more degrees of freedom are taller and have thinner tails. All three t-distributions have "heavier tails" than the z-distribution.

Obtained from: scribbr.

*A common rule of thumb is that for a sample size of at least 30, one can use the z-distribution in place of a t-distribution.

Usage

Describes the standardized distances of sample means to the population mean when the population standard deviation is not known. jmp.com. Used when: Estimating mean of a normally distributed population, where sample variance is unknow and for a small sample size.

The t-distribution plays a role in a number of widely used statistical analyses, including:

- Student's t-test for assessing the statistical significance of the difference between two sample means,
- the construction of confidence intervals for the difference between two population means,
- and in linear regression analysis.

The sample mean and sample variance

"For anyone pursuing study in Data Science, Statistics, or Machine Learning, stating that "The Central Limit Theorem (CLT) is important to know" is an understatement. Particularly from a Mathematical Statistics perspective, in most cases the CLT is what makes recovery of valid inferential coverage around parameter estimates a tractable and solvable problem." - towardsdatascience.com.

Although it is a website providing the proof of the CLT, it did provide the following useful information:

Let $X_1, X_2, X_3, \ldots, X_n$ be a sequence of independent $N(\mu, \sigma^2)$ random variables – a sample from a normal distribution. The mean of the sequence $(\bar{X})$ of the can be defined as:

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

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

Example 1: Theorem A Proof

The random variable $\bar{X}$ and the vector of random variables $(\boldsymbol{X_1} - \bar{\boldsymbol{X}}, \boldsymbol{X_2} - \bar{\boldsymbol{X}}, \dots, \boldsymbol{X_n} - \bar{\boldsymbol{X}})$ are independent.

If the joint MGF factors we can assume the MGF factors are independent.

$$M_{\bar{X}, X_1-\bar{X}, \dots, X_n-\bar{X}}(s, t_1, \dots, t_n) = E\left(e^{s\bar{X}+t_1(X_1-\bar{X})+\cdots+t_n(X_n-\bar{X})}\right)$$

We will show that this MGF equals the MGF of $\bar{X}$ times the mgf of $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ and thus that $\bar{X}$ and $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ are independent.

$$s\bar{X} + t_1(X_1 - \bar{X}) + \cdots + t_n(X_n - \bar{X})$$

Can be rewritten as:

$$= \sum_{i=1}^{n} \frac{s}{n} + \sum_{i=1}^{n} t_i X_i - \sum_{i=1}^{n} t_i \bar{X}$$

$$= \sum \frac{s}{n} + \sum t_i X_i - n\bar{t}\bar{X}$$

$\bar{X} = \frac{1}{n}\sum X_i$ so $n\bar{X} = \sum X_i$

Similarly, $n\bar{t} = \sum t_i$

$$= \sum \frac{s}{n} X_i + \sum t_i X_i - \bar{t} \sum X_i$$

$$= \sum \left[\frac{s}{n} + (t_i - \bar{t})\right] X_i$$

Take out a common X_i and condense to a single summation.

$$= \sum a_i X_i$$

$$\text{where } a_i = \frac{s}{n} + (t_i - \bar{t})$$

It is easy to show that $\sum a_i = s$ and $\sum a_i^2 = \frac{s^2}{n} + \sum(t_i - \bar{t})^2$

$$M_{\bar{X}, X_1-\bar{X}, \dots, X_n-\bar{X}}(s, t_1, \dots, t_n) = E\left(e^{\sum a_i X_i}\right)$$

$$= E(e^{a_1X_1+\cdots+a_nX_n}) = M_{X_1,\dots,X_n}(a_1, \dots, a_n) = M_{X_1}(a_1) \times \dots \times M_{X_n}(a_n)$$

The expectation is equal to the MGFs of the functions as X_i's are independent.

