CS 70 CLT Review Session

Central Limit Theorem:

Suppose we take n iid observations of a RV, and denote the outcomes as $X_1, X_2, \dots X_n$, each having common mean μ and common variance σ^2.

$$A_n = \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} = \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

"average" of the sample

divide by n in both numerator and denominator

$\hat{\mu}$ = sample mean estimator = $\dfrac{X_1 + \dots + X_n}{n}$

μ : true population mean parameter, unknown

$P(A_n \le c) = \Phi(c)$

some constant

the standard normal CDF

$X \sim N(\mu, \sigma^2)$

a, b are constants

$aX + b \sim N(\mu + b, a^2\sigma^2)$

$$A_n = \frac{\hat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}} = N(0,1)$$

$$\hat{\mu} - \mu = \frac{\sigma}{\sqrt{n}} N(0,1)$$

$$\hat{\mu} = \mu + \frac{\sigma}{\sqrt{n}} N(0,1)$$

As $n \to \infty$, the variance

$$\boxed{\hat{\mu} = N\left(\mu, \frac{\sigma^2}{n}\right)}$$

in the distribution of $\hat{\mu}$ goes to zero.

the sample mean estimator

$\boxed{1}$

$$\Phi(1.96) = 0.975$$

μ = true avg. number of dots per roll

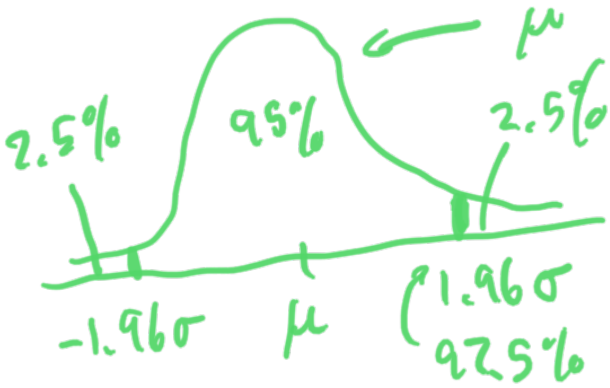
$\hat{\mu}$: empirical estimate, derived from our sample, for μ

Let $X_1, \ldots X_{10,000}$ be the outcomes of each dice roll

$$\hat{\mu} = \frac{X_1 + \ldots + X_{10,000}}{10^4} \qquad \hat{\mu} = N\left(\mu, \frac{\sigma^2}{n}\right)$$

sqrt

stdev of the distribution for $\hat{\mu}$ is thus $\frac{\sigma}{\sqrt{n}}$



$$\left(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

$\sigma^2 \le 16 \qquad n = 10^4 \qquad \hat{\mu} = 2.8$

$\sigma \le 4 \qquad \sqrt{n} = 10^2$

$$\Phi(x) = \int_{-\infty}^{x} \text{pdf}\, dx$$

$$\left(2.8 - 1.96\left(\frac{4}{10^2}\right) \le \mu \le 2.8 + 1.96\left(\frac{4}{100}\right)\right)$$

$$\boxed{2.31 \le \mu \le 3.29}$$

δ% confidence interval

$$\Phi^{-1}\left(\delta + \frac{1-\delta}{2}\right)$$

$\gamma = .95 \quad \Phi^{-1}(.975)$

$.95 + \frac{1-.95}{2} = .975$

$.90 + \frac{1-.90}{2} = .95$

$\Phi^{-1}(.95)$

$\boxed{2}$ $X \sim \text{Poisson}(\lambda)$

$\mathbb{E}[X] = \lambda \quad Var(X) = \lambda$

$A_n = \frac{X_1 + \dots + X_n}{n}$ — X_n is the number of radioactive particles emitted during the n^{th} hour

$$\mathbb{E}[A_n] = \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n}\mathbb{E}[X_1 + \dots + X_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \lambda$$

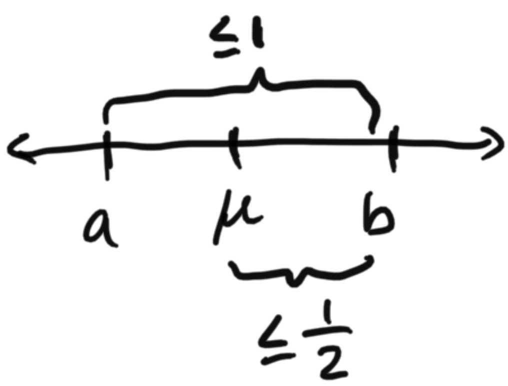
$$Var(A_n) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}Var(X_1 + \dots + X_n) = \frac{1}{n^2}(n)(\lambda) = \frac{\lambda}{n}$$

$$\text{stdev}(A_n) = \sqrt{\frac{\lambda}{n}}$$

95% CI

$$\left(\hat{\mu} - 1.96\sqrt{\frac{\lambda}{n}} \le \mu \le \hat{\mu} + 1.96\sqrt{\frac{\lambda}{n}}\right)$$

$$1.96\sqrt{\frac{\lambda}{n}} \le \frac{1}{2}$$



$$(1.96)^2\left(\frac{\lambda}{n}\right) \le \frac{1}{4}$$

$$n \cdot \frac{\lambda}{n} \le \frac{1}{4(1.96)^2} \cdot n$$

$$\lambda \le \frac{n}{4(1.96)^2}$$

$$n \ge \lambda \cdot 4(1.96)^2 \qquad \lambda \le 5$$

$$\boxed{n \geq 5.4(1.96)^2} \geq 76.8 \text{ hours}$$

$\boxed{3}$ Let $X_1, \ldots X_n$ be iid Bernoulli RVs indicating support amongst n citizens polled.

$A_n = \dfrac{X_1 + \ldots + X_n}{n}$ is an estimator for p.

$E[A_n] = \frac{1}{n}(n)(p) = p \quad Var(A_n) = \frac{1}{n^2}(n)\,p(1-p) = \dfrac{p(1-p)}{n}$

$\leq \dfrac{1}{4n}$

$$.02 \geq 1.96\left(\frac{1}{\sqrt{4n}}\right)$$

$$\frac{.02}{1.96} \geq \frac{1}{\sqrt{4n}}$$

$$\left(\frac{.02}{1.96}\right)^2 \geq \frac{1}{4n}$$

$$n \geq \frac{1}{4}\left(\frac{1.96}{.02}\right)^2$$

$$\boxed{n \geq 2401}$$

$\boxed{4}$ Again, let X_i denote the outcome of the i^{th} exponential random variable. $E[X_i] = \frac{1}{\lambda}$, $Var(X_i) = \frac{1}{\lambda^2}$

$A_n = \dfrac{X_1 + \ldots + X_n}{n} \qquad E[A_n] = \frac{1}{\lambda}$

$$Var(A_n) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{1}{n^2}(n)\left(\frac{1}{\lambda^2}\right) = \frac{1}{\lambda^2 n}$$

$\text{stdev}(A_n) = \frac{1}{\lambda\sqrt{n}}$

$$1.96\left(\frac{1}{\lambda\sqrt{n}}\right) \leq 0.05\mu$$

but the mean μ for an exponential RV is $= \frac{1}{\lambda}$

$$1.96\left(\frac{1}{\lambda\sqrt{n}}\right) \leq \frac{1}{20\lambda}$$

$$\frac{1.96}{\sqrt{n}} \leq \frac{1}{20}$$

$$1.96 \leq \frac{1}{20}\sqrt{n}$$

$$\sqrt{n} \geq 20(1.96)$$

$$\boxed{n \geq 400(1.96^2) = 1536.64}$$