Interval Estimation

Let, $X_1, X_2, \cdots, X_n$ be 'n' observable random variables whose joint distribution function 'F' is known except for some unknown parameters (or parameters) denoted by $\underset{\sim}{\theta}$ & varying over a parametric space Θ.

Suppose, our problem is to estimate the parametric function $\gamma(\theta)$ by means of an interval, say, $[t_1, t_2]$, where, t_1 & t_2 are the observed values of two statistics, say, T_1 & T_2 $(T_1 \le T_2)$.

If the statistics T_1 & T_2 are so chosen that

$P_\theta\{T_1 \le \gamma(\theta) \le T_2\} = 1-\alpha \cdots$ ①, then, statement ① means that if repeated observations are taken on $X_1, X_2, \cdots, X_n$ & if for each set of observations, $x_1, x_2, \cdots, x_n$, the values of T_1 & T_2 are computed (denoted by t_1 & t_2), then, in the long run, the interval between t_1 & t_2 will contain $\gamma(\theta)$ in $100(1-\alpha)\%$ of the cases & will fail to do so in the remaining $100\alpha\%$ of the cases.

As such, if $(1-\alpha)$ is a high value (ie a value close to 1), then the one can claim ~~that~~ with a good deal of confidence that the interval $[t_1, t_2]$ will include $\gamma(\theta)$, ie claim that $t_1 \le \gamma(\theta) \le t_2$.

The interval is called a confidence interval and t_1 & t_2 are called the lower & the upper confidence limits respectively.

The value $(1-\alpha)$ serves as a measure of the confidence interval with which one claims that $t_1 \le \gamma(\theta) \le t_2$ and is called the confidence coefficient associated with the interval.

(1) Confidence limits for mean of a Univariate normal population

Let, $x_1, x_2, \cdots, x_n$ be a random sample drawn from an $N(\mu, \sigma^2)$ distribution. Here, $\theta = \mu$.

Case → (1): σ^2 is known

We know that, $\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \sim N(0,1)$.

$\therefore P_\theta\left\{\left|\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\right| \le \tau_{\alpha/2}\right\} = 1-\alpha$, $\tau_{\alpha/2}$ being the upper $\frac{\alpha}{2}$ point of a standard normal distribution.

ie, $P_\theta\left\{-\tau_{\alpha/2} \le \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \le \tau_{\alpha/2}\right\} = 1-\alpha$.

ie, $P_\theta\left\{-\bar{x} - \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2} \le -\mu \le -\bar{x} + \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2}\right\} = 1-\alpha$.

ie, $P_\theta\left\{\bar{x} - \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2} \le \mu \le \bar{x} + \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2}\right\} = 1-\alpha$, where α is the desired level of significance.

~~wh~~

Thus, $\left(\bar{x} - \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2}\right)$ and $\left(\bar{x} + \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2}\right)$ are, respectively, the lower & the upper confidence limits to μ with confidence coefficient $(1-\alpha)$.

& $\left[\bar{x} - \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}}\tau_{\alpha/2}\right]$ will be the $100(1-\alpha)\%$ confidence interval.

Case → (2) : σ^2 is unknown

Let us write $s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

In this case, we know that $\frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t_{n-1}$.

$\therefore P_\theta\left\{\left|\frac{\sqrt{n}(\bar{x}-\mu)}{s}\right| \leq t_{\frac{\alpha}{2}, n-1}\right\} = 1-\alpha$, $t_{\alpha/2, n-1}$ being the upper $\frac{\alpha}{2}$ point of a 'T' dist^n with $(n-1)$ d.f.

ie, $P_\theta\left\{-t_{\alpha/2, n-1} \leq \frac{\sqrt{n}(\bar{x}-\mu)}{s} \leq t_{\alpha/2, n-1}\right\} = 1-\alpha$.

ie, $P_\theta\left\{-\bar{x} - t_{\alpha/2, n-1}\frac{s}{\sqrt{n}} \leq -\mu \leq \bar{x} - t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}\right\} = 1-\alpha$.

ie, $P_\theta\left\{\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right\} = 1-\alpha$, where α's the desired level of significance.