We manipulate the expression in order to show that it equals the product of the MGF of $\bar{X}$ and the MGF of the vector $(\boldsymbol{X_1} - \bar{\boldsymbol{X}}, \boldsymbol{X_2} - \bar{\boldsymbol{X}}, \dots, \boldsymbol{X_n} - \bar{\boldsymbol{X}})$.

$$= \prod M_{X_i}(a_i) = \prod e^{\mu a_i + \frac{\sigma^2}{2} a_i^2}$$

$$= e^{\mu a_1 + \frac{\sigma^2}{2} a_1^2} \times \dots \times e^{\mu a_n + \frac{\sigma^2}{2} a_n^2}$$

$$= e^{\mu a_1 + \frac{\sigma^2}{2} a_1^2 + \cdots + \mu a_n + \frac{\sigma^2}{2} a_n^2}$$

Simplify.

$$= e^{\sum\left(\mu a_i + \frac{\sigma^2}{2} a_i^2\right)}$$

$$= e^{\mu \sum a_i + \frac{\sigma^2}{2} \sum a_i^2} = e^{\mu s + \frac{\sigma^2}{2}\left[\frac{s^2}{n} + \sum(t_i - \bar{t})^2\right]} = e^{\mu s + \frac{\sigma^2}{2}\frac{s^2}{n} + \frac{\sigma^2}{2}\sum(t_i-\bar{t})^2} = e^{\mu s + \frac{\sigma^2}{2n}s^2 + \frac{\sigma^2}{2}\sum(t_i-\bar{t})^2}$$

$$= e^{\mu s + \frac{\frac{\sigma^2}{n}}{2} s^2} e^{\frac{\sigma^2}{2}\sum(t_i - \bar{t})^2}$$

$$= M_{\bar{X}}(s) h(\underline{t})$$

$M_{X_1-\bar{X}, \dots, X_n-\bar{X}}(t_1, \dots, t_n) = M_{\bar{X}}(0) h(\underline{t}) = (1) h(\underline{t}) = h(\underline{t})$

$$M_{\bar{X}, X_1-\bar{X}, \dots, X_n-\bar{X}}(s, t_1, \dots, t_n) = M_{\bar{X}}(s) M_{X_1-\bar{X}, \dots, X_n-\bar{X}}(t_1, \dots, t_n)$$

Thus $\bar{X}$ and $(X_1 - \bar{X}, \dots, X_n - \bar{X})$ are independent.

Back to Table of Contents

Example 2: Corollary A Proof

$\bar{X}$ and S^2 are independently distributed.

$\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ are independent (from the previous theorem).

$S^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})$ is a function of $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$. Thus $\bar{X}$ and S^2 are independent.

Theorem B

Kindly see assignment at the end of this topic for the proof of theorem B.

Example 3: Corollary B Proof

Given $\bar{X}$ and S^2 then:

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

We simply express the given ratio in a different form:

Let $Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$,

Express Z as an $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ as a normal distribution with mean zero and variance 1.

And $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Call the sample distribution U that is a chi-squared random variable with $n-1$ degrees of freedom (from theorem A).

and $T = \frac{Z}{\sqrt{\frac{U}{n-1}}}$

Express the distributions in another way.

But $T = \frac{Z}{\sqrt{\frac{U}{n-1}}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{S}{\sigma}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$

The $n-1$'s cancel and both square root σ^2 of and S^2 are taken.

Thus $T \sim t_{n-1}$

Rice Examples

Example 4: Question 3

Let X be the average of a sample of 16 independent normal random variables with mean 0 and variance 1. Determine c such that:

$$P(|X| < c) = .5$$

$$\bar{X} = \sum_{i=1}^{16} X_i$$

$\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ according to the CLT

$$\bar{X}_n \sim N(0, \tfrac{1}{16})$$

$$Z_n = \frac{\bar{X}_n - 0}{\frac{1}{16}}$$

$$= 4\bar{X}_n \sim N(0,1)$$

$$P(|\bar{X}_n| < c) = 0.5$$

$$P(-c < \bar{X}_n < c) = P(-4c < 4\bar{X} < 4c)$$

$$= \phi(4c) - \phi(-4c)$$

$$= -1 + \phi(4c) + \phi(4c)$$

$$0.5 = 2\phi(4c) - 1$$

$$0.75 = \phi(4c)$$

The value closest to 0.75 is 0.7486

$$c \approx \frac{0.67}{4} = 0.1675$$

Example 6: Question 5

If T follows a t_7 distribution, find t_0 such that: (a) $P(|T| < t_0) = .9$,

$T \sim t_7$

(a) $P(|T| < t_0) = P(-t_0 < T < t_0) = P(T < t_0) - P(T < -t_0)$

$= 2P(T < t_0) - 1$

$= 0{,}9$

and (b) $P(T > t_0) = .05$.

