NPTEL-MOOC Course

TOTAL QUALITY MANAGEMENT – II

Raghu Nandan SENGUPTA

IME Department, IIT Kanpur

SAMPLING DISTRIBUTIONS

We are aware that for any population distribution, *N*, our main concern is to know the parametric form of the distribution and also to have some knowledge about the different paramerts, like scale, shape and location, which will help us to know the general properties about the underlying population distribution. But the problem which invariably arises is the fact that due to paucity of time, or fund or even due to practicality we need to choose a corresponding sample, *n*, from the population, and then use the sample characteristics to comment intelligently about the population parameters.

Hence given *n* random variables with a specified joint distribution, one is interested in finding the distribution function of these variables. $F_{X_1}(x_1)$The distribution function $T = T(X_1, X_2, \ldots, X_n)$ based on the fact that the random variables are $(X_1, X_2, \ldots, X_n)$ is called the *statistic*. So in general we have a situation in which the distribution of X is denoted by $F_X(x)$, and in case $(X_1, X_2, \ldots, X_n)$ are i.i.d, which would be the case in maximum of the examples we have $F_{X_i}(x_i)$ as the distribution funcition of each X_i, $i = 1,2,\ldots,n$. Moreover for convenience we consider the joint distribution of $(X_1, X_2, \ldots, X_n)$ as $G_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots, x_n)$,

i.e., $G_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1) \times F_{X_2}(x_2) \times \ldots \times F_{X_n}(x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$

Three new distributions

Chi-square distribution (χ_n^2)

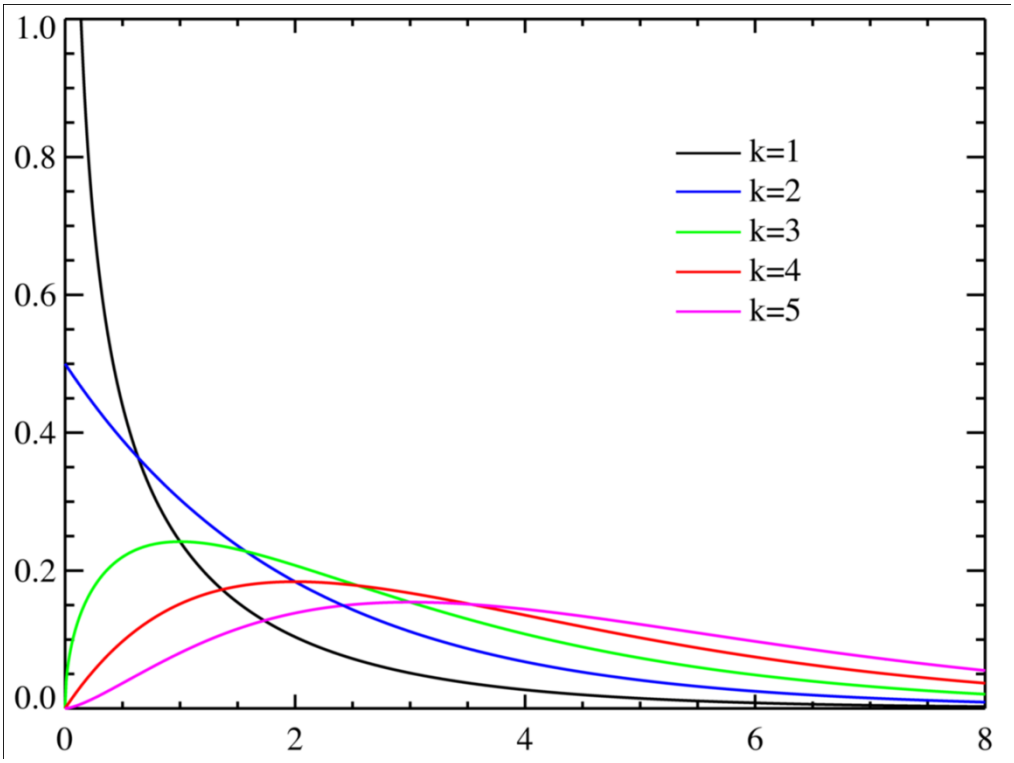
In probability theory and statistics, the chi-square distribution (χ_n^2) is one of the most widely used theoretical probability distributions in inferential statistics, e.g., in statistical significance

tests. It is useful because, under reasonable assumptions, easily calculated quantities can be proven to have distributions that approximate to the chi-square distribution if the null hypothesis is true. The best-known situations in which the chi-square distribution is used are the common chi-square tests for goodness of fit of an observed distribution to a theoretical one, and of the independence of two criteria of classification of qualitative data. Many other statistical tests also lead to a use of this distribution, like Friedman's analysis of variance by ranks.

The general form of the pdf for the chi-square (χ_n^2) distribution with *n* degrees of freedom (dof) is given as

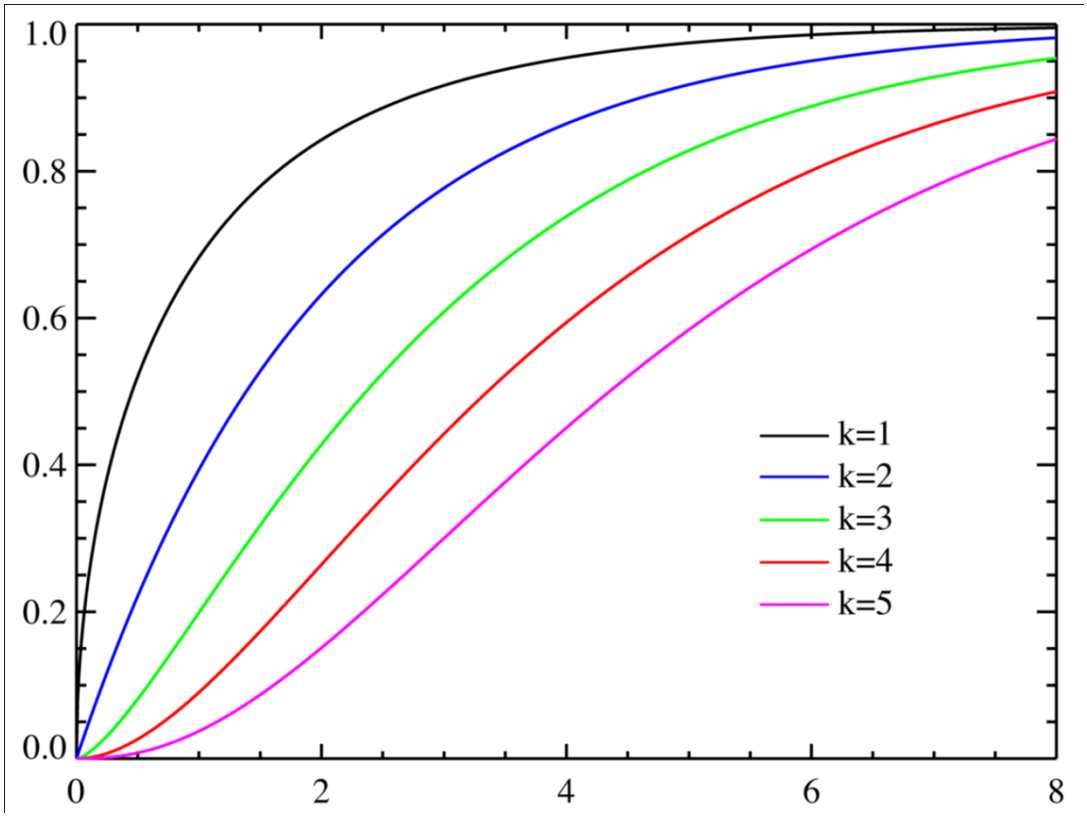
$$f_X(x) = \begin{cases} \dfrac{1}{\Gamma\left(\dfrac{n}{2}\right) 2^{\frac{n}{2}}} x^{\left(\frac{n}{2}-1\right)} \exp\left(-\dfrac{x}{2}\right) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Here *n* is the parameter, also called the degrees of freedom such that $n \in \Re^+$.



Probability density function of χ_k^2 with *k* dof

Its distribution function F(x) is given as $F_X(x)=\frac{\gamma\left(\frac{n}{2},\frac{x}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$, where $\gamma(m,x)$ is the lower incomplete Gamma function given by the following formulae, i.e., $\gamma(m,x)=\int_0^x t^{m-1}e^{-t}dt$.



Distribution function of χ_k^2 with *k* dof

One can easily calculate that the expected value and variance for (χ_n^2) are $E(X)=n$ and $V(X)=2n$, while the m.g.f. is $\frac{1}{(1-2t)^{\frac{n}{2}}}$ for 2t < 1.

Now the question is the chi-square distribution generated and what is its significance? Consider you have *n* number of *i.i.d.* which are standard normal distribution, Z_i, i = 1, 2,…, n and $Z_i \sim N(0,1)$. Then $Z_1^2+Z_2^2+\ldots+Z_n^2=\sum_{i=1}^{n}Z_i^2=\chi_n^2$. From this it is evident that

$$E\left(\chi_n^2\right)=E\left(\sum_{i=1}^{n}Z_i^2\right)=E\left[Z_1^2+Z_2^2+\ldots+Z_n^2\right]=E\left[Z_1^2\right]+E\left[Z_2^2\right]+\ldots+E\left[Z_n^2\right]$$

$$=V\left(Z_1\right)+\left\{E\left(Z_1\right)\right\}^2+V\left(Z_2\right)+\left\{E\left(Z_2\right)\right\}^2+\ldots+V\left(Z_n\right)+\left\{E\left(Z_n\right)\right\}^2=n$$

$$V\left(\chi_n^2\right)=V\left(\sum_{i=1}^{n}Z_i^2\right)=V\left[Z_1^2+Z_2^2+\ldots+Z_n^2\right]=V\left[Z_1^2\right]+V\left[Z_2^2\right]+\ldots+V\left[Z_n^2\right]+\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}\operatorname{cov}\left(Z_i^2,Z_j^2\right)=2n$$

Student t-distribution

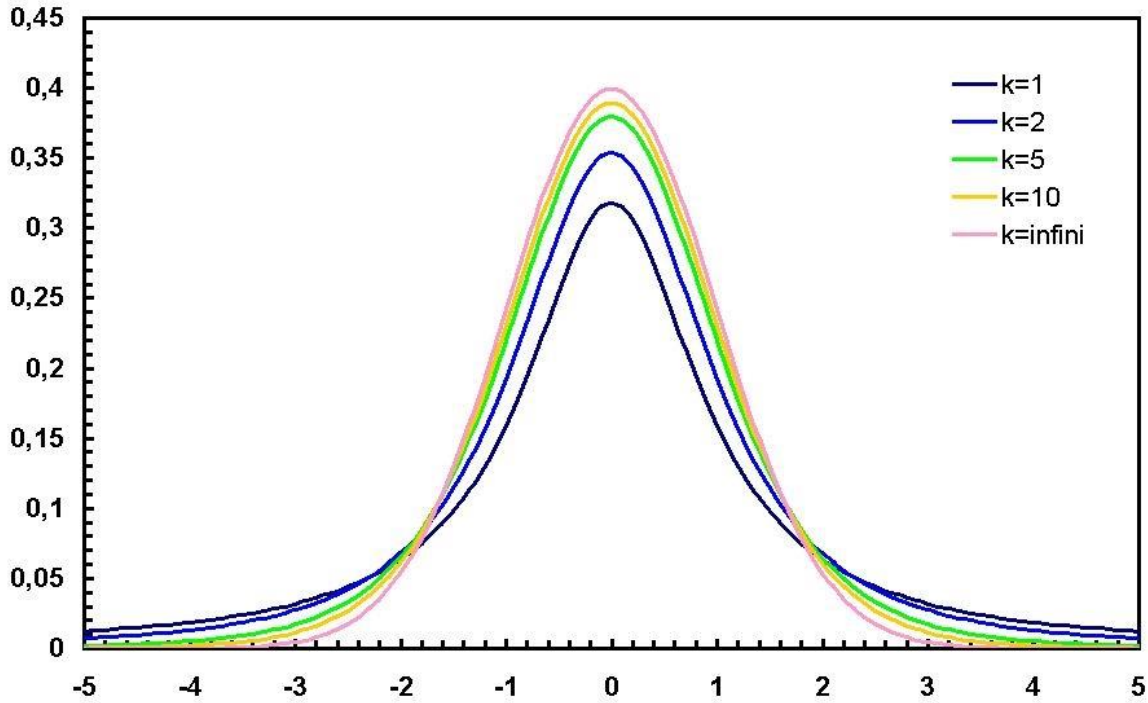
The derivation of the Student t-distribution was first published in 1908 by William Sealy Gosset, while he worked at a Guinness Brewery in Dublin. Due to proprietary issues, the paper was written under the pseudonym Student, hence the name. The t-test and the associated theory became well-known through the work of R.A. Fisher, who called the distribution "Student's distribution". The student's t-distribution (or simply the t-distribution) arises in the problem of estimating the mean of a normally distributed population when the sample size is small. It is the basis of the popular Student's t-tests for the statistical significance of the difference between two sample means, and for confidence intervals for the difference between two population means. The Student's t-distribution is a special case of the generalised hyperbolic distribution.

Student's distribution arises when (as in nearly all practical statistical work) the population standard deviation is unknown and has to be estimated from the data. Quite often, however, textbook problems will treat the population standard deviation as if it were known and thereby avoid the need to use the Student's t-test. These problems are generally of two kinds: (i) those in which the sample size is so large that one may treat a data-based estimate of the variance as if it were certain, and (ii) those that illustrate mathematical reasoning, in which the problem of estimating the standard deviation is temporarily ignored because that is not the point that the author or instructor is then explaining.

The general form of the pdf for the Student t-distribution with *n* degrees of freedom (dof) is given as

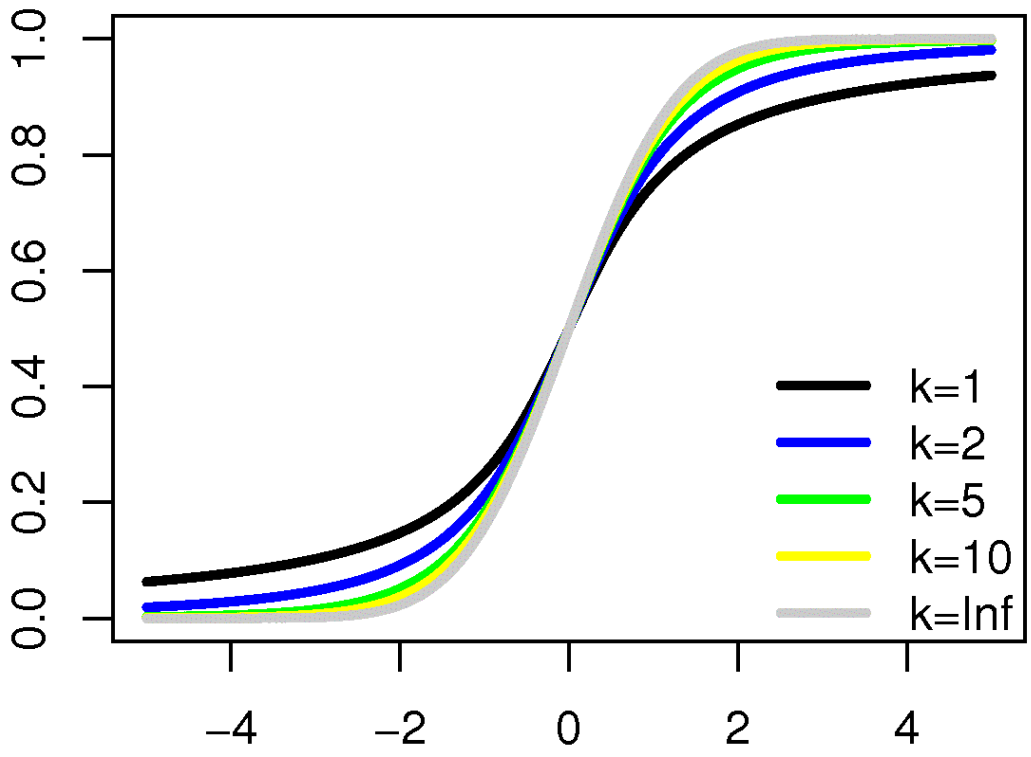
$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\left\{\Gamma\left(\frac{n}{2}\right)\right\}}\left(1+\frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} \qquad x \in (-\infty,+\infty)$$

Here *n* is the parameter, also called the degrees of freedom such that $n \in \Re^+$.



Probability density function of Student t-distribution with *k* dof

Its distribution function $F_X(x)$ is given as $\frac{{}_2F_1\left(\frac{1}{2},\frac{n+1}{2};\frac{3}{2};\frac{x^2}{n}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}$, where ${}_2F_1$ is the hyper geometric function.



Density function of Student t-distribution with *k* dof

One can easily calculate that the expected value for Student t-distribution as $E(X)=0$ for n > 1, otherwise it is undefined, while and $V(X)=\frac{n}{(n-2)}$, for n >2, ∞ for n = 2, and otherwise undefined. We must remember the m.g.f. is not defined for this distribution.

Now again coming back to the question, of how the Student t-distribution with *n* dof can be generated? Consider you have Z as a standard normal distribution, Z ~ N(0, 1) and Y ~ χ_n^2 with *n* dof, and more over Z and Y are independent, then $\frac{Z}{\sqrt{\frac{Y}{n}}} \sim f_T(t)=t_n$. Now if n is large (> 30), then $t_n \cong Z_\alpha$.

