Recall: Flip coin 100 times, comes up heads 40 times. Is it fair? I.e. Is $\theta = 1/2$?

A coin is a Bernoulli RV with $\mathbb{P}(X_i=1)=\theta$.

$Var(X_i)=\theta(1-\theta)$

$\hat{\theta} = \bar{X} = .4$ ← Point estimation.

Interval estimate: CLT (deMoivre-Laplace):

$$\mathbb{P}\left(\left|\frac{n\bar{X}-n\theta}{\sqrt{n\theta(1-\theta)}}\right| < \underset{\uparrow\atop 1.96}{z_{\alpha/2}}\right) \approx 1-\underset{\uparrow\atop 0.05}{\alpha}.$$

$$\boxed{\begin{aligned} Var(\bar{X}) &= \frac{\theta(1-\theta)}{n}. \\ Var(n\bar{X}) &= n^2\left(\frac{\theta(1-\theta)}{n}\right) \end{aligned}}$$

$$\mathbb{P}\left(\left|\frac{\bar{X}-\theta}{\cancel{\sqrt{\theta(1-\theta)/n}}}\right| < z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}\right)$$

$$\approx \mathbb{P}\left(\bar{X}-z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} < \theta < \underset{\uparrow\atop .4}{\bar{X}} + \underbrace{\underset{\uparrow\atop 1.96}{z_{\alpha/2}}\underbrace{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}}_{\frac{.5}{10}}}_{.196\times\frac{1}{2} = .098}\right)$$

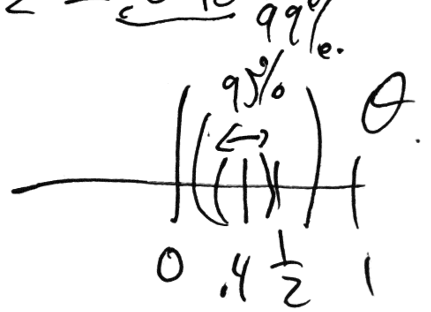
$\underbrace{.4(.6)}_{\sqrt{.24}\approx .5}$

With approx. 95% confidence $.302 < \theta < .498$

i.e. $\theta \neq 1/2$.

With 99% confidence: $z_{0.005} = 2.57$.

$\alpha = 0.01$



So: Cannot be $\approx 99\%$ certain that the coin is unfair (biased).

Good (optional) exercise: Don't approx $\sqrt{\theta(1-\theta)} \simeq \sqrt{\bar{x}(1-\bar{x})}$. What is the true numerical interval at 95%. (Does it contain 1/2?)

Note: not being that lazy in computing

$$\frac{\bar{X}-\theta}{\sqrt{\theta(1-\theta)/n}} = \underbrace{\frac{n\bar{X}-n\theta}{\sqrt{\theta(1-\theta)n}}}_{\text{depends on } \theta!} \simeq N(0,1).$$

(Out of necessity).

Next: Comparing proportions (i.e. difference of binomials/Bernoullis).

A: 132 out of 200 male voters prefer $\cancel{A}$ to $\cancel{B}$.

B: 90 out of 150 female " " " "

$\bar{X}_A = \hat{\theta}_A = \frac{132}{200} \simeq 0.66$. $\hat{\theta}_B = \bar{X}_B = \frac{90}{150} = .6$.

Want: Confidence interval for $\theta_A - \theta_B$.

$$\frac{\bar{X}_A - \theta_A}{\cancel{\sqrt{\theta_A(1-\theta_A)/n_A}}} \simeq N(0, \theta_A(1-\theta_A)/n_A).$$

$\uparrow$ CLT.

$$\bar{X}_B - \theta_B \simeq N(0, \theta_B(1-\theta_B)/n_B).$$

so: $(\bar{X}_A - \theta_A) - (\bar{X}_B - \theta_B) \approx N\left(0, \ \theta_A(1-\theta_A)/n_A + \theta_B(1-\theta_B)/n_B\right)$