(b) $P(T > t_0) = 0{,}05$

$P(T < t_0) = 9{,}95$

$t_0 = 1{,}895$

Example 7: Question 6

Show that if $X \sim F_{n,m}$, then $X^{-1} \sim F_{m,n}$.

(b) $P(T > t_0) = 0{,}05$

$P(T < t_0) = 9{,}95$

$t_0 = 1{,}895$

Suppose $X \sim F_{n,m}$

Then $X = \dfrac{\frac{T}{n}}{\frac{Y}{m}}$, where $T \sim \chi_n^2$ and $Y \sim \chi_m^2$ (independent)

$\dfrac{1}{X} = \dfrac{\frac{Y}{m}}{\frac{T}{n}}$, thus $X^{-1} \sim F_{m,n}$

Example 8: Question 7

Show that if $T \sim t_n$, then $T^2 \sim F_{1,n}$.

Suppose $T \sim t_n$

Then $T = \dfrac{Z}{\sqrt{\frac{U}{n}}}$, where $Z \sim N(0,1)$ and $U \sim \chi_n^2$ (independent)

$T^2 = \dfrac{Z^2}{\frac{U}{n}} = \dfrac{\frac{Z^2}{1}}{\frac{U}{n}}$, where $Z^2 \sim \chi_1^2$ and $U \sim \chi_n^2$ (independent)

Thus $T^2 \sim F_{1,n}$

Example 9: Question 8

Show that the Cauchy distribution and the t distribution with 1 degree of freedom are the same.

the density function of a cauchy X is :

$$f_X(x) = \frac{1}{\pi(1+x^2)}, x \in R$$

the density of a t-distribution, T with 1 degrees of freedom is :

$$f_T(x) = \left(\frac{1}{\sqrt{(1)(\pi)}}\right)\left(\frac{\Gamma(\frac{1+1}{2})}{\Gamma(\frac{1}{2})}\right)\left(\frac{1}{(1+\frac{x^2}{1})^{\frac{1+1}{2}}}\right)$$

$$= \frac{1}{\sqrt{\pi}}\left(\frac{\Gamma(1)}{\Gamma(\frac{1}{2})}\right)\left(\frac{1}{1+x^2}\right)$$

$$= \frac{1}{\sqrt{\pi}}\left(\frac{0!}{\sqrt{\pi}}\right)\left(\frac{1}{1+x^2}\right)$$

$$= \frac{1}{\pi(1+x^2)}, x \in R$$

$\therefore f_T(x) = f_X(x)$

T and X have the same density function

Example 9: Question 9

Find the mean and variance of S^2, where S^2 is as in Section 6.3.

$X \sim \chi^2_n$ with $E(X) = n$ and $Var(X) = 2n$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ with $E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1$

$$E(S^2) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$

and $Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$

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

Example 10: Assignment 2

Brief and Mark allocation:

The most difficult/interesting problem I encountered in Chapter 5:

Find a problem based on distributions derived from the Normal and type it out in Word (using Word's Equation Editor) or in LaTeX. Solve the problem (or retype the solution).

- The problem is interesting or practical, and well explained. [2, 2]
- Solution is well structured, easy to understand. [2, 2]
- Problem and solution are provided and are complete. [6]
- Write a small explanation of why your problem is interesting. [4]
- Explain what you learned by typing up these problems and solutions (critical cross-field outcomes are also relevant). [2]

[20]

I chose Theorem B as my assignment because it contained very useful discoveries that are used in every day statistical inference. It allows us to make inference about a population using a sampling distribution as it is equal to a chi-square distribution. That and it was a very long and difficult theorem.

This theory states that a sampling distribution, is a chi-square distribution. This is important because we hardly ever know the true distribution of the data, the true population mean or the true population variance. Thus, we can make use of sample means and sample variances and make inference about the population using a chi-square distribution. – complied using information from towardsdatascience.com.

I had to apply information learned in all the chapters we have covered so far, including MGF, Limit theorems and the chi-squared distribution. This proved more difficult that I had initially thought and while making my annotations I found myself constantly going to in my portfolio to revisit information. I learned how to compile information from different chapters in a single comprehensive solution.

During this theorem I had to develop my word equation editor skills. I watched the following video that showed me how to include summations in equation editor as well as how to use shortcuts when using equation editor. During this theorem, I could really practice my equation editor skills and I feel a lot more confident using it now.