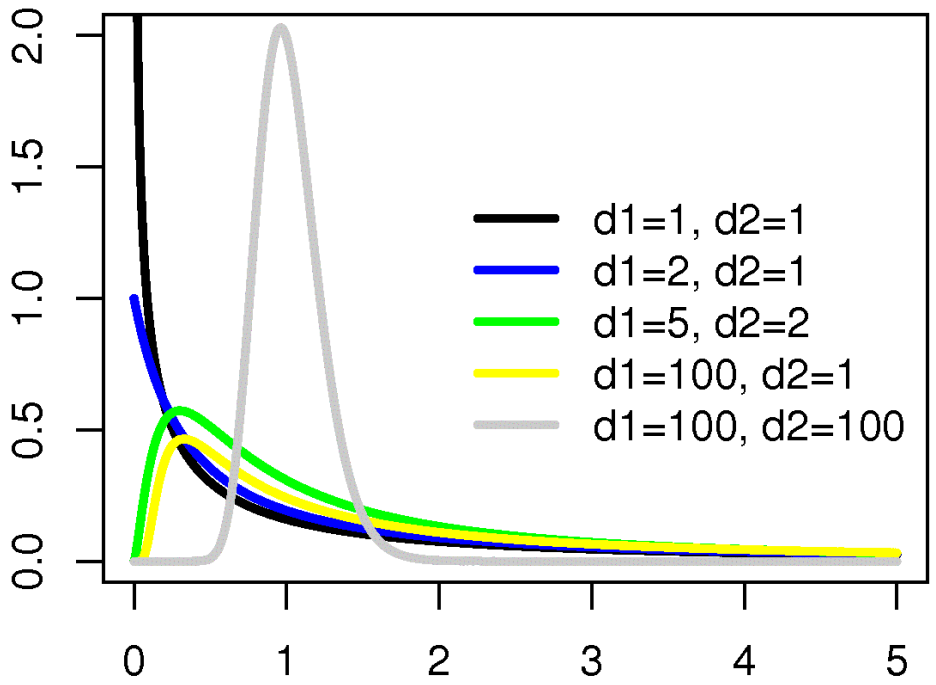
F distribution with (m, n) dof

In probability theory and statistics, the F-distribution is a continuous probability distribution, and it is also known as Snedecor's F distribution or the Fisher-Snedecor distribution (after R.A. Fisher and George W. Snedecor). This distribution arises frequently as the null distribution of a test statistic, especially in likelihood-ratio tests, perhaps most notably in the analysis of variance.

On a similar lines we state here the general form of the pdf for the $F_{m,n}$ distribution with *m, n* degrees of freedom (dof) and it is given as

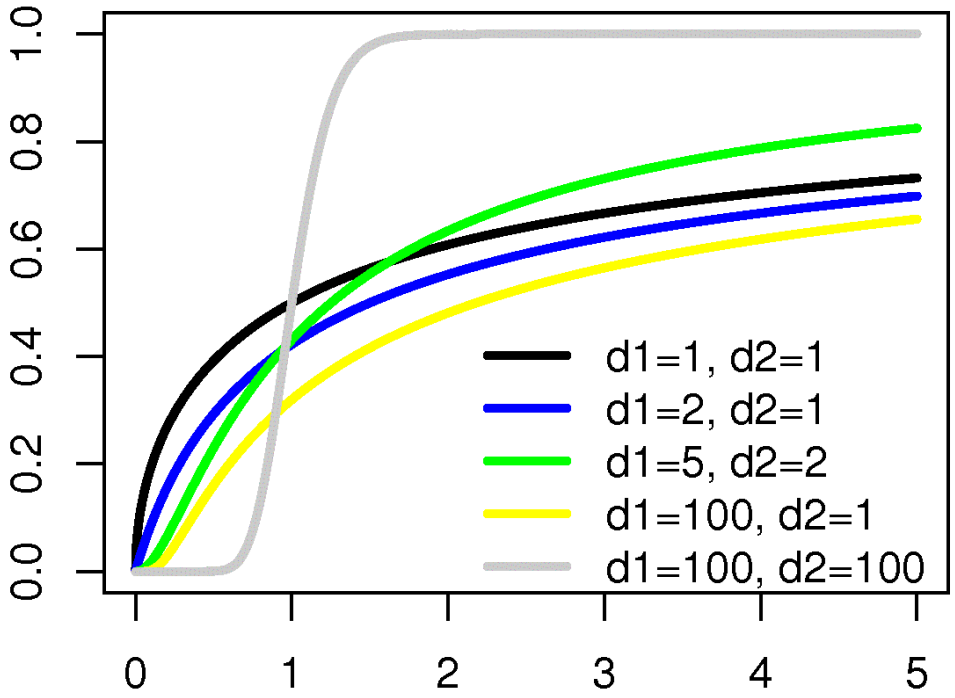
$$f_X(x) = \frac{m^{\frac{m}{2}} n^{\frac{n}{2}} \Gamma\left(\frac{m+n}{2}\right) x^{\left(\frac{m}{2}-1\right)}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(mx+n)^{\left(\frac{m+n}{2}\right)}} \qquad x \in [0,+\infty)$$

Here *m, n* are the parameters, also called the degrees of freedom such that, $m, n \in \Re^+$.



Probability density function of F distribution with d_1 and d_2 dof

The distribution function, $F_X(x)$, for $F_{m,n}$ is $F_X(x) = I_{\left(\frac{mx}{mx+n}\right)}\left(\frac{m}{2}, \frac{n}{2}\right)$, where I is the regularized incomplete beta function of the form, $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$.



Distribution of Student t-distribution with d_1 and d_2 dof

The expected value and variance for F are $E(X) = \left(\frac{n}{n-2}\right)$ for $n > 2$ and, $V(X) = \left\{\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\right\}$ for $n > 4$, respectively. The m.g.f equation does not exist.

The distribution is non-asymmetric and we must also remember that $F_{m,n} \neq F_{n,m}$.

Now suppose that $X \sim \chi^2_m$ and $Y \sim \chi^2_n$, and X and Y are independent, then $\frac{\left(\frac{X}{m}\right)}{\left(\frac{Y}{n}\right)} \sim F_{m,n}$

Now we state few interesting results which will be used quite frequently and universally in the stud of interval estimation as well as hypothesis testing.

Now for result 1 and result 2 consider that if $(X_1, X_2, \ldots, X_n)$ are n number of r.v. observations such that each time we have the realized values as $(x_{1,j}, x_{2,j}, \ldots, x_{n,j})$ for j being the number of such samples we take, from $X \sim N(\mu, \sigma^2)$. Then $X_1 + X_2 + \ldots + X_n \sim N(n\mu, n\sigma^2)$ and $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, i.e., $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Now let us define (i) $s_n'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ and $s_n^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

Result 1

Then we have

$\frac{ns_n'^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$ and

$\frac{(n-1)s_n^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi_{n-1}^2$ (which can be proved easily)

Here we see that it losses one degree of freedom due to the fact that $\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) = \bar{X}_n$.

Note: For a normal distribution $\bar{X}_n$ and s_n^2 are the sample mean and the sample variance and they are distributed normal and χ_n^2 or χ_{n-1}^2 respectively depending on whether population mean, μ, is known or unknown.

Result 2

Then we have

$$\frac{\left\{\frac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{\sigma}\right\}}{\sqrt{\frac{\left(\frac{(n-1)s_n^2}{\sigma^2}\right)}{(n-1)}}}=\left\{\frac{\sqrt{n}\left(\overline{X}_n-\mu\right)}{s_n}\right\}\sim t_{n-1}.$$ In case *n* is large, then we know that $t_{n-1}\sim N(0,1).$

Result 3

Assume $\left(X_1,X_2,\ldots,X_m\right)$ are *m* number of r.v. observations drawn from X ~ N(μ_1, σ_1^2) such that each time we have the realized values as $\left(x_{1,j},x_{2,j},\ldots,x_{n,j}\right)$ for *j* being the number of such samples and $\left(Y_1,Y_2,\ldots,Y_n\right)$ be *n* number of r.v. observations drawn from Y ~ N(μ_2, σ_2^2) such that each time we have the realized values as $\left(y_{1,k},y_{2,k},\ldots,y_{n,k}\right)$ for *k* being the number of such samples. Then $\overline{X}_m=\frac{1}{m}\sum_{i=1}^{m}X_i=\frac{1}{m}\left(X_1+X_2+\ldots+X_m\right)$ and $\overline{Y}_n=\frac{1}{n}\sum_{j=1}^{n}Y_i=\frac{1}{n}\left(Y_1+Y_2+\ldots+Y_n\right)$ are the respectively sample means for the two different samples of size *m* and *n* respectively. While we define, $s_m^{/2}=\frac{1}{m}\sum_{i=1}^{m}\left(X_i-\mu_1\right)^2$, $s_m^2=\frac{1}{(m-1)}\sum_{i=1}^{m}\left(X_i-\overline{X}_m\right)^2$, $s_n^{/2}=\frac{1}{n}\sum_{j=1}^{n}\left(Y_j-\mu_2\right)^2$ and $s_n^2=\frac{1}{(n-1)}\sum_{j=1}^{n}\left(Y_j-\overline{Y}_n\right)^2$. Then

$$\frac{ms_m^{/2}}{\sigma_1^2}=\sum_{i=1}^{m}\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2\sim\chi_m^2 \text{ and } \frac{(m-1)s_m^2}{\sigma_1^2}=\sum_{i=1}^{m}\left(\frac{X_i-\overline{X}_m}{\sigma_1}\right)^2\sim\chi_{m-1}^2$$

$$\frac{ns_n^{/2}}{\sigma_2^2}=\sum_{j=1}^{n}\left(\frac{Y_j-\mu_2}{\sigma_2}\right)^2\sim\chi_n^2 \text{ and } \frac{(n-1)s_n^2}{\sigma_2^2}=\sum_{j=1}^{n}\left(\frac{Y_j-\overline{Y}_n}{\sigma_2}\right)^2\sim\chi_{n-1}^2$$

Then $\dfrac{\left(\dfrac{ms_m'^2}{\sigma_1^2}\right)}{\left(\dfrac{ns_n'^2}{\sigma_2^2}\right)} = \left(\dfrac{\sigma_2^2}{\sigma_1^2}\right)\left(\dfrac{s_m'^2}{s_n'^2}\right) \sim F_{m,n}$ and $\dfrac{\left\{\dfrac{(m-1)s_m^2}{\sigma_1^2}\right\}}{\left\{\dfrac{(n-1)s_n^2}{\sigma_2^2}\right\}} = \left(\dfrac{\sigma_2^2}{\sigma_1^2}\right)\left(\dfrac{s_m^2}{s_n^2}\right) \sim F_{m-1,n-1}$ are true

Statistical Inference: Introduction

Estimator

Consider you take a set of *n* observations denoted by $(X_1, X_2, \ldots, X_n)$ and these are the set of observations which come from a random experiment. Let us also consider θ as the population parameter, such that, $\theta \in \Theta$, i.e., the parameter space. So for example in the one parameter exponential distribution case $\theta \in [0, +\infty)$, while for the normal distribution example we now have θ as a vector, i.e., **θ**, such that **θ** = (θ_1, θ_2) = (μ, σ^2), in which case the parameter space for θ_1 and θ_2 are different and denoted by $(\Re, \Re^+)$. So our job is to choose the sample of size, *n*, in such a way so that we have some information about the population parameter(s) Suing certain rule of combining the sample information we generate a single value or an element of information such that, that sample characteristics, thus obtained is know as the estimator for the population. As we already know the sample characteristics are the statistics and the these statistic are the estimators which have their respective estimated values, called the estimates. So if θ is the population parameter, then choosing observations, $(X_1, X_2, \ldots, X_n)$, would give us the statistic, $T(X_1, X_2, \ldots, X_n)$, and as we choose, say for example J number of such samples, j = 1, 2,…, J, we would have from them the estimators as $T(X_{1,1}, X_{2,1}, \ldots, X_{n,1})$, …, $T(X_{1,J}, X_{2,J}, \ldots, X_{n,J})$, and the corresponding estimates would be $T(x_{1,1}, x_{2,1}, \ldots, x_{n,1})$,…, $T(x_{1,J}, x_{2,J}, \ldots, x_{n,J})$. Hence for example in case for the sample mean we have $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$, where $\bar{X}_n$ is the estimator and $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ is the corresponding estimate. So generally the estimate eis

the function of the sample and usually denoted by $T(X_1, X_2, \ldots, X_n)$. Again without repetition let us state another example to make the things clear. In case $X_1 = 15$, $X_2 = 9$ and $X_3 = 7$, then $\overline{X}_3 = \frac{1}{3}\sum_{i=1}^{3} X_i = \frac{1}{3}(X_1 + X_2 + X_3) = \frac{1}{3}(15+9+7)$, then $\overline{X}_3$ is the estimator while $\frac{1}{3}(15+9+7)$ is the corresponding estimate. Why we use the word corresponding will become relevant as we proceed with the definition and explanations.

Statistical Inference

Method of point estimation

1. Method of maximum likelihood
2. Method of moments

Generally any estimator or set of estimators should have some set of desirable properties and we will discuss then later, the two most important being (i) unbiasedness and (ii) consistency.

Method of maximum likelihood estimation (MLE)

Consider you have $X_1, X_2, \ldots, X_n$ have the same distribution function or say the pmf/pdf as their population distribution or pmf/pdf and these observations are what constitutes a sample of size *n*. Now assume that $x_1, x_2, \ldots, x_n$ are the corresponding realized values for these observations. Let us assume the pmf/pdf is denoted by $f(x;\theta)$, then let us first consider the discrete case to show how the method of maximum likelihood is sued to find the estimator. The likelihood function is defined by the concept that signifies the chance or the likeliness of the function taking a value based on the realized values hence the main focus of this method is to find the estimator such that, the likelihood function is maximized. So if we denote the likelihood function as $L(X_1, X_2, \ldots, X_n;\theta)$, then the realized value of the likelihood function

is $L(x_1, x_2, \ldots, x_n; \theta)$, which is equal to $P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta] = \prod_{i=1}^{n} P(X_i = x_i)$, (as $X_i's$ are all *i.i.d*)which is the probability of observing a given sample, $(x_1, x_2, \ldots, x_n)$. Here L is the joint distribution of $X_i's$, i.e., the joint probability of observing the sample that we have observed, i.e., $(x_1, x_2, \ldots, x_n)$.

Let us consider examples for the discrete case, which are given below for an easy illustration of the methodology.

Example 1

Let us consider the example here, where $X \sim P(\theta)$, and if a is known we want to estimate θ.

So $e^{-\theta}\left(\frac{\theta^x}{x!}\right)$, $x = 0,1,\ldots$, then $L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} e^{-\theta}\left(\frac{\theta^x}{x!}\right)$, which is the likelihood function of observing a given sample. After we have observed the values $x_i's$ we know the value of the r.v.'s $X_i's$, hence L is now only a function of θ, i.e., we have to find L(θ). Thus given that $(x_1, x_2, \ldots, x_n)$ are actually observed we have to take that value of θ, which maximizes L(θ). This value of θ is called as the maximum estimator of θ. Thus

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} e^{-\theta}\left(\frac{\theta^x}{x!}\right) = e^{-n\theta}\left(\frac{\theta^{\sum_{\forall i} x_i}}{x_1! \ldots x_n!}\right)$$

$$\therefore \log_e \{L(x_1, \ldots, x_n; \theta)\} = -n\theta + \left(\sum_{\forall i} x_i\right) \log_e \theta - \log_e (x_1! \ldots x_n!)$$

$$\therefore \frac{d}{d\theta} \log_e \{L(x_1, \ldots, x_n; \theta)\} = -n + \frac{1}{\theta} \sum_{\forall i} x_i \text{ and}$$

$$\frac{d^2}{d\theta^2} \log_e \{L(x_1, \ldots, x_n; \theta)\} = -\frac{1}{\theta^2} \sum_{\forall i} x_i < 0, \text{ hence } \hat{\theta} = \frac{1}{n} \sum_{\forall i} x_i = \overline{X}_n$$

Example 2

$$X_i = \begin{cases} 0 & \text{if the } i^{th} \text{ student in the sample is a smoker} \\ 1 & \text{otherwise} \end{cases}$$

Now the observed values for $X_i's$ are $x_i's$ and consider the probability in the population is θ, i.e., $P[X=1]=\theta$ and $P[X=0]=(1-\theta)$, hence we have the pmf as $P[X=i]=\theta^i(1-\theta)^{(1-i)}$, i = 0, 1. So the likelihood function is of the form

$$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n}\theta^i(1-\theta)^{(1-i)}=\theta^{\sum_{\forall i} i}(1-\theta)^{\left(n-\sum_{\forall i} i\right)}$$

$$\log_e\{L(x_1,\ldots,x_n;\theta)\}=(\log_e\theta)\sum_{\forall i}x_i+\left(n-\sum_{\forall i}x_i\right)\log_e(1-\theta)$$

$$\therefore \frac{d}{d\theta}\log_e\{L(x_1,\ldots,x_n;\theta)\}=\frac{1}{\theta}\sum_{\forall i} i-\frac{1}{(1-\theta)}\left(n-\sum_{\forall i} i\right)$$

$$\therefore \frac{d^2}{d\theta^2}\log_e\{L(x_1,\ldots,x_n;\theta)\}=-\frac{1}{\theta^2}\sum_{\forall i} i-\frac{1}{(1-\theta)^2}\left(n-\sum_{\forall i} i\right)<0,\ \text{hence}\ \ \hat{\theta}=\frac{1}{n}\sum_{\forall i} i=p_n,\ \text{where}$$

p_n is the proportion of students in the sample who smoke

Let us now consider the continuous case, such that $f(x;\theta)$, such that $P[X=x]=0$.

$\therefore P[x\le X\le x+h]\cong Area=h\times f(x;\theta)$, provided *h* is small.

Now $\therefore P[x_1\le X_1\le x_1+h,\ldots,x_1\le X_1\le x_1+h]=\prod_{i=1}^{n}P[x_i\le X_i\le x_i+h]$

$$\cong h_1\times\ldots\times h_n\underbrace{\{f(x_1;\theta)\times\ldots\times f(x_n;\theta)\}}_{L(\theta)}$$

Thus the estimator of θ, i.e., $\hat{\theta}$, which maximizes $L(\theta)$ is the what we want to find.