Thus, $\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right)$ & $\left(\bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right)$ are respectively the lower & the upper confidence limits to μ with confidence coefficient $(1-\alpha)$.

Also, $\left[\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}, \ \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right]$ will be the $100(1-\alpha)\%$ confidence interval.

② Confidence interval for the variance of a normal population

Let, $x_1, x_2, \ldots, x_n$ be a random sample drawn from an $N(\mu, \sigma^2)$ population. Let us write $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Here, $\theta = \sigma^2$.

Case ①: μ is Known

In this case, we know that $\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 = \frac{n s_0^2}{\sigma^2} \sim \chi^2_n$.

$\therefore P_\theta\left\{ \chi^2_{1-\alpha/2, n} \leq \frac{n s_0^2}{\sigma^2} \leq \chi^2_{\alpha/2, n} \right\} = 1-\alpha$, where, $\chi^2_{1-\alpha/2, n}$ & $\chi^2_{\alpha/2, n}$ are respectively the lower & upper $\frac{\alpha}{2}$-point of a χ^2_n dist^n.

ie, $P_\theta\left\{ \frac{1}{n s_0^2}\chi^2_{1-\alpha/2, n} \leq \frac{1}{\sigma^2} \leq \frac{1}{n s_0^2}\chi^2_{\alpha/2, n} \right\} = 1-\alpha.$

ie, $P_\theta\left\{ \frac{n s_0^2}{\chi^2_{\alpha/2, n}} \leq \sigma^2 \leq \frac{n s_0^2}{\chi^2_{1-\alpha/2, n}} \right\} = 1-\alpha$, where α is the desired level of significance.

Thus, $\frac{n s_0^2}{\chi^2_{\alpha/2, n}}$ & $\frac{n s_0^2}{\chi^2_{1-\alpha/2, n}}$ are, respectively the lower & the upper confidence limits to σ^2 with confidence coefficient $(1-\alpha)$. Also, $\left(\frac{n s_0^2}{\chi^2_{\alpha/2, n}}, \frac{n s_0^2}{\chi^2_{1-\alpha/2, n}}\right)$ is the $100(1-\alpha)\%$ confidence interval.

coefficient $(1-\alpha)$. $\chi^2_{\alpha/2,n}$ $\chi^2_{1-\alpha/2,n}$

Case → 2 : μ is unknown

Let us write, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

In this case, we know that $\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$.

$$\therefore P_\theta\left\{\chi^2_{1-\frac{\alpha}{2},n-1} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha.$$

i.e., $P_\theta\left\{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right\} = 1-\alpha$, where '$\alpha$' is the desired level of significance.

Thus, $\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}$ & $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$ are, respectively, the lower & the upper confidence limits to σ^2 with confidence coefficient $(1-\alpha)$. Also $\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$ is the $100(1-\alpha)\%$ confidence interval.

(3) Confidence interval for the means of two univariate Normal popl^{n}s.

Suppose, we have two populations characterised by $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distributions respectively.

Let us draw two random samples of sizes n_1 and n_2 independently from the popl^n. Let, $x_{11}, x_{12}, \ldots, x_{1n_1}$ be the observations of the 1st sample & $x_{21}, x_{22}, \ldots, x_{2n_2}$ be the observations of the 2nd sample.

Let us write, $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, $\forall i=1,2$; &

$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$, $i=1,2$; Here, $\theta = (\mu_1 - \mu_2)$.

Case: I: σ_1 & σ_2 are known.

In this case, we know that $\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$.

$\therefore P_\theta\left\{\left|\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right| \le \tau_{\alpha/2}\right\} = 1 - \alpha$, $\tau_{\alpha/2}$ being the upper $\frac{\alpha}{2}$ point of a $N(0,1)$ dist^n & 'α' is the desired level of significance.

