Interval Estimation

In previous few lectures we estimated unknown parameters using Point Estimators. In Point estimation we provide a single value to the unknown parameter which we expect to be close to the parameter. Sometimes it is desirable to have a range of values for the unknown parameter. That is, we try to find a interval about which we can be highly confident that unknown parameter lies in it. Such a method is known as interval Estimation method. "An interval estimator of a population parameter is an interval that is supposed to contain the parameter."

Confidence Interval: Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution $f(x; \theta)$ where $\theta \in \Theta$ is the unknown parameter. Suppose $T_1(\underset{\sim}{X})$ and $T_2(\underset{\sim}{X})$ be any two statistics such that

$$P\left(T_1(\underset{\sim}{X}) \le g(\theta) \le T_2(\underset{\sim}{X})\right) = 1-\alpha \quad \forall\ \theta \in \Theta, \quad 0<\alpha<1$$

The interval $[T_1(\underline{x}), T_2(\underline{x})]$ is called $100(1-\alpha)\%$ Confidence interval of $g(\theta)$. Note that $g(\theta)$ can be θ itself. Here $(1-\alpha)$ is known as confidence coefficient of the interval. If observed value of rvs $\underline{X} = (X_1, X_2, \cdots, X_n)$ is $\underline{x} = (x_1, x_2, \cdots, x_n)$ then we say that $(T_1(\underline{x}), T_2(\underline{x}))$ is a $100(1-\alpha)\%$ confidence interval of $g(\theta)$.

Next we obtain confidence intervals (CI) for the parameters of a normal distribution.

Result: Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal $N(\mu, \sigma^2)$ distribution. ~~Find confidence interval of (i) μ~~ Find $100(1-\alpha)\%$ confidence interval of (i) μ when σ^2 is known

(ii) σ^2 when μ is known.

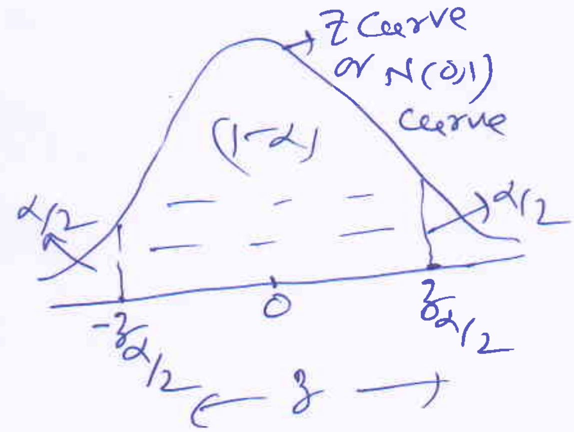
Ans: (i) Interval estimator of μ when σ^2 known.

For the given problem

$$\bar{X} \sim N(\mu, \sigma^2/n)$$ where $\bar{X}$ is the sample mean.

Let us form a pivotal quantity Z containing $\bar{X}$ and μ as follows:

$$Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$$



Thus $Z \sim N(0,1)$ and then we have

$$P\left(-z_{\alpha/2} < Z \le z_{\alpha/2}\right) = 1-\alpha \quad \text{(see figure also)}$$

$$\Rightarrow P\left(-z_{\alpha/2} < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \mu \le \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) = 1-\alpha$$

So, $\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$ is a $100(1-\alpha)\%$ confidence interval of μ. (this is interval estimator of μ)

When observed value of $\bar{X}$, say $\bar{x}$ is available then we get the interval estimate of μ as

$$\left[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right].$$

Note: In practice α is choosen as 0.05, 0.01, 0.1.
$\alpha = 0.05 \Rightarrow 1-\alpha = 0.95$ so we get CI with 95% confident.

Ex: Suppose length of life of a 60-watt bulb follow $N(\mu, 1296)$ distribution. If a random sample of size $n = 27$ bulbs were tested until they burned out, yielding a sample mean of $\bar{x} = 1478$ hours. Find a 95% confidence interval of μ.

