§6 Hypothesis Testing

§6.1 Introduction

6.1.1 $\boldsymbol{X}$: observed dataset $\sim$ probability function $f(\cdot|\theta)$
θ: unknown parameter

A *hypothesis test* of

$$\textit{(null hypothesis) } H_0 : \theta \in \Theta_0 \quad \text{vs} \quad \textit{(alternative hypothesis) } H_1 : \theta \in \Theta_1$$

seeks from empirical data evidence that supports rejection of H_0 in favour of H_1.

6.1.2 **Definition.** A hypothesis test may take the form of a *test function* $\varphi(\boldsymbol{X})$, such that

H_0 is rejected with probability $\varphi(\boldsymbol{X})$.

<u>Special case:</u> Given a *critical region* $\mathcal{C} \subset$ sample space,

$$\varphi(\boldsymbol{X}) = \mathbf{1}\{\boldsymbol{X} \in \mathcal{C}\} \quad \longrightarrow \quad \text{reject } H_0 \text{ if observe } \boldsymbol{X} \in \mathcal{C}.$$

6.1.3 **Definition.** If a critical region has the form $\{\boldsymbol{X} : T(\boldsymbol{X}) > c\}$ for some constant c and statistic $T(\boldsymbol{X})$, then $T(\boldsymbol{X})$ is called a *test statistic* and c is called a *critical value*.

6.1.4 **Definition.** The *power function* of the test φ is

$$w(\theta) = \mathbb{E}_\theta\big[\varphi(\boldsymbol{X})\big] = \mathbb{E}_\theta\big[\mathbb{P}(\text{reject } H_0|\boldsymbol{X})\big] = \mathbb{P}_\theta(\text{reject } H_0).$$

- for $\theta \in \Theta_0$, $w(\theta) =$ probability of rejecting **true** H_0 (i.e. *type I error*) at θ
- for $\theta \notin \Theta_0$, $1 - w(\theta) =$ probability of accepting **false** H_0 (i.e. *type II error*) at θ

6.1.5 **Definition.** The *size* of a test φ is defined to be $\sup_{\theta\in\Theta_0} w(\theta)$.
Definition. The *power* of the test φ at $\theta \in \Theta_1 \setminus \Theta_0$ is defined to be $w(\theta)$.

The size of a test gives its maximum *Type I error* probability, while the power of a test equals $1 -$ *Type II error* probability at a particular $\theta \in \Theta_1 \setminus \Theta_0$.

6.1.6 In hypothesis testing, one typically seeks a test φ whose <u>**size**</u> is kept below some prescribed level α and whose <u>**power**</u> is as large as possible for $\theta \in \Theta_1 \setminus \Theta_0$.

6.1.7 Let $\mathscr{T}$ be a collection of test functions such that

- $\mathscr{T}$ contains the trivial tests $\varphi \equiv 0$ and $\varphi \equiv 1$,
- for any $\varphi, \varphi^* \in \mathscr{T}$, we have either $\varphi(\boldsymbol{x}) \leq \varphi^*(\boldsymbol{x})\ \forall \boldsymbol{x}$ or $\varphi^*(\boldsymbol{x}) \leq \varphi(\boldsymbol{x})\ \forall \boldsymbol{x}$.

The above conditions imply that we can rank the tests in $\mathscr{T}$ by their degree of conservativeness ("reluctance to reject"), with $\varphi \equiv 0$ and $\varphi \equiv 1$ being the most and least conservative, respectively.

Definition. Suppose that $\boldsymbol{X} = \boldsymbol{x}$ is observed. With respect to the test function collection $\mathscr{T}$, we may define the p-value $pv(\boldsymbol{x})$ to be the smallest size of the test $\varphi \in \mathscr{T}$ with $\varphi(\boldsymbol{x}) > 1/2$, i.e.

$$pv(\boldsymbol{x}) = \inf\Big\{ \sup_{\theta\in\Theta_0} \mathbb{E}_\theta[\varphi(\boldsymbol{X})] : \varphi(\boldsymbol{x}) > 1/2,\ \varphi \in \mathscr{T}\Big\}.$$

The **smaller** the value of $pv(\boldsymbol{x})$, the **stronger** is the evidence contained in $\boldsymbol{x}$ against H_0.

Special case:

If $\mathcal{T} = \big\{\mathbf{1}\{T(\cdot) > \tau\} : -\infty \leq \tau \leq \infty\big\}$ for some test statistic $T = T(\boldsymbol{X})$, then

$$pv(\boldsymbol{x}) = \inf_{\tau < T(\boldsymbol{x})} \sup_{\theta\in\Theta_0} \mathbb{P}_\theta\big(T(\boldsymbol{X}) > \tau\big) = \sup_{\theta\in\Theta_0} \mathbb{P}_\theta\big(T(\boldsymbol{X}) \geq T(\boldsymbol{x})\big).$$

§6.2 Likelihood ratio test

6.2.1 **Definition.** The *likelihood ratio* for the test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, given data $\boldsymbol{X}$, is defined to be

$$\Lambda_{\boldsymbol{X}}(H_0, H_1) = \frac{\sup_{\theta\in\Theta_1} \ell_{\boldsymbol{X}}(\theta)}{\sup_{\theta\in\Theta_0} \ell_{\boldsymbol{X}}(\theta)},$$

where $\ell_{\boldsymbol{X}}(\theta)$ is the likelihood function.

The likelihood ratio may be loosely interpreted as the *odds* of H_1 against H_0.

6.2.2 **Definition.** A *likelihood ratio test* uses $\Lambda_{\boldsymbol{X}}(H_0, H_1)$ as the test statistic, with large values of $\Lambda_{\boldsymbol{X}}(H_0, H_1)$ being evidence against H_0 in favour of H_1.

6.2.3 Based on observed data $\boldsymbol{x}$, the likelihood ratio test yields the p-value

$$pv(\boldsymbol{x}) = \sup_{\theta\in\Theta_0} \mathbb{P}\big(\Lambda_{\boldsymbol{X}}(H_0, H_1) \geq \Lambda_{\boldsymbol{x}}(H_0, H_1)\big).$$

6.2.4 **Example §6.2.1** Let $\boldsymbol{X} = (M, N)$ be a pair of independent random variables such that $N - \beta \sim \text{Poisson}\,(\lambda)$ and $(M - \beta)|N \sim \text{binomial}\,(N - \beta, \rho)$, for unknown parameters $\lambda > 0$, $\rho \in [0, 1]$ and $\beta \in \{0, 1\}$.

Mass function of $\boldsymbol{X} = (M, N)$:

$$f(m, n|\lambda, \rho, \beta) = \frac{e^{-\lambda}\lambda^{n-\beta}}{(n-\beta)!}\binom{n-\beta}{m-\beta}\rho^{m-\beta}(1-\rho)^{n-m}\,\mathbf{1}\{n \geq m \geq \beta\}.$$

We wish to test

$$H_0 : \beta = 0 \quad \text{against} \quad H_1 : \beta = 1.$$

Notes:

- 0^0 is interpreted as 1 in the following expressions.
- Elementary probability calculations show that $M - \beta \sim$ Poisson $(\lambda\rho)$ unconditionally.

Likelihood: $\ell_{\boldsymbol{X}}(\lambda, \rho, \beta) \propto \mathbf{1}\{\beta \leq M\}\, e^{-\lambda}\lambda^{N-\beta}\rho^{M-\beta}(1-\rho)^{N-M}/(M-\beta)!$

$\rightarrow$ likelihood ratio test statistic:

$$\begin{aligned}
\Lambda_{\boldsymbol{X}}(H_0, H_1) &= \frac{\sup\limits_{\lambda>0,\, \rho\in[0,1]} \ell_{\boldsymbol{X}}(\lambda, \rho, 1)}{\sup\limits_{\lambda>0,\, \rho\in[0,1]} \ell_{\boldsymbol{X}}(\lambda, \rho, 0)} \\
&= \frac{\mathbf{1}\{M \geq 1\}\, e^{-(N-1)}(N-1)^{N-1}\left(\frac{M-1}{N-1}\right)^{M-1}\left(\frac{N-M}{N-1}\right)^{N-M}/(M-1)!}{e^{-N}N^N\left(\frac{M}{N}\right)^M\left(\frac{N-M}{N}\right)^{N-M}/M!} \\
&= e\,\mathbf{1}\{M \geq 1\}\,(1-1/M)^{M-1}\begin{cases} = 0, & M = 0, \\ \downarrow \text{ strictly}, & \text{as } M \uparrow \text{ on } [1, \infty), \\ 1, & M = \infty. \end{cases}
\end{aligned}$$

Based on observed data $\boldsymbol{X} = (m, n)$, the p-value is given by

$$\begin{aligned}
pv(m, n) &= \sup_{\lambda>0,\, \rho\in[0,1],\, \beta=0} \mathbb{P}_{\lambda,\rho,\beta}\big(\mathbf{1}\{M \geq 1\}\,(1-1/M)^{M-1} \geq \mathbf{1}\{m \geq 1\}\,(1-1/m)^{m-1}\big) \\
&= \begin{cases} 1, & m = 0, \\ \sup_{\lambda>0,\, \rho\in[0,1]} \mathbb{P}_{\lambda,\rho,0}\big(1 \leq M \leq m\big), & m \geq 1. \end{cases}
\end{aligned}$$

Under $\beta = 0$, we have $M \sim$ Poisson $(\lambda\rho)$, so that for $m \geq 1$,

$$\sup_{\lambda>0,\, \rho\in[0,1]} \mathbb{P}_{\lambda,\rho,0}\big(1 \leq M \leq m\big) = \sup_{\lambda>0,\, \rho\in[0,1]} \sum_{j=1}^{m} \frac{e^{-\lambda\rho}(\lambda\rho)^j}{j!} = \sup_{\theta>0} \sum_{j=1}^{m} \frac{e^{-\theta}\theta^j}{j!}.$$

To maximise $\sum_{j=1}^m e^{-\theta}\theta^j/j!$ over $\theta > 0$, consider

$$\frac{\partial}{\partial\theta}\ln\Big(\sum_{j=1}^m \frac{e^{-\theta}\theta^j}{j!}\Big) = \frac{1-\theta^m/m!}{\sum_{j=1}^m \theta^j/j!}\begin{cases} > 0, & \theta < (m!)^{1/m}, \\ < 0, & \theta > (m!)^{1/m}. \end{cases}$$

It follows that $\sum_{j=1}^m e^{-\theta}\theta^j/j!$ is maximised at $\theta = (m!)^{1/m}$, so that

$$pv(m,n) = e^{-(m!)^{1/m}} \sum_{j=1}^m \frac{(m!)^{j/m}}{j!}, \quad \text{for } m \geq 1.$$

§6.3 Criteria for optimal tests

6.3.1 **Definition.** The test φ is *unbiased of size* α if

$$\sup_{\theta\in\Theta_0} \mathbb{E}_\theta\big[\varphi(\boldsymbol{X})\big] = \alpha \quad \text{and} \quad \mathbb{E}_\theta\big[\varphi(\boldsymbol{X})\big] \geq \alpha \ \ \forall\,\theta \in \Theta_1 \setminus \Theta_0.$$

An unbiased test of size α has *Type II error* probability $\leq 1-\alpha$.

6.3.2 **Definition.** A test φ_0 is *uniformly most powerful* (UMP) among all tests φ of size $\leq \alpha$ if

$$\begin{cases} \mathbb{E}_\theta\big[\varphi_0(\boldsymbol{X})\big] \leq \alpha \ \ \forall\,\theta \in \Theta_0, \quad \text{and} \\ \mathbb{E}_\theta\big[\varphi_0(\boldsymbol{X})\big] \geq \mathbb{E}_\theta\big[\varphi(\boldsymbol{X})\big] \ \ \forall\,\theta \in \Theta_1 \setminus \Theta_0 \text{ and } \forall \text{ tests } \varphi \text{ of size } \leq \alpha. \end{cases}$$

6.3.3 Sometimes, for rather general Θ_0, Θ_1, we cannot find the UMP test. But if we restrict consideration to **unbiased** tests, we may be able to find the UMP test.