$\Rightarrow P_\theta\left\{-(\bar{x}_1 - \bar{x}_2) - \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\,\tau_{\alpha/2} \le -(\mu_1 - \mu_2) \le -(\bar{x}_1 - \bar{x}_2) + \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\,\tau_{\alpha/2}\right\} = 1 - \alpha$.

$\Rightarrow P_\theta\left\{(\bar{x}_1 - \bar{x}_2) - \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\,\tau_{\alpha/2} \le (\mu_1 - \mu_2) \le (\bar{x}_1 - \bar{x}_2) + \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\,\tau_{\alpha/2}\right\} = 1 - \alpha$

Then, $\left\{(\bar{x}_1 - \bar{x}_2) - \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right\}$ & $\left\{(\bar{x}_1 - \bar{x}_2) + \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right\}$ are respectively the lower & the upper confidence limits to $(\mu_1 - \mu_2)$ with confidence coefficient $(1-\alpha)$. Also $\left\{(\bar{x}_1 - \bar{x}_2) - \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\,\tau_{\alpha/2}, (\bar{x}_1 - \bar{x}_2) + \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_1}\,\tau_{\alpha/2}\right\}$ is the $100(1-\alpha)\%$ confidence interval.

known

Case → 2 : σ_1 & σ_2 are unknown

In this case, since, σ_1^2 & σ_2^2 are unknown, we replace them by their unbiased estimators s_1^2 & s_2^2 respectively. But,

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s_1^2/n + s_2^2/n_2}}$$ does not have a simple distribution.

In order to avoid complications, we shall assume that σ_1^2 and σ_2^2, although unknown individually, are known to be equal.

We, thus, make the so called homo-scedasticity assumption:

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ (say). We now have,

$V(\bar{x}_1 - \bar{x}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$. Further we replace σ^2 by it's pooled estimator

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}{n_1+n_2-2}$$

Now, we know that $\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

$\therefore P_\theta\left\{\left|\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| \le t_{\alpha/2, n_1+n_2-2}\right\} = 1-\alpha$, where, $t_{\alpha/2, n_1+n_2-2}$ is the upper $\frac{\alpha}{2}$-point of a t dist^n with (n_1+n_2-2) d.f. & 'α' is the desired level of significance.

$$\Rightarrow P_\theta\left\{-t_{\alpha/2, n_1+n_2-2} \le \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \le t_{\alpha/2, n_1+n_2-2}\right\} = 1-\alpha.$$

$$\Rightarrow P_\theta\left\{-(\bar{x}_1 - \bar{x}_2) - s\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}\, t_{\alpha/2, n_1+n_2-2} \le -(\mu_1 - \mu_2) \le -(\bar{x}_1 - \bar{x}_2) + s\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}\, t_{\alpha/2, n_1+n_2-2}\right\} = 1-\alpha.$$

$$\Rightarrow P_\theta\left\{(\bar{x}_1 - \bar{x}_2) - s\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}\, t_{\alpha/2, n_1+n_2-2} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + s\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}\, t_{\alpha/2, n_1+n_2-2}\right\} = 1-\alpha$$

Thus, $\left\{(\bar{x}_1 - \bar{x}_2) - s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2, n_1+n_2-2}\right\}$ & $\left\{(\bar{x}_1 - \bar{x}_2) + s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2, n_1+n_2-2}\right\}$ are, respectively the ~~upper~~ lower & the ~~lower~~ upper confidence limits to $(\mu_1 - \mu_2)$ with confidence coefficient $(1-\alpha)$.

Also, $\left\{(\bar{x}_1 - \bar{x}_2) - s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right\}$ is the $100(1-\alpha)\%$ confidence interval.

4) Confidence interval for variances of two univariate Normal populations

Suppose, we have two populations characterised by $N(\mu_1, \sigma_1^2)$ & $N(\mu_2, \sigma_2^2)$ distns respectively. Let us draw two random samples of sizes n_1 & n_2 independently from the popln.