Example

Suppose X~ N(μ, σ^2), i.e., $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with the values of, $x\in\Re$, $\mu\in\Re$ and $\sigma^2>0$, with μ unknown and σ known. So let us consider $(X_1,X_2,\ldots,X_n)$ is the set of observations where their corresponding realized values are $(x_1,x_2,\ldots,x_n)$, hence the likelihood function for μ is

$$L(x_1,\ldots,x_n;\mu)=\prod_{i=1}^{n}f(x_i;\mu)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\therefore \log_e\{L(x_1,\ldots,x_n;\mu)\}=-\frac{n}{2}\log_e\sigma^2-n\log_e\sqrt{2\pi}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \frac{d}{d\mu}\log_e\{L(x_1,\ldots,x_n;\mu)\}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\therefore \frac{d^2}{d\mu^2}\log_e\{L(x_1,\ldots,x_n;\mu)\}=-\frac{1}{\sigma^2}\sum_{i=1}^{n}\ <0\text{, hence }\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}_n$$

Example

Suppose X~ N(μ, σ^2), i.e., $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with the values of, $x\in\Re$, $\mu\in\Re$ and $\sigma^2>0$, with μ unknown and σ unknown. So let us consider $(X_1,X_2,\ldots,X_n)$ is the set of observations where their corresponding realized values are $(x_1,x_2,\ldots,x_n)$, hence the likelihood function for μ is

$$L(x_1,\ldots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}f(x_i;\mu)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\therefore \log_e\{L(x_1,\ldots,x_n;\mu,\sigma^2)\}=-\frac{n}{2}\log_e\sigma^2-n\log_e\sqrt{2\pi}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \frac{d}{d\mu}\log_e\{L(x_1,\ldots,x_n;\mu)\}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \text{ and}$$

$$\therefore \frac{d}{d\sigma^2}\log_e\{L(x_1,\ldots,x_n;\mu,\sigma^2)\}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \frac{d^2}{d\mu^2}\log_e\{L(x_1,\ldots,x_n;\mu)\}=-\frac{1}{\sigma^2}\sum_{i=1}^{n} <0 \text{ and}$$

$$\therefore \frac{d^2}{d\sigma^2}\log_e\{L(x_1,\ldots,x_n;\mu,\sigma^2)\}=\frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2$$

From these two we have $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}_n$ and $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$. Just note here that

$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$ may not be the best estimator and we will proof that later.

Example

Suppose X~ N(μ, σ²), i.e., $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with the values of, $x\in\Re$, $\mu\in\Re$ and $\sigma^2>0$, with μ known and σ unknown. So let us consider $(X_1,X_2,\ldots,X_n)$ is the set of observations where their corresponding realized values are $(x_1,x_2,\ldots,x_n)$, hence the likelihood function for σ is

$$L(x_1,\ldots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}f(x_i;\mu)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\therefore \log_e\{L(x_1,\ldots,x_n;\mu,\sigma^2)\}=-\frac{n}{2}\log_e\sigma^2-n\log_e\sqrt{2\pi}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \frac{d}{d\mu}\log_e\{L(x_1,\ldots,x_n;\mu)\}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\therefore \frac{d}{d\sigma^2}\log_e\{L(x_1,\ldots,x_n;\mu,\sigma^2)\}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \frac{d^2}{d\sigma^2}\log_e\left\{L\left(x_1,\ldots,x_n;\mu,\sigma^2\right)\right\}=\frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 \text{, hence } \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

Say for example θ is the unknown parameter and we are interested in computing $h(\theta)$, a function of θ. So in order to make this clear, consider X ~ E(λ), i.e., X is distributed exponentially with mean value of λ, such that $h(\lambda)=P[X\geq 5]=e^{-\left(\frac{5}{\lambda}\right)}$. Now for the above case, $h(\lambda)=e^{-\left(\frac{5}{\lambda}\right)}$, as λ is unknown we have to replace that with its best estimate, i.e, $\hat{\lambda}$, which is its MLE, but would that serve the purpose is the main question which is of concern to us. So in order to obtain the MLE of $h(\lambda)$, i.e., $\hat{h}(\lambda)$ we need the following to be true, i.e., $\frac{d}{dh(\lambda)}\log_e[L(x_1,\ldots,x_n;\lambda)]=\frac{d}{d\lambda}\log_e[L(x_1,\ldots,x_n;\lambda)]\times\frac{d\lambda}{dh(\lambda)}=0$. As $\frac{d\lambda}{dh(\lambda)}\neq 0$, hence it will be only possible when $\frac{d}{d\lambda}\log_e[L(x_1,\ldots,x_n;\lambda)]=0$, which is what we want to achieve and this will give us the MLE of λ, i.e., $\hat{\lambda}$. So replacing λ with $\hat{\lambda}$ in $h(\lambda)=e^{-\left(\frac{5}{\lambda}\right)}$ we have the estimator as $\hat{h}(\lambda)=e^{-\left(\frac{5}{\hat{\lambda}}\right)}=e^{-\left(\frac{5}{\bar{X}_n}\right)}=\exp\left\{-\left(\frac{5}{\frac{1}{n}\sum_{i=1}^{n}X_i}\right)\right\}$. So if $\hat{\sigma}^2$ is the best estimator for σ^2, then $\hat{\sigma}$ is the best estimator for σ

Example

If X the r.v. be exponential distribution, such that X ~ E(λ), i.e., $f(x)=\frac{1}{\lambda}e^{-\left(\frac{x}{\lambda}\right)}$,with $x\in[0,+\infty)$. Here we measure the health of a patient in the infectious ward of the hospital, and X denotes the r.v. which depicts the time till the patient develops the disease. So as a

doctor we are interested to estimate the average time before which a person is affected by the disease. Hence

x_1: denotes the time when we detect the disease in the first patient

x_2: denotes the time when we detect the disease in the second patient

.

.

X_n: denotes the time when we detect the disease in the n^{th} patient

So $L(x_1,\ldots,x_n;\lambda)=\prod_{i=1}^{n} f(x_i;\lambda)=\prod_{i=1}^{n}\frac{1}{\lambda}e^{-\left(\frac{x}{\lambda}\right)}=\frac{1}{\lambda^n}\exp\left\{-\left(\frac{1}{\lambda}\sum_{i=1}^{n}x_i\right)\right\}$

$$\therefore \log_e L(x_1,\ldots,x_n;\lambda)=-n\log_e\lambda-\left(\frac{1}{\lambda}\sum_{i=1}^{n}x_i\right)$$

$$\therefore \frac{d}{d\lambda}\log_e L(x_1,\ldots,x_n;\lambda)=-\frac{n}{\lambda}+\left(\frac{1}{\lambda^2}\sum_{i=1}^{n}x_i\right)$$

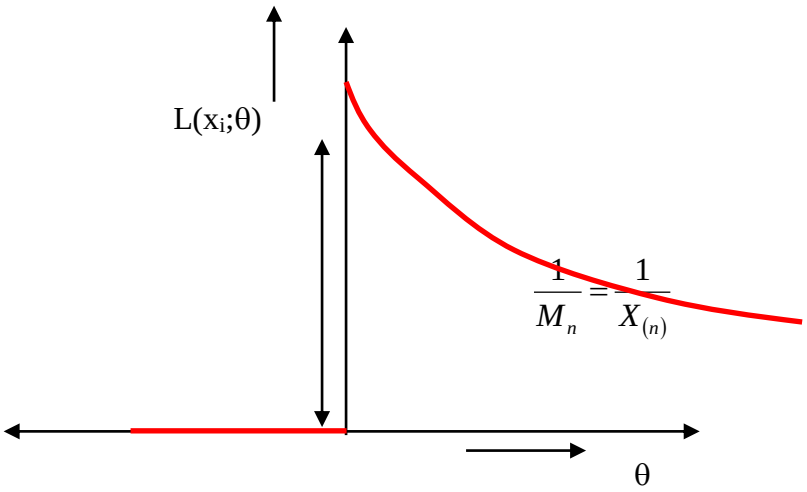
$$\therefore \frac{d^2}{d\lambda^2}\log_e L(x_1,\ldots,x_n;\lambda)=\frac{n}{\lambda^2}-\left(\frac{2}{\lambda^3}\sum_{i=1}^{n}x_i\right), \text{ hence } \hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}_n$$

Example

Let us now consider the uniform distribution (the continuous case), i.e., X~ U(0, θ), i.e., $f(x;0,\theta)=\frac{1}{\theta}$, with the values of $x\in\Re$, i.e., $x\in[0,\theta]$ and $\theta\in\Re$. As an example consider of a particular product are sold by a shop, where θ is the maximum amount/quantity sold of the product. Hence

$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\prod_{i=1}^{n}\frac{1}{\theta}=\frac{1}{\theta^n}$, $0\le x_i\le\theta$ and $i=1,2,\ldots,n$ If you look carefully then calculating the maximum would not solve the problem of finding the MLE of θ, so to

circumvent the problem we use graphical method to solve the problem which is as shown, where $M_{\max, n} = X_{(n)} = Max\{x_1, x_2, \ldots, x_n\}$



Hence $\hat{\theta} = M_{\max,n} = X_{(n)}$. Similarly if we have X~ U(a, b), i.e., $f(x;a,b) = \frac{1}{(b-a)}$, with the values of $x \in \Re$, i.e., $x \in [a,b]$, $\theta \in \Re$ and $a < b$, then $\hat{a} = X_{(1)} = Min\{X_1, X_2, \ldots, X_n\}$ and $\hat{b} = X_{(n)} = Max\{X_1, X_2, \ldots, X_n\}$.

Desirable properties: We have been deriving many of the estimators using MLE method, but have not thought about whether the estimators found out really fit our bill in the sense whether they do have some desirable properties which is wanted of them. More over if you have more than one estimator for the same case, then which is the so called best is of concern to us. In other words given an estimator we do not know how close that estimator is to the population parameter and here if where we sue sampling distribution. The two most important property which is wanted of the estimators are

1. **Property of unbiasedness**: Now if we take the estimated value of the estimator $\hat{\theta}$, then we should have $E_\theta(\hat{\theta}) = \theta$ (where θ is the actual value of the population paranmeter) if the

estimator is unbiased, else $E_\theta(\hat{\theta}) = \theta + b(\theta)$ is true, where $b(\theta)$ is the biased factor. So if $\hat{\theta}$ is the unbiased estimator of θ, then $b(\theta) = 0$. Let us make this clear with a simple example.

Suppose $(X_1, X_2, \ldots, X_n)$ are a sample of *n* number of random observations and also suppose $E(X) = \mu$ and $V(X) = \sigma^2$, then $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. So $E_\mu(\overline{X}_n) = \mu$ as $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}(X_1 + X_2 + \ldots + X_n)$. Now let us suppose we have another estimator, t_n, such that $t_n = \frac{1X_1 + 2X_2 + \ldots + nX_n}{1 + 2 + \ldots + n}$, then $E_\mu(t_n) = E_\mu\left[\frac{1X_1 + 2X_2 + \ldots + nX_n}{1 + 2 + \ldots + n}\right] = \mu$, then t_n is also an unbiased estimator of μ. Furthermore let us consider another estimator, $t_n^* = \frac{1X_1 + 2X_2 + \ldots + nX_n}{2 + 3 + \ldots + n}$, such that $E_\mu(t_n^*) = E_\mu\left[\frac{1X_1 + 2X_2 + \ldots + nX_n}{2 + 3 + \ldots + n}\right] = \left\{\frac{n(n+1)}{n(n+1) - 2}\right\}\mu$, which means that $E_\mu(t_n^*) \neq \mu$, only when *n* is large or $\mu = 0$. Hence t_n^* is not an unbiased estimator of μ.

Example

Let X be the r.v. which denotes the age of a student picked up at random in a day class and Y is the be r.v., which denotes the age of the a student picked up at random in a evening class, then if μ_X and μ_Y are the average age of the students in the day and evening class respectively, then $\hat{\mu}_X = \frac{1}{n_X}\sum_{i=1}^{n_X}\left[X_1 + X_2 + \ldots + X_{n_X}\right]$ and $\hat{\mu}_Y = \frac{1}{n_Y}\sum_{i=1}^{n_Y}\left[Y_1 + Y_2 + \ldots + Y_{n_Y}\right]$ are the estimator for μ_X and μ_Y respectively.

One should remember that it is not easy to construct an unbiased estimator for the population parameter, θ. So suppose that $(X_1, X_2, \ldots, X_n)$ and we also have $E(X) = \mu$ and $V(X) = \sigma^2$,

such that $E(X_i)=\mu$ and $V(X_i)=\sigma^2$ are true. Then the sample variances, can be given by

$s_n^{/2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ and $s_n^2=\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$, hence

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

$$\therefore E\{(n-1)s_n^2\}=E\left(\sum_{i=1}^{n}X_i^2\right)-E(n\bar{X}_n^2)=\sum_{i=1}^{n}\left[V(X_i)+\{E(X_i)\}^2\right]-n\left[V(\bar{X}_n)+\{E(\bar{X}_n)\}^2\right]$$

$$=\sum_{i=1}^{n}(\sigma^2+\mu^2)-n\left[\frac{\sigma^2}{n}+\mu^2\right]=(n-1)\sigma^2$$

$\therefore E(s_n^2)=\sigma^2$, hence s_n^2 is an unbiased estimator of σ^2 provided if the expected value of the population, i.e., μ is unknown.

Now consider $s_n^{/2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$, we can easily see and prove that $E(s_n^{/2})=\left\{\frac{(n-1)}{n}\right\}\sigma^2$, hence $s_n^{/2}$ is not an unbiased estimator of σ^2 if the expected value of the population, i.e., μ is known. It is only true that as $n\to\infty$, then only $E(s_n^{/2})\cong\sigma^2$ as

$$\underset{n\to\infty}{Lt}\left\{\frac{(n-1)}{n}\right\}=\underset{n\to\infty}{Lt}\left(1-\frac{1}{n}\right)\cong 1$$

Example

Let us assume the continuous uniform distribution, such that X ~ UD (0, θ), i.e., $f(x;\theta)=\frac{1}{\theta}$, with the values of $x\in\Re$, i.e., $x\in[0,\theta]$ and $\theta\in\Re$. Then we know that $E(X_i)=\frac{\theta}{2}$ and $V(X_i)=\frac{\theta^2}{12}$, $\forall i=1,2,\ldots,n$, considering we have a sample size of *n*. Consider $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}(X_1+X_2+\ldots+X_n)$, then $E(\bar{X}_n)=\frac{\theta}{2}$, hence

$t_n = 2 \times \overline{X}_n = 2 \times \frac{1}{n}\sum_{i=1}^{n}\left(X_1 + X_2 + \ldots + X_n\right)$ is an unbiased estimator of θ. Also

$X_{(n)} = Max\left[X_1, X_2, \ldots, X_n\right]$ and $X_{(1)} = Min\left[X_1, X_2, \ldots, X_n\right]$, hence

$$P\left[X_{(n)} \le x\right] = P\left[X_1 \le x, X_2 \le x, \ldots, X_n \le x\right] = P\left[X_1 \le x\right] \times \ldots \times P\left[X_n \le x\right] = \frac{x}{\theta} \times \ldots \times \frac{x}{\theta}$$

$$= \left(\frac{x}{\theta}\right)^n$$

$$\therefore f_{X_{(n)}}(x) = \frac{dF_{X_{(n)}}(x)}{dx} = \frac{d}{dx}\left\{\left(\frac{x}{\theta}\right)^n\right\} = n\left(\frac{x^{n-1}}{\theta^n}\right) \text{ for } 0 \le x \le \theta$$

$$\therefore E\left\{X_{(n)}\right\} = \int_0^\theta x f_{X_{(n)}}(x)dx = \int_0^\theta xn\left(\frac{x^{n-1}}{\theta^n}\right)dx = \int_0^\theta n\left(\frac{x^n}{\theta^n}\right)dx = \left(\frac{n}{n+1}\right)\theta,$$ hence $X_{(n)}$ is not the unbiased estimator of θ.

Now what if we have a new estimator $t_n^* = \left(\frac{n+1}{n}\right)X_{(n)}$, then $E\left(t_n^*\right) = E\left\{\left(\frac{n+1}{n}\right)X_{(n)}\right\} = X_{(n)}$ is an unbiased estimator of θ.

Note if we have X ~ UD (a, b), i.e., $f(x;a,b) = \frac{1}{(b-a)}$, with the values of $x \in \Re$, i.e., $x \in [a,b]$, $a < b$ and $a, b \in \Re$. We have already stated $X_{(n)} = \hat{b}$ and $X_{(1)} = \hat{a}$, but they are not the unbiased estimators. A practical example of this can be where we have the following systems shown in Fig below

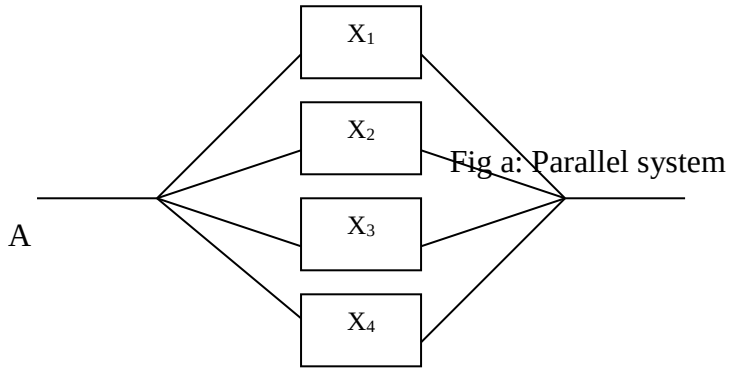


Fig a: Parallel system

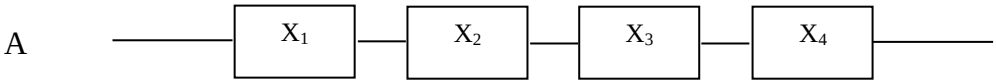


Fig b: Series system

For the first case, i.e., parallel system the system will fail iff all of the components $X_1,\ldots, X_4$ fail, such that we should have the best estimate of the working life of the system as $X_{(4)} = Max(X_1, X_2, X_3, X_4)$, while for the series system the system will fail iff the moment one of the component fails, which is the minimum of the lot, such that $X_{(1)} = Min(X_1, X_2, X_3, X_4)$

Example

Consider X~ B(2,p), i.e., $f(x;n,p) = {}^2C_x p^x q^{(2-x)}$, where x= 0, 1, 2, $0 \le p, q \le 1$ and (p + q) = 1. Here p = probability of success and $\theta = \left(\frac{1}{p}\right)$ is the average number of trials required to get the first success. Now we make one observation, X, based on which we estimate $\theta = \left(\frac{1}{p}\right)$. So now if the estimator is $t(X)$, then for unbiasedness we must have $E\{t(X)\} = \theta$.

Hence: $E_\theta\{t(X)\} = \sum_{x=0}^{2} t(x)P[X = x] = t(0)(1-p)^2 + 2t(1)(1-p)^2 p + t(2)p^2$. Now for unbiasedness we should have $pE_\theta\{t(X)\} = 1$, such that $p\left[t(0)(1-p)^2 + 2t(1)(1-p)^2 p + t(2)p^2\right] = 1$, is true for $0 \le p \le 1$.

If p = 0, then LHS is zero, but RHS is not zero but 1, hence we need to equate the coefficients of *p* on both sides, i.e., t(0) = 0, {-2t(0) + 2t(1)} = 0, {t(0) -2t(1) + t(2)} = 0, such that t(0) = t(1) = t(2) =0. Hence ∄ a function t(X) for which the equation is true, so $\theta = \left(\frac{1}{p}\right)$ cannot be estimated unbiasedly.