6.3.4 **Definition.** A test φ_0 is *uniformly most powerful unbiased* (UMPU) among all unbiased tests φ of size $\leq \alpha$ if

$$\begin{cases} \mathbb{E}_\theta\big[\varphi_0(\boldsymbol{X})\big] \leq \alpha \ \ \forall\,\theta \in \Theta_0, \quad \text{and} \\ \mathbb{E}_\theta\big[\varphi_0(\boldsymbol{X})\big] \geq \mathbb{E}_\theta\big[\varphi(\boldsymbol{X})\big] \ \ \forall\,\theta \in \Theta_1 \setminus \Theta_0 \text{ and } \forall \text{ unbiased tests } \varphi \text{ of size } \leq \alpha. \end{cases}$$

6.3.5 A UMP (or UMPU) test among all (or all unbiased) tests of size $\leq \alpha$ is necessarily unbiased.

Proof: The test $\varphi \equiv \alpha$ has $\mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \alpha\ \forall\,\theta$, so that it has size $\leq \alpha$ and is unbiased. Thus, the UMP (or UMPU) test φ_0 satisfies $\mathbb{E}_\theta[\varphi_0(\boldsymbol{X})] \geq \mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \alpha \geq \sup_{\theta\in\Theta_0}\mathbb{E}_\theta[\varphi_0(\boldsymbol{X})]\ \ \forall\,\theta \in \Theta_1 \setminus \Theta_0$. ∎

§6.4 UMP test under monotone likelihood ratio

6.4.1 **Lemma.** *(Neyman-Pearson)*

For testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, the likelihood ratio test of size α is UMP among all tests of size $\leq \alpha$.

. . . .

Proof:

Let $\varphi_0(\boldsymbol{X}) = \mathbf{1}\{f(\boldsymbol{X}|\theta_1)/f(\boldsymbol{X}|\theta_0) > k\}$ *be the test function of the likelihood ratio test. Then* $k > 0$ *necessarily. Take any test* φ *of size* $\leq \alpha$ *such that*

$$\mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})] \leq \alpha = \mathbb{E}_{\theta_0}[\varphi_0(\boldsymbol{X})].$$

Consider

$$[\varphi_0(\boldsymbol{X}) - \varphi(\boldsymbol{X})]\,[f(\boldsymbol{X}|\theta_1) - kf(\boldsymbol{X}|\theta_0)] \geq 0,$$

so that

$$\int [\varphi_0(\boldsymbol{x}) - \varphi(\boldsymbol{x})]\,[f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)]\,d\boldsymbol{x} \geq 0$$

$$\Rightarrow \quad \mathbb{E}_{\theta_1}[\varphi_0(\boldsymbol{X})] - \mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})] \geq k\,\{\mathbb{E}_{\theta_0}[\varphi_0(\boldsymbol{X})] - \mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})]\} \geq 0 \quad \Rightarrow \quad \mathbb{E}_{\theta_1}[\varphi_0(\boldsymbol{X})] \geq \mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})]. \quad \blacksquare$$

Note: Since H_1 consists of only a singleton $\{\theta_1\}$, we may drop the word "uniformly" and simply say the likelihood ratio test is *most powerful* among all tests of size $\leq \alpha$.

6.4.2 **Example.**

$\boldsymbol{X} = (X_1, \ldots, X_n)$ iid $\sim \exp(\theta)$

Likelihood $\ell_{\boldsymbol{X}}(\theta) \propto \theta^n \exp\left(-\theta \sum_{i=1}^n X_i\right)$

Test $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ $\quad$ [$\theta_1 > \theta_0 > 0$ fixed]

Likelihood ratio: $\Lambda_{\boldsymbol{X}}(H_0, H_1) = (\theta_1/\theta_0)^n \exp\left\{-(\theta_1 - \theta_0)\sum_{i=1}^n X_i\right\}$.

Since $\Lambda_{\boldsymbol{X}}(H_0, H_1)$ is decreasing in $\sum_{i=1}^n X_i$, a likelihood ratio test has critical region equivalent to $\left\{\boldsymbol{X} : \sum_{i=1}^n X_i < c\right\}$, for some $c > 0$. It has size

$$\alpha = \mathbb{P}_{\theta_0}\Big(\sum_{i=1}^n X_i < c\Big) = \mathbb{P}_{\theta_0}\big(\text{Gamma }(n,1) < \theta_0 c\big) = \int_0^{\theta_0 c} \frac{y^{n-1}e^{-y}}{(n-1)!}\,dy,$$

since $\theta \sum_{i=1}^n X_i \sim$ Gamma $(n, 1)$. Thus,

<u>*likelihood ratio test*</u>*: reject* H_0 *if* $\sum_{i=1}^n X_i < c$

is most powerful among all tests of size $\leq \alpha = \displaystyle\int_0^{\theta_0 c} \frac{y^{n-1}e^{-y}}{(n-1)!}\,dy$.

6.4.3 **Definition.**

Data $\boldsymbol{X} \sim f(\cdot|\theta)$ (probability function)
Test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$

If there exists a statistic $T = T(\boldsymbol{X})$ such that for any $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$, the ratio $f(\boldsymbol{X}|\theta_1)/f(\boldsymbol{X}|\theta_0)$ is **nondecreasing** in $T(\boldsymbol{X})$, then the parametric family $\big\{f(\cdot|\theta) : \theta \in \Theta\big\}$ has *monotone likelihood ratio* (mlr) in $T(\boldsymbol{X})$ (w.r.t. the test of H_0 vs H_1).

6.4.4 **Examples.**

(i) Data $\boldsymbol{X} \sim$ probability function $f(\boldsymbol{x}|\theta) = c(\theta)h(\boldsymbol{x})\exp\big\{\theta\, t(\boldsymbol{x})\big\}$
Test $H_0 : \theta \le \theta^*$ vs $H_1 : \theta > \theta^*$ (some given constant θ^*)
For any $\theta_0 \le \theta^*$ and $\theta_1 > \theta^*$, we have

$$\frac{f(\boldsymbol{X}|\theta_1)}{f(\boldsymbol{X}|\theta_0)} = \frac{c(\theta_1)}{c(\theta_0)} \exp\big\{(\theta_1 - \theta_0)t(\boldsymbol{X})\big\},$$

which is nondecreasing in the natural statistic $t(\boldsymbol{X})$ (since $\theta_1 > \theta_0$). Thus

the model has mlr in $t(\boldsymbol{X})$.

Clearly, the model also has mlr in $t(\boldsymbol{X})$ w.r.t. the test of $H_0 : \theta \le \theta^*$ vs $H_1 : \theta \ge \theta^*$.

(ii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid $\sim U[0, \theta]$
Test $H_0 : \theta \ge c$ vs $H_1 : \theta < c$ (some given constant $c > 0$)
For any $\theta_0 \ge c$ and $\theta_1 < c$, we have

$$f(\boldsymbol{X}|\theta_1)/f(\boldsymbol{X}|\theta_0) = (\theta_0/\theta_1)^n \mathbf{1}\{\max_i(X_i) \le \theta_1\}/\mathbf{1}\{\max_i(X_i) \le \theta_0\},$$

which is nondecreasing in $T(\boldsymbol{X}) = -\max_i(X_i)$ (since $\theta_1 < \theta_0$). Thus

the model has mlr in $T(\boldsymbol{X}) = -\max_i(X_i)$.

Clearly, the model also has mlr in $-\max_i(X_i)$ w.r.t. the test of $H_0 : \theta \ge c$ vs $H_1 : \theta \le c$.

(iii) Test $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ (some given constants θ_0, θ_1)
It is trivial that

$$f(\boldsymbol{X}|\theta_1)/f(\boldsymbol{X}|\theta_0) = \Lambda_{\boldsymbol{X}}(H_0, H_1)$$

is nondecreasing in $T(\boldsymbol{X}) = \Lambda_{\boldsymbol{X}}(H_0, H_1)$. Thus, w.r.t. any test of simple against simple hypotheses, it is always true that

the model has mlr in $T(\boldsymbol{X}) = \Lambda_{\boldsymbol{X}}(H_0, H_1)$.

6.4.5 The mlr property enables us to construct UMP tests for θ easily.

Theorem.

Consider the test of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$.

Suppose $\{f(\cdot|\theta) : \theta \in \Theta\}$ has mlr in T w.r.t. the test. Define, for some constant t_0, the test function

$$\varphi_0(\boldsymbol{X}) = \mathbf{1}\{T(\boldsymbol{X}) > t_0\}.$$

Then the following results hold.

(i) φ_0 is a likelihood ratio test.

(ii) $\sup_{\theta\in\Theta_0} \mathbb{E}_\theta[\varphi_0(\boldsymbol{X})] \leq \inf_{\theta\in\Theta_1} \mathbb{E}_\theta[\varphi_0(\boldsymbol{X})]$.

(iii) φ_0 is **UMP** among all tests of size $\leq \sup_{\theta\in\Theta_0} \mathbb{E}_\theta[\varphi_0(\boldsymbol{X})]$.

. . . .

Proof:

(i) Fix $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $T(\boldsymbol{x}) \geq T(\boldsymbol{y})$. For any $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$, we have

$$\begin{aligned} \sup_{\theta\in\Theta_1} f(\boldsymbol{x}|\theta)/f(\boldsymbol{x}|\theta_0) &\geq f(\boldsymbol{x}|\theta_1)/f(\boldsymbol{x}|\theta_0) \\ &\geq f(\boldsymbol{y}|\theta_1)/f(\boldsymbol{y}|\theta_0). \quad \textit{(by mlr property)} \end{aligned}$$

The above relation holds for all $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$, and so

$$\Lambda_{\boldsymbol{x}}(H_0, H_1) = \inf_{\theta_0\in\Theta_0} \sup_{\theta\in\Theta_1} f(\boldsymbol{x}|\theta)/f(\boldsymbol{x}|\theta_0) \geq \inf_{\theta_0\in\Theta_0} \sup_{\theta_1\in\Theta_1} f(\boldsymbol{y}|\theta_1)/f(\boldsymbol{y}|\theta_0) = \Lambda_{\boldsymbol{y}}(H_0, H_1),$$

which implies that $\Lambda_{\boldsymbol{x}}(H_0, H_1)$ is nondecreasing in $T(\boldsymbol{x})$.

(ii) Take any $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$. By mlr, φ_0 defines a likelihood ratio test of $\theta = \theta_0$ against $\theta = \theta_1$, with

$$\textit{size} = \mathbb{E}_{\theta_0}[\varphi_0(\boldsymbol{X})] = \alpha_0 \textit{ say}, \quad \textit{power} = \mathbb{E}_{\theta_1}[\varphi_0(\boldsymbol{X})].$$

By Neyman-Pearson Lemma, φ_0 is most powerful among all tests of size $\leq \alpha_0$. The trivial test $\varphi_r(\boldsymbol{X}) \equiv \alpha_0$ has size = power = α_0. Neyman-Pearson Lemma implies that

$$\mathbb{E}_{\theta_1}[\varphi_0(\boldsymbol{X})] \geq \mathbb{E}_{\theta_1}[\varphi_r(\boldsymbol{X})] = \alpha_0 = \mathbb{E}_{\theta_0}[\varphi_0(\boldsymbol{X})].$$

The result (ii) follows since $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$ are arbitrary.

