I. NOTATIONS:

Let $l(\theta)$ be the log liklihood with $\theta \in \Theta \subseteq \mathbb{R}$

Hypothesis: $H_0: \theta = \theta_0 \qquad H_1: \theta \neq \theta_0$

Assume $\hat{\theta} = \arg\max_{\theta} l(\theta)$ to exist and be unique.

Let $\mu(\theta) = l'(\theta)$ be the score

and $I(\theta) = -E[l''(\theta)]$ be the expected Fisher information.

and $J(\theta) = -l''(\theta)$ be the observed information

with $E[\mu(\theta;Y)] = 0$

and $I(\theta) = E[J(\theta;Y)]$

and $I(\theta) = Var[l'(\theta)] = Var[\mu(\theta)]$

II. FORMAL DEF OF TESTS.

0. LRT (likelihood ratio test)

Test Statistic

$$\lambda(x) = \frac{\sup_{\Theta_0} \mathcal{L}(\theta)}{\sup_{\Theta} \mathcal{L}(\theta)} \qquad \begin{aligned} H_0&: \theta \in \Theta_0 \\ H_1&: \theta \in \Theta_0^c \end{aligned}$$

1 Wilks LRT

$$-2\log\lambda(X_n) \xrightarrow{d} \chi_d^2$$

Suppose Θ contains an open set in $\mathbb{R}^p$

Θ_0 contains an open set in $\mathbb{R}^q$

$d = p - q$

and the true parameter resides in the open set in Θ_0.

2 Asymptotic Estimators

both Wald Test and score test base on the observation that

$$\frac{\hat{\theta}_n - \theta_0}{a_n} \xrightarrow{d} N(0,1)$$

thus if we have a consistent estimator of a_n i.e.

$$\frac{S_n}{a_n} \xrightarrow{P} 1$$

the by Slutsky's Theorem

$$\frac{\hat{\theta}_n - \theta}{S_n} \xrightarrow{d} \mathcal{N}(0,1)$$

2.1 Wald Test

General Wald Test for M-estimator

$$Z_{GW} = \sqrt{n}\,\frac{\hat{\theta}_M - \theta_0}{\sqrt{\widehat{Var}_{\theta_0}(\hat{\theta}_M)}} \xrightarrow{d} \mathcal{N}(0,1)$$

2.2 Score Test

General Score Test for M-estimator

$$Z_{GS} = \sqrt{n}\,\frac{\hat{\theta}_M - \theta_0}{\sqrt{Var_{\theta_0}(\hat{\theta}_M)}} \xrightarrow{d} \mathcal{N}(0,1)$$

In case of MLE

$$\hat{\theta}_{mle} = \arg\max_{\theta} \ell(\theta)$$

$$\forall \theta, \quad Var[\ell'(\theta)] = -E[\ell''(\theta)] = I(\theta)$$

Define the Score statistic

$$S(\theta) = \frac{\partial \ell}{\partial \theta}(\theta)$$

$Var[S(\theta)] = I_n(\theta)$ Fisher Information.
$= nI(\theta)$

$$Z_S = \frac{S(\theta_0)}{\sqrt{n}\sqrt{I(\theta_0)}} \xrightarrow{d} \mathcal{N}(0,1)$$

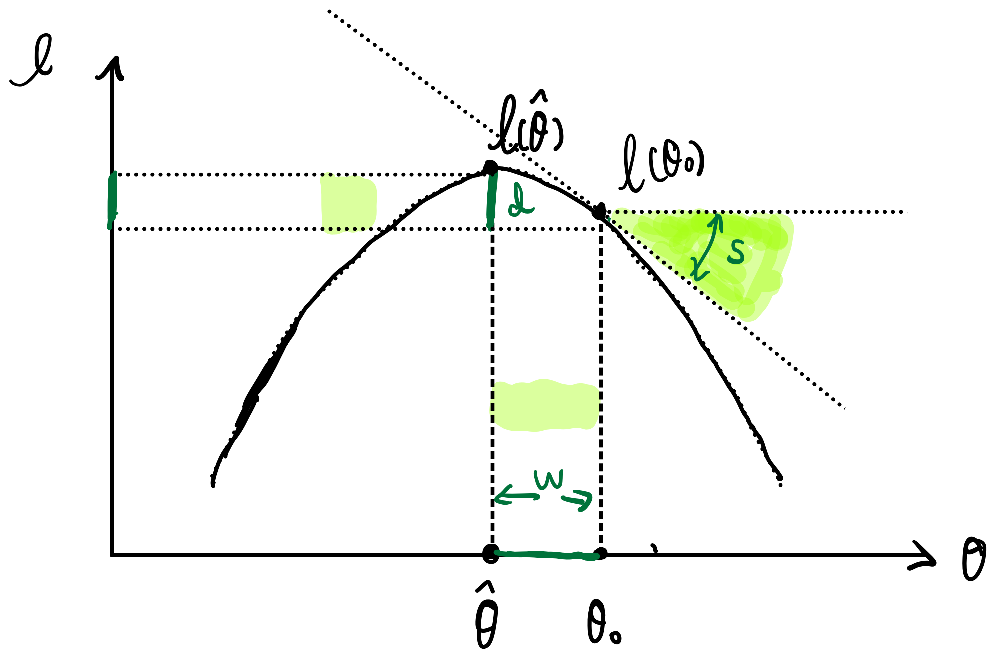
III GRAPHICAL INTERPRETATION

Consider the Holy Trinity   we ignore the difference of $J(\theta)$ and $I(\theta)$.