Example (an extension of the earlier example)

Now consider X~ B(n,p), i.e., $f(x;n,p) = {}^{n}C_{x}\, p^{x} q^{(n-x)}$, where x= 0, 1,…, n, 0 ≤ p, q ≤ 1 and (p + q) = 1. Here p = probability of success and $\theta = \left(\frac{1}{p}\right)$ is the average number of trials required to get the x^{th} success. Now we make *n* number of observations, X_1, X_2,…, X_n, , based on which we estimate $\theta = \left(\frac{1}{p}\right)$. So now if the estimator is again considered as $t(X)$, then for unbiasedness we must have $E\{t(X)\} = \theta$.

Hence:

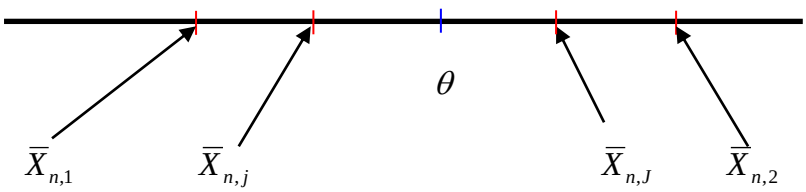
$$E_{\theta}\{t(X)\} = \sum_{x=0}^{n} \{t(x)P(X=x)\} = {}^{n}C_{0}\, t(0) p^{0}(1-p)^{n} + {}^{n}C_{1}\, t(1) p^{1}(1-p)^{n-1} + \ldots + {}^{n}C_{n}\, t(n) p^{n}(1-p)^{0}.$$

Now for unbiasedness we should have $pE_{\theta}\{t(X)\} = 1$, such that $p\{{}^{n}C_{0}\, t(0) p^{0}(1-p)^{n} + {}^{n}C_{1}\, t(1) p^{1}(1-p)^{n-1} + \ldots + {}^{n}C_{n}\, t(n) p^{n}(1-p)^{0}\} = 1$, is true for 0 ≤ p ≤ 1.

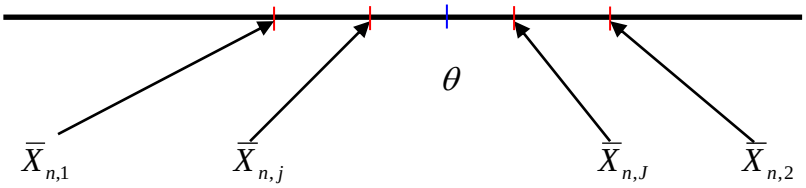
If p = 0, then LHS is zero, but RHS is not zero but 1, hence we need to equate the coefficients of *p* on both sides and comments whether there exists a function t(X) for which the equation is true, so that $\theta = \left(\frac{1}{p}\right)$ can be estimated unbiasedly.

2. Property of consistency: Suppose t_n is the estimator of the population parameter, θ, then the error in the estimation process is given as $|t_n - \theta|$. Now if $\underset{n\to\infty}{Lt}\left[P\{|t_n - \theta| \geq \varepsilon\}\right] \to 0$ for every specified margin of error ε, then t_n is said to be consistent estimator of θ. Note that $\underset{n\to\infty}{Lt}\left[P\{|t_n - \theta| \geq \varepsilon\}\right] \to 0$ can also be equivalently written as $\underset{n\to\infty}{Lt}\left[P\{|t_n - \theta| \leq \varepsilon\}\right] \to 1$. Now this asymptotic propert is achieved as we take more and more number of such samples, where in each set of sample we take *n* number of observations, ie., to say we take the set of samples as $(X_{1j}, X_{2j}, \ldots, X_{nj})$, where j = 1, 2,…. Now we need to be careful and note that $E_\theta(t_n) = \theta$ may mean that the observations are far dispersed in and around the value of θ, such that we want to achieve maximum amount of concentration of the readings in and around the population parameter, θ, such that dispersion is as low as possible. The illustrations below may help us to understand this concept better and it is important to note that the observations are symmetrically dispersed about θ.

Case I



Case II



Example

Consider $(X_1, X_2, \ldots, X_n)$ are n number of random observations, such that $E(X) = \mu$ and $V(X) = \sigma^2$, also consider $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, hence we already know that $E(\bar{X}_n) = \mu$ and $V(\bar{X}_n) = \frac{\sigma^2}{n}$. Using Tchebychev′s inequality we have:

$$P\left\{\left|\bar{X}_n - \mu\right| \geq \varepsilon\right\} \leq \frac{V(\bar{X}_n)}{\varepsilon^2}$$

$$\therefore P\left\{\left|\bar{X}_n - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ as } n \to \infty, \text{ so } \bar{X}_n \text{ is an consistent estimator of } \mu$$

Example

Consider $(X_1, X_2, \ldots, X_n)$ are n number of random observations, such that $E(X) = \mu$ and $V(X) = \sigma^2$, also consider $t_n = X_n$, i.e., we estimate the average of the population by looking at $t_n = X_n$. Now $E(t_n) = E(X_n) = \mu$, but it is not consistent as we see below. Now $P\left\{\left|X_n - \mu\right| \geq \varepsilon\right\} \leq \frac{V(X_n)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}$, hence $\underset{n\to\infty}{Lt}\left[P\left\{\left|X_n - \mu\right| \geq \varepsilon\right\}\right] \leq \frac{V(X_n)}{\varepsilon^2} = \underset{n\to\infty}{Lt}\left(\frac{\sigma^2}{\varepsilon^2}\right) \neq 0$, hence $t_n = X_n$ is not consistent though it is unbiased.

Example

Let us consider another example where π is the proportion of a category of goods in the particular population, so in order to estimate π, we take a random sample of size, *n*, such that the observations are as follows, $(X_1, X_2, \ldots, X_n)$. Then if we pick up an observation, i.e., at random, then X_i = 1 if the i^{th} observation belongs to the above mentioned category and 0 otherwise. Then $P[X_i = k] = \pi^k (1-\pi)^{1-k}$, where k = 0, 1, hence expected value is given by $E(X) = \sum_{k=0}^{1} \{x_k \times P[X_k = k]\} = 0(1-\pi) + 1\pi$ and variance is

$V(X)=\sum_{k=0}^{1}\left[\{x_k - E(X)\}^2 \times P(X_k = k)\right]=\pi(1-\pi)$. Furthermore $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n}(X_1 + \ldots + X_n)$, whci is technically equal to the following ration which is $\frac{1}{n}\{\#\ of\ category\ product\ in\ sample\} = p_n$ (say) and as we can easily understand p_n is the proportion of that particular category of goods in the sample. Now we can easily prove that $E(p_n)=\pi$ and $V(p_n)=\left\{\frac{\pi(1-\pi)}{n}\right\}$, thus p_n is both an unbiased as well as a consistent estimator of π.

We can recall here that if we have a population distribution where the random variable is denoted by X, and we have $E(X)=\mu$ and $V(X)=\sigma^2$, then have the following which is true:

Metric of measure	Methodology of sampling SRSWR	Methodology of sampling SRSWOR
Mean	μ	μ
Variance	$\left(\frac{\sigma^2}{n}\right)$	$\left\{\frac{N-n}{n(N-1)}\right\}\sigma^2$
Covariance	0	$\left(\frac{1}{N-1}\right)\sigma^2$

Consider $(X_1, X_2, \ldots, X_n)$ is a set of *n* random observations, where X denotes the r.v., and assume that $t_n = t(X_1, X_2, \ldots, X_n)$ is the estimator for the parameter, θ, then we should have $\underset{n\to\infty}{Lt}\ P\{|t_n - \theta| \geq \varepsilon\}=0$ in case t_n is a consistent estimator of θ. Now by Tchebychev's inequality we have $P\{|t_n - \theta| \geq \varepsilon\} \leq \frac{V(t_n)}{\varepsilon^2} = \frac{E(t_n - \theta)^2}{\varepsilon^2}$ and the term $E(t_n - \theta)^2 = 0$ if t_n is consistent as $n \to \infty$. However if $E(t_n - \theta)^2$ does not tend to zero, then we cannot say that t_n is not consistent.

Example

Let us continue with our earlier example, where we have the continuous uniform distribution, such that X ~ UD (0, θ), i.e., $f(x;\theta)=\frac{1}{\theta}$, with the values of $x\in\Re$, i.e., $x\in[0,\theta]$ and $\theta\in\Re$.

We already know that $\hat{\theta}=X_{(n)}=Max[X_1,X_2,\ldots,X_n]$ and $E(\hat{\theta})=E\{X_{(n)}\}=\left(\frac{n}{n+1}\right)\theta$. Also pdf of $X_{(n)}$ is given by

$$f_{X_{(n)}}(x)=\begin{cases} n\left(\frac{x^{n-1}}{\theta^n}\right) & 0\le x\le\theta \\ 0 & otherwise \end{cases}$$

Now $E\{X_{(n)}^k\}=\int_0^\theta x^k f_{X_{(n)}}(x)dx=\int_0^\theta x^k\left\{n\left(\frac{x^{n-1}}{\theta^n}\right)\right\}dx=\left(\frac{n\theta^k}{n+k}\right)$ hence $E\{X_{(n)}\}=\left(\frac{n\theta}{n+1}\right)$,

$$V\{X_{(n)}\}=\left[E\{X_{(n)}^2\}-E\{X_{(n)}\}^2\right]=\left\{\left(\frac{n\theta^2}{n+2}\right)-\left(\frac{n\theta}{n+1}\right)^2\right\}=\left\{\frac{n}{(n+1)(n+2)}\right\}\theta^2$$

Also $E[X_{(n)}-\theta]^2=E\left[\left\{X_{(n)}-\left(\frac{n\theta}{n+1}\right)\right\}+\left\{\left(\frac{n\theta}{n+1}\right)-\theta\right\}\right]^2=E\left[\left\{X_{(n)}-\left(\frac{n\theta}{n+1}\right)\right\}-\left(\frac{\theta}{n+1}\right)\right]^2$

$$=E\left\{X_{(n)}-\left(\frac{n\theta}{n+1}\right)\right\}^2-2\left(\frac{\theta}{n+1}\right)E\left\{X_{(n)}-\left(\frac{n\theta}{n+1}\right)\right\}+\left(\frac{\theta}{n+1}\right)^2$$

$$=V\{X_{(n)}\}+\left(\frac{\theta}{n+1}\right)^2=\left\{\frac{n}{(n+1)(n+2)}\right\}\theta^2+\left(\frac{\theta}{n+1}\right)^2$$

Thus as $n\to\infty$, we have $X_{(n)}$ as consistent ($\underset{n\to\infty}{Lt}\left\{\frac{n}{(n+1)(n+2)}\right\}\theta^2+\underset{n\to\infty}{Lt}\left(\frac{\theta}{n+1}\right)^2=0$), but biased.

Can you find whether $\left(\frac{n+1}{n}\right)X_{(n)}$ which is already unbiased is consistent or not. Similarly one can find that for X ~ UD (a, b), i.e., $f(x;a,b)=\frac{1}{(b-a)}$, with the values of $x\in\Re$, i.e., $x\in[a,b]$, a ≤ b and $a,b\in\Re$, whether $\hat{a}=Min\{X_1,\ldots,X_n\}$ and $\hat{b}=Max\{X_1,\ldots,X_n\}$ are unbiased and consistent estimator of *a* and *b* respectively. In case $\hat{a}=Min\{X_1,\ldots,X_n\}$ and $\hat{b}=Max\{X_1,\ldots,X_n\}$ are not unbiased/consistent, then what are the corresponding estimators which are both unbiased as well as consistent.

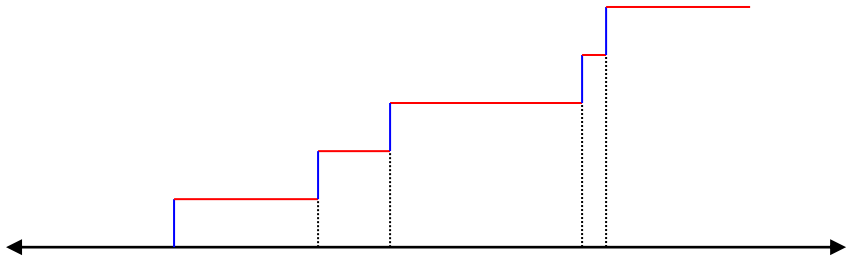
Till now we have considered the distribution function of the population such that we were required to find the parameters, $\theta=(\theta_1,\ldots,\theta_k)$, but one question which is of considerable interest to us to study is what are the corresponding distributions of the parameters, $\theta=(\theta_1,\ldots,\theta_k)$, i.e., what are the functional forms of $F_{\theta_1}\{t_{n,1}(X_1,X_2,\ldots X_n)\}, F_{\theta_2}\{t_{n,2}(X_1,X_2,\ldots X_n)\},\ldots, F_{\theta_k}\{t_{n,k}(X_1,X_2,\ldots X_n)\}$, where $t_{n,1}(X_1,X_2,\ldots X_n), t_{n,2}(X_1,X_2,\ldots X_n),\ldots, t_{n,k}(X_1,X_2,\ldots X_n)$ are the corresponding estimators for $\theta_1,\ldots,\theta_k$. For example we know that if we consider the normal distribution such that X ~ N(μ,σ^2), and take a sample $(X_1,X_2,\ldots X_n)$, then $\overline{X}_n$ is the estimator of μ, such that $\overline{X}_n \sim N\left(\mu,\frac{\sigma^2}{n}\right)$. Else if $s_n^{/2}$ or s_n^2 are the estimators of σ^2, as the case may be, depending on whether μ is known or unknown, then $s_n^{/2}$ and s_n^2 will have the χ_n^2 and χ_{n-1}^2 distribution respectively.

Consider $(X_1,X_2,\ldots X_n)$ are the random set of *n* observations which constitutes the sample, and the population distribution is denoted by $F_X(x)$, i.e., $F_X(x)=P[X\le x]$. Now consider

$F_n(x)$ as the empirical distribution of the population, then

$F_n(x) = \frac{1}{n}\{\#\ of\ observations\ amongst\ X_1, X_2, \ldots, X_n\ that\ do\ not\ exceed\ x\}$.

Now $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, then we have the following way of illustration which is



One should understand that the step function is of equal heights but of unequal lengths/width.

Now if $E[F_n(x)] = F_X(x)$, then $F_n(x)$ is an unbiased estimator of $F_X(x)$ and

$V\{F_n(x)\} = E\{F_n(x) - F_X(x)\}^2 = \left[\frac{F_X(x)\{1 - F_X(x)\}}{n}\right]$, so that we have

$P[|F_n(x) - F_X(x)| \ge \varepsilon] \le V\{F_X(x)\} = \left[\frac{F_X(x)\{1 - F_X(x)\}}{n\varepsilon^2}\right]$, thus

$\underset{n\to\infty}{Lt}\ P[|F_n(x) - F_X(x)| \ge \varepsilon] \le \underset{n\to\infty}{Lt}\left[\frac{F_X(x)\{1 - F_X(x)\}}{n\varepsilon^2}\right] = 0$. Hence, (i) $F_n(x)$ is an unbiased estimator of $F_X(x)$ as well as (ii) consistent

Suppose we have two unbiased estimators given by $t_n(X_1, X_2, \ldots, X_n) = \bar{X}_n$ and $t_n^*(X_1, X_2, \ldots, X_n) = \frac{1X_1 + 2X_2 + \ldots + nX_n}{1 + 2 + \ldots + n}$, then $E(t_n) = E(\bar{X}_n) = E(t_n^*) = \mu$, while $V(t_n) = \frac{\sigma^2}{n}$

and $V(t_n^*) = \left\{\frac{(1^2 + 2^2 + \ldots + n^2)}{(1 + 2 + \ldots + n)^2}\right\}\sigma^2 = \left[\frac{\frac{n(n+1)(2n+1)}{6}}{\left\{\frac{n(n+1)}{2}\right\}^2}\right]\sigma^2$, now if $V(t_n) \le V(t_n^*)$, i.e.,

$\frac{\frac{n(n+1)(2n+1)}{6}}{\left\{\frac{n(n+1)}{2}\right\}^2} \geq 1$, i.e., $n \geq 1$, which is always true hence $V(t_n) \leq V(t_n^*)$. So we can say that t_n is a better estimator than t_n^* for the population parameter, θ, then $V(t_n) = E(t_n - \theta)^2 \leq V(t_n^*) = E(t_n^* - \theta)^2$ irrespective of the true value of θ, which implies that t_n is an uniformly better estimator than t_n^*. Thus if we have two or more estimators, say, $t_{n,1}, t_{n,2}, \ldots, t_{n,j}, \ldots, t_{n,J}$, and all are unbiased, then we say that $t_{n,j}$ is the uniformly minimum variance unbiased estimator (UMVUE) of θ iff (i) $E_\theta(t_{n,1}) = E_\theta(t_{n,2}) = \ldots = E_\theta(t_{n,j}) = \ldots = E_\theta(t_{n,J}) = \theta$ and (ii) $V_\theta(t_{n,j}) \leq \underset{\forall i=1,\ldots,J \neq j}{V_\theta}(t_{n,i})$, and remember that $t_{n,i}$ ($\forall\, i = 1, \ldots, J \neq j$) is any unbiased estimator of θ. In many cases it may be true that we have no UMVUE for any $\theta \in \Theta$, or for a certain value of θ we get an UMVUE, while for other range of θ we d not get an UMVUE.

Rao-Cràmer′s Inequality

Consider f(x;θ) is any pmf/pdf and the corresponding distribution function is $F_X(x)$, then we have $\int_{-\infty}^{+\infty} \left\{\frac{\partial \log_e f(x;\theta)}{\partial \theta}\right\}^2 f(x;\theta)dx = \int_{-\infty}^{+\infty} \left\{\frac{\partial \log_e f(x;\theta)}{\partial \theta}\right\}^2 dF_X(x;\theta) = I(\theta)$ implies that $E\left[\frac{\partial \log_e f(x;\theta)}{\partial \theta}\right]^2 = I(\theta)$, whci means the information contained in a single observation. According to Rao-Cràmer′s inequality under some mild constraints (which are satisfied automatically by the distribution functions we are already familiar with, excepting those for which the range of the random variable depends on θ) the following inequality holds

$V_\theta(t_n) \geq \left[\dfrac{\left\{1+\dfrac{\partial}{\partial\theta}b(x;\theta)\right\}^2}{nI(\theta)} \right]$. Hence if t_n is an unbiased estimator and its variance is equal to $\dfrac{1}{nI(\theta)}$, then t_n must be UMVUE. Also remember that $I(\theta) = \int_{-\infty}^{+\infty} \left\{ \dfrac{\partial}{\partial\theta} \log_e f(x;\theta) \right\}^2 f(x;\theta)dx = \int_{-\infty}^{+\infty} \left\{ \dfrac{\partial}{\partial\theta} \log_e f(x;\theta) \right\}^2 dF_X(x;\theta)$. Now we know from simply calculus that $\dfrac{\partial}{\partial\theta} \log_e f(x;\theta) = \left\{ \dfrac{\partial}{\partial f(x;\theta)} \log_e f(x;\theta) \right\} \times \left\{ \dfrac{\partial}{\partial\theta} f(x;\theta) \right\}$, hence $\dfrac{\partial}{\partial\theta} \log_e f(x;\theta) = \left\{ \dfrac{\dfrac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} \right\}$.

