Confidence Intervals

Suppose you have a MLE $\hat{\theta}$ of θ, the parameter of a continuous distribution.
What is $P(\hat{\theta}=\theta)$? It's 0, because $\hat{\theta}$ is a continuous r.v.
Can we find some Δ such that

$$P(\theta \in [\hat{\theta}-\Delta, \hat{\theta}+\Delta]) \geq 0.95$$

$[\hat{\theta}-\Delta, \hat{\theta}+\Delta]$ is called the <u>95% confidence interval</u>.

<u>Ex:</u> Recall MLE for $X \sim N(\theta_1, \theta_2)$.

$\hat{\theta}_1 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$, where $x_1, x_2, \ldots, x_n$ are independent samples of X.

$\hat{\theta}_1$ is a r.v. with

$$E(\hat{\theta}_1) = \theta_1$$

$$\mathrm{Var}(\hat{\theta}_1) = \theta_2/n.$$

$\hat{\theta}_1 \sim N(\theta_1, \theta_2/n)$.

$$\boxed{\frac{\hat{\theta}_1-\theta_1}{\sqrt{\theta_2/n}} \sim N(0,1)}$$

$$P\left(-z < \frac{\hat{\theta}_1-\theta_1}{\sqrt{\theta_2/n}} < z\right) = \Phi(z)-(1-\Phi(z)) = 2\Phi(z)-1$$

$$P\left(-z < \frac{\theta_1-\hat{\theta}_1}{\sqrt{\theta_2/n}} < z\right) = 2\Phi(z)-1$$

$$P\left(\hat{\theta}_1 - z\sqrt{\theta_2/n} < \theta_1 < \hat{\theta}_1 + z\sqrt{\theta_2/n}\right) = 2\Phi(z)-1 = 0.95$$

$$\Phi(z) = 0.975$$

$$\Delta \approx 1.96\sqrt{\theta_2/n} \approx 1.96\sqrt{\hat{\theta}_2'/n} \qquad z \approx 1.96$$

$$P(A < B) = P(-B < -A)$$

If θ_2 is known, we know Δ and we're done.
If θ_2 is unknown, we can use the estimator

$$\hat{\theta}_2' = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2,$$ the unbiased estimator of θ_2.

It is no longer true that

$$\frac{\hat{\theta}_1 - \theta_1}{\sqrt{\hat{\theta}_2' / n}}$$ is normal. But as n increases, ~~CLT says~~ this approaches $N(0,1)$.

What if the samples came from a distribution other than normal? Then $\hat{\theta}_1$ the sample mean, would not be normal. But, the CLT says it approaches normal as n increases. So everything we've done works approximately.