(iii) Define $\alpha = \sup_{\theta\in\Theta_0} \mathbb{E}_\theta\big[\varphi_0(\boldsymbol{X})\big]$. Then for any $\alpha' < \alpha$, there exists $\theta^* \in \Theta_0$ with $\alpha' < \mathbb{E}_{\theta^*}[\varphi_0(\boldsymbol{X})] \le \alpha$. Let φ be an arbitrary test of size $\le \alpha$. Then necessarily $\mathbb{E}_{\theta^*}[(\alpha'/\alpha)\varphi(\boldsymbol{X})] \le \alpha' < \mathbb{E}_{\theta^*}[\varphi_0(\boldsymbol{X})]$. Fix any $\theta_1 \in \Theta_1$. Regarding φ_0 and $(\alpha'/\alpha)\varphi$ as tests of size $\le \mathbb{E}_{\theta^*}[\varphi_0(\boldsymbol{X})]$ for testing $\theta = \theta^*$ vs $\theta = \theta_1$, Neyman-Pearson Lemma implies that $\mathbb{E}_{\theta_1}[\varphi_0(\boldsymbol{X})] \ge \mathbb{E}_{\theta_1}[(\alpha'/\alpha)\varphi(\boldsymbol{X})]$, as φ_0 defines a likelihood ratio test of $\theta = \theta^*$ against $\theta = \theta_1$. Letting $\alpha' \uparrow \alpha$, we have $\mathbb{E}_{\theta_1}[\varphi_0(\boldsymbol{X})] \ge \mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})]$. Since θ_1 is arbitrary, we have $\mathbb{E}_\theta[\varphi_0(\boldsymbol{X})] \ge \mathbb{E}_\theta[\varphi(\boldsymbol{X})]$ for all $\theta \in \Theta_1$, so that φ_0 is UMP among all tests of size $\le \alpha$. ▍

6.4.6 **Remark.** If $\Theta_0 \cap \Theta_1 \neq \emptyset$, then it follows from Theorem §6.4.5 that for any $\theta^\dagger \in \Theta_0 \cap \Theta_1$,

$$\mathbb{E}_{\theta^\dagger}\big[\varphi_0(\boldsymbol{X})\big] = \sup_{\theta\in\Theta_0} \mathbb{E}_\theta\big[\varphi_0(\boldsymbol{X})\big] = \text{ size of } \varphi_0.$$

6.4.7 **Examples §6.4.4.** (cont'd)

(i) Data $\boldsymbol{X} \sim$ probability function $f(\boldsymbol{x}|\theta) = c(\theta)h(\boldsymbol{x})\exp\big\{\theta\, t(\boldsymbol{x})\big\}$.
Test $H_0 : \theta \le \theta^*$ vs $H_1 : \theta > \theta^*$.
The model has mlr in $T = t(\boldsymbol{X})$ w.r.t. the above test, as well as the test of $H_0 : \theta \le \theta^*$ vs $H_1 : \theta \ge \theta^*$.
<u>Likelihood ratio test</u>: *reject H_0 if* $t(\boldsymbol{X}) > t_0$
is UMP among all tests of size $\le \alpha = \sup_{\theta\le\theta^*} \mathbb{P}_\theta\big(t(\boldsymbol{X}) > t_0\big) = \mathbb{P}_{\theta^*}\big(t(\boldsymbol{X}) > t_0\big)$.
Based on the observation $\boldsymbol{X} = \boldsymbol{x}$, the p-value is given by

$$pv(\boldsymbol{x}) = \sup_{\theta\le\theta^*} \mathbb{P}_\theta\big(t(\boldsymbol{X}) \ge t(\boldsymbol{x})\big) = \mathbb{P}_{\theta^*}\big(t(\boldsymbol{X}) \ge t(\boldsymbol{x})\big).$$

(ii) $\boldsymbol{X} = (X_1, \ldots, X_n)$ iid $\sim U[0, \theta]$
Test $H_0 : \theta \ge c$ vs $H_1 : \theta < c$
The model has mlr in $T = -\max_i(X_i)$ w.r.t. the above test, as well as the test of $H_0 : \theta \ge c$ vs $H_1 : \theta \le c$.
<u>Likelihood ratio test</u>: *reject H_0 if* $\max_i(X_i) < t_0$
is UMP among all tests of size

$$\le \alpha = \sup_{\theta\ge c} \mathbb{P}_\theta\big(\max_i X_i < t_0\big) = \mathbb{P}_c\big(\max_i X_i < t_0\big) = \min\big\{(t_0/c)^n, 1\big\}.$$

Based on the observation $\boldsymbol{X} = \boldsymbol{x}$, the p-value is given by

$$pv(\boldsymbol{x}) = \sup_{\theta\ge c} \mathbb{P}_\theta\big(\max_i X_i \le \max_i x_i\big) = \mathbb{P}_c\big(\max_i X_i \le \max_i x_i\big) = \min\Big\{\big(\max_i x_i/c\big)^n, 1\Big\}.$$

6.4.8 **Example.**

Let X be a discrete random variable distributed over the set $\{1, 2, \ldots, r\}$ with probabilities $\boldsymbol{\theta} = (p_1, \ldots, p_r)$, such that

$$\mathbb{P}_{\boldsymbol{\theta}}(X = i) = p_i, \quad i = 1, 2, \ldots, r.$$

Let $\boldsymbol{a} = (a_1, \ldots, a_r)$ be a given sequence of probabilities such that $\sum_{i=1}^r a_i = 1$.

Define $\Theta_1 = \left\{\boldsymbol{\theta} = (p_1, \ldots, p_r) : \dfrac{p_1}{p_2} \geq \dfrac{a_1}{a_2}, \dfrac{p_2}{p_3} \geq \dfrac{a_2}{a_3}, \ldots, \dfrac{p_{r-1}}{p_r} \geq \dfrac{a_{r-1}}{a_r}\right\}$.

Consider the test of

$$H_0 : \boldsymbol{\theta} = \boldsymbol{a} \qquad \text{vs} \qquad H_1 : \boldsymbol{\theta} \in \Theta_1.$$

Note that for any $\boldsymbol{\theta} = (p_1, \ldots, p_r) \in \Theta_1$, we have

$$\frac{p_1}{a_1} \geq \frac{p_2}{a_2} \geq \cdots \geq \frac{p_r}{a_r},$$

so that $\ell_X(\boldsymbol{\theta})/\ell_X(\boldsymbol{a}) = p_X/a_X$ is nondecreasing in $-X$. Thus the model has mlr in $-X$ w.r.t. the test.

<u>*Likelihood ratio test*</u>: *reject* H_0 *if* $X < t_0$

is UMP among all tests of size $\leq \alpha = \mathbb{P}_{\boldsymbol{a}}\big(X < t_0\big) = \sum_{\{i:\, i<t_0\}} a_i$.

Based on the observation $X = x$, the p-value is given by

$$pv(x) = \mathbb{P}_{\boldsymbol{a}}\big(X \leq x\big) = a_1 + \cdots + a_x.$$

We see that the significance of an observation x depends not only on its actual value (the smaller the more significant) but also on the probability of observing a value at least as small as x under H_0.

§6.5 Two-sided UMPU test under exponential family

6.5.1 No UMP test exists for testing $H_0 : \theta \in [\theta_1, \theta_2]$ vs $H_1 : \theta \notin [\theta_1, \theta_2]$, in which case one might need to weaken the optimality criterion to UMPU.

6.5.2 Data $\boldsymbol{X} \sim$ probability function $f(\boldsymbol{x}|\theta) = c(\theta)h(\boldsymbol{x})\exp\big\{\theta\, t(\boldsymbol{x})\big\}$.
Natural parameter: θ, natural statistic: $t(\boldsymbol{X})$.

Test $H_0 : \theta \in [\theta_1, \theta_2]$ vs $H_1 : \theta \notin [\theta_1, \theta_2]$.

We assume without loss of generality that $[\theta_1, \theta_2]$ is contained in the natural parameter space. Define test function

$$\varphi(\boldsymbol{X}) = \mathbf{1}\big\{t(\boldsymbol{X}) \notin [t_1, t_2]\big\},$$

where t_1, t_2 satisfy

$$\mathbb{E}_{\theta_1}\big[\varphi(\boldsymbol{X})\big] = \mathbb{E}_{\theta_2}\big[\varphi(\boldsymbol{X})\big] = \alpha.$$

Theorem. The test φ is the UMPU size α test of H_0 against H_1.

. . . .

Proof: *We start by proving the following lemma.*

Lemma. *Let $\tilde{\varphi}$ be an arbitrary test function such that $\mathbb{E}_{\theta_1}\big[\tilde{\varphi}(\boldsymbol{X})\big] = \mathbb{E}_{\theta_2}\big[\tilde{\varphi}(\boldsymbol{X})\big] = \alpha$. Then*

$$\mathbb{E}_{\theta}\big[\varphi(\boldsymbol{X})\big] \begin{cases} \geq \mathbb{E}_{\theta}\big[\tilde{\varphi}(\boldsymbol{X})\big], & \theta \notin [\theta_1, \theta_2], \\ \leq \mathbb{E}_{\theta}\big[\tilde{\varphi}(\boldsymbol{X})\big], & \theta \in [\theta_1, \theta_2]. \end{cases}$$

Proof of lemma:

Let $u < v < w$ be fixed in the natural parameter space. For any $K_1, K_2 > 0$, consider the function

$$\begin{aligned} g(\boldsymbol{X}) &\equiv K_1 \frac{f(\boldsymbol{X}|u)}{f(\boldsymbol{X}|v)} + K_2 \frac{f(\boldsymbol{X}|w)}{f(\boldsymbol{X}|v)} - 1 \\ &= K_1 \frac{c(u)}{c(v)} e^{(u-v)t(\boldsymbol{X})} + K_2 \frac{c(w)}{c(v)} e^{(w-v)t(\boldsymbol{X})} - 1, \end{aligned}$$

which is convex in $t(\boldsymbol{X})$, diverges to ∞ as $t(\boldsymbol{X}) \to \pm\infty$, and negative at $t(\boldsymbol{X}) = 0$ for sufficiently small $K_1, K_2 > 0$. We can find constants $K_1, K_2 > 0$ such that $g(\boldsymbol{X}) > 0$ if and only if $t(\boldsymbol{X}) \notin [t_1, t_2]$. Thus

$$\varphi(\boldsymbol{X}) = \mathbf{1}\{g(\boldsymbol{X}) > 0\} = \mathbf{1}\left\{K_1 f(\boldsymbol{X}|u) + K_2 f(\boldsymbol{X}|w) > f(\boldsymbol{X}|v)\right\}.$$

Consider

$$[\varphi(\boldsymbol{X}) - \tilde{\varphi}(\boldsymbol{X})]\,[K_1 f(\boldsymbol{X}|u) + K_2 f(\boldsymbol{X}|w) - f(\boldsymbol{X}|v)] \geq 0,$$

so that

$$\begin{aligned} &\int [\varphi(\boldsymbol{x}) - \tilde{\varphi}(\boldsymbol{x})]\,[K_1 f(\boldsymbol{x}|u) + K_2 f(\boldsymbol{x}|w) - f(\boldsymbol{x}|v)]\, d\boldsymbol{x} \geq 0 \\ \Rightarrow \quad & K_1 \{\mathbb{E}_u \varphi(\boldsymbol{X}) - \mathbb{E}_u \tilde{\varphi}(\boldsymbol{X})\} + K_2 \{\mathbb{E}_w \varphi(\boldsymbol{X}) - \mathbb{E}_w \tilde{\varphi}(\boldsymbol{X})\} \geq \mathbb{E}_v \varphi(\boldsymbol{X}) - \mathbb{E}_v \tilde{\varphi}(\boldsymbol{X}). \end{aligned}$$

If $\theta < \theta_1$, setting $u = \theta$, $v = \theta_1$ and $w = \theta_2$ in the above gives that

$$K_1 \{\mathbb{E}_\theta \varphi(\boldsymbol{X}) - \mathbb{E}_\theta \tilde{\varphi}(\boldsymbol{X})\} + K_2(\alpha - \alpha) \geq \alpha - \alpha \quad \Rightarrow \quad \mathbb{E}_\theta \varphi(\boldsymbol{X}) \geq \mathbb{E}_\theta \tilde{\varphi}(\boldsymbol{X}).$$

Similarly, if $\theta > \theta_2$, setting $u = \theta_1$, $v = \theta_2$ and $w = \theta$ gives also that $\mathbb{E}_\theta \varphi(\boldsymbol{X}) \geq \mathbb{E}_\theta \tilde{\varphi}(\boldsymbol{X})$.

