

Chapter 6: Likelihood-based testing: JAW Ch.4 ctd

6.1 The test statistics for simple null P_{θ_0} .

(i) **Definitions: (JAW 4.11)**

(a) $2\log(\tilde{\lambda}_n) \equiv 2(\ell_n(\tilde{\theta}_n) - \ell_n(\theta_0))$

(b) $W_n \equiv n(\tilde{\theta}_n - \theta_0)^t I(\widehat{\theta}_0)(\tilde{\theta}_n - \theta_0)$ where $I(\widehat{\theta}_0)$ is any of the three consistent estimators of 5.5.

(c) $R_n \equiv Z_n^t I^{-1}(\theta_0) Z_n$ where $Z_n = n^{-\frac{1}{2}} \nabla \ell(\theta_0; X^{(n)})$.

(ii) **Dsn under simple null. (JAW 4.9)**

If $\theta = \theta_0$, each converges in dsn to χ_k^2 .

(iii) **For W_n : From 5.4 (ii) and 5.5.1**

$$\begin{aligned} n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)^t (n^{-1} J_n(\tilde{\theta}_n)) n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) &\rightarrow_d D^t I(\theta_0) D \sim \chi_k^2 \\ n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)^t I(\tilde{\theta}_n) n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) &\rightarrow_d D^t I(\theta_0) D \sim \chi_k^2 \end{aligned}$$

(iv) **For $2\log(\tilde{\lambda}_n)$: On G_n , $\nabla \ell_n(\tilde{\theta}_n) = 0$ so**

$$\begin{aligned} \ell_n(\theta_0) &= \ell_n(\tilde{\theta}_n) - \frac{1}{2}(\tilde{\theta}_n - \theta_0)^t (J_n(\theta_n^*)) (\tilde{\theta}_n - \theta_0) \\ 2\log(\tilde{\lambda}_n) &= 2(\ell_n(\tilde{\theta}_n) - \ell_n(\theta_0)) \\ &= 2\frac{1}{2}n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)^t (n^{-1} J_n(\theta_n^*)) n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \\ &\rightarrow_d D^t I(\theta_0) D \sim \chi_k^2 \end{aligned}$$

where $D \sim N_k(0, I^{-1}(\theta_0))$

(v) **For R_n , result follows trivially, since we already know $Z_n \rightarrow_d N_k(0, I(\theta_0))$.**

6.2 Distributions under fixed alternatives

(i) **Fixed alternatives: θ is true; θ_0 hypothesized.**

$$\begin{aligned}
 \text{(ii)} \quad n^{-1}2\log(\tilde{\lambda}_n) &= 2n^{-1}(\ell_n(\tilde{\theta}_n) - \ell_n(\theta_0)) \\
 &= 2n^{-1}(\ell_n(\theta) - \ell_n(\theta_0)) + 2n^{-1}(\ell_n(\tilde{\theta}_n) - \ell_n(\theta)) \\
 &\rightarrow_p 2E_\theta\left(\log \frac{f_\theta(X)}{f_{\theta_0}(X)}\right) + 0.\chi_k^2 = 2K(P_\theta, P_{\theta_0})
 \end{aligned}$$

by WLLN and 6.1 (iv).

(iii) $\tilde{\theta}_n \rightarrow_p \theta$, $\hat{I}(\tilde{\theta}_n) \rightarrow_p I(\theta)$ so, by cts mapping thm,
 $n^{-1}W_n \rightarrow_p (\theta - \theta_0)^t I(\theta)(\theta - \theta_0)$

(iv) **A6:** $E_\theta(\nabla \ell(\theta_0; X_i)) < \infty$

Then $n^{-\frac{1}{2}}Z_n = n^{-1} \nabla \ell(\theta_0; X^{(n)}) \rightarrow_p E_\theta(\nabla \ell(\theta_0; X_i))$ so
 $n^{-1}R_n \rightarrow_p E_\theta(\nabla \ell(\theta_0; X_i))^t I^{-1}(\theta_0) E_\theta(\nabla \ell(\theta_0; X_i))$

(v) **Assuming, in the case of R_n that $E_\theta(\nabla \ell(\theta_0; X_i)) \neq 0$,**
then $n^{-1} \times$ each statistic \rightarrow_p to something strictly positive
if $\theta \neq \theta_0$.

