The proof is on page 472 of C&B.

Recall Taylor series formula for f(y) "centered at" constant a.

$$f(y) = f(a) + (y-a) f'(a) + (y-a)^2 \frac{f''(y)}{2} + \dots$$

let $f = \ell'$, $y = \hat\theta^{MLE}$, $a = \theta$

$$\ell'(\hat\theta^{MLE}; X_1,\dots,X_n) = \ell'(\theta; X_1,\dots,X_n) + (\hat\theta^{MLE}-\theta)\,\ell''(\theta; X_1,\dots,X_n) + (\hat\theta^{MLE}-\theta)^2 \frac{\ell'''(\theta; X_1,\dots,X_n)}{2} + \dots$$

Assuming technical conditions on p516 of C&B and a large enough sample size n, we can use the first-order Taylor approx:

$$\ell'(\hat\theta^{MLE}; X_1,\dots,X_n) = \ell'(\theta; X_1,\dots,X_n) + (\hat\theta^{MLE}-\theta)\,\ell''(\theta; X_1,\dots,X_n)$$

Recall the defintion of the MLE, $\hat\theta^{MLE} := \operatorname{argmax}\{\ell(\theta; X_1,\dots,X_n)\}$

which is usually the solution to the equation: $\ell'(\theta; X_1,\dots,X_n) \overset{set}{=} 0$

$\Rightarrow \ell'(\hat\theta^{MLE}; X_1,\dots,X_n) = 0$ Amazing!

$$\Rightarrow 0 = \ell'(\theta; X_1,\dots,X_n) + (\hat\theta^{MLE}-\theta)\,\ell''(\theta; X_1,\dots,X_n)$$

move $\div$

$$\Rightarrow \hat\theta^{MLE} - \theta = \frac{-\ell'(\theta; X_1,\dots,X_n)}{\ell''(\theta; X_1,\dots,X_n)} \frac{\frac{1}{n}}{\frac{1}{n}}$$

Mult lhs, rhs by $\frac{1}{\sqrt{\frac{I(\theta)^{-1}}{n}}}$

$$\Rightarrow \frac{\hat\theta^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \frac{\frac{1}{n}\ell'(\theta; X_1,\dots,X_n)}{-\frac{1}{n}\ell''(\theta; X_1,\dots,X_n)\sqrt{\frac{I(\theta)^{-1}}{n}}} \frac{I(\theta)}{\sqrt{I(\theta)^2}}$$

$$\Rightarrow \frac{\hat\theta^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} = \underbrace{\frac{I(\theta)}{-\frac{1}{n}\ell''(\theta; X_1,\dots,X_n)}}_{\hat A} \underbrace{\frac{\frac{1}{n}\ell'(\theta; X_1,\dots,X_n)}{\sqrt{\frac{I(\theta)}{n}}}}_{\hat B}$$

If $\hat A \xrightarrow{p} 1$, $\hat B \xrightarrow{d} N(0,1) \Rightarrow$ l.h.s $\xrightarrow{d} N(0,1)$ by Slutsky's thm.

Pf. $\hat A \xrightarrow{p} 1$

def 7, def 8

$$S = \ell'(\theta; X_1,\dots,X_n) = \sum_{i=1}^n \ell'(\theta; X_i)$$

$$\Rightarrow \ell''(\theta; X_1,\dots,X_n) = \sum_{i=1}^n \ell''(\theta; X_i)$$ mult lhs, rhs by $-\frac{1}{n}$

$$-\frac{1}{n}\ell''(\theta; X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^n -\ell''(\theta; X_i) = \frac{1}{n}\sum Y_i = \bar Y \xrightarrow{p} E[Y] = I(\theta)$$

let $Y_i := -\ell''(\theta; X_i)$, $E[Y_i] \overset{HW}{=} \dots I(\theta)$

Law of Large #'s (MATH 368)

$$\hat A = \frac{I(\theta)}{-\frac{1}{n}\ell''(\theta; X_1,\dots,X_n)} \xrightarrow{p} 1$$ by corr 5.5.4b

Pf. $\hat B \xrightarrow{d} N(0,1)$ We use a CLT.

$$\frac{1}{n}\ell'(\theta; X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^n \ell'(\theta; X_i) = \frac{1}{n}\sum W_i = \bar W$$

let $W_i = \ell'(\theta; X_i)$, $E[W_i] = E[\ell'(\theta; X_i)] = 0$ (Fact 1b, Lec 9) $\Rightarrow E[\bar W] = 0$

Def of $I(\theta)$

$$SE[\bar W] = \sqrt{\frac{Var[W]}{n}} = \sqrt{\frac{E[\ell'(\theta;X_i)^2] - \cancel{E[\ell'(\theta;X_i)]^2}}{n}} = \sqrt{\frac{I(\theta)}{n}}$$

$$\frac{\frac{1}{n}\ell'(\theta; X_1,\dots,X_n)}{\sqrt{\frac{I(\theta)}{n}}} = \frac{\bar W}{\sqrt{\frac{I(\theta)}{n}}} = \frac{\bar W - E[\bar W]}{SE[\bar W]} \xrightarrow{d} N(0,1)$$ by CLT

i.e. $\hat B \xrightarrow{d} N(0,1)$

We're done, we've proven the MLE is asymptotically normal and asymptotically efficient (i.e. its variance gets close to the minimum possible, the CRLB, as n grows larger),

