STAT 417. Lecture NOTE 9

§6.3. Confidence interval.

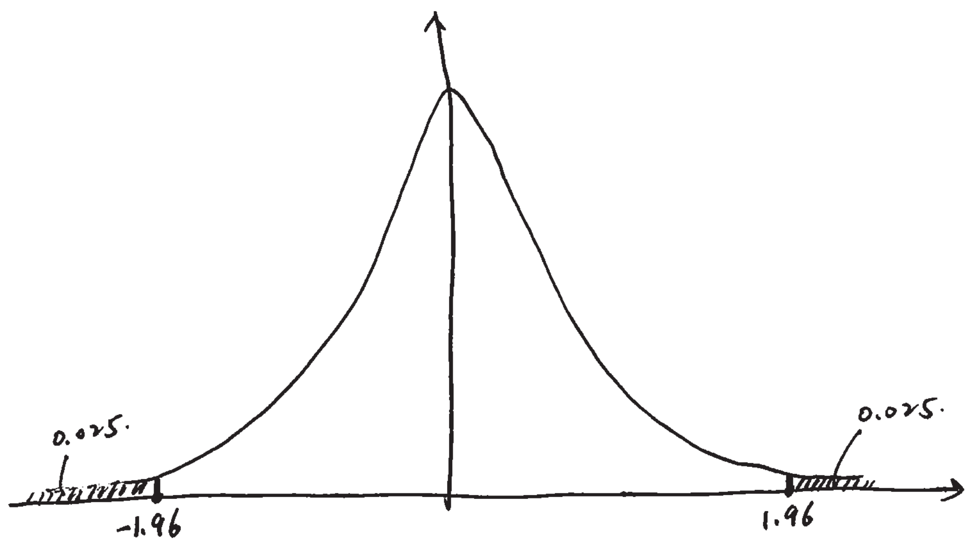
Suppose $X_1, X_2, \ldots, X_{36}$ are i.i.d samples from $N(\mu, \sigma^2)$. $\sigma^2 = 1$.

If we observe that $\bar{X} = 2$. then $\hat{\mu}_{ML} = \bar{X} = 2$.

Today we are going to provide an alternative way to quantify the scope of μ, i.e. the confidence interval.

From central limit theorem,

$$\frac{\sqrt{36}\,(\bar{X} - \mu)}{\sigma} \approx Z \sim N(0,1).$$



$$P\left(-1.96 < \sqrt{36}\left(\frac{\bar{X}-\mu}{\sigma}\right) < 1.96 \right) = 1 - 0.025 - 0.025 = 95\%.$$

$$-1.96 < \sqrt{36}\left(\frac{\bar{X}-\mu}{\sigma}\right) < 1.96.$$

$$\Downarrow$$

$$\bar{X} - \frac{1.96\sigma}{\sqrt{36}} < \mu < \bar{X} + \frac{1.96\sigma}{\sqrt{36}}$$

So $\left[\bar{X} - \frac{1.96}{\sqrt{36}}\sigma,\ \bar{X} + \frac{1.96}{\sqrt{36}}\sigma\right]$. $= [1.673, 2.327]$

is called 95% confidence interval for μ.

Z- confidence intervals. (Large sample).

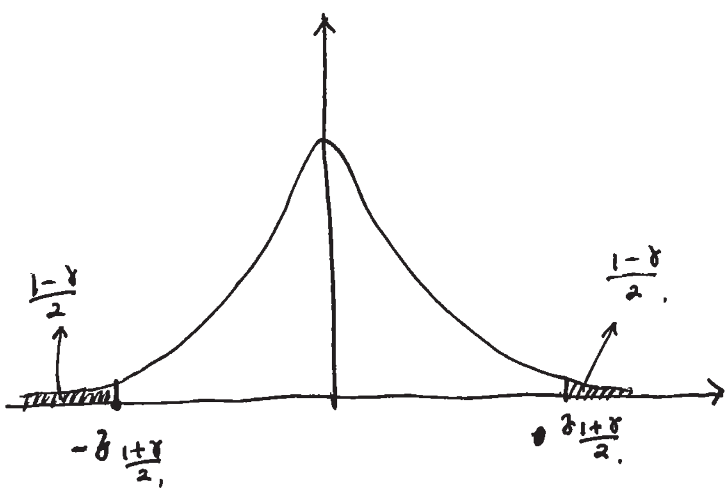
suppose $X_1, X_2, \ldots, X_n$ are i.i.d samples from $N(\mu, \sigma^2)$. σ^2 is known.

~~then~~. and we want to establish a γ-confidence interval.

From CLT,

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

so $\sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right)$ approximately follows the following Z-curve



so.

$$P\left(-z_{\frac{1+\gamma}{2}} < \sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) < z_{\frac{1+\gamma}{2}}\right) = \gamma.$$

From $-z_{\frac{1+\gamma}{2}} < \sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) < z_{\frac{1+\gamma}{2}}$. we get an interval.

$$\bar{X} - z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

so. the γ-confidence interval for μ is

$$\left[\bar{X} - z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right].$$

E.g. 1. Suppose $X_1, X_2, \ldots, X_{100}$ are i.i.d $N(\mu, 4)$ samples. establish an 90% - confidence interval, if $\bar{X} = 1$.

Sol: when $\gamma = .90\%$ $z_{\frac{1+\gamma}{2}} = 1.645$.

so, the 90% - confidence interval for μ is

$$\left[1 - 1.645 \cdot \frac{2}{\sqrt{100}} ,\ 1 + 1.645 \cdot \frac{2}{\sqrt{100}} \right] = [0.671, 1.329].$$

EX 1. In E.g. 1. establish an 99% - confidence interval.