So $I(\theta) = \int_{-\infty}^{+\infty} \left\{ \dfrac{\dfrac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} \right\}^2 f(x;\theta)dx = \int_{-\infty}^{+\infty} \left\{ \dfrac{\dfrac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} \right\}^2 dF_X(x;\theta)$

Thus if $b(x;\theta) = 0$ and $V_\theta(t_n)$ equals the RHS, then this t_n is the UMVUE of θ.

<u>Example</u>

Say the population proportion is π and hence we make use of the sample information to find π, thus

$$X_i = \begin{cases} 1 & \textit{if the } i^{th} \textit{ unit belongs to the specified category} \\ 0 & \textit{otherwise} \end{cases}$$

$\therefore f(x;\pi) = \pi^x (1-\pi)^{(1-x)}$ for $x = 0, 1$

$\therefore \log_e f(x;\pi) = x \log_e \pi + (1-x)\log_e(1-\pi)$, i.e., $\dfrac{d}{d\pi} \log_e f(x;\pi) = \dfrac{x}{\pi} - \dfrac{(1-x)}{(1-\pi)}$, hence

$I(\pi)=\sum_{x=0}^{1}\left\{\frac{x-\pi}{\pi(1-\pi)}\right\}^2\pi^x(1-\pi)^{(1-x)}=\frac{1}{\pi(1-\pi)}$. Thus the RHS of Rao-Cràmer′s inequality is $\frac{1}{\pi(1-\pi)}$ and hence the sample proportion p_n is an unbiased estimator of π

Example

Let us consider the X~ P(λ), i.e., $f(x;\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}$, which is the Poisson distribution, with the values of $x=0,1,\ldots$, where $\lambda\in\Re^+$, then we already know that $E_\lambda\left[\overline{X}_n\right]=\lambda$. Hence we have

$$\log_e f(x;\lambda)=\log_e\left[\frac{\lambda^x e^{-\lambda}}{x!}\right]=x\log_e\lambda-\lambda-\log_e x!$$

$$\therefore\frac{\partial}{\partial\lambda}\log_e f(x;\lambda)=\frac{x-\lambda}{\lambda}$$

$$\therefore I(x;\lambda)=\sum_{x=0}^{+\infty}\left(\frac{x-\lambda}{\lambda}\right)^2\frac{\lambda^x e^{-\lambda}}{x!}=\frac{1}{\lambda^2}\sum_{x=0}^{+\infty}(x-\lambda)^2\frac{\lambda^x e^{-\lambda}}{x!}=\frac{1}{\lambda}$$

Hence the RHS of Rao-Cràmer′s inequality bound is $\frac{\lambda}{n}$. On the other hand the sample mean, i.e., $\overline{X}_n$ is unbiased for λ and $V_\lambda\left[\overline{X}_n\right]$ is the RHS of Rao-Cràmer′s inequality, hence $\overline{X}_n$ is UMVUE.

Example

Consider the Normal distribution, i.e., X~ N(μ, σ^2), i.e., $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with the values of, $x\in\Re$, $\mu\in\Re$ and $\sigma^2>0$. We also consider that $(X_1,X_2,\ldots,X_n)$ is the sample set consisting of *n* number of observations and in that case we have $E_\mu\left[\overline{X}_n\right]=\mu$.

Moreover:

$$\log_e f(x;\mu,\sigma) = -\log_e \sigma - \log_e \sqrt{2\pi} - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\therefore \frac{\partial}{\partial \mu} \log_e f(x;\mu,\sigma) = \left(\frac{x-\mu}{2\sigma^2} \right)$$

$$\therefore I(x;\mu) = \int_{-\infty}^{+\infty} \left(\frac{x-\mu}{2\sigma^2} \right)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma^4} \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma^2}$$

Therefore the RHS of Rao-Cràmer's inequality is $\frac{\sigma^2}{n}$, while on the other hand the sample mean, $\overline{X}_n$ is unbiased estimator for μ and $V_\mu\left[\overline{X}_n\right]$ is equal to the RHS of the Rao-Cràmer's inequality. Hence $\overline{X}_n$ is UMVUE estimator for μ, both for the case when σ^2 is ***known*** or ***unknown***.

Example

Consider the Normal distribution, i.e., X~ N(μ, σ²), i.e., $f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with the values of, $x \in \Re$, $\mu \in \Re$ and $\sigma^2 > 0$. We also consider that $(X_1, X_2, \ldots, X_n)$ is the sample set consisting of *n* number of observations and we also know that $s_n'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

Moreover:

$$\log_e f(x;\mu,\sigma) = -\frac{1}{2}\log_e \sigma^2 - \frac{1}{2}\log_e 2\pi - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\therefore \frac{\partial}{\partial \sigma^2} \log_e = \frac{1}{2\sigma^2} \left\{ \left(\frac{x-\mu}{\sigma} \right)^2 - 1 \right\}$$

$$\therefore I(x;\sigma^2) = \int_{-\infty}^{+\infty} \left[\frac{1}{2\sigma^2} \left\{ \left(\frac{x-\mu}{\sigma} \right)^2 - 1 \right\} \right]^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$\therefore I\left(x;\sigma^2\right)=\int_{-\infty}^{+\infty}\frac{1}{4\sigma^2}\left\{\left(\frac{x-\mu}{\sigma}\right)^4-2\left(\frac{x-\mu}{\sigma}\right)^2+1\right\}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

Now we already know that $\int_{-\infty}^{+\infty}(x-\mu)^2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\sigma^2$ and

$$\int_{-\infty}^{+\infty}(x-\mu)^4\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=3\sigma^4$$

Also the following is true which is

$$\int_{-\infty}^{+\infty}(x-\mu)^{2k}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=(k-1)\sigma^{2k}\text{ and}$$

$$\int_{-\infty}^{+\infty}(x-\mu)^{2k-1}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=0$$

$$\therefore I\left(x;\sigma^2\right)=\int_{-\infty}^{+\infty}\frac{1}{4\sigma^2}\left\{\left(\frac{x-\mu}{\sigma}\right)^4-2\left(\frac{x-\mu}{\sigma}\right)^2+1\right\}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\frac{1}{4\sigma^4}\left(\frac{3\sigma^4}{\sigma^4}-\frac{2\sigma^2}{\sigma^2}+1\right)=\frac{1}{2\sigma^4}$$

Therefore the RHS of Rao-Cràmer′s inequality is $\frac{2\sigma^4}{n}$, while on the other hand remember that $s_n^{/2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$, hence $E_{\sigma^2}\left(s_n^{/2}\right)=\frac{\sigma^2}{n}E_{\sigma^2}\left\{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}=\sigma^2$.

Also $V_{\sigma^2}\left(s_n^{/2}\right)=\frac{\sigma^4}{n^2}V_{\sigma^2}\left\{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right\}=\frac{\sigma^4}{n^2}2n=\frac{2\sigma^4}{n}$. So as the RHS of Rao-Cràmer′s inequality is $\frac{2\sigma^4}{n}$, and $V_{\sigma^2}\left(s_n^{/2}\right)=\frac{2\sigma^4}{n}$, hence $s_n^{/2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ is the UMVUE of σ^2 when the population mean, i.e., μ is ***known***.

Example

Consider the Normal distribution, i.e., X~ N(μ, σ^2), i.e., $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with the values of, $x\in\Re$, $\mu\in\Re$ and $\sigma^2>0$. We also consider that $(X_1,X_2,\ldots,X_n)$ is the sample set consisting of *n* number of observations and we also know that $s_n^2=\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2$.

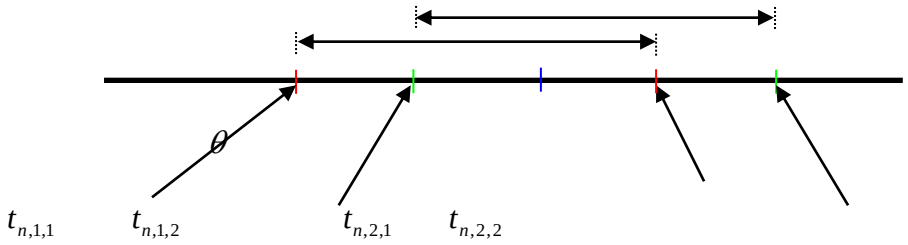
Moreover we know that $\left(\frac{n-1}{\sigma^2}\right)s_n^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}_n}{\sigma}\right)^2\sim\chi^2_{n-1}$, hence $E\left\{\left(\frac{n-1}{\sigma^2}\right)s_n^2\right\}=E\left\{\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}_n}{\sigma}\right)^2\right\}=E\left(\chi^2_{n-1}\right)=(n-1)$ and $V\left\{\left(\frac{n-1}{\sigma^2}\right)s_n^2\right\}=V\left(\chi^2_{n-1}\right)=2(n-1)$, hence $V\left(s_n^2\right)=\left(\frac{2\sigma^4}{n-1}\right)$. We can prove that the Rao-Cràmer′s inequality is $\left(\frac{2\sigma^4}{n-1}\right)$, hence $s_n^2=\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2$ is the UMVUE of σ^2 when the population mean, i.e., μ is ***unknown***.

Remember the Rao-Cràmer′s inequality may not hold for which the range of random variable is unknown and it depends on the parameter, θ∈Θ. The technique we generally use is Blackwellization through which we can construct the relevant UMVUIE in case the Rao-Cràmer′s inequality does not hold.

Note we can easily prove that for X ~ U(0, θ), if $X_{(n)}=Max\{X_1,X_2,\ldots,X_n\}$, then $\left(\frac{n+1}{n}\right)X_{(n)}$ is the UMVUE of θ.

Statistical Inference: Interval Estimation

We already have some idea about interval estimation, where if we consider a set of *n* observations, $(X_1, X_2, \ldots, X_n)$, then say we find two estimators, $t_{n,1}(X_1, X_2, \ldots, X_n)$ and $t_{n,2}(X_1, X_2, \ldots, X_n)$ in order to estimate the population parameter, θ. Moreover we are interested in (1-α) being the level of confidence that we are sure that the value of the parameter, whci his by itself constant lies in the random interval between $t_{n,1}(X_1, X_2, \ldots, X_n)$ and $t_{n,2}(X_1, X_2, \ldots, X_n)$, i.e., as shown



Thus we should have $P\left[t_{n,1}(X_1, X_2, \ldots, X_n) \le \theta \le t_{n,2}(X_1, X_2, \ldots, X_n)\right] = (1-\alpha)$, this means that $t_{n,2}(X_1, X_2, \ldots, X_n) - t_{n,1}(X_1, X_2, \ldots, X_n)$ is the confidence interval (C.I.) which has the parameter, θ, inside it with probability $(1-\alpha)$

Confidence Interval for μ when σ is ***known***

We already know that $\dfrac{\overline{X}_n - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$, hence the C.I. is given by

$$P\left[Z_{1-\frac{\alpha}{2}} \le \frac{\overline{X}_n - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} \le Z_{\frac{\alpha}{2}} \right] = (1-\alpha)$$

$$\therefore P\left[-Z_{\frac{\alpha}{2}} \le \frac{\bar{X}_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \le Z_{\frac{\alpha}{2}}\right] = (1-\alpha)$$

$\therefore P\left[\bar{X}_n - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = (1-\alpha)$. We see the interval between $\left(\bar{X}_n - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = t_{n,1}(X_1, X_2, \ldots, X_n)$ and $\left(\bar{X}_n + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = t_{n,2}(X_1, X_2, \ldots, X_n)$ is random, but the probability that this interval will contain the population parameter, in this case the expected value, μ is given by $(1-\alpha)$. Remember this distance between $t_{n,1}$ and $t_{n,2}$ is the shortest distance. Such that it contains the parameter μ with probability $(1-\alpha)$.

Confidence Interval for μ when σ is ***unknown***

We have the following, i.e., facts which are:

1. $s_n^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

2. $\bar{X}_n$ and s_n^2 are independent

3. $\frac{\bar{X}_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$

4. $(n-1)\frac{s_n^2}{\sigma^2} \sim \chi^2_{n-1}$, such that $\frac{\left\{\frac{\bar{X}_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right\}}{\sqrt{\frac{(n-1)\frac{s_n^2}{\sigma^2}}{(n-1)}}} \sim t_{n-1}$, i.e., $\frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \sim t_{n-1}$

Hence the C.I is given as

$$P\left[t_{n-1,1-\frac{\alpha}{2}} \le \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{s_n} \le t_{n-1,\frac{\alpha}{2}} \right] = (1-\alpha)$$

$$\therefore P\left[-t_{n-1,\frac{\alpha}{2}} \le \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{s_n} \le t_{n-1,\frac{\alpha}{2}} \right] = (1-\alpha)$$

$$\therefore P\left[\bar{X}_n - \left(\frac{s_n}{\sqrt{n}}\right) t_{n-1,\frac{\alpha}{2}} \le \mu \le \bar{X}_n + \left(\frac{s_n}{\sqrt{n}}\right) t_{n-1,\frac{\alpha}{2}} \right] = (1-\alpha)$$

Confidence Interval for σ when μ is ***known***

We have the following, i.e., facts which are:

1. $s_n^{/2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

2. $n\frac{s_n^{/2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$

Hence we would have $P\left[\chi^2_{n,1-\alpha_1} \le n\frac{s_n^{/2}}{\sigma^2} \le \chi^2_{n,\alpha_2} \right] = (1-\alpha)$, i.e.,

$P\left[\frac{ns_n^{/2}}{\chi^2_{n,\alpha_2}} \le \sigma^2 \le \frac{ns_n^{/2}}{\chi^2_{n,1-\alpha_1}} \right] = (1-\alpha)$ is the C.I. for this example, such that [(1-α) + (1-α_1) + α_2 = 1] is true and for the shortest interval we would have $\alpha_1 = \alpha_2 = \alpha/2$, such that the C.I. is now

$P\left[\frac{ns_n^{/2}}{\chi^2_{n,\frac{\alpha}{2}}} \le \sigma^2 \le \frac{ns_n^{/2}}{\chi^2_{n,1-\frac{\alpha}{2}}} \right] = (1-\alpha)$ holds

Confidence Interval for σ when μ is ***unknown***

We have the following, i.e., facts which are:

1. $s_n^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$, hence 2. $(n-1)\left(\frac{s_n^2}{\sigma^2}\right) = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2_{n-1}$

Hence we would have $P\left[\chi^2_{n-1,1-\alpha_1} \le (n-1)\left(\frac{s_n^2}{\sigma^2}\right) \le \chi^2_{n-1,\alpha_2}\right] = (1-\alpha)$, i.e.,

$P\left[\frac{(n-1)s_n^2}{\chi^2_{n-1,\alpha_2}} \le \sigma^2 \le \frac{(n-1)s_n^2}{\chi^2_{n-1,1-\alpha_1}}\right] = (1-\alpha)$ is the C.I. for this example, such that [(1-α) + (1-α_1) + α_2 = 1] is true and for the shortest interval we would have $\alpha_1 = \alpha_2 = \alpha/2$, such that

$P\left[\frac{(n-1)s_n^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)s_n^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right] = (1-\alpha)$ is holds

Example

Suppose $(X_1, X_2, \ldots, X_n)$ are set of *n* random observations, which constitutes the sample. Now these random set of observations are take from any population. So if we are interested in finding the estimator $\hat{\theta}$ for θ, then it depends on the sample $(X_1, X_2, \ldots, X_n)$ and it itself is random, i.e., $\hat{\theta} \sim N\left\{\theta, \frac{1}{nI(\theta)}\right\}$, whenever n ≥ 25, such that $P\left[\frac{\hat{\theta}-\theta}{\sqrt{\frac{1}{nI(\theta)}}} \le x\right] \cong \Phi(x)$, i.e., the sample mean is normally distributed with a certain mean and variance.

Example

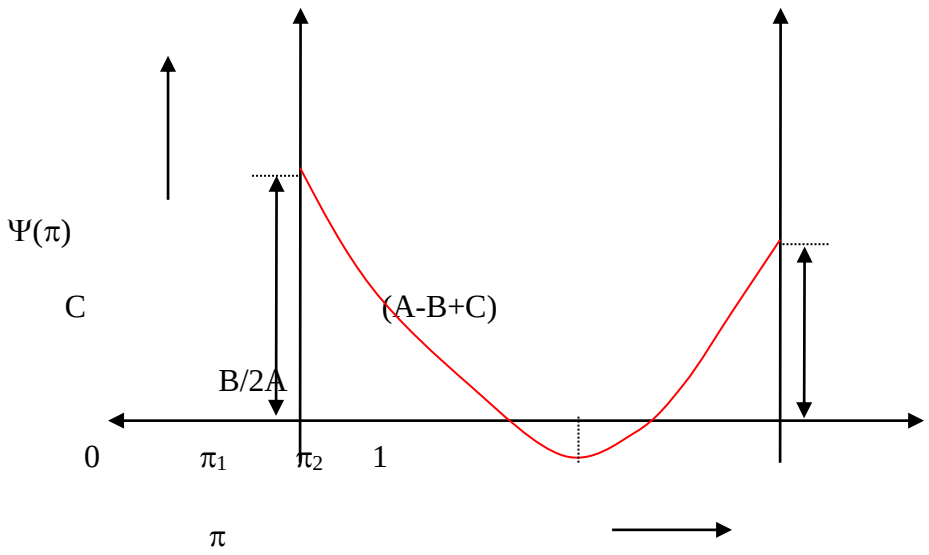
Let us consider the proportion problem, such that $E_\pi(p_n)=\pi$, $V_\pi(p_n)=\frac{\pi(1-\pi)}{n}$ and $I(\pi)=\frac{1}{\pi(1-\pi)}$, and when n is large, then we have $P\left[-Z_{\frac{\alpha}{2}} \le \frac{p_n-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \le Z_{\frac{\alpha}{2}}\right] \cong (1-\alpha)$, i.e.,

$$P\left[\left|\frac{\sqrt{n}(p_n-\pi)}{\sqrt{\pi(1-\pi)}}\right| \le Z_{\frac{\alpha}{2}}\right] \cong (1-\alpha)$$

$$\therefore P\left[\frac{n(p_n-\pi)^2}{\pi(1-\pi)} \le Z_{\frac{\alpha}{2}}^2\right] \cong (1-\alpha), \quad \text{i.e.,} \quad P\left[n(p_n-\pi)^2 \le Z_{\frac{\alpha}{2}}^2\{\pi(1-\pi)\}\right] \cong (1-\alpha), \quad \text{i.e.,}$$

$$P\left[n(p_n-\pi)^2 \le Z_{\frac{\alpha}{2}}^2\{\pi(1-\pi)\}\right] \cong (1-\alpha). \quad \text{Hence} \quad P\left[A\pi^2-B\pi+C \le 0\right] \cong (1-\alpha), \quad \text{i.e.,}$$

$P[\Psi(\pi) \le 0] \cong (1-\alpha)$ where $A=\left(n+Z_{\frac{\alpha}{2}}^2\right)$, $B=\left(2np_n+Z_{\frac{\alpha}{2}}^2\right)$ and $C=np_n^2$. So let us plot the curve $\Psi(\pi) \le 0$ and we have the following illustration to make it more clear.