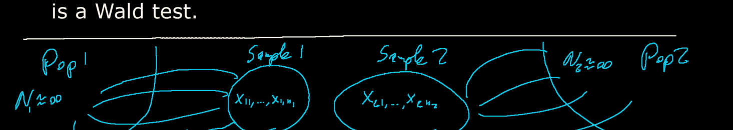
$$\frac{\hat\theta^{MLE} - \theta}{\sqrt{\frac{I(\theta)^{-1}}{n}}} \xrightarrow{d} N(0,1)$$

However, the denominator has I(θ) which is a function of θ and we don't know θ! So we can employ the thm. we proved last class. If you plug in a consistent estimator of θ, you preserve the convergence in distribution (albeit it goes slower). By monster thm, conclusion (1), the MLE itself is consistent, so:

$$\frac{\hat\theta^{MLE} - \theta}{\sqrt{\frac{I(\hat\theta^{MLE})^{-1}}{n}}} \xrightarrow{d} N(0,1)$$

This is a very useful thm. Now we can do an approximate z test using *any* MLE!

The approx z test employing the above two formulas is called a "Wald Test". You've seen one, the 1-prop z-test is a Wald test.



Assume $X_{11},\dots,X_{1n_1} \overset{iid}{\sim} Bern(\theta_1)$ indep. of $X_{21},\dots,X_{2n_2} \overset{iid}{\sim} Bern(\theta_2)$.

There are three likely testing scenarios:
(1) $H_a: \theta_1 > \theta_2$, $H_0: \theta_1 \le \theta_2$ (2) $H_a: \theta_1 < \theta_2$, $H_0: \theta_1 \ge \theta_2$ (3) $H_a: \theta_1 \ne \theta_2$, $H_0: \theta_1 = \theta_2$

Which test statistic measured departure from the null hypothesis?

$\hat\theta_1 - \hat\theta_2 = \bar x_1 - \bar x_2$

Then we ask is the departure "significant". To assess this, we need to know the distribution of $\hat\theta_1 - \hat\theta_2$ i.e. the process that realizes our test statistic / estimate.

From Math 368, $X_1,\dots,X_{n_1} \overset{iid}{\sim}$ DGP mean μ_1, variance σ_1^2 indep of $Y_1,\dots,Y_{n_2} \overset{iid}{\sim}$ DGP mean μ_2, variance σ_2^2 then,

$$\frac{(\bar X - \bar Y) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{d} N(0,1) \text{ as } n_1, n_2 \to \infty$$

Using this fact we obtain:

$$\frac{(\hat\theta_1 - \hat\theta_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \xrightarrow{d} N(0,1)$$

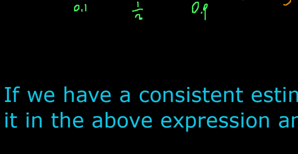
"shared"

Under the null hypothesis, $\theta_1 = \theta_2 = \theta_S$, so we get:

$$\frac{\hat\theta_1 - \hat\theta_2}{\sqrt{\frac{\theta_S(1-\theta_S)}{n_1} + \frac{\theta_S(1-\theta_S)}{n_2}}} = \frac{\hat\theta_1 - \hat\theta_2}{\sqrt{\theta_S(1-\theta_S)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{d} N(0,1)$$

$$= \frac{\hat\theta_1 - \hat\theta_2}{SE[\hat\theta_1 - \hat\theta_2](\theta_S)}$$

Can we run our test now? No! Because θ_S is not known! And its value really matters:



Based on the value of the shared θ, the std. error of your sampling distribution changed by quite a lot. So, we really can't just employ 0.5, we should get it right!

If we have a consistent estimator for the shared θ, we can employ it in the above expression and retain the asymptotic normality.

$$\hat\theta_S = \frac{X_{11} + \dots + X_{1n_1} + X_{21} + \dots + X_{2n_2}}{n_1 + n_2} = \frac{n_1 \bar x_1 + n_2 \bar x_2}{n_1 + n_2}$$ weighted avg

we know this is consistent since we know $\bar X \xrightarrow{p} \mu$.

Now, we can finally run tests since we have a sampling distr:

$$\frac{\hat\theta_1 - \hat\theta_2}{\sqrt{\hat\theta_S(1-\hat\theta_S)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{\hat\theta_1 - \hat\theta_2}{\sqrt{\frac{\sum X_{1i} + \sum X_{2i}}{n_1 + n_2}\left(1 - \frac{\sum X_{1i} + \sum X_{2i}}{n_1 + n_2}\right)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{d} N(0,1)$$

And this is the 2-prop z-test!

Note: you may see a less efficient / powerful expression in many textbooks:

$$\frac{\hat\theta_1 - \hat\theta_2}{\sqrt{\frac{\bar x_1(1-\bar x_1)}{n_1} + \frac{\bar x_2(1-\bar x_2)}{n_2}}} \xrightarrow{d} N(0,1)$$

Please don't use this. It's not as good as the one we derived.