$\parallel \quad (\bar{X}_A - \bar{X}_B) - (\theta_A - \theta_B).$

$$\Rightarrow \mathbb{P}\left(\left| \frac{(\bar{X}_A - \bar{X}_B) - (\theta_A - \theta_B)}{\sqrt{\frac{\theta_A(1-\theta_A)}{n_A} + \frac{\theta_B(1-\theta_B)}{n_B}}} \right| < z_{\alpha/2} \right) \approx 1-\alpha.$$

$\theta_A \rightsquigarrow \bar{X}_A$

$$\approx \boxed{\mathbb{P}\left(\left| (\bar{X}_A - \bar{X}_B) - (\theta_A - \theta_B) \right| < z_{\alpha/2} \sqrt{\frac{\bar{X}_A(1-\bar{X}_A)}{n_A} + \frac{\bar{X}_B(1-\bar{X}_B)}{n_B}} \right)}$$

<u>Estimating variances:</u> $X_1, \dots, X_n$ iid Gaussian.

Point estimator for σ^2 is $\hat{\sigma}^2 = S^2$.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad \boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2} \text{ with } \underline{n-1} \text{ deg.}$$

<u>pdf:</u> $$f_{\chi^2}(t) = \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} \cdot e^{-t/2} \underbrace{t^{\frac{n-1}{2}} \cdot \frac{1}{t}}, \quad t > 0.$$

$t^{-1/2}$ if $n-1=1$,

if $n-1=2$, $t^0 = 1$.