Let, $x_{11}, x_{12}, \ldots, x_{1n_1}$ be the observations of the 1^{st} sample & $x_{21}, x_{22}, \ldots, x_{2n_2}$ be the observations of the 2^{nd} sample. Let us write

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad i=1,2. \; ; \quad s_{i0}^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2, \; i=1,2. \; [\mu_i \text{ is known}]$$

$$s_{i0}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \; i=1,2. \; [\mu_i \text{ is unknown}].$$

Case → 1 : μ_1 & μ_2 are known

We know that $\frac{s_{10}^2}{s_{20}^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1,n_2}$

$\therefore$ $P_\theta\left\{F_{1-\frac{\alpha}{2};n_1,n_2} \le \frac{s_{10}^2}{s_{20}^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} \le F_{\frac{\alpha}{2};n_1,n_2}\right\} = 1-\alpha$, where $F_{\frac{\alpha}{2};n_1,n_2}$ is the upper $\frac{\alpha}{2}$- point of a F dist^n with parameters n_1 & n_2.

$\Rightarrow P_\theta\left\{\frac{s_{20}^2}{s_{10}^2}\frac{1}{F_{\frac{\alpha}{2};n_1,n_2}} \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{s_{20}^2}{s_{10}^2}F_{\frac{\alpha}{2};n_1,n_2}\right\} = 1-\alpha.$

$\Rightarrow P_\theta\left\{\frac{s_{10}^2}{s_{20}^2}\cdot\frac{1}{F_{\alpha/2;n_1,n_2}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_{10}^2}{s_{20}^2}\cdot F_{\alpha/2;n_1,n_2}\right\} = 1-\alpha.$

Thus, $\left\{\frac{s_{10}^2}{s_{20}^2}\cdot\frac{1}{F_{\alpha/2;n_1,n_2}}\right\}$ & $\left\{\frac{s_{10}^2}{s_{20}^2}\cdot F_{\alpha/2;n_1,n_2}\right\}$ are respectively the lower & upper confidence limits to $\frac{\sigma_1^2}{\sigma_2^2}$ with confidence coefficient $(1-\alpha)$.

Also, $\left\{\frac{s_{10}^2}{s_{20}^2}\cdot\frac{1}{F_{\alpha/2;n_1,n_2}}, \frac{s_{10}^2}{s_{20}^2}\cdot F_{\alpha/2;n_1,n_2}\right\}$ is the $100(1-\alpha)\%$ confidence interval.

Case → 2 : μ_1 & μ_2 are unknown

s_{20} ... $F_{\alpha/2, n_1, n_2}$ ... s_{20}^2 ... $\alpha/2, n_1, n_2\}$ is the ... confidence interval.

Case → 2: μ_1 & μ_2 are unknown.

we know that $\frac{s_1^2}{s_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1, n_2-1}$.

$\therefore P_\theta\left\{\frac{1}{F_{\alpha/2, n_1-1, n_2-1}} \leq \frac{s_1^2}{s_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq F_{\alpha/2, n_1-1, n_2-1}\right\} = 1-\alpha$, where, $F_{\alpha/2, n_1, n_2}$ &

$\frac{1}{F_{\alpha/2, n_1, n_2}}$ are respectively the upper & lower $\frac{\alpha}{2}$ point of a F dist^n with parameters (n_1-1) & (n_2-1).

$\Rightarrow P_\theta\left\{\frac{s_2^2}{s_1^2} \cdot \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{s_2^2}{s_1^2} \cdot F_{\alpha/2, n_1-1, n_2-1}\right\} = 1-\alpha$.

$\Rightarrow P_\theta\left\{\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2, n_1-1, n_2-1}\right\} = 1-\alpha$.

Thus, $\left\{\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}}\right\}$ & $\left\{\frac{s_1^2}{s_2^2} \cdot F_{\alpha/2, n_1-1, n_2-1}\right\}$ are respectively the lower & upper confidence limits to $\frac{\sigma_1^2}{\sigma_2^2}$ with confidence coefficient $(1-\alpha)$. Also $\left(\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\alpha/2, n_1-1, n_2-1}}, \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2, n_1-1, n_2-1}\right)$ is the $100(1-\alpha)\%$ confidence interval.