If $\theta \in (\theta_1, \theta_2)$, setting $u = \theta_1$, $v = \theta$ and $w = \theta_2$ gives that

$$K_1(\alpha - \alpha) + K_2(\alpha - \alpha) \geq \mathbb{E}_\theta\, \varphi(\boldsymbol{X}) - \mathbb{E}_\theta\, \tilde{\varphi}(\boldsymbol{X}) \;\; \Rightarrow \;\; \mathbb{E}_\theta\, \varphi(\boldsymbol{X}) \leq \mathbb{E}_\theta\, \tilde{\varphi}(\boldsymbol{X}).$$

∎

Taking $\tilde{\varphi}(\boldsymbol{X}) \equiv \alpha$ in the above lemma, we have

$$\mathbb{E}_\theta[\varphi(\boldsymbol{X})] \begin{cases} \geq \alpha, & \theta \notin [\theta_1, \theta_2], \\ \leq \alpha, & \theta \in [\theta_1, \theta_2]. \end{cases}$$

Thus φ has size α and is unbiased.

Let φ^* be an arbitrary unbiased test function of size α. The exponential family form ensures continuity of its power function, so that $\mathbb{E}_{\theta_1}[\varphi^*(\boldsymbol{X})] = \mathbb{E}_{\theta_2}[\varphi^*(\boldsymbol{X})] = \alpha$. Taking $\tilde{\varphi} \equiv \varphi^*$ in the lemma, we have

$$\mathbb{E}_\theta[\varphi(\boldsymbol{X})] \geq \mathbb{E}_\theta[\varphi^*(\boldsymbol{X})] \quad \forall\, \theta \notin [\theta_1, \theta_2].$$

The above results together imply that φ is the UMPU size α test. ∎

6.5.3 Data $\boldsymbol{X} \sim$ probability function $f(\boldsymbol{x}|\theta) = c(\theta)h(\boldsymbol{x}) \exp\big\{\theta\, t(\boldsymbol{x})\big\}$.
Natural parameter: θ, natural statistic $t(\boldsymbol{X})$.

Test $\quad H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Define test function

$$\varphi(\boldsymbol{X}) = \mathbf{1}\big\{t(\boldsymbol{X}) \notin [t_1, t_2]\big\},$$

where t_1, t_2 satisfy

$$\mathbb{E}_{\theta_0}\big[\varphi(\boldsymbol{X})\big] = \alpha, \qquad \frac{d}{d\theta}\, \mathbb{E}_\theta\big[\varphi(\boldsymbol{X})\big]\bigg|_{\theta_0} = 0.$$

Theorem. The test φ is the UMPU size α test of H_0 against H_1.

. . . .

Proof: (outline)

Let $\delta > 0$ be fixed and small. Then the UMPU size α test of $\theta \in [\theta_0 - \delta, \theta_0 + \delta]$ against $\theta \notin [\theta_0 - \delta, \theta_0 + \delta]$ has the form $\varphi_\delta(\boldsymbol{X}) = \mathbf{1}\{t(\boldsymbol{X}) \notin [t_1(\delta), t_2(\delta)]\}$, which satisfies

$$\mathbb{E}_{\theta_0 \pm \delta}[\varphi_\delta(\boldsymbol{X})] = \alpha \leq \mathbb{E}_\theta[\varphi_\delta(\boldsymbol{X})] \;\; \forall\, \theta \notin [\theta_0 - \delta, \theta_0 + \delta].$$

Letting $\delta \downarrow 0$, we have $\varphi_\delta(\boldsymbol{X}) \to \varphi(\boldsymbol{X}) = \mathbf{1}\{t(\boldsymbol{X}) \notin [t_1, t_2]\}$ and

$$\mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})] = \alpha \leq \mathbb{E}_\theta[\varphi(\boldsymbol{X})] \;\; \forall\, \theta \neq \theta_0.$$

Note that φ is UMPU for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$, and that the function $\theta \mapsto \mathbb{E}_\theta[\varphi(\boldsymbol{X})]$ attains its minimum α at $\theta = \theta_0$, so that t_1, t_2 can be found by solving the equations

$$\mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})] = \alpha, \qquad \frac{d}{d\theta}\,\mathbb{E}_\theta[\varphi(\boldsymbol{X})]\bigg|_{\theta_0} = 0. \qquad \blacksquare$$

6.5.4 **Examples.** Denote by Φ and ϕ the cdf and pdf of $N(0,1)$ respectively.

$\boldsymbol{X} = (X_1, \ldots, X_n)$ iid $\sim N(\theta, 1)$ $\quad\rightarrow\quad$ Natural statistic $t(\boldsymbol{X}) = \sum_{i=1}^n X_i$.
Test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Define $w(\theta) = \mathbb{P}_\theta\big(t(\boldsymbol{X}) \notin [t_1, t_2]\big)$.
Since $t(\boldsymbol{X}) = \sum_{i=1}^n X_i \sim N(n\theta, n)$, we have

$$w(\theta) = 1 - \Phi\left(\frac{t_2 - n\theta}{\sqrt{n}}\right) + \Phi\left(\frac{t_1 - n\theta}{\sqrt{n}}\right).$$

Fix t_1, t_2 by

$$w(\theta_0) = \alpha, \quad w'(\theta_0) = 0.$$

Solving $w'(\theta_0) = 0$ gives $\theta_0 = (t_1 + t_2)/(2n)$. Note that $w(\theta)$ has a minimum at and is symmetric about θ_0, which is the mid-point of $[t_1/n, t_2/n]$. Thus, we have

$$t_1/n = \theta_0 - c, \quad t_2/n = \theta_0 + c, \quad c > 0,$$

where c is given by

$$w(\theta_0) = \alpha \;\Leftrightarrow \Phi(\sqrt{n}\,c) = 1 - \alpha/2 \;\Leftrightarrow\; c = n^{-1/2}\Phi^{-1}(1 - \alpha/2).$$

$\rightarrow$ *UMPU test of size α:*

reject H_0 if $\bar{X} \notin \big[\, \theta_0 - n^{-1/2}\Phi^{-1}(1-\alpha/2),\ \theta_0 + n^{-1/2}\Phi^{-1}(1-\alpha/2)\,\big]$.

The above critical region is equivalent to $\big\{\sqrt{n}|\bar{X} - \theta_0| > \Phi^{-1}(1-\alpha/2)\big\}$.

The p-value $pv(\boldsymbol{x})$ based on the observation $\boldsymbol{X} = \boldsymbol{x}$ is given by the smallest value of α that leads to rejection of H_0, which satisfies

$$\sqrt{n}|\bar{x} - \theta_0| = \Phi^{-1}(1 - \alpha/2) \quad \Rightarrow \quad pv(\boldsymbol{x}) = 2\big\{1 - \Phi\big(\sqrt{n}|\bar{x} - \theta_0|\big)\big\}.$$

§6.6 Conditional test under exponential family

6.6.1 General exponential family: one-sided case —

Data $\boldsymbol{X} \sim$ exponential family $f(\boldsymbol{x}|\boldsymbol{\pi}) = C(\boldsymbol{\pi})h(\boldsymbol{x})\exp\left\{\sum_{j=1}^{k} \pi_j t_j(\boldsymbol{x})\right\}$.

Test $H_0 : \pi_1 \le \pi_1^*$ against $H_1 : \pi_1 > \pi_1^*$.

[Assume that the natural parameter space contains an open rectangle in $\mathbb{R}^k$ around $(\pi_1^*, \pi_2^*, \ldots, \pi_k^*)$, for some $(\pi_2^*, \ldots, \pi_k^*)$.]

It is known that the conditional distribution of $t_1(\boldsymbol{X})$ given $\big(t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big)$ has an exponential family form with natural parameter π_1 and natural statistic $t_1(\boldsymbol{X})$ — hence free of the unknown *nuisance parameters* $(\pi_2, \ldots, \pi_k)$.

Define test function

$$\varphi_0(\boldsymbol{X}) = \mathbf{1}\big\{t_1(\boldsymbol{X}) > c\big\},$$

where $c = c(\alpha, \pi_1^*, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ satisfies

$$\mathbb{E}_{\pi_1^*}\left[\varphi_0(\boldsymbol{X})\middle|t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\right] = \mathbb{P}_{\pi_1^*}\left(t_1(\boldsymbol{X}) > c\middle|t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\right) = \alpha.$$

This is called a *conditional test*.

Theorem. The test φ_0 is the UMPU size α test of H_0 against H_1.

6.6.2 General exponential family: two-sided case —

Data $\boldsymbol{X} \sim$ exponential family $f(\boldsymbol{x}|\boldsymbol{\pi}) = C(\boldsymbol{\pi})h(\boldsymbol{x})\exp\left\{\sum_{j=1}^{k} \pi_j t_j(\boldsymbol{x})\right\}$.

Test $H_0 : \pi_1 \in [\pi_1^*, \pi_1^{**}]$ against $H_1 : \pi_1 \notin [\pi_1^*, \pi_1^{**}]$.

[Assume that the natural parameter space contains open rectangles in $\mathbb{R}^k$ around $(\pi_1^*, \pi_2^*, \ldots, \pi_k^*)$ and $(\pi_1^{**}, \pi_2^{**}, \ldots, \pi_k^{**})$, respectively, for some $(\pi_2^*, \ldots, \pi_k^*)$ and $(\pi_2^{**}, \ldots, \pi_k^{**})$.]

Define test function

$$\varphi_0(\boldsymbol{X}) = \mathbf{1}\big\{t_1(\boldsymbol{X}) \notin [c^*, c^{**}]\big\},$$

where $c^* = c^*(\alpha, \pi_1^*, \pi_1^{**}, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ and $c^{**} = c^{**}(\alpha, \pi_1^*, \pi_1^{**}, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X}))$ satisfy

$$\begin{cases} \mathbb{P}_{\pi_1^*}\big(t_1(\boldsymbol{X}) \notin [c^*, c^{**}] \,\big|\, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big) = \alpha, \\ \mathbb{P}_{\pi_1^{**}}\big(t_1(\boldsymbol{X}) \notin [c^*, c^{**}] \,\big|\, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big) = \alpha, \end{cases}$$

or, if $\pi_1^* = \pi_1^{**}$,

$$\begin{cases} \mathbb{P}_{\pi_1^*}\big(t_1(\boldsymbol{X}) \notin [c^*, c^{**}] \,\big|\, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big) = \alpha, \\ \dfrac{d}{d\pi_1}\,\mathbb{P}_{\pi_1}\big(t_1(\boldsymbol{X}) \notin [c^*, c^{**}] \,\big|\, t_2(\boldsymbol{X}), \ldots, t_k(\boldsymbol{X})\big)\Big|_{\pi_1=\pi_1^*} = 0. \end{cases}$$

Theorem. The test φ_0 is the UMPU size α test of H_0 against H_1.

6.6.3 *Outline of proof of Theorems §6.6.1 and §6.6.2*

Let φ be an arbitrary unbiased test of size α.

Write $\boldsymbol{\eta} = (\pi_2, \dots, \pi_k)$. Continuity of the power function of φ and its unbiasedness imply

$$\mathbb{E}_{\pi_1^*, \boldsymbol{\eta}}[\varphi(\boldsymbol{X})] = \alpha \quad \textit{(and, for two-sided case, } \mathbb{E}_{\pi_1^{**}, \boldsymbol{\eta}}[\varphi(\boldsymbol{X})] = \alpha) \quad \forall \boldsymbol{\eta}.$$

For fixed $\pi_1 = \pi_1^*$ or π_1^{**}, the statistic $\boldsymbol{T}_2 = (t_2(\boldsymbol{X}), \dots, t_k(\boldsymbol{X}))$ has a distribution of the exponential family form with natural parameter $\boldsymbol{\eta}$, and that $\boldsymbol{T}_2$ is complete sufficient for $\boldsymbol{\eta}$ under the conditions assumed on the natural parameter space.