Each test has χ_k^2 dsn if $\theta = \theta_0$.

Hence if $\theta \neq \theta_0$, each test rejects $H_0 : \theta = \theta_0$ with prob.
 $\rightarrow 1$ as $n \rightarrow \infty$.

That is, each test is consistent.

(vi) To obtain non-degenerate limit dsn, we need
alternatives $\theta_n \rightarrow \theta_0$ as $n \rightarrow \infty$. In fact, local alternatives
 $\theta_n = \theta_0 + sn^{-\frac{1}{2}}$.

6.3 Distributions under alternatives $\theta_n = \theta_0 + sn^{-\frac{1}{2}}$

(i) First generalization of 5.4 Theorem for P_{θ_n}

$$\begin{aligned} (a) \quad & n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_n) \rightarrow_d D \sim N_k(0, I^{-1}(\theta_0)) \\ (b) \quad & Z_n(\theta_n) \equiv n^{-\frac{1}{2}} \nabla \ell_n(\theta_n; X^{(n)}) \rightarrow_d Z \sim N_k(0, I(\theta_0)) \end{aligned}$$

(ii) Second generalization of 5.4 Thm for P_{θ_n}

$$\begin{aligned} (a) \quad & n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \rightarrow_d N_k(s, I^{-1}(\theta_0)) \\ (b) \quad & Z_n(\theta_0) \equiv n^{-\frac{1}{2}} \nabla \ell_n(\theta_0; X^{(n)}) \rightarrow_d N_k(I(\theta_0)s, I(\theta_0)) \end{aligned}$$

(iii) Using LeCam's 3 rd Lemma, can prove (ii) and hence deduce (i) — see JAW notes 3.35,4.13. Alternatively, Ibramigov & Has'minskii (1981), can prove (i) and hence deduce (ii). This proof of (i) is as for Thm of 5.4, EXCEPT expectations are taken at $\theta_n \approx \theta_0$, and limits have to work as $\theta_n \rightarrow \theta_0$. WE ASSUME THIS IS OK!!

(iv) So ASSUME (i): now we show (ii).

$$\begin{aligned} n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) &= n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_n) + n^{\frac{1}{2}}(\theta_n - \theta_0) \\ &\rightarrow_d D + s \sim N_k(s, I^{-1}(\theta_0)) \\ Z_n(\theta_0) &= Z_n(\theta_n) + (-n^{-1}J_n(\theta_n^*))n^{\frac{1}{2}}(\theta_0 - \theta_n) \\ &\rightarrow_d Z + I(\theta_0)s \sim N_k(I(\theta_0)s, I(\theta_0)) \end{aligned}$$

(v) Under prev. assumptions, and with $\theta_n = \theta_0 + sn^{-\frac{1}{2}}$

$$\begin{aligned} 2\log \tilde{\lambda}_n &\rightarrow_d (D + s)^t I(\theta_0)(D + s) \quad \text{see 6.1} \\ W_n &\rightarrow_d (D + s)^t I(\theta_0)(D + s) \\ R_n &\rightarrow_d (Z + I(\theta_0)s)^t I^{-1}(\theta_0)(Z + I(\theta_0)s) \\ &= {}_d (D + s)^t I(\theta_0)I^{-1}(\theta_0)I(\theta_0)(D + s) \end{aligned}$$

and $(D + s)^t I(\theta_0)(D + s) \sim \chi_k^2(\delta)$ where $\delta = s^t I(\theta_0)s$.

6.4 Defns and dsns: composite null hypothesis (JAW 4.14)

(i) Partition $\theta = (\theta_1, \theta_2)$ of dim m and $k - m$.

