$$I(\theta) = 2\int_{\mathbb{R}} \frac{\left(e^{-(x-\theta)}\right)^2}{\left(1+e^{-(x-\theta)}\right)^2} \cdot \frac{1}{\left(1+e^{-(x-\theta)}\right)^2}\, dx$$

let $u = \frac{1}{1+e^{-(x-\theta)}} \Rightarrow 1-u = \frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}}$ $\quad x \to -\infty \Rightarrow u \to 0$, $x \to +\infty \Rightarrow u \to 1$

$$\Rightarrow \frac{du}{dx} = -\frac{1}{\left(1+e^{-(x-\theta)}\right)^2}\left(-e^{-(x-\theta)}\right) = \frac{e^{-(x-\theta)}}{\left(1+e^{-(x-\theta)}\right)^2} = (1-u)u \Rightarrow dx = \frac{1}{(1-u)u}du$$

$$= 2\int_0^1 (1-u)^{\cancel{2}} u^{\cancel{2}} \frac{1}{\cancel{(1-u)u}}\, du = 2\int_0^1 (u-u^2)\, du = 2\left[\frac{u^2}{2} - \frac{u^3}{3}\right]_0^1$$

$$= 2\left(\frac{1}{2} - \frac{1}{3}\right) = 2 \cdot \frac{1}{6} = \frac{1}{3}$$

$$\frac{S(\theta_0, X_1, \ldots, X_n)}{\sqrt{n I(\theta_0)}} = \underbrace{\frac{n - 2\sum_{i=1}^{n} \frac{e^{-(X_i-\theta)}}{1+e^{-(X_i-\theta)}}}{\sqrt{n/3}}}_{\text{score z statistic}} \overset{\cdot}{\sim} N(0,1)$$

Data is <1.85, 4.18, 2.63, 3.41, 4.23, 4.24, 3.17, 3.69, 3.13, 1.04>
n = 10. Test θ ≠ 0 at alpha = 5%.

$$z = \frac{10 - 2\left(\frac{e^{-1.85}}{1+e^{-1.85}} + \ldots + \frac{e^{-1.04}}{1+e^{-1.04}}\right)}{\sqrt{10/3}} = 4.77 \notin [-1.96, 1.96]$$

Reject $H_0: \theta = 0$.

Likelihood Ratio Test (LRT) which is another means of testing the alternative hypothesis of $\theta \neq \theta_0$. Recall that test statistics gauged "departure" from the null hypothesis. Consider:

$$\hat{LR} := \frac{\mathcal{L}(\hat{\theta}^{MLE}; X_1, \ldots, X_n)}{\mathcal{L}(\theta_0; X_1, \ldots, X_n)} \overset{iid}{=} \frac{\prod_{i=1}^{n} \mathcal{L}(\hat{\theta}^{MLE}; X_i)}{\prod_{i=1}^{n} \mathcal{L}(\theta_0; X_i)} = \prod_{i=1}^{n} \frac{\mathcal{L}(\hat{\theta}^{MLE}; X_i)}{\mathcal{L}(\theta_0; X_i)}$$

This ratio is always ≥ 1 since the numerator is the maximum value of the likelihood function. If it's much larger than 1 what does that say about the null hypothesis? It should be rejected.

You can prove that:

$$\hat{\Lambda} := 2\ln(\hat{LR}) \xrightarrow{d} \chi^2_1$$

$$2\left(\ell(\hat{\theta}^{MLE}; X_1, \ldots, X_n) - \ell(\theta_0; X_1, \ldots, X_n)\right)$$

And $\hat{\Lambda}$ can be used to then test by computing the cutoff threshold (which is 3.84 if alpha = 5%) and checking if $\hat{\Lambda}$ is less than it (retain null) or greater (reject null).

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta) \Rightarrow \hat{\theta}^{MLE} = \bar{X}$

$$\hat{LR} = \prod_{i=1}^{n} \frac{\mathcal{L}(\hat{\theta}^{MLE}; X_i)}{\mathcal{L}(\theta_0; X_i)} = \prod_{i=1}^{n} \frac{\bar{X}^{X_i}(1-\bar{X})^{1-X_i}}{\theta_0^{X_i}(1-\theta_0)^{1-X_i}} = \left(\frac{\bar{X}}{\theta_0}\right)^{\sum X_i} \left(\frac{1-\bar{X}}{1-\theta_0}\right)^{n-\sum X_i}$$

$$\Rightarrow \hat{\Lambda} = 2\ln(\hat{LR}) = 2\left(\sum X_i \ln\left(\frac{\bar{X}}{\theta_0}\right) + (n - \sum X_i)\ln\left(\frac{1-\bar{X}}{1-\theta_0}\right)\right)$$

let O_1 = # of observed $X_i = 1$, $\bar{X} = O_1/n$
let E_1 = # of expected $X_i = 1$ under H_0, $\theta_0 = \frac{E_1}{n}$

$$= 2\left(O_1 \ln\left(\frac{O_1}{E_1}\right) + O_2 \ln\left(\frac{O_2}{E_2}\right)\right)$$

let O_2 = # of observed $X_i = 0$, $1 - \bar{X} = O_2/n$
let E_2 = # of expected $X_i = 0$, $1 - \theta_0 = E_2/n$

this is a completely different test than the Wald or Score test.