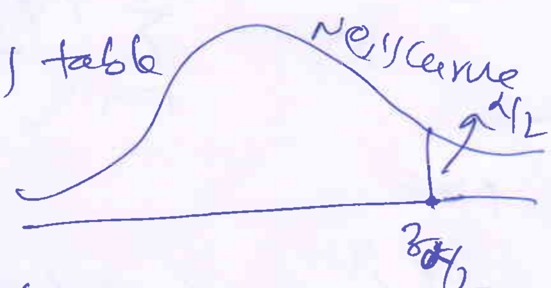
$\Rightarrow$ Here $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$. Furthermore following the interval obtained on page (3) we have

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right] \text{ is a } 100(1-\alpha)\% \text{ CI}$$

of μ. ~~Here we~~ The corresponding estimated interval is $\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$

$$\Rightarrow \left[1478 - \frac{36}{\sqrt{27}} \cdot z_{\alpha/2},\ 1478 + \frac{36}{\sqrt{27}} z_{\alpha/2}\right] \text{ is a } 100(1-\alpha)\%$$

CI of μ. But it is given $\alpha = 0.05$, so what is $z_{\alpha/2}$. Note that from $N(0,1)$ table



$$P(Z > z_{\alpha/2}) = 0.025$$

From the table $z_{\alpha/2} = 1.16$.

So required 95% CI of μ is

$$\left[1478 - \frac{36}{\sqrt{27}} \cdot (1.96),\ 1478 + \frac{36}{\sqrt{27}} (1.96)\right]$$

After simplification, the 95% CI of μ is given by $[1464.42, 1491.58]$.

So the interpretation is that: "average life of a bulb will be between $[1464.42, 1491.58]$ hours with 95% confident.

$$\left(P(1464.42 < \mu < 1491.58) = 0.95 \right).$$

<u>Note</u>: For this problem you can find 90%, 99% CI of μ.

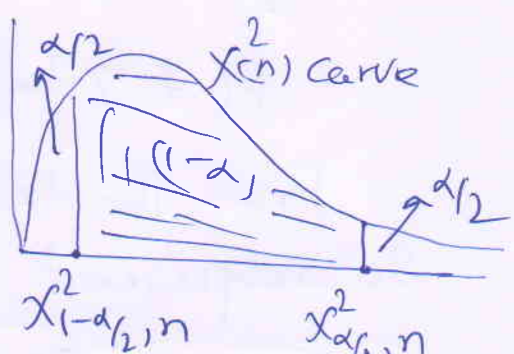
(ii) Next we find interval estimator of σ^2 when μ is known.

Like part (i) we first ~~a~~ form a pivotal quantity containing a point estimator of σ^2 and σ^2.

Recall if $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ then

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

Now



$$P\left(\chi^2_{1-\alpha/2, n} < W \le \chi^2_{\alpha/2, n}\right) = 1-\alpha$$

or, $$P\left(\chi^2_{1-\alpha/2, n} < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}, n}\right) = 1-\alpha$$

or, $$P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\alpha/2,n}} < \sigma^2 \le \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right) = 1-\alpha.$$

So, $\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2},n}}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right]$ is a $100(1-\alpha)\%$ CI of σ^2.

<u>Note</u>: $$P\left(\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2},n}}} < \sigma \le \sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}}\right) = 1-\alpha$$

So $\left(\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)}{\chi^2_{\alpha/2,n}}}, \sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\alpha/2,n}}}\right)$ is a $100(1-\alpha)\%$ CI of σ.

<u>Ex</u>: (Previous problem) Suppose length of life of 60 watt bulb follow $N(1478, \sigma^2)$ distn. If a random sample of size $n=27$ ~~is taken~~ bulbs are tested until they burned out yielding the following data: 1481, 1537, 1513, 1583, 1453, 1310, 1370, 1500, 1257, 1333, 1418, 1227, 1326, 1243, 1499, 1427, 1421, 1337, 1299, 1178, 1210, 1499, 1478, 1483, 1468, 1500, 1497.