Define $g(\boldsymbol{T}_2) = \mathbb{E}_{\pi_1^*}[\varphi(\boldsymbol{X})|\boldsymbol{T}_2] - \alpha$. Then

$$\mathbb{E}_{\pi_1^*, \boldsymbol{\eta}}[g(\boldsymbol{T}_2)] = \mathbb{E}_{\pi_1^*, \boldsymbol{\eta}}[\varphi(\boldsymbol{X})] - \alpha = 0 \quad \forall \boldsymbol{\eta},$$

so that by completeness of $\boldsymbol{T}_2$ for $\boldsymbol{\eta}$,

$$\mathbb{P}_{\pi_1^*, \boldsymbol{\eta}}\{g(\boldsymbol{T}_2) = 0\} = 1 \quad \forall \boldsymbol{\eta}.$$

Similar results hold with π_1^* replaced by π_1^{**}. Thus, almost surely,

$$\mathbb{E}_{\pi_1^*}[\varphi(\boldsymbol{X})|\boldsymbol{T}_2] = \mathbb{E}_{\pi_1^{**}}[\varphi(\boldsymbol{X})|\boldsymbol{T}_2] = \alpha.$$

Conditional on $\boldsymbol{T}_2$, $\boldsymbol{X}$ has distribution of exponential family form with natural parameter π_1 and natural statistic $t_1(\boldsymbol{X})$. Under this "one-parameter" conditional model:

(a) The test φ_0 has size α and is unbiased according to results stated in §6.4 and §6.5. Thus

$$\mathbb{E}_{\pi_1}[\varphi_0(\boldsymbol{X})|\boldsymbol{T}_2] \le \alpha \quad \textit{under } H_0 \qquad \textit{and} \qquad \mathbb{E}_{\pi_1}[\varphi_0(\boldsymbol{X})|\boldsymbol{T}_2] \ge \alpha \quad \textit{under } H_1.$$

Taking expectations on both sides, we have

$$\mathbb{E}_{\pi_1, \boldsymbol{\eta}}[\varphi_0(\boldsymbol{X})] \le \alpha \quad \textit{under } H_0 \qquad \textit{and} \qquad \mathbb{E}_{\pi_1, \boldsymbol{\eta}}[\varphi_0(\boldsymbol{X})] \ge \alpha \quad \textit{under } H_1.$$

(b) According to Theorem §6.4.5(iii) (for the one-sided case) and the lemma in §6.5.2 (for the two-sided case), we see that

$$\mathbb{E}_{\pi_1}[\varphi_0(\boldsymbol{X})|\boldsymbol{T}_2] \ge \mathbb{E}_{\pi_1}[\varphi(\boldsymbol{X})|\boldsymbol{T}_2] \quad \textit{under } H_1.$$

Taking expectations on both sides, we have

$$\mathbb{E}_{\pi_1, \boldsymbol{\eta}}[\varphi_0(\boldsymbol{X})] \ge \mathbb{E}_{\pi_1, \boldsymbol{\eta}}[\varphi(\boldsymbol{X})] \quad \textit{under } H_1.$$

Results (a) and (b) above together imply that φ_0 is UMPU of size α. ∎

6.6.4 **Example.**

$\boldsymbol{X} = (X_1, \ldots, X_n)$ iid $N(\mu, \sigma^2)$.
Test $H_0 : \sigma = \sigma_0$ vs $H_1 : \sigma \neq \sigma_0$ (for a specified constant $\sigma_0 > 0$).

Natural parameter: $(\pi_1, \pi_2) = \big(-1/(2\sigma^2),\ \mu/\sigma^2\big)$, natural statistic: $\big(\sum_{i=1}^n X_i^2,\ \sum_{i=1}^n X_i\big)$.

Equivalent problem $\longrightarrow$ *test* $H_0 : \pi_1 = -1/(2\sigma_0^2)$ *vs* $H_1 : \pi_1 \neq -1/(2\sigma_0^2)$.

Given $\bar{X} = \bar{x}$, choose constants c_1, c_2 such that

$$\begin{cases} \mathbb{P}\Big(\sum_{i=1}^n X_i^2 \notin [c_1, c_2] \,\Big|\, \bar{X} = \bar{x},\ \pi_1 = -1/(2\sigma_0^2)\Big) = \alpha, \\ \dfrac{d}{d\pi_1} \mathbb{P}_{\pi_1}\Big(\sum_{i=1}^n X_i^2 \notin [c_1, c_2] \,\Big|\, \bar{X} = \bar{x}\Big)\Big|_{\pi_1 = -1/(2\sigma_0^2)} = 0. \end{cases}$$

Basic facts: under normal distribution,

$$S_{xx} = \textstyle\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \text{ and } \bar{X} \text{ are } \textbf{independent}, \text{ and } S_{xx}/\sigma^2 \sim \chi^2_{n-1}.$$

Let F_d and f_d denote the χ^2_d cdf and pdf, respectively. Then c_1, c_2 satisfy

$$\begin{cases} \mathbb{P}\Big(\chi^2_{n-1} \in \big[\dfrac{c_1 - n\bar{x}^2}{\sigma_0^2}, \dfrac{c_2 - n\bar{x}^2}{\sigma_0^2}\big]\Big) = 1 - \alpha, \\ \dfrac{d}{d\pi_1} \mathbb{P}_{\pi_1}\Big(\chi^2_{n-1} \in \big[-2\pi_1(c_1 - n\bar{x}^2),\, -2\pi_1(c_2 - n\bar{x}^2)\big]\Big)\Big|_{\pi_1 = -1/(2\sigma_0^2)} = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} F_{n-1}\left(\dfrac{c_2 - n\bar{x}^2}{\sigma_0^2}\right) - F_{n-1}\left(\dfrac{c_1 - n\bar{x}^2}{\sigma_0^2}\right) = 1 - \alpha, \\ -2\left(\dfrac{c_2 - n\bar{x}^2}{\sigma_0^2}\right) f_{n-1}\left(\dfrac{c_2 - n\bar{x}^2}{\sigma_0^2}\right) = -2\left(\dfrac{c_1 - n\bar{x}^2}{\sigma_0^2}\right) f_{n-1}\left(\dfrac{c_1 - n\bar{x}^2}{\sigma_0^2}\right). \end{cases}$$

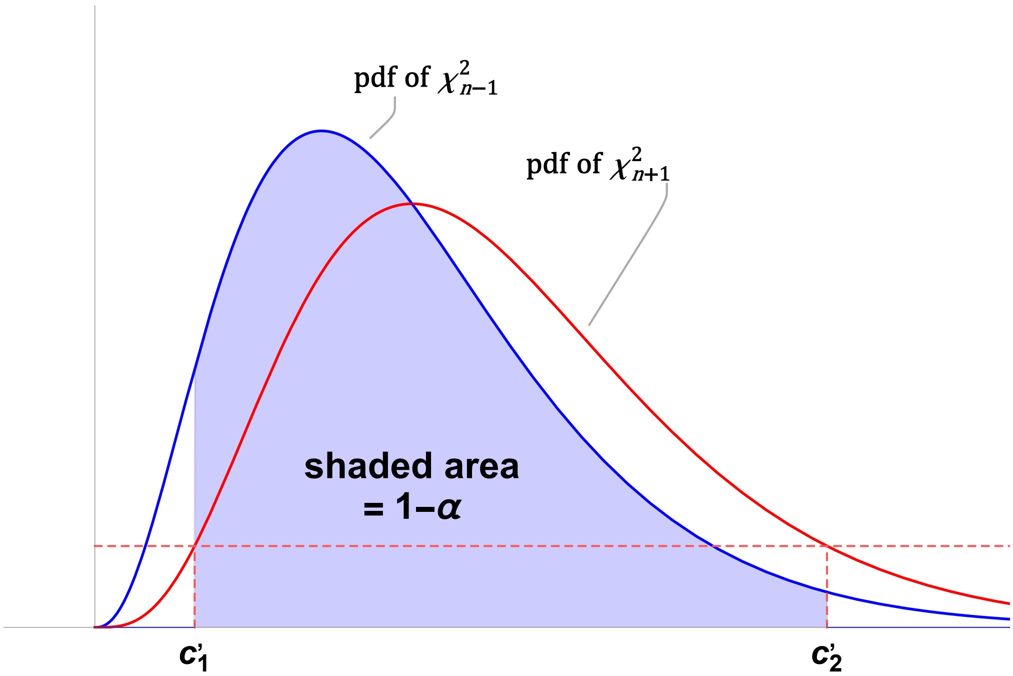
Writing $c_i' = \dfrac{c_i - n\bar{x}^2}{\sigma_0^2}$, $i = 1, 2$, and noting that

$$f_{n+1}(x) = \frac{x^{(n-1)/2} e^{-x/2}}{2^{(n+1)/2}\Gamma((n+1)/2)} = \left(\frac{x}{n-1}\right) \frac{x^{(n-3)/2} e^{-x/2}}{2^{(n-1)/2}\Gamma((n-1)/2)} = \frac{1}{n-1}\{x f_{n-1}(x)\},\ x > 0,$$

we have

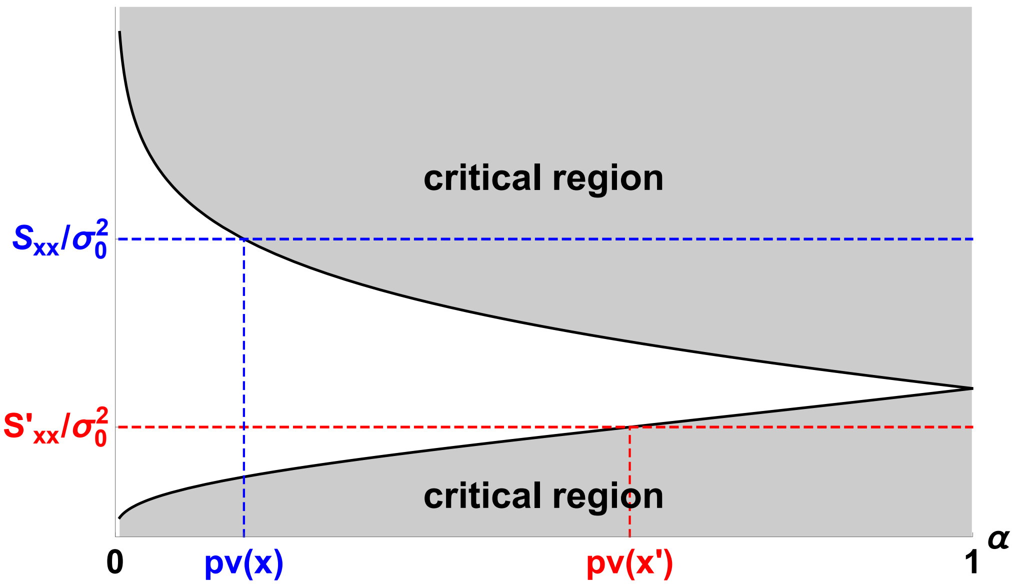
$$\begin{cases} F_{n-1}(c_2') - F_{n-1}(c_1') = 1 - \alpha, \\ f_{n+1}(c_2') = f_{n+1}(c_1'), \end{cases}$$

which can be solved numerically for c_1', c_2' as illustrated in the diagram below.



UMPU test of size α: *reject H_0 if $\sum_{i=1}^n X_i^2 \notin [c_1, c_2]$, i.e. $S_{xx}/\sigma_0^2 \notin [c_1', c_2']$.*

The following diagram displays the critical region $[0, c_1') \cup (c_2', \infty)$ under different values of α. It also shows the p-values $pv(\boldsymbol{x})$ and $pv(\boldsymbol{x}')$ based on two possible datasets $\boldsymbol{x}$ and $\boldsymbol{x}'$, respectively, which are given by the smallest values of α that lead to rejection of H_0.