Sol. $\gamma = 0.99$. then. $z_{\frac{1+\gamma}{2}} = 2.58$.

so the 99% - confidence interval for μ is

$$\left[1 - 2.58 \cdot \frac{2}{\sqrt{100}} ,\ 1 + 2.58 \cdot \frac{2}{\sqrt{100}} \right] = [0.484, 1.516].$$

Confidence interval for general model:

If $X_1, X_2, \ldots, X_n$ are samples from general models, how to establish CI?

E.g. 3. Suppose $X_1, X_2, \ldots, X_{100}$ are samples from Exponential(θ), if we know $\bar{X} = 3$. then how to establish 95% - confidence interval for θ?

Sol: From CLT, $\left(\mu = \frac{1}{\theta},\ \sigma^2 = \frac{1}{\theta^2} \right)$

$$\sqrt{100} \left(\frac{\bar{X} - \frac{1}{\theta}}{\frac{1}{\theta}} \right) \approx Z \sim N(0,1).$$

so. For 95%-confidence. solve.

$$-1.96 < \sqrt{100}\left(\frac{\bar{X}-\frac{1}{\theta}}{\frac{1}{\theta}}\right) < 1.96$$

which gives $0.268 < \theta < 0.399$. so the 95%-confidence interval is

$$[0.268,\ 0.399].$$

<u>Ex2</u>. Suppose that $X_1, X_2, \ldots, X_{100}$ are iid $Unif(0, \theta)$.

Suppose. $\bar{X} = 1$. Find 95%-CI for θ.

Sol:
$$-1.96 < \sqrt{100}\left(\frac{\bar{X}-\frac{\theta}{2}}{\frac{\theta}{\sqrt{12}}}\right) < 1.96$$

then $1.797 < \theta < 2.255$

<u>Eg 3</u>. Binomial model: ~~Consider~~ toss a coin 100 times. $P(H) = \theta$, $P(T) = 1-\theta$.
Let X = # of Heads among 100 tosses. Find 95%-CI for θ. if we
Observe $X = 42$.

Sol: we employ a trick in STAT 416, ~~define~~ for $i = 1, 2, \ldots, 100$. define.

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th toss is H.} \\ 0. & \text{if the } i\text{-th toss is T.} \end{cases}$$

then $X = X_1 + X_2 + \cdots + X_{100}$. $\bar{X} = \dfrac{X_1 + \cdots + X_{100}}{100}$.

Now $\theta = E(X_i)$

E.g. 2. (Normal model with unknown σ)

If $X_1, X_2, \ldots, X_{100}$ are i.i.d samples from $N(\mu, \sigma^2)$, where σ^2 is unknown. We use the following "plug-in" type γ-confidence interval.

$$\left[\bar{X} - z_{\frac{1+\gamma}{2}} \cdot \frac{S}{\sqrt{n}} \ , \ \bar{X} + z_{\frac{1+\gamma}{2}} \cdot \frac{S}{\sqrt{n}} \right]$$

where $\bar{X}$ is the sample mean, $S = \sqrt{S^2}$ is the sample standard deviation.

If $\bar{X} = 1$, $S^2 = 2$, then the 95% - CI is $[0.72, 1.28]$

E.g. 4. Suppose $X_1, X_2, \cdots, X_{100}$ are 100 samples from the following Bernoulli distribution

$$P(X_i=1)=\theta, \quad P(X_i=0)=1-\theta. \qquad 0\le\theta\le 1,$$

We know that $E(X_i)=\theta$, $\mathrm{Var}(X_i)=\theta(1-\theta)$.

If we observe that $\bar{X}=0.42$, Find an 95% - CI for θ.

Sol: By CLT.

$$-1.96 < \sqrt{100}\left(\frac{\bar{X}-\theta}{\sqrt{\theta(1-\theta)}}\right) < 1.96$$

~~then~~. directly solving the above inequality is hard, we play the following trick. Since the MLE of θ is $\hat{\theta}_{ML}=\bar{X}$. (check ?).

we replace the bottom $\sqrt{\theta(1-\theta)}$ by $\sqrt{\bar{x}(1-\bar{x})}$.

then the above inequality becomes

$$-1.96 < \sqrt{100}\left(\frac{\bar{x}-\theta}{\sqrt{\bar{x}(1-\bar{x})}}\right) < 1.96.$$

then by $\bar{x}=0.42$ and solve it, we have the 95% CI is

$$[0.323,\ 0.517].$$

EX3. Verify that in Eg3. $\hat{\theta}_{ML}=\bar{X}$. $\boxed{f_\theta(x)=\theta^x(1-\theta)^{1-x}.}$

$$L(\theta\,|\,X_1 \dots X_{100}) = \theta^{X_1+X_2+\dots+X_{100}}(1-\theta)^{100-(X_1+X_2+\dots+X_{100})}$$

$$=\theta^{100\bar{X}}(1-\theta)^{100(1-\bar{X})}$$

so $l(\theta\,|\,X_1 \cdots X_{100}) = 100\bar{X}\log\theta + 100(1-\bar{X})\log(1-\theta)$.

$$\frac{\partial}{\partial\theta} l(\theta|\cdots)=0 \Rightarrow \frac{100\bar{X}}{\theta}-\frac{100(1-\bar{X})}{1-\theta}=0. \Rightarrow \hat{\theta}_{ML}=\bar{X}.$$