Now $\Psi(\pi)=\left(A\pi^2-B\pi+C\right)$, i.e., $\frac{d}{d\pi}\Psi(\pi)=2A\pi-B$, which means $2A\pi-B>0$ if $2A\pi-B=+ve$, i.e., $\pi=\left(\frac{B}{2A}\right)$, as shown in the diagram. Also $\frac{d^2}{d\pi^2}\Psi(\pi)=2A$. Now

$$\Psi\left(\frac{B}{2A}\right)=-\frac{B^2}{4A}+C=\left(\frac{-B^2+4AC}{4A}\right)=\left[\frac{\left\{-\left(4n^2p_n+4np_nZ_{\frac{\alpha}{2}}^2+Z_{\frac{\alpha}{2}}^4\right)+4n^2p_n+4np_nZ_{\frac{\alpha}{2}}^2\right\}}{\left(4n+4Z_{\frac{\alpha}{2}}^2\right)}\right].$$

Also $\Psi\left(\frac{B}{2A}\right)<0$, hence $\Psi(\pi)$ will have two roots π_1 and π_2 as shown where $\pi_1<\pi_2$. Here

π_1 = LCL and π_2 = UCL

Confidence Interval for (μ_1 - μ_2) when both σ_1 and σ_2 are ***known***

Consider we have two population where $X\sim N\left(\mu_1,\sigma_1^2\right)$ and $Y\sim N\left(\mu_2,\sigma_2^2\right)$ and from these two populations we take m and n number of observations respectively, such that $\frac{\bar{X}_m-\mu_1}{\left(\frac{\sigma_1}{\sqrt{m}}\right)}\sim N(0,1)$ and $\frac{\bar{Y}_n-\mu_2}{\left(\frac{\sigma_2}{\sqrt{n}}\right)}\sim N(0,1)$ are true, and one wants to find the C.I for the case of

(μ_1 - μ_2), hence $\left\{\frac{\left(\bar{X}_m-\bar{Y}_n\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)}}\right\}\sim N(0,1)$, so that the C.I. can finally be written as

$$P\left[\left\{\left(\bar{X}_m-\bar{Y}_n\right)-Z_{\frac{\alpha}{2}}\sqrt{\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)}\right\}\le\left(\mu_1-\mu_2\right)\le\left\{\left(\bar{X}_m-\bar{Y}_n\right)+Z_{\frac{\alpha}{2}}\sqrt{\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)}\right\}\right]=(1-\alpha).$$

Confidence Interval for (μ_1 - μ_2) when both σ_1 and σ_2 are ***unknown*** but ***equal***

As in the previous example, again consider we have two population where $X\sim N\left(\mu_1,\sigma_1^2\right)$ and $Y\sim N\left(\mu_2,\sigma_2^2\right)$ and from these two populations we take m and n number of observations

respectively, such that $\frac{\overline{X}_m - \mu_1}{\left(\frac{s_{m,1}}{\sqrt{m}}\right)} \sim t_{m-1}$ and $\frac{\overline{Y}_n - \mu_2}{\left(\frac{s_{m,2}}{\sqrt{n}}\right)} \sim t_{n-1}$, hence the C.I. can finally be written

as

$$P\left[\left\{\left(\overline{X}_m - \overline{Y}_n\right) - t_{m+n-2,\frac{\alpha}{2}} \times s_p \frac{1}{\sqrt{m+n}}\right\} \le \left(\mu_1 - \mu_2\right) \le \left\{\left(\overline{X}_m - \overline{Y}_n\right) + t_{m+n-2,\frac{\alpha}{2}} \times s_p \frac{1}{\sqrt{m+n}}\right\}\right] = \left(1-\alpha\right)$$

Where $s_p^2 = \left\{\frac{(m-1)s_{m,1}^2 + (n-1)s_{n,2}^2}{(m+n-2)}\right\}$ is the pooled sample variance

Confidence Interval for (μ_1 - μ_2) when both σ_1 and σ_2 are ***unknown*** but ***unequal***

As in the previous example, again consider we have two population where $X \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y \sim N\left(\mu_2, \sigma_2^2\right)$ and from these two populations we take *m* and *n* number of observations respectively, such that $\frac{\overline{X}_m - \mu_1}{\left(\frac{s_{m,1}}{\sqrt{m}}\right)} \sim t_{m-1}$ and $\frac{\overline{Y}_n - \mu_2}{\left(\frac{s_{m,2}}{\sqrt{n}}\right)} \sim t_{n-1}$, hence the C.I. can finally be written

as

$$P\left[\left\{\left(\overline{X}_m - \overline{Y}_n\right) - t_{m+n-2,\frac{\alpha}{2}} \times \sqrt{\frac{s_{m,1}^2}{m} + \frac{s_{n,1}^2}{n}}\right\} \le \left(\mu_1 - \mu_2\right) \le \left\{\left(\overline{X}_m - \overline{Y}_n\right) + t_{m+n-2,\frac{\alpha}{2}} \times \sqrt{\frac{s_{m,1}^2}{m} + \frac{s_{n,1}^2}{n}}\right\}\right] = \left(1-\alpha\right)$$

Confidence Interval for $\frac{\sigma_1^2}{\sigma_2^2}$ when both μ_1 and μ_2 are ***known***

Consider we have two population where $X \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y \sim N\left(\mu_2, \sigma_2^2\right)$ and from these two populations we take *m* and *n* number of observations respectively, such that $m\frac{s_{m,1}^{/2}}{\sigma_m^2} = \sum_{i=1}^{m}\left(\frac{X_i - \mu_1}{\sigma_m^2}\right)^2 \sim \chi_m^2$ and $n\frac{s_{n,2}^{/2}}{\sigma_n^2} = \sum_{i=1}^{n}\left(\frac{Y_i - \mu_2}{\sigma_n^2}\right)^2 \sim \chi_n^2$. Now formulate the interval

accordingly

Confidence Interval for $\frac{\sigma_1^2}{\sigma_2^2}$ when both μ_1 and μ_2 are ***unknown***

Consider we have two population where $X \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y \sim N\left(\mu_2, \sigma_2^2\right)$ and from these two populations we take m and n number of observations respectively, such that $(m-1)\frac{s_{m,1}^2}{\sigma_m^2} = \sum_{i=1}^{m}\left(\frac{X_i - \bar{X}_m}{\sigma_m^2}\right)^2 \sim \chi_{m-1}^2$ and $(n-1)\frac{s_{n,2}^2}{\sigma_n^2} = \sum_{i=1}^{n}\left(\frac{Y_i - \bar{Y}_n}{\sigma_n^2}\right)^2 \sim \chi_{n-1}^2$. Hence formulate the interval accordingly

Statistical Inference: Hypotheses Testing

Before considering what are the aspects of hypothesis testing and the statistical rules for the same depending on how we formulate our hypothesis we would like to initiate the discussion of this topic through few simple examples.

Example 1

A manufacturer of a particular type of electrical motor has come up with a better hp rating motor then its existing competitors and wants to market that. As is the norm for any manufacturing product, a certain warranty life is to be specified by the manufacturer and the company under our consideration specifies a warranty life of 1 year instead of 8 months given for such rating products. Now you as an engineer are quite skeptical on hearing that the warranty time is 1 years and want to test the validity of this statement which the manufacturer is making.

Example 2

The food and beverage company which manufactures jelly and jams sells them in bottles of 100 gms, 250 gms and ½ kg sizes and you are the marketing general manager of that firm. In order to meet the growing market demand for these products your company has installed a new high productive automatic jelly/jam filling machine, but there has been complains afterwards that on an average the weight of the 100 gms bottles for the jams are never exactly the same as they have been found to be either more or less than 100 gms. So in order to answer this complain and monitor the productivity of the new machine the company has entrusted you the responsibility to solve the problem and hence you would like to test whether the weights on an average for the bottles coming out are about 100 gms (with some

errors) or is there a significant difference in the weights of the bottles, which may be a major concern for the company and hence may necessitate the implementation of some corrective action in order to first identify and then rectify the problem.

Example 3

The traffic flow at the main market road in the city of Amritsar is highest in the each day in the morning from 1000 to 1200 hours, and you as the commissioner of police of Amritsar, want to see whether it is really needed to have the road closed to motor cars and commercial vehicles for that part of the day in order avoid any accidents, which has been reported at a rate of 3 per week with a certain distribution.

Example 4

In the city of Guwahati a new internet service provider (ISP) has just opened its service and is providing high speed internet services and claims the speed of its services are 500 Mbps. You are the resident of that city are interested to get a new internet connection but would like to verify this fact and then take a decision whether to take the connection from this new ISP or continue with the old ISP.

In all these four examples discussed above there are some statements/hypothesis or to be prices they are basically statistical statements about one or more unknown characteristics/facts about the population which needs to be verified. So when we make a claim about a particular hypothesis, at the same time a complementary hypothesis also crops up, such that the original hypothesis is called or termed as the null hypothesis (H_0), while its competitor/negative/opposite hypothesis is the alternative hypothesis (H_A) respectively. So the problem in hypothesis testing, very simply sated is to test the validity of H_0 against H_A.

For the general implementation of hypothesis testing the following steps are generally followed which are

1. As is the norm for sampling we first consider a sample of size *n*, such that the set of observations are $(X_1, X_2, \ldots, X_n)$ which are taken from the population.

2. Then one sets up a decision rule, based on which one tests and makes a decision and chooses one of the decisions, i.e., one of the hypothesis. One should also remember that the decision rule is a technique for making a decision and the decision is generally described by a statistic. Hence we can have any two of the outcomes which are (i) H_0: Accepted and H_A: Rejected and (ii) H_0: Not Accepted and H_A: Accepted.

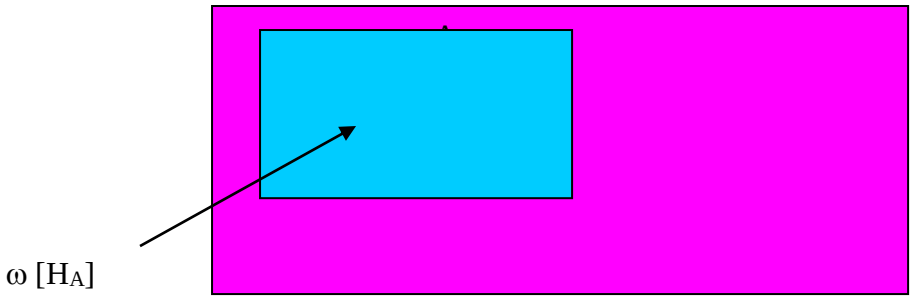
One should remember that to come to any conclusion we use both the sample information and the decision rule. Generally the hypothesis which is a-priori not believed to be true is considered as H_0, but these is no hard and fast rule for choosing between H_0 and H_A. When we accept H_0, it is not that we totally reject H_A or for that fact H_0 is universally true. It only says that based on the sample information and as per evidence it is only though that H_0 is most probable.

To make this clear let us consider an example where we want to test the hypothesis that the population proportion for a particular type of good is not equal to 0.25. Then our hypotheses are: H_0: $\pi = 0.25$ versus H_A: $\pi \neq 0.25$. So in order to prove or disprove the hypothesis and take a decision we take a sample of *n* observations, i.e., $(X_1, X_2, \ldots, X_n)$, which are random in nature and find the proportion of that particular good in the sample, and let that be denoted by p_n. Then based on the sample thus chosen we would deduce that whether $p_n = \pi = 0.25$ or not and depending on our answer we would either accept H_0 (i.e., reject H_A) or not accept H_0 (i.e., accept H_A). So what we do is find the following probability, which is $P[p_n | \pi = 0.25]$. Now say based on the sample, $(X_1, X_2, \ldots, X_n)$, we find $P[p_n = 0.20 | \pi = 0.25] = 0.90$, which

would mean that considering the population proportion as 0.25, the probability of finding the sample proportion as 0.20 is 0.90, which is highly likely. So we would accept H_0 or on the other hand reject H_A. Now say if had found that $P\left[p_n = 0.20 \middle| \pi = 0.25\right] = 0.005$, then this happening would have been highly unlikely, in which case we would not accept H_0, i.e., accept H_A.

We must remember that in statistical hypothesis testing, a test statistic would be obtained from the set of sample observations, and it signifies the numerical summary of a set of data, i.e., $(X_1, X_2, \ldots, X_n)$, that reduces the data to one or a small number of values that can be used to perform a hypothesis test. So given a null hypothesis, H_0, and a test statistic, t_n, we specify a null value, t_{n,H_0}, which would mean that, that being the value of the test statistic under the condition of H_0 being true. Hence if t_n, is close to t_{n,H_0}, then it would present to us evidence in support of H_0, while on the other hand if t_n is different from t_{n,H_0} then there is obvious reason to not accept H_0. An important property of a test statistic is that one must be able to determine its sampling distribution under H_0 being true.

Let us make this matter a little more explicit with the diagram below



Now in the set of all test statistic, there are certain values marked as A (in this diagram) which are most likely to be chosen for testing H_0, while the other values excluding A are most likely to be chosen to validate H_A. One should remember that a *test criterion* is the set

of statistic which will be used for making the decision. Moreover one will also denote ω as the set of values of statistic or a test criterion which is to be selected for the validation of H_0 and this ω is called the *critical region*. Thus the *critical region* of a hypothesis test is the set of all outcomes which, if they occur, will lead us to decide that there is a difference, i.e., cause the null hypothesis, H_0, to be rejected in favor of the alternative hypothesis, H_A. Hence the general steps for hypothesis testing are

1. Take a sample of certain size, say *n*.
2. Decide on a test criterion.
3. Demarcate the critical region (ω).
4. Compute the value of the test criterion for the given sample.
5. Reject H_0 if the computed value of the test criterion belongs to ω.

Now in all the cases when ever we do a hypotheses testing we have the following situation which is denoted very simply by the following situation which is as follows

ACTION	NATURE		
		H_0: True	H_A: True
	H_0: Reject	Type I error (α)	
	H_A: Reject		Type II error (β)

As the sample does not give the exact information about the population hence one cannot get rid of the errors and one also does not know which one of the outcome will come out to be true after the characteristics of the sample is analyzed. Hence we have

Probability of type I error is the probability that the value of the test criterion (TC) will belong to the critical region, ω, assuming H_0 to be true, while

Probability of type II error is the probability that the value of the test criterion (TC) will not belong to the critical region, ω, assuming H_A to be true

i.e.,

$$P[Value\ of\ TC \in \omega | H_0\ is\ true] = P_{H_0}[Value\ of\ TC \in \omega] = P[Type\ I\ error]$$
$$P[Value\ of\ TC \notin \omega | H_A\ is\ true] = P_{H_A}[Value\ of\ TC \notin \omega] = P[Type\ II\ error]$$

and as these probabilities are called he level of significance, hence we also denote them as $P_{H_0}[Value\ of\ TC \in \omega] = \alpha$ and $P_{H_A}[Value\ of\ TC \notin \omega] = \beta$. In general $(1-\beta) = \gamma$ is also called the power of the test, which means it is the probability of the value of $TC \notin \omega$, given that H_A is not true. So we should find such a value of ω, for which α and β are minimum. But the problem is that trying to minimize both of them simultaneously is not possible, as decreasing α, would lead to the increase in the value of β, and vice versa, hence choosing ω is not easy.

So to sum it up we say

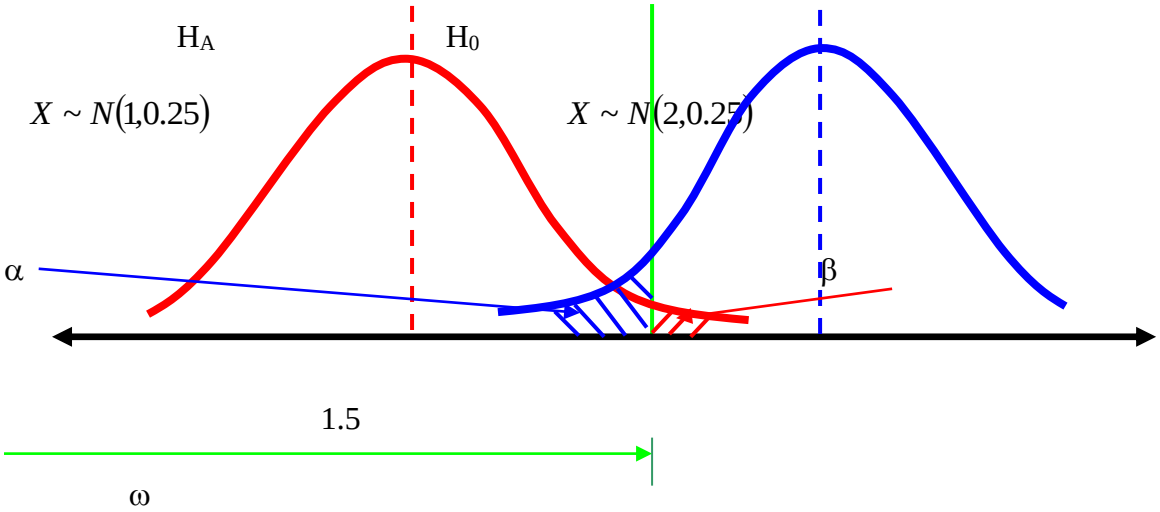
1. α is the probability of rejecting H_0, i.e., accepting H_A pretending that nature says H_0 is true.