6.6.5 Data $\boldsymbol{X} \sim$ exponential family $f(\boldsymbol{x}|\pi) = C(\pi)h(\boldsymbol{x})\exp\left\{\sum_{j=1}^{k}\pi_j t_j(\boldsymbol{x})\right\}$.
Parameter of interest: $\xi = \sum_{j=1}^{k} c_j\pi_j$ (e.g. $\pi_1 - \pi_2$).
Suppose, w.l.o.g., $c_1 > 0$. Putting $\pi_1 = \left(\xi - \sum_{j=2}^{k} c_j\pi_j\right)/c_1$, then

$$\sum_{j=1}^{k}\pi_j t_j(\boldsymbol{X}) = \xi t_1(\boldsymbol{X})/c_1 + \sum_{j=2}^{k}\pi_j\left\{t_j(\boldsymbol{X}) - c_j t_1(\boldsymbol{X})/c_1\right\}.$$

We can test hypotheses about ξ by considering

conditional distribution of

$$t_1(\boldsymbol{X}) \quad \textit{given} \quad \left(t_2(\boldsymbol{X}) - c_2 t_1(\boldsymbol{X})/c_1\,,\ \ldots\ ,\ t_k(\boldsymbol{X}) - c_k t_1(\boldsymbol{X})/c_1\,\right),$$

which is free of nuisance parameters.

6.6.6 **Example.**

Two independent samples: $(X_1, \ldots, X_m)$ i.i.d. $\sim$ Poisson (λ), $(Y_1, \ldots, Y_n)$ i.i.d. $\sim$ Poisson (μ)

Test: $H_0 : \lambda \leq \mu$ vs $H_1 : \lambda > \mu$.

Joint mass function of the two samples:

$$\begin{aligned} f(x_1, \ldots, x_m, y_1, \ldots, y_n|\lambda, \mu) &\propto \Big(\prod_i x_i! \prod_j y_j!\Big)^{-1} \exp\Big\{(\ln\lambda)\sum_i x_i + (\ln\mu)\sum_j y_j\Big\} \\ &= \Big(\prod_i x_i! \prod_j y_j!\Big)^{-1} \exp\Big\{\xi\sum_i x_i + \eta\Big(\sum_i x_i + \sum_j y_j\Big)\Big\}, \end{aligned}$$

where $\xi = \ln\lambda - \ln\mu$ and $\eta = \ln\mu$. Equivalently, we wish to

test $H_0 : \xi \leq 0$ vs $H_1 : \xi > 0$.

Given that $\sum_i X_i + \sum_j Y_j$ is observed to be t, let $\alpha \in (0, 1)$ and c satisfy

$$\mathbb{P}\Big(\sum_i X_i > c \,\Big|\, \sum_i X_i + \sum_j Y_j = t,\, \xi = 0\Big) = \alpha.$$

[Note: Above conditional probability does **not** depend on nuisance parameter η and can therefore be computed numerically.]

At $\xi = 0$, i.e. $\lambda = \mu$,

$$\mathbb{P}\Big(\sum_i X_i = x \,\Big|\, \sum_i X_i + \sum_j Y_j = t,\, \xi = 0\Big) = \frac{\{e^{-m\lambda}(m\lambda)^x/x!\}\{e^{-n\lambda}(n\lambda)^{t-x}/(t-x)!\}}{e^{-(m+n)\lambda}\{(m+n)\lambda\}^t/t!}$$
$$= \binom{t}{x}\Big(\frac{m}{m+n}\Big)^x\Big(\frac{n}{m+n}\Big)^{t-x} = \mathbb{P}\Big(\text{binomial}\,\big(t, m/(m+n)\big) = x\Big).$$

Hence α and c are related by the equation

$$\Big(\frac{n}{m+n}\Big)^{\sum_i X_i + \sum_j Y_j} \sum_{c < x \le \sum_i X_i + \sum_j Y_j} \binom{\sum_i X_i + \sum_j Y_j}{x}\Big(\frac{m}{n}\Big)^x = \alpha.$$

UMPU test of size α: *reject H_0 if* $\sum_i X_i > c$.

Similarly, the p-value can be calculated as the smallest α which leads to rejection of H_0, given by

$$\Big(\frac{n}{m+n}\Big)^{\sum_i X_i + \sum_j Y_j} \sum_{x = \sum_i X_i}^{\sum_i X_i + \sum_j Y_j} \binom{\sum_i X_i + \sum_j Y_j}{x}\Big(\frac{m}{n}\Big)^x.$$

§6.7 Test based on large-sample theory

6.7.1 Problem setting:

Data $\boldsymbol{X} = (X_1, \ldots, X_n) \sim$ joint probability function $p_1(x_1|\boldsymbol{\theta}) \times \cdots \times p_n(x_n|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^r$.
Test $H_0 : \boldsymbol{\theta} \in \Theta_0$ vs $H_1 : \boldsymbol{\theta} \in \Theta_1$, where

$$\Theta_0 = \left\{\boldsymbol{g}(\boldsymbol{\psi}) = \begin{bmatrix} g_1(\boldsymbol{\psi}) \\ \vdots \\ g_r(\boldsymbol{\psi}) \end{bmatrix} : \boldsymbol{\psi} \in \Psi \subset \mathbb{R}^s\right\} \subset \Theta_1 \subset \Theta$$

for continuously differentiable functions $g_1, \ldots, g_r : \Psi \to \mathbb{R}$ and $s < r$.

Define mle: $\hat{\boldsymbol{\theta}}_n$ [maximises $\ell_{\boldsymbol{X}}(\boldsymbol{\theta}) \propto \prod_{i=1}^n p_i(X_i|\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \Theta_1$],
constrained mle: $\hat{\boldsymbol{\psi}}_n$ [maximises $\ell_{\boldsymbol{X}}(\boldsymbol{g}(\boldsymbol{\psi}))$ over $\boldsymbol{\psi} \in \Psi$].

6.7.2 Recall: *likelihood ratio test statistic*

$$\Lambda_{\boldsymbol{X}}(H_0, H_1) = \ell_{\boldsymbol{X}}(\hat{\boldsymbol{\theta}}_n)/\ell_{\boldsymbol{X}}(\boldsymbol{g}(\hat{\boldsymbol{\psi}}_n)).$$

Theorem. *(Wilks)*
Under regularity conditions[2] and if H_0 is true, then as $n \to \infty$,

$$2 \ln \Lambda_{\boldsymbol{X}}(H_0, H_1) \longrightarrow \chi^2_{r-s} \quad \text{in distribution.}$$

. . . .

Proof: (outline)

Denote by $\mathbb{I}_k$ the $k \times k$ identity matrix, by $\boldsymbol{U}(\boldsymbol{\theta})$ the score function and by $I(\boldsymbol{\theta})$ the Fisher information matrix. Suppose H_0 is true, and let $\boldsymbol{\theta}_0 = \boldsymbol{g}(\boldsymbol{\psi}_0) \in \Theta_0$ be the true value of $\boldsymbol{\theta}$. Large-sample theory of mle implies that $\hat{\boldsymbol{\psi}}_n$ and $\hat{\boldsymbol{\theta}}_n$ are consistent estimators of $\boldsymbol{\psi}_0$ and $\boldsymbol{\theta}_0$, respectively.
Define $G(\boldsymbol{\psi}) = \dfrac{\partial \boldsymbol{g}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}^\top}$ to be the $r \times s$ matrix with (a,b)th entry equal to $\dfrac{\partial g_a(\boldsymbol{\psi})}{\partial \psi_b}$, and

$$Q_1 = \left\{G(\boldsymbol{\psi}_0)^\top I(\boldsymbol{\theta}_0) G(\boldsymbol{\psi}_0)\right\}^{-1/2} G(\boldsymbol{\psi}_0)^\top I(\boldsymbol{\theta}_0)^{1/2}, \quad \textit{(which is an } s \times r \textit{ matrix).}$$

Since $Q_1 Q_1^\top = \mathbb{I}_s$, we can find an $(r-s) \times r$ matrix Q_2 such that the rows of Q_1 and Q_2 together form an orthonormal basis for $\mathbb{R}^r$. Thus, we have

$$Q_2 Q_2^\top = \mathbb{I}_{r-s} \quad \textit{and} \quad Q_1^\top Q_1 + Q_2^\top Q_2 = \mathbb{I}_r.$$

Since $\hat{\boldsymbol{\psi}}_n$ satisfies $G(\hat{\boldsymbol{\psi}}_n)^\top \boldsymbol{U}(\boldsymbol{g}(\hat{\boldsymbol{\psi}}_n)) = \boldsymbol{0}$, Taylor expanding $\boldsymbol{U}$ and $\boldsymbol{g}$ gives

$$\begin{aligned}
\boldsymbol{0} = G(\hat{\boldsymbol{\psi}}_n)^\top \boldsymbol{U}(\boldsymbol{g}(\hat{\boldsymbol{\psi}}_n)) &\approx G(\hat{\boldsymbol{\psi}}_n)^\top \boldsymbol{U}(\boldsymbol{\theta}_0) - G(\hat{\boldsymbol{\psi}}_n)^\top I(\boldsymbol{\theta}_0) G(\boldsymbol{\psi}_0)\big(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0\big) \\
\Rightarrow \quad G(\boldsymbol{\psi}_0)^\top \boldsymbol{U}(\boldsymbol{\theta}_0) &\approx G(\boldsymbol{\psi}_0)^\top I(\boldsymbol{\theta}_0) G(\boldsymbol{\psi}_0)\big(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0\big) \\
\Rightarrow \quad \hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0 &\approx \left\{G(\boldsymbol{\psi}_0)^\top I(\boldsymbol{\theta}_0) G(\boldsymbol{\psi}_0)\right\}^{-1/2} Q_1 I(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{U}(\boldsymbol{\theta}_0) \\
\Rightarrow \quad \boldsymbol{g}(\hat{\boldsymbol{\psi}}_n) - \boldsymbol{\theta}_0 &\approx G(\boldsymbol{\psi}_0)\big(\hat{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0\big) \approx I(\boldsymbol{\theta}_0)^{-1/2} Q_1^\top Q_1 I(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{U}(\boldsymbol{\theta}_0).
\end{aligned}$$

Similarly,

$$-\boldsymbol{U}(\boldsymbol{\theta}_0) = \boldsymbol{U}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{U}(\boldsymbol{\theta}_0) \approx -I(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \quad \Rightarrow \quad \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \approx I(\boldsymbol{\theta}_0)^{-1} \boldsymbol{U}(\boldsymbol{\theta}_0).$$

It follows that

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{g}(\hat{\boldsymbol{\psi}}_n) \approx I(\boldsymbol{\theta}_0)^{-1} \boldsymbol{U}(\boldsymbol{\theta}_0) - I(\boldsymbol{\theta}_0)^{-1/2} Q_1^\top Q_1 I(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{U}(\boldsymbol{\theta}_0) = I(\boldsymbol{\theta}_0)^{-1/2} Q_2^\top Q_2 I(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{U}(\boldsymbol{\theta}_0).$$

Taylor expanding $\ln \ell_{\boldsymbol{X}}(\cdot)$ implies

$$\begin{aligned}
2 \ln\left\{\ell_{\boldsymbol{X}}(\hat{\boldsymbol{\theta}}_n)/\ell_{\boldsymbol{X}}(\boldsymbol{g}(\hat{\boldsymbol{\psi}}_n))\right\} &= -2\left\{\ln \ell_{\boldsymbol{X}}(\boldsymbol{g}(\hat{\boldsymbol{\psi}}_n)) - \ln \ell_{\boldsymbol{X}}(\hat{\boldsymbol{\theta}}_n)\right\} \\
&\approx \big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{g}(\hat{\boldsymbol{\psi}}_n)\big)^\top I(\boldsymbol{\theta}_0)\big(\hat{\boldsymbol{\theta}}_n - \boldsymbol{g}(\hat{\boldsymbol{\psi}}_n)\big) \approx \big\|Q_2 I(\boldsymbol{\theta}_0)^{-1/2} \boldsymbol{U}(\boldsymbol{\theta}_0)\big\|_2^2.
\end{aligned}$$

[2] For details see Bradley, R.A. and Gart, J.J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika*, **49**, 205–214.