Partition $I(\theta)$, D , Z_n , Z etc. similarly.

Let $H_0 : \theta \in \Theta_0 \subset \Theta$ where $\Theta_0 \equiv \{\theta; \theta_1 = \theta_{1,0}\}$. $H_1; \theta \in \Theta \setminus \Theta_0$

Let $\tilde{\theta}_n$ and $\tilde{\theta}_n^0 = (\theta_{1,0}, \theta_{2,n}^0)$ be consistent roots of the likelihood eqn under H_1 and H_0 .

(ii) $(I^{-1})_{11} = (I_{11.2})^{-1}$; $I_{11.2} = I_{11} - I_{12}I_{22}^{-1}I_{21}$ and

$(I^{-1})_{12} = -I_{11}^{-1}I_{12}(I^{-1})_{22}$ with I symm., and $1 \leftrightarrow 2$ eqns also.

(iii) $2 \log \tilde{\lambda}_n \equiv 2 \log(L(\Theta; X^{(n)})/L(\Theta_0; X^{(n)})) = 2(\ell_n(\tilde{\theta}_n) - \ell_n(\tilde{\theta}_n^0))$

$$W_n \equiv n(\tilde{\theta}_{n1} - \theta_{1,0})^t \hat{I}_{11.2}(\tilde{\theta}_{n1} - \theta_{1,0})$$

$$R_n \equiv Z_n^t(\tilde{\theta}_n^0)I^{-1}(\tilde{\theta}_n^0)Z_n(\tilde{\theta}_n^0).$$

(iv) Suppose $\theta = \theta_0 = (\theta_1^0, \theta_2^0)$ where $\theta_1^0 = \theta_{1,0}$ so H_0 is true.

Now $n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \rightarrow_d D \sim N_k(0, I^{-1}(\theta_0))$ so $n^{\frac{1}{2}}(\tilde{\theta}_{1,n} - \theta_{1,0}) \rightarrow_d D_1 \sim N_m(0, (I^{-1})_{11}) \equiv N_m(0, I_{11.2}^{-1})$

(v) $Z_n(\tilde{\theta}_n^0) = (Z_{n,1}(\tilde{\theta}_n^0), Z_{n,2}(\tilde{\theta}_n^0))^t$. Note, by defn,

$$Z_{n,2}(\tilde{\theta}_n^0) = n^{-\frac{1}{2}} \frac{\partial}{\partial \theta_2} \ell(\theta_{1,0}, \theta_2) \Big|_{\theta_2 = \theta_{2,n}^0} = 0, \text{ and } (\tilde{\theta}_n^0 - \theta_0) =$$

$(0, \theta_{2,n}^0 - \theta_2^0)$. Also, $\tilde{\theta}_n^0$ is estimated with $\theta_1 = \theta_1^0$ fixed and true, so $n^{\frac{1}{2}}(\theta_{2,n}^0 - \theta_2^0) \rightarrow_d I_{22}^{-1}Z_2$. (NOT $I^{22} = (I^{-1})_{22}$).

(vi) Then, with $|\theta_n^* - \theta_0| < |\tilde{\theta}_n^0 - \theta_0|$

$$\begin{aligned} Z_{n,1}(\tilde{\theta}_n^0) &= n^{-\frac{1}{2}} \frac{\partial \ell_n}{\partial \theta_1} \Big|_{\tilde{\theta}_n^0} \\ &= Z_{n,1}(\theta_0) + (-n^{-1}J_{12}(\theta_n^*))n^{\frac{1}{2}}(\theta_{2,n}^0 - \theta_2^0) \\ &\rightarrow_d Z_1 - I_{12}(\theta_0)I_{22}^{-1}(\theta_0)Z_2 \equiv Z^* \equiv Z_{1.2} \end{aligned}$$

where $Z = (Z_1, Z_2)^t = I(\theta_0)D \sim N_k(0, I(\theta_0))$.