Proof of the asymptotic convergence of $2\ln(\hat{LR})$ to χ^2_1.

$$\hat{\Lambda} = 2\left(\ell(\hat{\theta}^{MLE}; X_1 \ldots X_n) - \ell(\theta_0; X_1 \ldots X_n)\right)$$

Consider the Taylor series expansion of $\ell(\theta_0; X_1, \ldots, X_n)$ centered at $\hat{\theta}^{MLE}$ at the second order:

$$\ell(\theta_0; X_1 \ldots X_n) = \ell(\hat{\theta}^{MLE}; X_1, \ldots X_n) + (\theta_0 - \hat{\theta}^{MLE})\underbrace{\ell'(\hat{\theta}^{MLE}; X_1, \ldots, X_n)}_{0} + (\theta_0 - \hat{\theta}^{MLE})^2 \frac{\ell''(\hat{\theta}^{MLE}; X_1, \ldots, X_n)}{2} + \ldots$$

$$\Rightarrow \ell(\theta_0; X_1, \ldots X_n) \approx \ell(\hat{\theta}^{MLE}; X_1, \ldots X_n) + \frac{1}{2}(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\hat{\theta}^{MLE}; X_1, \ldots X_n)$$

$$\Rightarrow \hat{\Lambda} \approx \cancel{2}\left(\cancel{\ell(\hat{\theta}^{MLE}; X_1, \ldots X_n)} - \left(\cancel{\ell(\hat{\theta}^{MLE}; X_1, \ldots X_n)} + \cancel{\tfrac{1}{2}}(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\hat{\theta}^{MLE}; X_1, \ldots X_n)\right)\right)$$

$$\Rightarrow \hat{\Lambda} = -(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\hat{\theta}^{MLE}; X_1, \ldots X_n) \cdot \frac{\ell''(\theta; X_1, \ldots X_n)}{\ell''(\theta; X_1, \ldots X_n)}$$

$$\Rightarrow \hat{\Lambda} = -(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\theta; X_1, \ldots X_n) \cdot \underbrace{\frac{\ell''(\hat{\theta}^{MLE}; X_1, \ldots X_n)}{\ell''(\theta; X_1, \ldots X_n)}}_{\hat{A}_1}$$

$$\Rightarrow \hat{\Lambda} = \frac{(\theta_0 - \hat{\theta}^{MLE})^2}{\frac{1}{-\ell''(\theta; X_1, \ldots, X_n)}} \hat{A}_1$$

$$\overset{iid}{=} \frac{(\theta_0 - \hat{\theta}^{MLE})^2}{\frac{1}{\frac{n}{n} \cdot \sum -\ell''(\theta; X_i)}} \hat{A}_1 \cdot \frac{I(\theta)}{I(\theta)} = \frac{(\theta_0 - \hat{\theta}^{MLE})^2}{\frac{1}{n I(\theta)}} \underbrace{\frac{\overbrace{\frac{1}{n}\sum -\ell''(\theta; X_i)}^{\approx E[-\ell''(\theta; X)]}}{I(\theta)}}_{\hat{A}_2} \hat{A}_1$$

$$= \Bigg(\underbrace{\frac{\theta_0 - \hat{\theta}^{MLE}}{\sqrt{\frac{I(\theta)^{-1}}{n}}}}_{\hat{B}}\Bigg)^2 \hat{A}_2 \hat{A}_1$$

$\hat{B} \xrightarrow{d} N(0,1)$ by monster MLE thm and thus $\hat{B}^2 \xrightarrow{d} \chi^2_1$

$$\frac{1}{n}\sum -\ell''(\theta; X) \xrightarrow{P} E[-\ell''(\theta; X_i)] = I(\theta)$$

$$\Rightarrow \frac{\frac{1}{n}\sum -\ell''(\theta; X)}{I(\theta)} = \hat{A}_2 \xrightarrow{P} 1 \quad \text{CMT}$$

$\hat{\theta}^{MLE} \xrightarrow{P} \theta$ by monster MLE $\Rightarrow \ell''(\hat{\theta}^{MLE}; X_1, \ldots X_n) \xrightarrow[\text{CMT}]{P} \ell''(\theta; X_1, \ldots X_n)$

$$\Rightarrow \hat{A}_1 \xrightarrow{P} 1 \quad \text{CMT}$$

New thm $\hat{A}_1 \xrightarrow{P} 1$, $\hat{A}_2 \xrightarrow{P} 1 \Rightarrow \hat{A}_1 \hat{A}_2 \xrightarrow{P} 1$

Then Slutsky's thm $\hat{B}^2 \hat{A}_1 \hat{A}_2 \xrightarrow{d} \chi^2_1$