Theorem §5.5.6(iii) asserts that $I(\boldsymbol{\theta}_0)^{-1/2}\boldsymbol{U}(\boldsymbol{\theta}_0)$ converges in distribution to the r-variate standard normal distribution $N(\mathbf{0},\mathbb{I}_r)$. Thus we have

$$Q_2 I(\boldsymbol{\theta}_0)^{-1/2}\boldsymbol{U}(\boldsymbol{\theta}_0) \to N\big(\mathbf{0}, Q_2Q_2^\top\big) \equiv N\big(\mathbf{0},\mathbb{I}_{r-s}\big) \quad \textit{in distribution.}$$

It follows that $2\ln\big\{\ell_{\boldsymbol{X}}(\hat{\boldsymbol{\theta}}_n)/\ell_{\boldsymbol{X}}(\boldsymbol{g}(\hat{\boldsymbol{\psi}}_n))\big\} \approx \big\|Q_2 I(\boldsymbol{\theta}_0)^{-1/2}\boldsymbol{U}(\boldsymbol{\theta}_0)\big\|_2^2$ converges in distribution to the sum of squares of $(r-s)$ i.i.d. standard normal random variables, i.e. to the χ^2_{r-s} distribution. ▮

6.7.3 Define $\chi^2_f(\alpha)$ to be α^{th} upper quantile of χ^2_f.

Wilks' Theorem suggests an approximate size α *generalised likelihood ratio test*.

Generalised likelihood ratio test: *reject H_0 iff* $2\ln\Lambda_{\boldsymbol{X}}(H_0,H_1) > \chi^2_{r-s}(\alpha)$.

6.7.4 Special case:

Test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs $H_1 : \boldsymbol{\theta} \in \Theta$ (subset of dimension k in $\mathbb{R}^k$) (hence $r = k$, $s = 0$).

mle: $\hat{\boldsymbol{\theta}}_n$ maximises $\ell_{\boldsymbol{X}}(\boldsymbol{\theta})$ over $\boldsymbol{\theta}\in\Theta$

Under H_0, $2\ln\Lambda_{\boldsymbol{X}}(H_0,H_1) = 2\big\{\ln\ell_{\boldsymbol{X}}(\hat{\boldsymbol{\theta}}_n) - \ln\ell_{\boldsymbol{X}}(\boldsymbol{\theta}_0)\big\}$ converges in distribution to χ^2_k.

6.7.5 **Example §6.6.4.** *(cont'd)*

$\boldsymbol{X} = (X_1,\ldots,X_n)$ iid $N(\mu,\sigma^2)$.
Test $H_0 : \sigma = \sigma_0$ vs $H_1 : \sigma \neq \sigma_0$ (for a specified constant $\sigma_0 > 0$).

Loglikelihood function: $S_{\boldsymbol{X}}(\mu,\sigma) = -n\ln\sigma - (1/2)\sigma^{-2}\sum_{i=1}^n (X_i-\mu)^2$

Score function: $\boldsymbol{U}(\mu,\sigma) = \begin{bmatrix} \sigma^{-2}\sum_{i=1}^n (X_i-\mu) \\ -n\sigma^{-1} + \sigma^{-3}\sum_{i=1}^n (X_i-\mu)^2 \end{bmatrix}$

Solving $\boldsymbol{U}(\mu,\sigma) = \mathbf{0}$ gives the (unconstrained) mle under H_1:

$$\hat{\mu}_n = \bar{X} = n^{-1}\sum_{i=1}^n X_i \quad \text{and} \quad \hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n (X_i-\bar{X})^2 = n^{-1}S_{xx}.$$

Solving $\sigma_0^{-2}\sum_{i=1}^n (X_i-\mu) = 0$ for μ gives the (constrained) mle under H_0:

$$\tilde{\mu}_n = \bar{X} \quad \text{and} \quad \tilde{\sigma}_n^2 = \sigma_0^2.$$

Generalised likelihood ratio (GLR) test statistic:

$$\begin{aligned} 2\ln\Lambda_{\boldsymbol{X}}(H_0,H_1) &= 2\big\{S_{\boldsymbol{X}}(\hat{\mu}_n,\hat{\sigma}_n) - S_{\boldsymbol{X}}(\tilde{\mu}_n,\tilde{\sigma}_n)\big\} \\ &= n\Big\{\Big(\frac{\hat{\sigma}_n}{\sigma_0}\Big)^2 - 2\ln\Big(\frac{\hat{\sigma}_n}{\sigma_0}\Big) - 1\Big\} = \frac{S_{xx}}{\sigma_0^2} - n\ln\Big(\frac{S_{xx}}{n\sigma_0^2}\Big) - n. \end{aligned}$$

Size α GLR test: *reject H_0 if* $2\ln\Lambda(H_0,H_1) > \chi^2_1(\alpha)$.

§6.8 Bootstrap test

6.8.1 Data: $\boldsymbol{X} \sim G$.

Given some subclasses of distributions $\mathscr{G}_0$ and $\mathscr{G}_1$, we wish to conduct a test of

$$H_0 : G \in \mathscr{G}_0 \qquad \text{vs} \qquad H_1 : G \in \mathscr{G}_1,$$

based on a test statistic $W = W(\boldsymbol{X}, \mathscr{G}_0, \mathscr{G}_1)$, extreme values of which suggest evidence against H_0 in favour of H_1.

In a typical real-life problem, the distributions of W under $H_0 : G \in \mathscr{G}_0$ are not explicitly available, making it difficult to construct a critical region for the test. The bootstrap method provides a possible solution to the problem.

6.8.2 Let $T^* = T(\boldsymbol{X}^*; \boldsymbol{X})$ be a bootstrap statistic for which

- $\boldsymbol{X}^*$ denotes a sample drawn in *some* way from $\boldsymbol{X}$;
  (e.g. $\boldsymbol{X}^*$ may be, but not necessarily, a conventional bootstrap sample drawn with replacement from $\boldsymbol{X} = (X_1, \ldots, X_n)$)
- the bootstrap distribution function $\hat{F}_T(\cdot) = \mathbb{P}\big(T^* \leq \cdot \big| \boldsymbol{X}\big)$ approximates $\mathbb{P}_G(W \leq \cdot)$ if $G \in \mathscr{G}_0$, and does not deviate much from the above null distribution if $G \in \mathscr{G}_1 \setminus \mathscr{G}_0$.

Note: To ensure the above properties, one possible approach is to draw $\boldsymbol{X}^*$ from an estimated distribution (likely to be different from the empirical distribution) which belongs to $\mathscr{G}_0$, i.e. respects the null hypothesis.

Denote by $\hat{F}_T^{-1}(\xi)$ the ξ^{th} quantile of the bootstrap distribution $\hat{F}_T(\cdot)$.

A size α *bootstrap test*

$$\textit{rejects } H_0 \textit{ if } \; W \notin \big[\hat{F}_T^{-1}\big((1-\lambda)\alpha\big), \hat{F}_T^{-1}(1-\lambda\alpha)\big],$$

for some pre-specified $\lambda \in [0, 1]$. As a general rule, λ is set to ensure that under $G \in \mathscr{G}_1 \setminus \mathscr{G}_0$, the distribution of W concentrates outside $\big[\hat{F}_T^{-1}\big((1-\lambda)\alpha\big), \hat{F}_T^{-1}(1-\lambda\alpha)\big]$, thus yielding a satisfactory power for the test.

Commonly, we set:

	λ	$\mathbb{P}_G(W \leq \cdot)$, $G \in \mathscr{G}_1 \setminus \mathscr{G}_0$
(i)	1/2	concentrate on either small or large values relative to H_0
(ii)	1	concentrate on large values relative to H_0
(iii)	0	concentrate on small values relative to H_0

Alternatively, based on observed values $\boldsymbol{X} = \boldsymbol{x}$ and $W = w = W(\boldsymbol{x}, \mathscr{G}_0, \mathscr{G}_1)$, calculate

$$\begin{aligned} p\text{-}value &= \inf\Big\{\alpha \in [0,1] : w \notin \big[\hat{F}_T^{-1}\big((1-\lambda)\alpha\big), \hat{F}_T^{-1}(1-\lambda\alpha)\big]\Big\} \\ &\approx \min\left\{\frac{\mathbb{P}\big(T^* \le w\big|\boldsymbol{x}\big)}{1-\lambda}, \frac{\mathbb{P}\big(T^* \ge w\big|\boldsymbol{x}\big)}{\lambda}, 1\right\}. \end{aligned}$$

<u>Informal justification of bootstrap test:</u> (denote by $J_G(\cdot)$ the cdf of W under G)

- At $G \in \mathscr{G}_0$ —
  the power function of the bootstrap test is given by
  $$\begin{aligned} \mathbb{P}_G\Big(W \notin \big[\hat{F}_T^{-1}\big((1-\lambda)\alpha\big), \hat{F}_T^{-1}(1-\lambda\alpha)\big]\Big) &\approx \mathbb{P}_G\Big(W \notin \big[J_G^{-1}\big((1-\lambda)\alpha\big), J_G^{-1}(1-\lambda\alpha)\big]\Big) \\ &\approx 1 - J_G\big(J_G^{-1}(1-\lambda\alpha)\big) + J_G\big(J_G^{-1}((1-\lambda)\alpha)\big) = \alpha. \end{aligned}$$

- At $G \in \mathscr{G}_1 \setminus \mathscr{G}_0$ —
  by the property of $\hat{F}_T$ and the choice of λ, J_G has a concentration outside the interval $\big[\hat{F}_T^{-1}\big((1-\lambda)\alpha\big), \hat{F}_T^{-1}(1-\lambda\alpha)\big]$, implying a high power for the bootstrap test.

6.8.3 Monte Carlo procedure for size α bootstrap test:

1. simulate a large number, B say, of independent replicates of $\boldsymbol{X}^*$;
2. calculate $T^* = T\big(\boldsymbol{X}^*; \boldsymbol{X}\big)$ for each $\boldsymbol{X}^*$ and rank their values as
   $$-\infty = t^*_{(0)} < t^*_{(1)} \le \cdots \le t^*_{(B)} < t^*_{(B+1)} = \infty,$$
3. reject H_0 if $W \notin \big[t^*_{(k_1)}, t^*_{(k_2)}\big]$, where $k_1 \approx (1-\lambda)\alpha(B+1)$ and $k_2 \approx (1-\lambda\alpha)(B+1)$,
   or compute p-value to be
   $$\min\left\{\frac{\sum_{b=1}^B \mathbf{1}\big\{t^*_{(b)} \le W\big\}}{(1-\lambda)B}, \frac{\sum_{b=1}^B \mathbf{1}\big\{t^*_{(b)} \ge W\big\}}{\lambda B}, 1\right\}.$$

6.8.4 **Example §6.8.1** Consider two independent random samples, $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$, drawn respectively from the distribution functions F and G. For given constants $c \in \mathbb{R}$ and $p_0 \in (0,1)$, we wish to test

$$H_0 : \mathbb{P}_{F,G}(X_1 > cY_1) = p_0 \quad \text{against} \quad H_1 : \mathbb{P}_{F,G}(X_1 > cY_1) \ne p_0.$$

Consider $\rightarrow$ parameter of interest: $\theta(F, G) = \mathbb{P}_{F,G}(X_1 > cY_1)$

nonparametric mle: $\hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) = \theta(\hat{F}_n, \hat{G}_m) = (mn)^{-1}\sum_{j=1}^m \sum_{i=1}^n \mathbf{1}\{X_i > cY_j\}$, where $\hat{F}_n$ and $\hat{G}_m$ denote the empirical cdf's of the samples $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.

The test amounts to setting

$$\mathscr{G}_0 = \{F \times G : \theta(F, G) = p_0\} \text{ and } \mathscr{G}_1 = \{F \times G : \theta(F, G) \neq p_0\}.$$

Based on the observed dataset, a natural choice of test statistic is $W = \hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) - p_0$, since an observed value of W too large or too small provides evidence against H_0 in favour of H_1.

Consider a bootstrap test as follows. Draw bootstrap samples $(\boldsymbol{X}^*, \boldsymbol{Y}^*)$ such that

- $\boldsymbol{X}^* = (X_1^*, \ldots, X_n^*)$ i.i.d. from $\hat{F}_n$ (sampling from $\boldsymbol{X}$ with replacement),
- $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_m^*)$ i.i.d. from $\hat{G}_m$ (sampling from $\boldsymbol{Y}$ with replacement).