(vii) Now $\text{var}(Z_1 - I_{12}I_{22}^{-1}Z_2) = I_{11} - I_{12}I_{22}^{-1}I_{21} = I_{11.2}$, so overall

$$Z_n(\tilde{\theta}_n^0) \rightarrow_d (Z^*, 0)^t \sim (N_m(0, I_{11.2}), 0)^t.$$

6.5 Dsns of text statistics, under null and local alternatives

(i) **Theorem 1:** If assumptions holds as previously, and $\theta_0 \in \Theta_0$ (H_0 true) then $T_n \rightarrow_d D_1^t I_{11.2} D_1 \sim \chi_{k-(k-m)}^2 \equiv \chi_m^2$, where T_n is any of $2 \log \tilde{\lambda}_n$, W_n , or R_n .

(ii) $W_n \rightarrow_d D_1^t I_{11.2} D_1 \sim \chi_m^2$ by (iv) above
 $R_n \rightarrow_d (Z^*, 0) I^{-1}(\theta_0) (Z^*, 0)^t = (Z^*)^t I_{11.2}^{-1}(\theta_0) Z^* \sim \chi_m^2$ by (vi), (vii) above.

$$\begin{aligned} 2 \log \tilde{\lambda}_n &= 2(\ell_n(\tilde{\theta}_n) - \ell_n(\tilde{\theta}_n^0)) \\ &= 2(\ell_n(\tilde{\theta}_n) - \ell_n(\theta_0)) - 2(\ell_n(\tilde{\theta}_n^0) - \ell_n(\theta_0)) \\ &\rightarrow_d Z^t I^{-1} Z - Z_2^t I_{22}^{-1} Z_2 \end{aligned}$$

(iii) Now recall $Z^* = Z_1 - I_{12} I_{22}^{-1} Z_2$, so

$$Z^t I^{-1} Z = (Z^* + I_{12} I_{22}^{-1} Z_2, Z_2)^t I^{-1} (Z^* + I_{12} I_{22}^{-1} Z_2, Z_2)$$

Quadr term in Z^* is $(Z^*)^t (I^{-1})_{11} Z^* \equiv (Z^*)^t I_{11.2}^{-1} Z^*$

Linear term in Z^* is $2Z_2^t (I_{22}^{-1} I_{21} (I^{-1})_{11} + (I^{-1})_{21}) Z^* \equiv 0$

Quadr term in Z_2 is $Z_2^t H Z_2$ with $H = \dots = \dots = I_{22}^{-1}$.

So, putting it together $2 \log \tilde{\lambda}_n \rightarrow_d (Z^*)^t I_{11.2}^{-1} Z^*$

(We will do this arithmetic – but no time to latex it here.)

(iv) Define $Z^* = Z_1 - I_{12} I_{22}^{-1} Z_2 \equiv Z_{1.2}$, then

(a) $D = (I(\theta_0))^{-1} Z = (I_{11.2}^{-1} Z_{1.2}, I_{22.1}^{-1} Z_{2.1})^t$

(b) $\text{Cov}(Z_{1.2}, Z_2) = \text{Cov}(Z_{2.1}, Z_1) = 0$

(v) **Theorem 2 (not proved):**

Under all previous assumptions and conditions, and $\theta_n = \theta_0 + sn^{-\frac{1}{2}}$, $\theta_0 \in \Theta_0$, then, under P_{θ_n} , $2 \log \tilde{\lambda}_n, W_n, R_n \rightarrow_d$

$(D_1 + s_1)^t I_{11.2} (D_1 + s_1) \sim \chi_m^2(\delta)$ where $\delta = s_1^t I_{11.2} s_1$.

6.6 Practicalities: a useful table

	$2\log \tilde{\lambda}_n$	R_n	W_n
(a) Simple $H_0; \theta = \theta_0$			
Max. (unconstr.)	Yes	No	Yes
Eval. $\ell_n(\theta)$	Yes	No	No
Eval. $\nabla \ell_n(\theta)$	No	Yes	No
Eval. $I(\theta)$	No	Yes	(Yes*)
(b) Composite $H_0: \theta_1 = \theta_{1,0}$			
Max. (unconstr.)	Yes	No	Yes
Max. (constrained)	Yes	Yes	No
Eval. $\ell_n(\theta)$	Yes	No	No
Eval. $\nabla \ell_n(\theta)$	No	Yes	No
Eval. $I(\theta)$	No	Yes	(Yes**)

*: or some variance estimate for MLE.