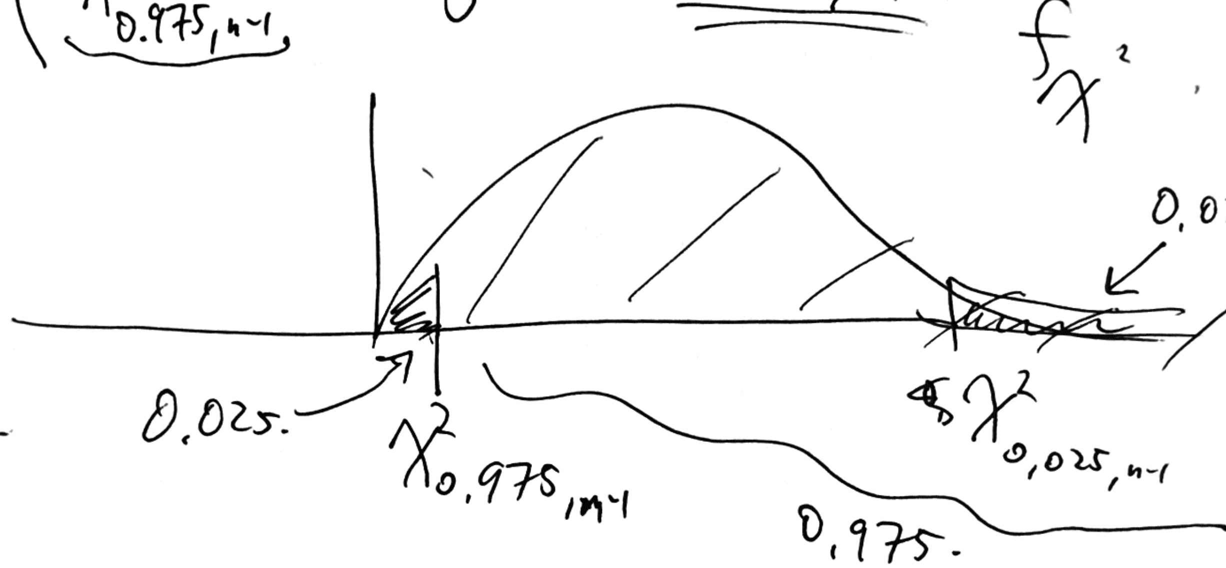
<u>Say:</u> $\chi^2_{\alpha, n-1}$

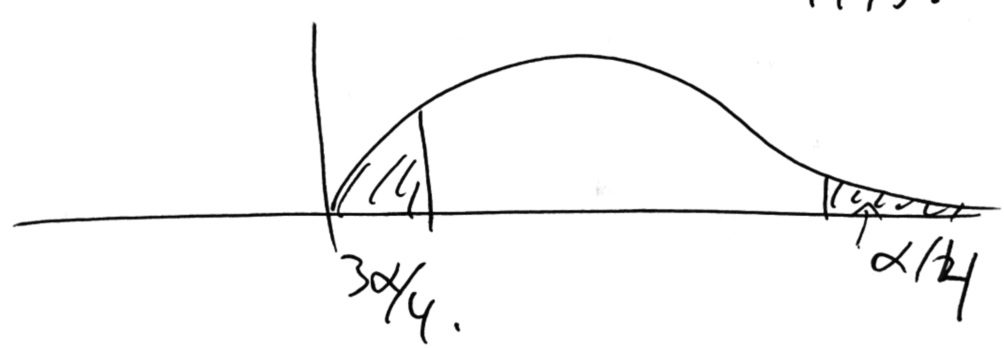
$$\mathbb{P}\left(\chi^2_{1-\frac{\alpha}{2}, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2, n-1}\right) = 1-\alpha.$$

Another $(1-\alpha)100\%$ confidence interval is:

$$\mathbb{P}\left(\chi^2_{1-\frac{3\alpha}{4}, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/4, n-1}\right) = 1-\alpha.$$

$$\mathbb{P}\left(\underline{\chi^2_{0.975, n-1}} < \frac{(n-1)S^2}{\sigma^2} < \underline{\chi^2_{0.025, n-1}}\right)$$





or:

A $(1-\alpha)100\%$ confidence interval for σ^2 is:

$$\mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,\, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},\, n-1}} \right) = 1-\alpha.$$

Recall: if we had two estimators, $\hat{\theta}_1, \hat{\theta}_2$.

Efficiency of one to ~~the~~ other is: $\dfrac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$

How can we get an interval estimate for ratio of two variances? Need an F-distribution:

(Sir Ronald Fisher)

Def. $F = \dfrac{U/n}{V/m}$, where $U \sim \chi^2$ with n deg, $V \sim \chi^2$ with m deg. & U & V are indep.

Derive its pdf. $f_F(t) = 0$ for $t \le 0$.

$$\mathbb{P}(F \le a) = \int_{v=0}^{\infty} \int_{u=0}^{a v \frac{n}{m}} f_{\chi^2_{,n}}(u)\, du \; f_{\chi^2_{,m}}(v)\, dv$$

$F = \dfrac{u/n}{v/m}$

$\dfrac{u/n}{v/m} \le a. \iff u \le a v \dfrac{n}{m}.$

$$f_{\chi^2_{,n}}(u) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-u/2} \underline{u^{n/2}} \frac{1}{u}.$$

(5)

$t = \dfrac{u}{v} \cdot \dfrac{m}{n}$. $\dfrac{dt}{t} = \dfrac{du}{u}$

$$\boxed{u = t v \tfrac{n}{m}}.$$

$$= \int_{v=0}^{\infty} \int_{0}^{a} \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-t v \frac{n}{m}/2} \underbrace{t^{n/2}}_{\left(v\frac{n}{m}\right)^{n/2}} \frac{dt}{t} f_{\chi^2_{1,m}}(v)\, dv$$

$$= \int_0^a \frac{\left(\frac{n}{m}\right)^{n/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \int_{v=0}^{\infty} \underbrace{e^{-tv\frac{n}{m}/2} \, v^{n/2} \frac{1}{2^{m/2}\Gamma\left(\frac{m}{2}\right)} e^{-v/2}}_{e^{-v\left(1+t\cdot\frac{n}{m}\right)/2}} v^{\frac{m}{2}} \frac{dv}{v} t^{n/2} \frac{dt}{t}$$

$$= \int_0^a \underbrace{\frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{n/2} \left(1+t\frac{n}{m}\right)^{-\left(\frac{n+m}{2}\right)} t^{n/2} \frac{dt}{t}}_{f_F(t)}.$$