Reasonable choice of bootstrap statistic:

$$\begin{aligned} T^* &= \hat{\theta}(\boldsymbol{X}^*, \boldsymbol{Y}^*) - \hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) \\ &= \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{1}\{X_i^* > cY_j^*\} - \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \mathbf{1}\{X_i > cY_j\}. \end{aligned}$$

Informal justification:

> The bootstrap distribution of T^* is always centred at 0, no matter which hypothesis is correct. Thus, it is similar to the null distribution of W, which is also centred at 0. However, if H_0 is wrong, then W has a distribution centred away from 0, rendering it different from that of T^*.

A size α bootstrap test can be set up to

$$\textit{reject } H_0 \textit{ if } \quad \hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) - p_0 \notin \left[\hat{F}_T^{-1}(\alpha/2), \hat{F}_T^{-1}(1-\alpha/2)\right],$$

where $\hat{F}_T$ denotes the cdf of T^* conditional on $(\boldsymbol{X}, \boldsymbol{Y})$.

The p-value is approximated by

$$2 \min\left\{\mathbb{P}\big(T^* \leq \hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) - p_0 \big| \boldsymbol{X}, \boldsymbol{Y}\big), \mathbb{P}\big(T^* \geq \hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) - p_0 \big| \boldsymbol{X}, \boldsymbol{Y}\big), 1/2\right\}.$$

Note: For a size α one-sided test of

$$H_0 : \mathbb{P}_{F,G}(X_1 > cY_1) \geq p_0 \qquad \text{against} \qquad H_1 : \mathbb{P}_{F,G}(X_1 > cY_1) < p_0,$$

we may follow a similar procedure and reject H_0 if $\hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) - p_0 < \hat{F}_T^{-1}(\alpha)$. The p-value is approximated by $\mathbb{P}\big(T^* \leq \hat{\theta}(\boldsymbol{X}, \boldsymbol{Y}) - p_0 \big| \boldsymbol{X}, \boldsymbol{Y}\big)$.

§6.9 Confidence set based on test inversion

6.9.1 **Definitions.**

$\boldsymbol{X}$ (data) $\sim f(\boldsymbol{x}|\theta)$ (probability function), $\theta \in \Theta$.

A *confidence set* for θ of *confidence level* $1-\alpha$ is a **random** set $\mathcal{S}_{1-\alpha}(\boldsymbol{X}) \subset \Theta$ such that

$$\mathbb{P}_\theta\big(\theta \in \mathcal{S}_{1-\alpha}(\boldsymbol{X})\big) = 1-\alpha \quad \text{for all } \theta \in \Theta.$$

6.9.2 *How to find a "good" confidence set $\mathcal{S}_{1-\alpha}(\boldsymbol{X})$ for θ?*

By *inverting* an appropriate hypothesis test of size α —

Idea:

Consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ with a size α test. Let $\mathcal{C}_\alpha(\theta_0)$ be the critical region, i.e.

$$\mathbb{P}_{\theta_0}\big(\boldsymbol{X} \in \mathcal{C}_\alpha(\theta_0)\big) = \alpha \text{ for all } \theta_0.$$

Define

$$\mathcal{S}_{1-\alpha}(\boldsymbol{X}) = \big\{\vartheta \in \Theta : \boldsymbol{X} \notin \mathcal{C}_\alpha(\vartheta)\big\},$$

i.e. those parameter values not conflicting with data $\boldsymbol{X}$ according to the above test. Then $\mathcal{S}_{1-\alpha}(\boldsymbol{X})$ is a level $(1-\alpha)$ confidence set for θ, since

$$\mathbb{P}_\theta\big(\theta \in \mathcal{S}_{1-\alpha}(\boldsymbol{X})\big) = 1 - \mathbb{P}_\theta\big(\boldsymbol{X} \in \mathcal{C}_\alpha(\theta)\big) = 1-\alpha.$$

Typically, an **optimal** test yields an **optimal** confidence set.

6.9.3 The above procedure produces a *two-sided* confidence interval in general. If we adopt a one-sided $H_1 : \theta < \theta_0$ (or $H_1 : \theta > \theta_0$), a *one-sided* confidence interval in the form of an *upper* (or *lower*) *confidence bound* will result instead.

6.9.4 **Example.** (cf. §6.6.4 and §6.7.5)

$\boldsymbol{X} = (X_1, \ldots, X_n)$ iid $N(\mu, \sigma^2)$

Suppose we want to find a level $(1-\alpha)$ *confidence set for* σ.

Consider inversion of a size α test of $H_0 : \sigma = \sigma_0$ vs $H_1 : \sigma \neq \sigma_0$.

§6.6.4 shows that a UMPU test of size α rejects H_0 if $S_{xx} \notin \big[\sigma_0^2 c_1',\, \sigma_0^2 c_2'\big]$, which leads to a level $(1-\alpha)$ confidence set for σ:

$$\Big\{\sigma : S_{xx} \in \big[\sigma^2 c_1',\, \sigma^2 c_2'\big]\Big\} = \left[\sqrt{\frac{S_{xx}}{c_2'}},\, \sqrt{\frac{S_{xx}}{c_1'}}\,\right].$$

§6.7.5 shows that a GLR test of size α rejects H_0 if $\dfrac{S_{xx}}{\sigma_0^2} - n\ln\left(\dfrac{S_{xx}}{n\sigma_0^2}\right) - n > \chi_1^2(\alpha)$, which leads to an *approximate* level $(1-\alpha)$ confidence set for σ:

$$\left\{\sigma : \frac{S_{xx}}{\sigma^2} - n\ln\left(\frac{S_{xx}}{n\sigma^2}\right) - n \le \chi_1^2(\alpha)\right\}.$$

6.9.5 **Exercise:** (continuing **Exercise** §5.6)

Let θ_0 be a fixed positive integer. Construct a size α UMP test of $H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$.

By inverting the above UMP test, construct a level $(1-\alpha)$ lower confidence bound for θ.

6.9.6 Bootstrap confidence intervals can be derived by inverting appropriate bootstrap tests.

Consider the problem of constructing a level $1-\alpha$ confidence interval for $\theta = \theta(F) \in \mathbb{R}$.
Data: $\boldsymbol{X} = (X_1, \ldots, X_n)$ i.i.d. from F Empirical cdf: $\hat{F}_n$
Parameter of interest: $\theta = \theta(F) \in \mathbb{R}$ Estimator: $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$
A test of $H_0 : \theta = \vartheta$, for some specified ϑ, can often be based on the test statistic

$$R(\boldsymbol{X}, \vartheta) = \hat{\theta} - \vartheta \quad \text{or} \quad R(\boldsymbol{X}, \vartheta) = \frac{\hat{\theta} - \vartheta}{\sqrt{\hat{v}}},$$

where $\hat{v} = \hat{v}(\boldsymbol{X})$ is a consistent estimator of $\mathrm{Var}_F\big(\hat{\theta}\big)$.

Let $\boldsymbol{X}^* = \big(X_1^*, \ldots, X_n^*\big)$ be i.i.d. from $\hat{F}_n$. Let $\hat{r}_\xi$ be the ξ^{th} bootstrap quantile which satisfies

$$\mathbb{P}_{\hat{F}_n}\Big(R\big(\boldsymbol{X}^*, \theta(\hat{F}_n)\big) \le \hat{r}_\xi\Big) = \xi.$$

Approximate level $1-\alpha$ *bootstrap confidence intervals* for θ are obtained generally as

$$\big\{\vartheta : \hat{r}_{(1-\lambda)\alpha} \le R(\boldsymbol{X}, \vartheta) \le \hat{r}_{1-\lambda\alpha}\big\},$$

for some pre-specified $\lambda \in [0, 1]$.

Method 1: ***Bootstrap percentile method***

Set $R(\boldsymbol{X}, \vartheta) = \hat{\theta} - \vartheta$. Write $\hat{\theta}^* = \hat{\theta}(\boldsymbol{X}^*)$. Then $R\big(\boldsymbol{X}^*, \theta(\hat{F}_n)\big) = \hat{\theta}^* - \theta(\hat{F}_n)$ and $\hat{r}_\xi$ solves

$$\mathbb{P}_{\hat{F}_n}\big(\hat{\theta}^* - \theta(\hat{F}_n) \le \hat{r}_\xi\big) = \xi.$$

The corresponding level $1-\alpha$ bootstrap confidence intervals are given generally by

$$\big[\hat{\theta} - \hat{r}_{1-\lambda\alpha},\ \hat{\theta} - \hat{r}_{(1-\lambda)\alpha}\big].$$

Special cases:

- $\lambda = 1/2 \;\rightarrow\;$ equal-tailed interval $\;\rightarrow\; \big[\,\hat\theta - \hat r_{1-\alpha/2},\, \hat\theta - \hat r_{\alpha/2}\,\big]$
- $\lambda = 1 \;\rightarrow\;$ lower confidence bound $\;\rightarrow\; \big[\,\hat\theta - \hat r_{1-\alpha},\, \infty\,\big)$
- $\lambda = 0 \;\rightarrow\;$ upper confidence bound $\;\rightarrow\; \big(-\infty,\, \hat\theta - \hat r_{\alpha}\,\big]$

Method 2: $\boxed{\textbf{\textit{Bootstrap-t method}}}$

Set $R(\boldsymbol{X}, \vartheta) = \dfrac{\hat\theta - \vartheta}{\sqrt{\hat v}}$. Write $\hat v^* = \hat v(\boldsymbol{X}^*)$. Then $R\big(\boldsymbol{X}^*, \theta(\hat F_n)\big) = \dfrac{\hat\theta^* - \theta(\hat F_n)}{\sqrt{\hat v^*}}$ and $\hat r_\xi$ solves

$$\mathbb{P}_{\hat F_n}\left(\frac{\hat\theta^* - \theta(\hat F_n)}{\sqrt{\hat v^*}} \le \hat r_\xi\right) = \xi.$$

The corresponding level $1-\alpha$ bootstrap confidence intervals are given generally by

$$\Big[\hat\theta - \sqrt{\hat v}\,\hat r_{1-\lambda\alpha},\, \hat\theta - \sqrt{\hat v}\,\hat r_{(1-\lambda)\alpha}\Big].$$

Special cases:

- $\lambda = 1/2 \;\rightarrow\;$ equal-tailed interval $\;\rightarrow\; \big[\,\hat\theta - \sqrt{\hat v}\,\hat r_{1-\alpha/2},\, \hat\theta - \sqrt{\hat v}\,\hat r_{\alpha/2}\,\big]$
- $\lambda = 1 \;\rightarrow\;$ lower confidence bound $\;\rightarrow\; \big[\,\hat\theta - \sqrt{\hat v}\,\hat r_{1-\alpha},\, \infty\,\big)$
- $\lambda = 0 \;\rightarrow\;$ upper confidence bound $\;\rightarrow\; \big(-\infty,\, \hat\theta - \sqrt{\hat v}\,\hat r_{\alpha}\,\big]$

Both bootstrap percentile and bootstrap-t intervals can be obtained by calculating $\hat r_\xi$ for suitably chosen ξ. Monte Carlo procedure for approximating $\hat r_\xi$:

1. calculate $\theta(\hat F_n)$ based on $\boldsymbol{X}$,
2. simulate a large number, B say, of bootstrap samples $\boldsymbol{X}^*$ from $\hat F_n$,
3. calculate $R\big(\boldsymbol{X}^*, \theta(\hat F_n)\big)$ for each $\boldsymbol{X}^*$ to obtain $R^{*1}, \ldots, R^{*B}$ say,
4. rank the R^{*b}'s into a sorted sequence $-\infty = R^*_{(0)} < R^*_{(1)} \le \cdots \le R^*_{(B)} < R^*_{(B+1)} = \infty$,
5. set $k \approx \xi(B+1)$, approximate $\hat r_\xi$ by $R^*_{(k)}$.

6.9.7 **Exercise.** (c.f. Example §6.8.1)

Consider two independent random samples, $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$, drawn respectively from the distribution functions F and G, which are defined on $(0, \infty)$.

Construct a level $(1-\alpha)$ upper confidence bound on the median of X_1/Y_1 by inverting an appropriate bootstrap test.