Find 95% CI of σ^2.

Result: Confidence interval of μ when σ^2 unknown.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find a $100(1-\alpha)\%$ CI of μ when σ^2 unknown.

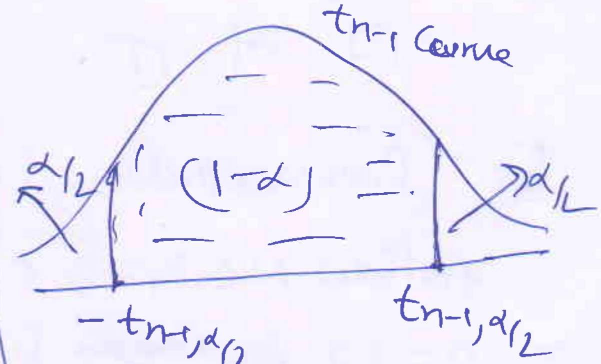
$\Rightarrow$ For the given problem we have to find a pivotal quantity containing some statistics and unknown parameter of interest which is μ here.

Note $\bar{X} \sim N(\mu, \sigma^2/n) \Rightarrow \dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$. — (1)

Also $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$. — (2)

Thus from (1) & (2) we have

$$T = \frac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t_{n-1}.$$



$$P\left(-t_{\alpha/2, n-1} < T \le t_{\alpha/2, n-1}\right) = 1-\alpha$$

$$\Rightarrow P\left(-t_{\alpha/2, n-1} < \frac{\sqrt{n}(\bar{X}-\mu)}{s} < t_{\frac{n}{2}, \alpha}\right) = 1-\alpha$$

After manipulation we get the following:

$$P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2, n-1} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}\right) = 1-\alpha$$

$\therefore \left[\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}, \ \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2, n-1}\right]$ is a $100(1-\alpha)\%$ CI of μ when σ is unknown.

<u>Ex</u>: Suppose that amount of buttorfat in pounds produced by cow milk production follows normal $N(\mu, \sigma^2)$ distn. To estimate μ a farmer measured the butterfat production for $n=20$ cows yielding the following data:

481, 537, 513, 583, 453, 510, ~~570~~, 500, 457
555, 618, 327, 350, 643, 499, 421, 505
637, 599, 392.

Find 95% CI for μ.

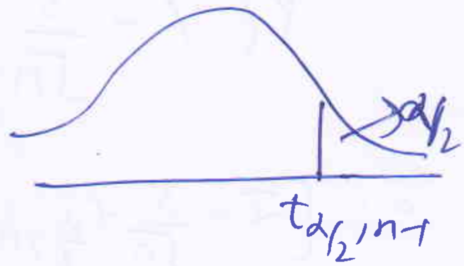
<u>Soln</u>: $\bar{x} = \frac{1}{n}\sum_{i=1}^{20} x_i = 507.50$, $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
$= 89.75$

a 95% CI for μ is (using previous result)

$\left[\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}, \ \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right]$ where ——— 1

$t_{\alpha/2, n-1}$ is such that

$$P(T > t_{\alpha/2, n-1}) = \alpha/2$$



In air case $\alpha = 0.05$, $n = 20$

so from table (T-table

$$P(T > t_{\alpha/2, n-1}) = 0.025$$

$$\Rightarrow t_{0.025, 19} = 2.10.$$

So required interval from (1) is

$$\left[507.50 - 2.10\left(\frac{89.75}{\sqrt{20}}\right), 507.50 + 2.10\,\frac{89.75}{\sqrt{20}}\right]$$

$$\Rightarrow \left[507.5 - \frac{188.475}{\sqrt{20}}, 507.5 + \frac{188.475}{\sqrt{20}}\right]$$

$$\Rightarrow \cancel{(507.277, 507.723)} \Rightarrow [507.5 - 42.16, 507.5 + 42.16]$$

$$\Rightarrow [465.34, 549.66]$$ is a 95% CI of μ.