** : Need to evaluate $I_{11,2}$, or some consistent estimator.

Note about dimensions and units:

(a) Only matrices of appropriate dimensions can be multiplied, for example $I_{12}I_{22}^{-1}(I^{-1})_{21}I_{11,2}I_{12}\dots$

(b) Vectors such as $Z_n, \tilde{\theta}_n$ etc are column vectors, transposes t should be interpreted appropriately (and may be wrong): for example $(Z_1, Z_2)^t$ denotes the $k \times 1$ column vector of Z_1 piled above Z_2 .

(c) Let units of θ be u , so also $\tilde{\theta}_n, \tilde{\theta}_n^0, D$.

Units of $\text{var}(\tilde{\theta}_n)$ would be u^2 , so of $I(\theta)$ is u^{-2} .

ℓ_n (and any standardized test statistic) should be u^0 .

Score = $\partial \ell_n / \partial \theta$, has dim u^{-1} , so also Z .

(d) Equations such as $(I^{-1})_{11} = (I_{11} - I_{12}I_{22}^{-1}I_{21})^{-1}$ must have matching units – here I -inverse units.

6.7 Reparametrization

(i) A hypothesis $\mu = \lambda$ could be parametrized as $\psi = (\mu - \lambda) = 0$, $\psi = (\mu/\lambda - 1) = 0$, $\psi = (\lambda/\mu - 1) = 0$, $\psi = \log(\mu/\lambda) = 0$ etc.

Does it matter which ψ we choose?

Let $q(\theta)$ be a 1-1 transf $\mathfrak{R}^k \rightarrow \mathfrak{R}^k$.

(ii) $2 \log \tilde{\lambda}_n \equiv 2(\ell_n(\tilde{\theta}_n) - \ell_n(\tilde{\theta}_n^0)) = 2(\ell_n(\tilde{q}_n) - \ell_n(\tilde{q}_n^0))$.

Maximized lhds unchanged by reparametrization – so $2 \log \tilde{\lambda}_n$ is invariant.

(iii) $R_n \equiv Z_n^t(\tilde{\theta}_n^0) I^{-1}(\tilde{\theta}_n^0) Z_n(\tilde{\theta}_n^0)$ is invariant, since

$Z_n(q) = (\nabla_q(\theta)) Z_n(\theta)$ and $I(q) = (\nabla_q(\theta)) I(\theta) (\nabla_q(\theta))^t$, so

$R_n(q) = (Z_n(\theta))^t (\nabla_q(\theta))^t ((\nabla_q(\theta))^t)^{-1} I^{-1}(\theta) (\nabla_q(\theta))^{-1} (\nabla_q(\theta)) Z_n(\theta) \equiv R_n(\theta)$. $((\nabla_q(\theta))_{ij} = \frac{d\theta_j}{dq_i})$.

(iv) Note the Rao statistic may be computed in any parametrization, giving the same statistic, but it is simplest to compute for q for which $H_0 : q_1 = q_{1,0}$ since then $Z_{n,2}(\tilde{q}_n^0) \equiv 0$.

(v) In θ parametrization: $W_n \equiv n(\tilde{\theta}_{n1} - \theta_{1,0})^t \hat{I}_{11 \cdot 2}(\tilde{\theta}_{n1} - \theta_{1,0})$ is quadratic form in $(\tilde{\theta}_{n1} - \theta_{1,0})$.

In $q(\theta)$ parametrization, W_n is quadratic form in $(\tilde{q}_n - q(\theta_{1,0}))$: which will not in general have corresponding 0 components. W_n is NOT invariant under non-linear transformations of the parameters.