$$d = 2\left[\,\ell(\hat{\theta}) - \ell(\theta_0)\,\right]$$ the log likelihood scale

$$w = (\hat{\theta} - \theta_0)^2 / J(\hat{\theta})^{-1}$$ the parameter scale

$$s = u(\theta_0)^2 / J(\theta_0)$$ the 1st derivative scale.



3.1 Log likelihood Scale

Second Order Approximation

Wald Test and Score Test are both based on 2nd order approximation of the log-likelihood, but at 2 different points.

$$P_w(\theta) = \ell(\hat{\theta}) + (\theta-\hat{\theta})\,\ell'(\hat{\theta}) + \frac{1}{2}(\theta-\hat{\theta})^2\,\ell''(\hat{\theta})$$

$$P_s(\theta) = \ell(\theta_0) + (\theta-\theta_0)\,\ell'(\theta_0) + \frac{1}{2}(\theta-\theta_0)^2\,\ell''(\hat{\theta})$$

$$W = 2\left(P_w(\hat{\theta}) - P_w(\theta_0)\right)$$

$$S = 2\left(P_s(\tilde{\theta}_0) - P_s(\theta_0)\right)$$

$$\tilde{\theta}_0 = \theta_0 - \frac{\ell'(\theta_0)}{\ell''(\theta_0)}$$

is the first Newton-Raphson Iteration.

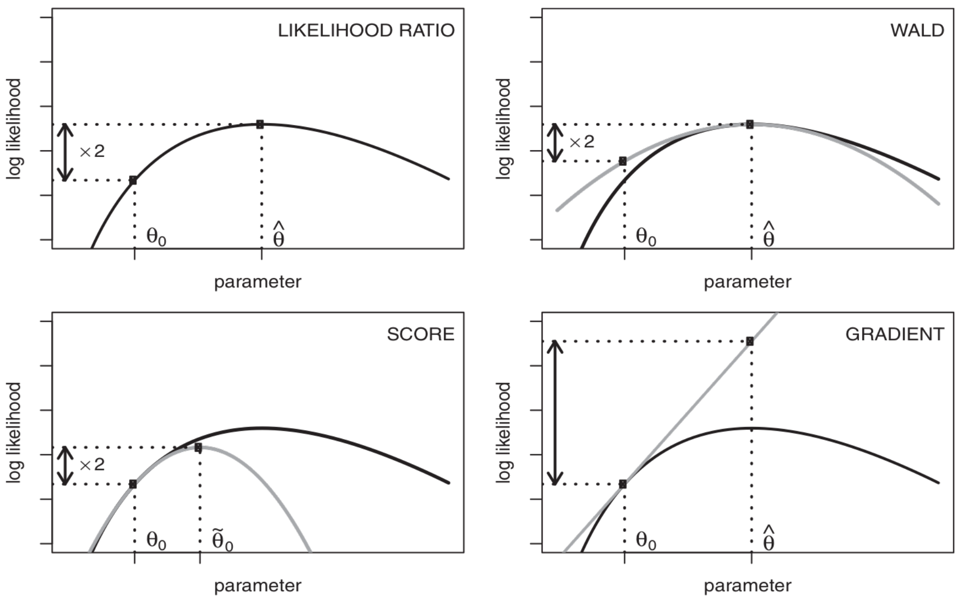


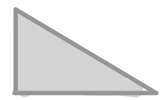
Figure 2. Comparing the four test statistics on the log-likelihood scale. On each plot the log-likelihood is illustrated (black line) along with the relevant approximation underlying the test statistic: in the Wald panel the gray line is $\mathcal{P}_w(\theta)$, in the score panel it is $\mathcal{P}_s(\theta)$, and in the gradient panel it is $\mathcal{P}_g(\theta)$. The arrows on the left side quantify the corresponding observed test statistic; the longer the arrow, the larger the evidence against H_0. Notice that for the likelihood ratio, Wald, and score, the arrow lengths have to be doubled to obtain the actual values comparable to those from the gradient statistic.