2. β is the probability of acceptiong H_0, i.e., rejecting H_A pretending that nature says H_A is true.

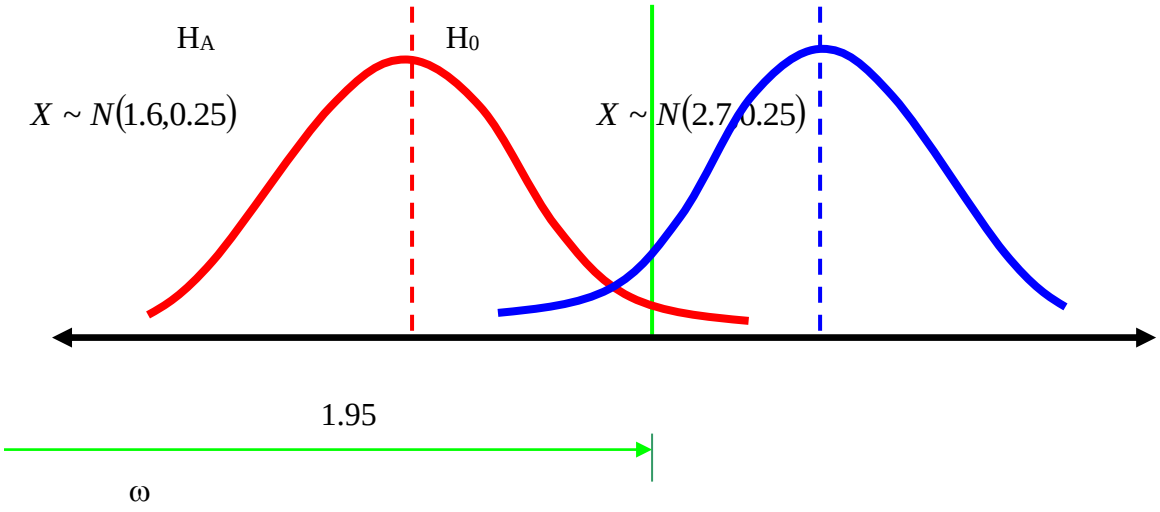
<u>Example</u>

Consider we have $X \sim N(\mu, \sigma^2)$, with μ being unknown while σ^2 is known, i.e., suppoe we have the following whci is $X \sim N(\mu, 0.25)$. Also assume that we have the two competing hypothesis which are: H_0: μ = 2 vs H_A: μ = 1 and we are to test the efficacy of the hypothesis statement. No if we ake a sample of size 10, such that we have $\hat{\mu} = \bar{X}_{10} = \frac{1}{10}(X_1 + X_2 + \ldots + X_{10})$ as we know the sample estimate is the UMVUE of the population mean, i.e., its expected value. Then it is also true that $\bar{X}_{10} \sim N\left(\mu, \frac{0.25}{10}\right)$ and

suppose that after choosing a sample of size 10, we get $\hat{\mu} = \overline{X}_{10} = 1.5$, then in that case we will definitely reject H_0. But remember the value of $\hat{\mu} = \overline{X}_{10} = 1.5$ is just random as taking samples will give us different value of $\hat{\mu} = \overline{X}_{10}$, so in the case when we obtain $\hat{\mu} = \overline{X}_{10} = 2$, we accept H_0.

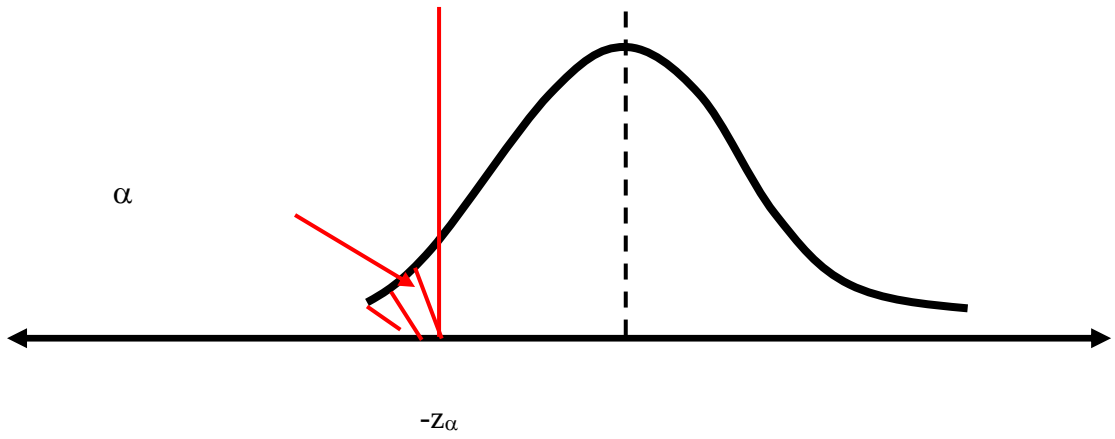


The hypothesis could have been framed like H_0: $\mu > 2$ vs H_A: $\mu \leq 2$, then in that case the decision to accept or not accept H_0 will depend on the value of the sample statistic which is the sample mean we obtain after we randomly select a sample of size, *n*.



Reducing α by choosing ω will increase the value of β, else trying to minimize β by choosing ω will increase the value of α, and this is very clearly demonstrated in the two figures above. Ence simulataneously minimizaing (α + β) for a choosen value of ω, will give us the result.

Thus we first select α whci is always at a very low value, and then try to maximize (1 - β), which is the same thing as minimizing β. This is also know as trying to maximize the power of the test, i.e., γ = (1 - β) and this would ensure minimization of variation. Remember this type of test is the most powerful decision rule for the level of significance which is already fixed by the value of α.



So the rule is reject H_0 if $\bar{X}_n < \mu_0 - z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right)$ holds

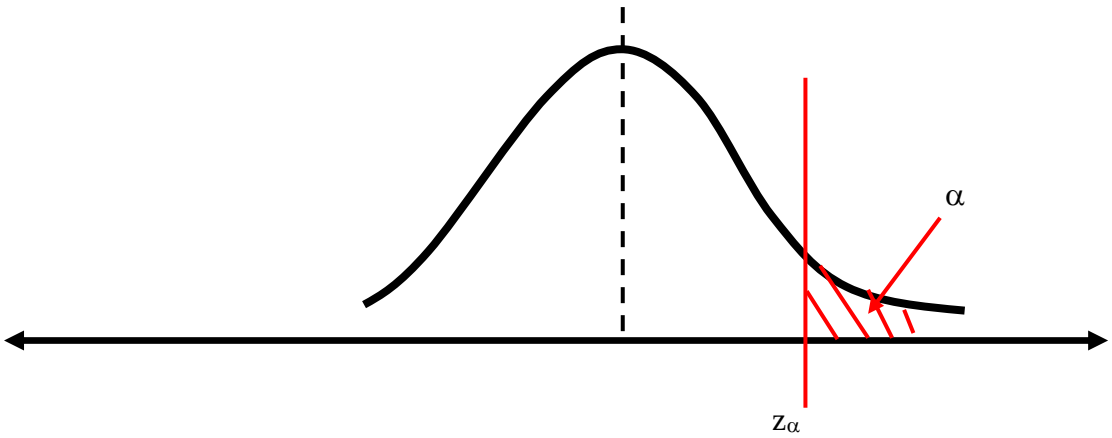
H_0: $\mu = \mu_0$ vs H_A: $\mu = \mu_A$ ($\mu_A > \mu_0$)

As required using Neyman-Pearson Lemma we must have the following true

$$\frac{L_A}{L_0} = \frac{\prod_i^n f(x_i;\theta_A)}{\prod_i^n f(x_i;\theta_0)} = \frac{\prod_i^n \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu_A)^2\right\}}{\prod_i^n \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu_0)^2\right\}} > c$$ such that we one can H_0, based on the fact the above inequality holds. Thus as $\mu_A > \mu_0$ is true we have

$\bar{X}_n > k\left\{= \dfrac{\log_e c + \dfrac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}$ is true, which mean we have

$$P_{H_0}\left[\bar{X}_n > k\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} > \frac{k - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} > z_\alpha\right] = 1 - \Phi\left\{\frac{k - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right\} = \alpha$$



So the rule is reject H_0 if $\bar{X}_n > \mu_0 + z_\alpha\left(\dfrac{\sigma}{\sqrt{n}}\right)$ holds

H_0: $\mu = \mu_0$ vs H_A: $\mu = \mu_A$ ($\mu_A \neq \mu_0$)

Then we see $\bar{X}_n > k\left\{= \dfrac{\log_e c + \dfrac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}$ or $\bar{X}_n < k\left\{= \dfrac{\log_e c + \dfrac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}$ is true,

which mean we have the following as true is both of the above inequalities are to be true,

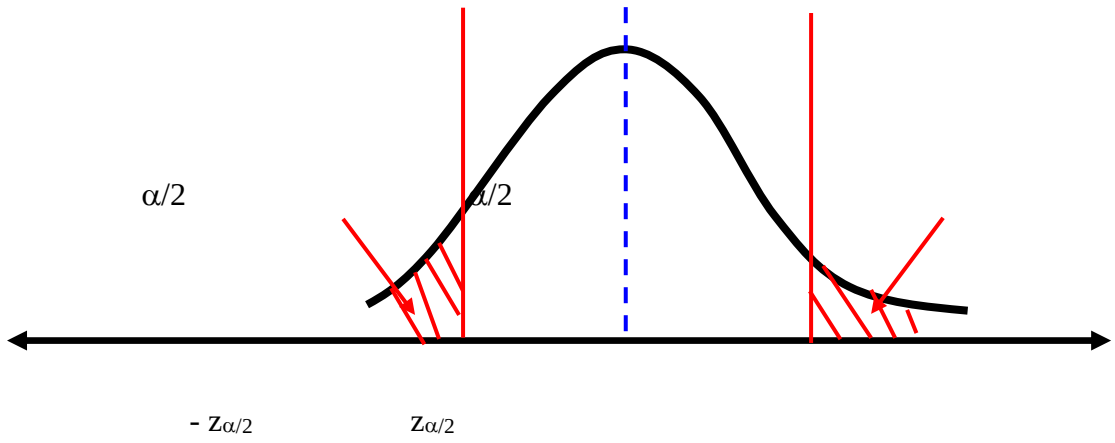
$$P_{H_0}\left[\bar{X}_n < k\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} < z_{1-\frac{\alpha}{2}}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} < -z_{\frac{\alpha}{2}}\right] = \Phi\left\{\frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right\} = \alpha$$

$$P_{H_0}\left[\bar{X}_n > k\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} > \frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} > z_{\frac{\alpha}{2}}\right] = 1 - \Phi\left\{\frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right\} = \alpha$$

i.e.,

$$P_{H_0}\left[\bar{X}_n < k\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} < -z_{\frac{\alpha}{2}}\right] = \Phi\left\{\frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right\} = \alpha$$

$$P_{H_0}\left[\bar{X}_n > k\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} > \frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} > z_{\frac{\alpha}{2}}\right] = 1 - \Phi\left\{\frac{k-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right\} = \alpha$$



So the rule is reject H_0 if $\left|\bar{X}_n - \mu_0\right| > z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ holds, i.e., $\bar{X}_n < \mu_0 - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ or

$\bar{X}_n > \mu_0 + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$ holds

Example

Suppose we draw *n* number of random observations from the normal distribution given by $X \sim N(\mu, \sigma^2)$, $x \in \Re$, $\mu \in \Re$ and $\sigma^2 \in \Re^+$. We also know that we have a set of random observatiosn given by $(X_1, X_2, \ldots, X_n)$. Also assume that μ is unknown (one has to hypothesize on this fact and draw meaningful conclusion about μ, based on the sample chosen) while σ^2 is is ***unknown***. And we are given the following hypotheses to test, which are

$H_0: \mu = \mu_0$ vs $H_A: \mu = \mu_A$ ($\mu_A < \mu_0$)

As done for the first example let us also proceed using the first principle, i.e., the Neyman-Pearson Lemma and arrive at the rule for rejecting H_0. So we have

$L_A = \prod_i^n f(x_i;\theta_A) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu_A)^2\right\}$, and

$L_0 = \prod_i^n f(x_i;\theta_0) = \prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu_0)^2\right\}$, hence using Neyman-Pearson Lemma we reject H_0 iff the following is true, i.e.,

$$\frac{L_A}{L_0} = \frac{\prod_i^n f(x_i;\theta_A)}{\prod_i^n f(x_i;\theta_0)} = \frac{\prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu_A)^2\right\}}{\prod_i^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu_0)^2\right\}} > c$$

$$\therefore \frac{L_A}{L_0} = \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^n (x_i - \mu_A)^2 - \sum_{i=1}^n (x_i - \mu_0)^2\right\}\right] > c$$

$$\therefore \frac{L_A}{L_0} = \exp\left[-\frac{1}{2\sigma^2}\left\{n\mu_A^2 - 2n\mu_A\overline{X}_n + nX_i^2 - n\mu_0^2 + 2n\mu_0\overline{X}_n - nX_i^2\right\}\right] > c$$

$\therefore \frac{L_A}{L_0} = -\frac{1}{2\sigma^2}\left\{n\left(\mu_A^2 - \mu_0^2\right) - 2n\bar{X}_n\left(\mu_A - \mu_0\right)\right\} > \log_e c$, and remember here all the variables values are known or given.

$$\therefore n\bar{X}_n\left(\frac{\mu_A - \mu_0}{\sigma^2}\right) > \log_e c + \frac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right), \text{ i.e., } \therefore \bar{X}_n < k\left\{= \frac{\log_e c + \frac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\frac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}, \text{ i.e.,}$$

we reject H_0 iff $\bar{X}_n < k$, where k is such that the level of significance is prespecified. Now also remember that $P_{H_0}\left[H_A\right] = \alpha$, while $P_{H_A}\left[H_0\right] = \beta$, so we have

$$P_{H_0}\left[H_A\right] = P_{H_0}\left[Type\ I\ error\right] = \alpha, \quad \text{i.e.,} \quad P_{H_0}\left[\bar{X}_n < k\right] = \alpha, \quad \text{i.e.,}$$

$$P_{H_0}\left[\bar{X}_n < k\left\{= \frac{\log_e c + \frac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\frac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}\right] = \alpha, \text{ i.e., } P_{H_0}\left[\bar{X}_n < k\left\{= \frac{\log_e c + \frac{n}{2s_n^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\frac{\mu_A - \mu_0}{s_n^2}\right)}\right\}\right] = \alpha$$

here we replace σ^2 with $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$. We already know the following to be true,

$$\text{i.e., } \frac{\left\{\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right\}}{\sqrt{\frac{\left(\frac{(n-1)s_n^2}{\sigma^2}\right)}{(n-1)}}} = \left\{\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{s_n}\right\} \sim t_{n-1}, \text{ so we obtain } P_{H_0}\left[\bar{X}_n < k\right] = \alpha, \text{ which is}$$

$$P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s_n^2}{n}}} < \frac{k - \mu_0}{\sqrt{\frac{s_n^2}{n}}}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s_n^2}{n}}} < t_{n-1,1-\alpha}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s_n^2}{n}}} < -t_{n-1,\alpha}\right] = \alpha$$

So the rule is reject H_0 if $\bar{X}_n < \mu_0 - t_{n-1,\alpha}\left(\frac{s_n}{\sqrt{n}}\right)$ holds

H_0: $\mu = \mu_0$ vs H_A: $\mu = \mu_A$ ($\mu_A > \mu_0$)

For this rule as usual $P_{H_0}[H_A] = \alpha$, so we have $P_{H_0}[H_A] = P_{H_0}[Type\ I\ error] = \alpha$, i.e.,

$P_{H_0}[\bar{X}_n > k] = \alpha$, i.e., $P_{H_0}\left[\bar{X}_n > k\left\{= \dfrac{\log_e c + \dfrac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}\right] = \alpha$, i.e.,

$P_{H_0}\left[\bar{X}_n > k\left\{= \dfrac{\log_e c + \dfrac{n}{2s_n^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{s_n^2}\right)}\right\}\right] = \alpha$ here as usual we replace σ^2 with

$s_n^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$, so we obtain $P_{H_0}[\bar{X}_n > k] = \alpha$, which is

$$P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s_n^2}{n}}} > \frac{k - \mu_0}{\sqrt{\frac{s_n^2}{n}}}\right] = P_{H_0}\left[\frac{\bar{X}_n - \mu_0}{\sqrt{\frac{s_n^2}{n}}} > t_{n-1,\alpha}\right] = \alpha .$$

So the rule is reject H_0 if $\bar{X}_n > \mu_0 + t_{n-1,\alpha}\left(\dfrac{s_n}{\sqrt{n}}\right)$ holds

H_0: $\mu = \mu_0$ vs H_A: $\mu = \mu_A$ ($\mu_A \neq \mu_0$)

We reject H_0, when $P_{H_0}[H_A] = \alpha$, i.e., $P_{H_0}[H_A] = P_{H_0}[Type\ I\ error] = \alpha$, i.e.,

$P_{H_0}[\bar{X}_n < k] = \alpha$, or $P_{H_0}[\bar{X}_n > k] = \alpha$ is true. This is the same if we say if (i)

$P_{H_0}\left[\bar{X}_n < k\left\{= \dfrac{\log_e c + \dfrac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}\right] = \alpha$, i.e., $P_{H_0}\left[\bar{X}_n < k\left\{= \dfrac{\log_e c + \dfrac{n}{2s_n^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{s_n^2}\right)}\right\}\right] = \alpha$

or (ii) $P_{H_0}\left[\bar{X}_n > k\left\{= \dfrac{\log_e c + \dfrac{n}{2\sigma^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{\sigma^2}\right)}\right\}\right] = \alpha$, i.e.,

$P_{H_0}\left[\bar{X}_n > k\left\{= \dfrac{\log_e c + \dfrac{n}{2s_n^2}\left(\mu_A^2 - \mu_0^2\right)}{n\left(\dfrac{\mu_A - \mu_0}{s_n^2}\right)}\right\}\right] = \alpha$ is true, whci is possible as we replace σ^2 with