Theorem B Proof

Proof that distribution of a sampling distribution / $\frac{(n-1)S^2}{\sigma^2}$ is the chi-square distribution with $n-1$ degrees of freedom.

Adapted from Dr. Morné Sjölander's slides.

Firstly, let U equal the distribution as mentioned above:

1. Rewrite the sample variance or S^2 as $\frac{1}{n-1}\sum(X_i - \bar{X})^2$.

$$U = \frac{(n-1)S^2}{\sigma^2} = \frac{(n-1)\frac{1}{n-1}\sum(X_i - \bar{X})^2}{\sigma^2}$$

2. $\frac{1}{n-1}$ and $n-1$ cancel thus leaving $\sum(X_i - \bar{X})^2$ and move $\frac{1}{\sigma^2}$ to the front of the equation.

$$= \frac{\sum(X_i - \bar{X})^2}{\sigma^2} = \frac{1}{\sigma^2}\sum(X_i - \bar{X})^2$$

3. Assign a value W to the sum of n normal random variables. In the expression, both $\frac{1}{\sigma^2}$ and $(X_i - \mu)^2$ contain squares. This can be rewritten as $\left(\frac{X_i - \mu}{\sigma}\right)$. This expression is a sequence of chi-square random variables with n degrees of freedom.

$$W = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim X_n^2$$

4. The expression $X_i - \mu$ can be rewritten as $(X_i - \bar{X}) + (\bar{X} - \mu)$.

$$W = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}[(X_i - \bar{X}) + (\bar{X} - \mu)]^2$$

5. Expand the square and multiply each term with $\frac{1}{\sigma^2}$.

$$= \frac{1}{\sigma^2}\sum(X_i - \bar{X})^2 + \frac{1}{\sigma^2}2(\bar{X} - \mu)\sum(X_i - \bar{X}) + \frac{1}{\sigma^2}\sum(\bar{X} - \mu)^2$$

Write the summation over both X_i and $\bar{X}$. Evaluate the summation over the terms in blue to obtain n.

$$= \frac{1}{\sigma^2}\sum(X_i - \bar{X})^2 + \frac{1}{\sigma^2}2(\bar{X} - \mu)\left(\sum X_i - \sum \bar{X}\right) + \frac{1}{\sigma^2}n(\bar{X} - \mu)^2$$

$$= \frac{1}{\sigma^2}\sum(X_i - \bar{X})^2 + \underbrace{\frac{1}{\sigma^2}2(\bar{X} - \mu)(n\bar{X} - n\bar{X})}_{0} + \frac{(\bar{X} - \mu)^2}{\frac{\sigma^2}{n}}$$

$$\sum_{i=0}^{n}(X_i - \bar{X}) = 0$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 = U + V \quad \text{where} \quad \mathrm{V} = \left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi_1^2$$

This is a relation of the form $W = U + V$. U is a function of S^2 and V is a function of $\bar{X}$. So U So U and V are independent by corollary A (that states that $\bar{X}$ and S^2 are independently distributed). We can say: $M_{W(t)} = M_U(t)\, M_V(t)$, where both W and V follow chi-square distributions, so:

We get the MGF of a chi-squared as follows:

As stated in the summaries above:
$\chi_n^2 = Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ and $\chi_1^2 = Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$

And knowing the MGF of a Gamma from our Formulae sheet:

- The MGF of $Gamma(\alpha, \lambda)$ is $\left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$
- Therefore, the MGF of $\chi_n^2 = Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ is $\left(\frac{\frac{1}{2}}{\frac{1}{2}-t}\right)^{\frac{n}{2}} = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}} = (1-2t)^{-\frac{n}{2}}$
- And the of $\chi_1^2 = Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ is $\left(\frac{\frac{1}{2}}{\frac{1}{2}-t}\right)^{\frac{1}{2}} = \left(\frac{1}{1-2t}\right)^{\frac{1}{2}} = (1-2t)^{-\frac{1}{2}}$

Therefore:

$$M_{U(t)} = \frac{M_W(t)}{M_V(t)}$$

$$= \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}}$$

$$= (1-2t)^{-\frac{(n-1)}{2}}$$

Which is the MGF of a random variable with a X_{n-1}^2 distribution.