Log Likelihood Scale

3.2 Score Scale.

$$\frac{1}{2} d = \int_{\theta_0}^{\hat{\theta}} \mu(s)\, d$$

$$\frac{1}{2} w = \frac{-1}{2}(\hat{\theta}-\theta_0)^2 \cdot \mu'(\hat{\theta}) = \frac{1}{2} \overbrace{(\hat{\theta}-\theta_0)}^{a} \cdot \overbrace{\left[-\mu'(\hat{\theta})(\hat{\theta}-\theta_0)\right]}^{b}$$



$$\frac{1}{2} S = \frac{1}{2} u^2(\theta_0)^2 / -\mu'(\theta_0)$$

$$= \frac{1}{2} (\mu(\theta_0) - u(\hat{\theta}))^2 / -\mu'(\theta_0) \qquad \mu(\hat{\theta}) \longrightarrow 0$$

$$= \frac{1}{2} \underbrace{(\mu(\theta_0) - \mu(\hat{\theta}))}_{b} \cdot \underbrace{\frac{\mu(\theta_0) - \mu(\hat{\theta})}{-\mu'(\theta_0)}}_{a}$$

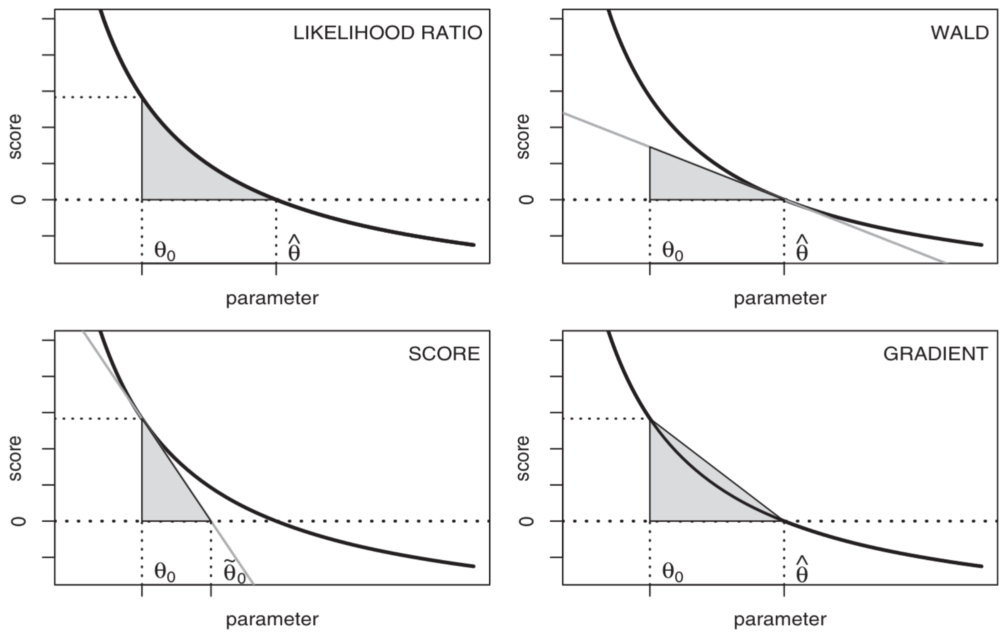
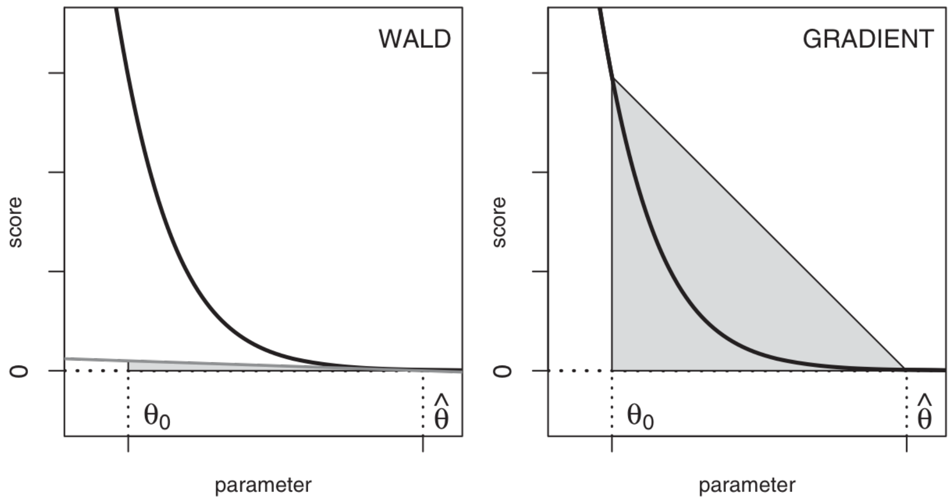


Figure 3. Illustrating the four test statistics on the score scale. In each plot the gray area represents half the actual empirical value of test statistic. The bigger the "triangle," the larger the evidence against H_0. In this example, clearly $g > d > w$ (or s), but in general the areas of triangles depend on the shape of the score $u(\cdot)$ (in turn depending on the assumed model), and on the locations of $\hat{\theta}$ and θ_0.

IV The Monotone Likelihood Problem.

The Monotone Problem is the case when the log likelihood increase monotonely but then goes in to a plateau.



Here the wald statistic is too small and thus it is powerless while the Gradient statistic is too large and thus increases the Type I error.

But the LRT and Score Test still holds.