$s_n^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$. So we would reject H_0 if we obtain either (i) $P_{H_0}\left[\bar{X}_n < k\right] = \alpha$, which

is $P_{H_0}\left[\dfrac{\bar{X}_n - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}} < \dfrac{k - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}}\right] = P_{H_0}\left[\dfrac{\bar{X}_n - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}} < t_{n-1,1-\frac{\alpha}{2}}\right] = P_{H_0}\left[\dfrac{\bar{X}_n - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}} < -t_{n-1,\frac{\alpha}{2}}\right] = \alpha$ or (ii)

$P_{H_0}\left[\bar{X}_n > k\right] = \alpha$, i.e., $P_{H_0}\left[\dfrac{\bar{X}_n - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}} > \dfrac{k - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}}\right] = P_{H_0}\left[\dfrac{\bar{X}_n - \mu_0}{\sqrt{\dfrac{s_n^2}{n}}} > t_{n-1,\frac{\alpha}{2}}\right] = \alpha$

So the rule is reject H_0 if $\left|\bar{X}_n - \mu_0\right| > t_{n-1,\frac{\alpha}{2}}\left(\dfrac{s_n}{\sqrt{n}}\right)$ holds. i.e., $\bar{X}_n < \mu_0 - t_{n-1,\frac{\alpha}{2}}\left(\dfrac{s_n}{\sqrt{n}}\right)$ or

$\bar{X}_n > \mu_0 + t_{n-1,\frac{\alpha}{2}}\left(\dfrac{s_n}{\sqrt{n}}\right)$ holds

Example

Suppose we draw n number of random observations from the normal distribution given by $X \sim N\left(\mu, \sigma^2\right)$, $x \in \Re$, $\mu \in \Re$ and $\sigma^2 \in \Re^+$. We also know that we have a set of random observatiosn given by $\left(X_1, X_2, \ldots, X_n\right)$. Also assume that μ is ***known*** while σ^2 is unknown (one has to hypothesize on this fact and draw meaningful conclusion about σ^2, based on the sample chosen), and we are given the following hypotheses to test, which are

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma = \sigma_A$ ($\sigma_A < \sigma_0$), i.e.,

H_0: $\sigma^2 - \sigma^2_0 = 0$ vs H_A: $\sigma^2 - \sigma^2_A < 0$ ($\sigma_A < \sigma_0$), i.e.,

H_0: $\frac{\sigma^2}{\sigma_0^2} = 1$ vs H_A: $\frac{\sigma_A^2}{\sigma^2} < 1$ ($\sigma_A < \sigma_0$), i.e.,

Thus ideally we should look at $\sigma - \sigma_0 = 0$, i.e., $\sigma^2 - \sigma^2_0 = 0$ under H_0 versus $\sigma - \sigma_A < 0$, i.e., $\sigma^2 - \sigma^2_A < 0$ ($\sigma_A < \sigma_0$) under H_A. So using Neyman-Pearson Lemma we have to reject H_0 if

$$P_{H_0}\left[s_n^{*2} - \sigma_0^2 < k\right] = \alpha \text{, i.e., } P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} < n\left\{\frac{\left(k + \sigma_0^2\right)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} < \chi^2_{n,1-\alpha}\right] = \alpha$$

So the rule is reject H_0 if $s_n^{*2} < \frac{\sigma_0^2 \chi^2_{n,1-\alpha}}{n}$ holds

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma = \sigma_A$ ($\sigma_A > \sigma_0$), i.e.,

H_0: $\sigma^2 - \sigma^2_0 = 0$ vs H_A: $\sigma^2 - \sigma^2_A > 0$ ($\sigma_A > \sigma_0$), i.e.,

H_0: $\frac{\sigma^2}{\sigma_0^2} = 1$ vs H_A: $\frac{\sigma_A^2}{\sigma^2} > 1$ ($\sigma_A > \sigma_0$), i.e.,

Thus ideally we should look at $\sigma - \sigma_0 = 0$, i.e., $\sigma^2 - \sigma^2_0 = 0$ under H_0 versus $\sigma - \sigma_A > 0$, i.e., $\sigma^2 - \sigma^2_A > 0$ ($\sigma_A > \sigma_0$) under H_A. So using Neyman-Pearson Lemma we have to reject H_0 if

$$P_{H_0}\left[s_n^{*2} - \sigma_0^2 > k\right] = \alpha \text{, i.e., } P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} > n\left\{\frac{\left(k + \sigma_0^2\right)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} > \chi^2_{n,\alpha}\right] = \alpha$$

So the rule is reject H_0 if $s_n^{*2} > \frac{\sigma_0^2 \chi^2_{n,\alpha}}{n}$ holds

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma \neq \sigma_A$ ($\sigma_A \neq \sigma_0$), i.e.,

H_0: $\sigma^2 - \sigma^2_0 = 0$ vs H_A: $\sigma^2 - \sigma^2_A \neq 0$ ($\sigma_A \neq \sigma_0$), i.e.,

H_0: $\frac{\sigma^2}{\sigma_0^2} = 1$ vs H_A: $\frac{\sigma^2}{\sigma_A^2} \neq 1$ *or* $\frac{\sigma_A^2}{\sigma^2} \neq 1$ ($\sigma_A \neq \sigma_0$)), i.e.,

Thus ideally we should look at $\sigma - \sigma_0 = 0$, i.e., $\sigma^2 - \sigma^2_0 = 0$ under H_0 versus $\sigma - \sigma_A \neq 0$, i.e., $\sigma^2 - \sigma^2_A \neq 0$ ($\sigma_A \neq \sigma_0$) under H_A. So using Neyman-Pearson Lemma we have to reject H_0 if $P_{H_0}\left[s_n^{*2} - \sigma_0^2 < k\right] = \alpha$ or $P_{H_0}\left[s_n^{*2} - \sigma_0^2 > k\right] = \alpha$ holds, i.e.,

$$P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} < n\left\{\frac{\left(k + \sigma_0^2\right)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} < \chi^2_{n,1-\frac{\alpha}{2}}\right] = \alpha \quad \text{or}$$

$$P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} > n\left\{\frac{\left(k + \sigma_0^2\right)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{ns_n^{*2}}{\sigma_0^2} > \chi^2_{n,\frac{\alpha}{2}}\right] = \alpha \text{ holds}$$

So the rule is reject H_0 if $s_n^{*2} < \frac{\sigma_0^2 \chi^2_{n,1-\frac{\alpha}{2}}}{n}$ or $s_n^{*2} > \frac{\sigma_0^2 \chi^2_{n,\frac{\alpha}{2}}}{n}$ holds

Example

Suppose we draw n number of random observations from the normal distribution given by $X \sim N\left(\mu, \sigma^2\right)$, $x \in \Re$, $\mu \in \Re$ and $\sigma^2 \in \Re^+$. We also know that we have a set of random observatiosn given by $\left(X_1, X_2, \ldots, X_n\right)$. Also assume that μ is ***unknown*** while σ^2 is unknown (one has to hypothesize on this fact and draw meaningful conclusion about σ^2, based on the sample chosen), and we are given the following hypotheses to test, which are

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma = \sigma_A$ ($\sigma_A < \sigma_0$), i.e.,

H_0: $\sigma^2 - \sigma^2_0 = 0$ vs H_A: $\sigma^2 - \sigma^2_A < 0$ ($\sigma_A < \sigma_0$), i.e.,

H_0: $\frac{\sigma^2}{\sigma_0^2} = 1$ vs H_A: $\frac{\sigma_A^2}{\sigma^2} < 1$ ($\sigma_A < \sigma_0$), i.e.,

Thus ideally we should look at $\sigma - \sigma_0 = 0$, i.e., $\sigma^2 - \sigma^2_0 = 0$ under H_0 versus $\sigma - \sigma_A < 0$, i.e., $\sigma^2 - \sigma^2_A < 0$ ($\sigma_A < \sigma_0$) under H_A. So using Neyman-Pearson Lemma we have to reject H_0 if

$$P_{H_0}\left[s_n^2 - \sigma_0^2 < k\right] = \alpha,$$ i.e.,

$$P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} < (n-1)\left\{\frac{\left(k+\sigma_0^2\right)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha}\right] = \alpha$$

So the rule is reject H_0 if $s_n^2 < \left\{\frac{\sigma_0^2 \chi^2_{n-1,1-\alpha}}{(n-1)}\right\}$ holds

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma = \sigma_A$ ($\sigma_A > \sigma_0$), i.e.,

H_0: $\sigma^2 - \sigma^2_0 = 0$ vs H_A: $\sigma^2 - \sigma^2_A > 0$ ($\sigma_A > \sigma_0$), i.e.,

H_0: $\frac{\sigma^2}{\sigma_0^2} = 1$ vs H_A: $\frac{\sigma_A^2}{\sigma^2} > 1$ ($\sigma_A > \sigma_0$), i.e.,

Thus ideally we should look at $\sigma - \sigma_0 = 0$, i.e., $\sigma^2 - \sigma^2_0 = 0$ under H_0 versus $\sigma - \sigma_A > 0$, i.e., $\sigma^2 - \sigma^2_A > 0$ ($\sigma_A > \sigma_0$) under H_A. So using Neyman-Pearson Lemma we have to reject H_0 if

$$P_{H_0}\left[s_n^2 - \sigma_0^2 > k\right] = \alpha,$$ i.e.,

$$P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} > (n-1)\left\{\frac{\left(k+\sigma_0^2\right)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} > \chi^2_{n-1,\alpha}\right] = \alpha$$

So the rule is reject H_0 if $s_n^2 > \left\{\frac{\sigma_0^2 \chi^2_{n-1,\alpha}}{(n-1)}\right\}$ holds

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma \neq \sigma_A$ ($\sigma_A \neq \sigma_0$), i.e.,

H_0: $\sigma^2 - \sigma^2_0 = 0$ vs H_A: $\sigma^2 - \sigma^2_A \neq 0$ ($\sigma_A \neq \sigma_0$), i.e.,

$H_0: \frac{\sigma^2}{\sigma_0^2} = 1$ vs $H_A: \frac{\sigma^2}{\sigma_A^2} \neq 1$ *or* $\frac{\sigma_A^2}{\sigma^2} \neq 1$ ($\sigma_A \neq \sigma_0$), i.e.,

Thus ideally we should look at $\sigma - \sigma_0 = 0$, i.e., $\sigma^2 - \sigma^2_0 \neq 0$ under H_0 versus $\sigma - \sigma_A > 0$, i.e., $\sigma^2 - \sigma^2_A \neq 0$ ($\sigma_A \neq \sigma_0$) under H_A. So using Neyman-Pearson Lemma we have to reject H_0 if $P_{H_0}\left[s_n^2 - \sigma_0^2 < k\right] = \alpha$ or $P_{H_0}\left[s_n^2 - \sigma_0^2 > k\right] = \alpha$ holds , i.e.,

$$P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} < (n-1)\left\{\frac{(k+\sigma_0^2)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right] = \alpha$$ or

$$P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} > (n-1)\left\{\frac{(k+\sigma_0^2)}{\sigma_0^2}\right\}\right] = P_{H_0}\left[\frac{(n-1)s_n^2}{\sigma_0^2} > \chi^2_{n-1,\frac{\alpha}{2}}\right] = \alpha$$ is true

So the rule is reject H_0 if $s_n^2 < \left\{\frac{\sigma_0^2 \chi^2_{n-1,1-\frac{\alpha}{2}}}{(n-1)}\right\}$ or $s_n^{*2} > \left\{\frac{\sigma_0^2 \chi^2_{n-1,\frac{\alpha}{2}}}{(n-1)}\right\}$ holds

Example

Suppose we draw *m* and *n* number of random observations from two normal distributions given by $X \sim N(\mu_X, \sigma_X^2)$, where $x \in \Re$, $\mu_X \in \Re$, $\sigma_X^2 \in \Re^+$ hold and $Y \sim N(\mu_Y, \sigma_Y^2)$, where $y \in \Re$, $\mu_Y \in \Re$, $\sigma_Y^2 \in \Re^+$. We also know that we have a set of random observatiosn given by $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$. Also assume that μ_X and μ_Y are unknown (one has to hypothesize on this fact and draw meaningful conclusion about μ_X and μ_Y, based on the samples chosen), while σ_X^2 and σ_Y^2 are ***known***, and we are given the following hypotheses to test, which are

$H_0: \mu_X - \mu_Y = 0$ vs $H_A: \mu_X - \mu_Y < 0$

So the rule is reject H_0 if $\left(\bar{X}_m - \bar{Y}\right)_n < \left(\mu_X - \mu_Y\right) - z_\alpha\left(\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}\right)$ holds

H_0: $\mu_X - \mu_Y = 0$ vs H_A: $\mu_X - \mu_Y > 0$

So the rule is reject H_0 if $\left(\bar{X}_m - \bar{Y}\right)_n > \left(\mu_X - \mu_Y\right) + z_\alpha\left(\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}\right)$ holds

H_0: $\mu_X - \mu_Y = 0$ vs H_A: $\mu_X - \mu_Y \neq 0$

So the rule is reject H_0 if $\left(\bar{X}_m - \bar{Y}\right)_n < \left(\mu_X - \mu_Y\right) - z_{\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}\right)$ or

$\left(\bar{X}_m - \bar{Y}\right)_n > \left(\mu_X - \mu_Y\right) + z_{\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}\right)$ holds

Example

Suppose we draw m and n number of random observations from two normal distributions given by $X \sim N\left(\mu_X, \sigma_X^2\right)$, where $x \in \Re$, $\mu_X \in \Re$, $\sigma_X^2 \in \Re^+$ hold and $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$, where $y \in \Re$, $\mu_Y \in \Re$, $\sigma_Y^2 \in \Re^+$. We also know that we have a set of random observatiosn given by $\left(X_1, X_2, \ldots, X_m\right)$ and $\left(Y_1, Y_2, \ldots, Y_n\right)$. Also assume that μ_X and μ_Y are ***known*** while σ_X^2 and σ_Y^2 are unknown (one has to hypothesize on this fact and draw meaningful conclusion about σ_X^2 and σ_Y^2, based on the samples chosen), and we are given the following hypotheses to test, which are

H_0: $\sigma = \sigma_0$ vs H_A: $\sigma = \sigma_A$ ($\sigma_A < \sigma_0$), i.e.,

$H_0: \sigma^2 - \sigma^2_0 = 0$ vs $H_A: \sigma^2 - \sigma^2_A < 0$ ($\sigma_A < \sigma_0$)

So the rule is reject H_0 if $\frac{s^{*2}_m}{s^{*2}_n} < F_{m,n,1-\alpha}$ holds

$H_0: \sigma = \sigma_0$ vs $H_A: \sigma = \sigma_A$ ($\sigma_A > \sigma_0$), i.e.,

$H_0: \sigma^2 - \sigma^2_0 = 0$ vs $H_A: \sigma^2 - \sigma^2_A > 0$ ($\sigma_A > \sigma_0$)

So the rule is reject H_0 if $\frac{s^{*2}_m}{s^{*2}_n} > F_{m,n,\alpha}$ holds

$H_0: \sigma = \sigma_0$ vs $H_A: \sigma = \sigma_A$ ($\sigma_A \neq \sigma_0$), i.e.,

$H_0: \sigma^2 - \sigma^2_0 = 0$ vs $H_A: \sigma^2 - \sigma^2_A \neq 0$ ($\sigma_A \neq \sigma_0$)

So the rule is reject H_0 if $\frac{s^{*2}_m}{s^{*2}_n} < F_{m,n,1-\frac{\alpha}{2}}$ or $\frac{s^{*2}_m}{s^{*2}_n} > F_{m,n,\frac{\alpha}{2}}$ holds

Example

Suppose we draw m and n number of random observations from two normal distributions given by $X \sim N(\mu_X, \sigma^2_X)$, where $x \in \Re$, $\mu_X \in \Re$, $\sigma^2_X \in \Re^+$ hold and $Y \sim N(\mu_Y, \sigma^2_Y)$, where $y \in \Re$, $\mu_Y \in \Re$, $\sigma^2_Y \in \Re^+$. We also know that we have a set of random observatiosn given by $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$. Also assume that μ_X and μ_Y are ***unknown*** while σ^2_X and σ^2_Y are unknown (one has to hypothesize on this fact and draw meaningful conclusion about σ^2_X and σ^2_Y, based on the samples chosen), and we are given the following hypotheses to test, which are

Example

Suppose we draw m and n number of random observations from two normal distributions given by $X \sim N\left(\mu_X, \sigma_X^2\right)$, where $x \in \Re$, $\mu_X \in \Re$, $\sigma_X^2 \in \Re^+$ hold and $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$, where $y \in \Re$, $\mu_Y \in \Re$, $\sigma_Y^2 \in \Re^+$. We also know that we have a set of random observatiosn given by $\left(X_1, X_2, \ldots, X_m\right)$ and $\left(Y_1, Y_2, \ldots, Y_n\right)$. Also assume that μ_X and μ_Y are unknown while σ_X^2 and σ_Y^2 are ***unknown (but both are equal)***, and we are given the following hypotheses to test, which are

H_0: $\mu_X - \mu_Y = 0$ vs H_A: $\mu_X - \mu_Y < 0$

So the rule is reject H_0 if $\left(\bar{X}_m - \bar{Y}_n\right) < \left(\mu_1 - \mu_2\right) - t_{m+n-2,\alpha} \times \sqrt{\frac{s_{m,1}^2}{m} + \frac{s_{n,1}^2}{n}}$

H_0: $\mu_X - \mu_Y = 0$ vs H_A: $\mu_X - \mu_Y > 0$

So the rule is reject H_0 if $\left(\bar{X}_m - \bar{Y}_n\right) > \left(\mu_1 - \mu_2\right) + t_{m+n-2,\alpha} \times \sqrt{\frac{s_{m,1}^2}{m} + \frac{s_{n,1}^2}{n}}$

H_0: $\mu_X - \mu_Y = 0$ vs H_A: $\mu_X - \mu_Y \neq 0$

So the rule is reject H_0 if $\left(\bar{X}_m - \bar{Y}_n\right) < \left(\mu_1 - \mu_2\right) - t_{m+n-2,\frac{\alpha}{2}} \times \sqrt{\frac{s_{m,1}^2}{m} + \frac{s_{n,1}^2}{n}}$ or

$\left(\bar{X}_m - \bar{Y}_n\right) > \left(\mu_1 - \mu_2\right) + t_{m+n-2,\frac{\alpha}{2}} \times \sqrt{\frac{s_{m,1}^2}{m} + \frac{s_{n,1}^2}{n}}$ holds