Theory of Statistical Inference - Lecture IV.3
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

- recall $\Theta = H_0 \cup H_a$ where $H_0 \cap H_a = \phi$ and based on observed data x decide if $\theta_{true} \in H_0$ or $\theta_{true} \in H_a$ using 0-1 loss

- this is based on a test function $\varphi : \mathcal{X} \to [0,1]$ where $\varphi(x) =$ probability of rejecting $H_0 = \delta(x, H_a)$

- optimal φ minimizes

$$R(\theta, \varphi) = I_{H_0}(\theta)E_\theta(\varphi) + I_{H_a}(\theta)(1 - E_\theta(\varphi))$$

for every $\theta \in \Theta$

- but we know that there doesn't exist an optimal φ

- so various restrictions are applied

- **size α restriction** - let $\mathcal{D}_\alpha(I^{dec}) \subset \mathcal{D}(I^{dec})$ be given by

$$\begin{aligned} \mathcal{D}_\alpha(I^{dec}) &= \{\varphi : I_{H_0}(\theta)E_\theta(\varphi) \le \alpha \text{ for all } \theta \in \Theta\} \\ &= \{\varphi : E_\theta(\varphi) \le \alpha \text{ for all } \theta \in H_0\} \end{aligned}$$

the set of size α tests

- now search for $\varphi \in \mathcal{D}_\alpha(I^{dec})$ that minimizes $I_{H_a}(\theta)(1 - E_\theta(\varphi))$ for all $\theta \in \Theta$ or equivalently maximizes the power function $\beta(\theta) = E_\theta(\varphi)$ for all $\theta \in H_a$

- recall that this is a nonconvex problem

Theorem IV.3.1 *(The Fundamental Lemma or Neyman-Pearson Theorem)*
Suppose $\Theta = \{\theta_0, \theta_1\}$ and $H_0 = \{\theta_0\}$. Then,

(i) there exists $\varphi \in \mathcal{D}_\alpha(I^{dec})$ satisfying $E_{\theta_0}(\varphi) = \alpha$ where φ is of the form

$$\varphi(x) = \begin{cases} 1, & f_{\theta_1}(x)/f_{\theta_0}(x) > k_0 \\ \gamma, & f_{\theta_1}(x)/f_{\theta_0}(x) = k_0 \\ 0, & f_{\theta_1}(x)/f_{\theta_0}(x) < k_0 \end{cases}$$

for some $k_0 \in [0, \infty], \gamma \in [0, 1]$.
(ii) φ in (i) is MP (most powerful) size α.
(iii) If φ_* is also MP size α, then the set $\{x : \varphi_*(x) \neq \varphi(x), f_{\theta_1}(x)/f_{\theta_0}(x) \neq k_0\}$ has (support) measure 0 and $E_{\theta_0}(\varphi_*) = \alpha$ unless $E_{\theta_1}(\varphi) = 1$.

Proof: See Lehmann and Romano (2005) Testing Statistical Hypotheses.

Corollary IV.3.1 If φ is MP size α, then $\alpha \leq E_{\theta_1}(\varphi)$ with the inequality strict when $0 < \alpha < 1$.
Proof of the Corollary: Consider $\varphi_\alpha(x) \equiv \alpha$. Then $E_{\theta_1}(\varphi) \geq E_{\theta_1}(\varphi_\alpha) = \alpha$. If $E_{\theta_1}(\varphi) = \alpha$, then φ_α is MP size α which, by (iii), implies that $\{x : f_{\theta_1}(x)/f_{\theta_0}(x) = k_0\}$ almost everywhere, which implies $k_0 = 1$, which implies $f_{\theta_1}(x) = f_{\theta_0}(x)$ almost everywhere, which is a contradiction. ■

- in practice, after observing x generate $u \sim U(0,1)$ and reject H_0 whenever $u \leq \varphi(x)$

- **note 1** - if $\{x : f_{\theta_1}(x)/f_{\theta_0}(x) = k_0\}$ has (support) measure 0, then randomized test is not required

- **note 2** - to find k_0 and γ, let

$$r(k) = P_{\theta_0}(f_{\theta_1}(x)/f_{\theta_0}(x) > k) = 1 - P_{\theta_0}(f_{\theta_1}(x)/f_{\theta_0}(x) \le k)$$

and find k_0 s.t. $r(k_0 - \epsilon) > \alpha$ for every $\epsilon > 0$ and $r(k_0) \le \alpha$ and put

$$\gamma = \begin{cases} \frac{\alpha - r(k_0)}{r(k_0 - 0) - r(k_0)}, & r(k_0) < \alpha \\ 0, & \text{otherwise} \end{cases}$$

and note $1 - r(k_0 - \epsilon) < 1 - \alpha$ for every $\epsilon > 0$ and $1 - r(k_0) \ge 1 - \alpha$ so $k_0 = \inf\{k : F(k; n, \theta_0) \ge 1 - \alpha\} = (1 - \alpha)$-th quantile of the distribution of $f_{\theta_1}(x)/f_{\theta_0}(x)$ under P_{θ_0}

- and so $\gamma = 0$ iff $F(k_0; n, \theta_0) = 1 - \alpha$ (there are continuous problems that require randomization)

- then

$$
\begin{aligned}
E_{\theta_0}(\varphi) &= P_{\theta_0}(f_{\theta_1}(x)/f_{\theta_0}(x) > k_0) + \gamma P_{\theta_0}(f_{\theta_1}(x)/f_{\theta_0}(x) = k_0) \\
&= r(k_0) + \frac{\alpha - r(k_0)}{r(k_0 - 0) - r(k_0)} \times \\
&\quad \left(P_{\theta_0}\left(\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} \leq k_0 \right) - P_{\theta_0}\left(\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} \leq k_0 - 0 \right) \right) \\
&= r(k_0) + \frac{\alpha - r(k_0)}{r(k_0 - 0) - r(k_0)}(1 - r(k_0) - 1 + r(k_0 - 0)) = \alpha
\end{aligned}
$$

Example IV.3.1 *binomial*

- $x = (x_1, \ldots, x_n)$ iid Bernoulli(θ) where $\theta \in \Theta = \{\theta_0, \theta_1\}$ with $\theta_0 < \theta_1$
- $T(x) = n\bar{x}$ is sufficient so we can base the test on T
- then

$$\frac{f_{\theta_1}(T(x))}{f_{\theta_0}(T(x))} = \left(\frac{\theta_1}{\theta_0}\right)^{T(x)} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-T(x)}$$

$$r(k) = P_{\theta_0 T}\left(\left(\frac{\theta_1}{\theta_0}\right)^{T(x)} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-T(x)} > k\right)$$

$$= \quad P_{\theta_0 T}\left(T(x) > \frac{k - n\log\left(\frac{1-\theta_1}{1-\theta_0}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right) - \log\left(\frac{1-\theta_1}{1-\theta_0}\right)}\right) \text{ since } \log\left(\frac{\theta_1}{\theta_0}\right) > 0$$

- k_0 is then specified so that

$$\frac{k_0 - n\log\left(\frac{1-\theta_1}{1-\theta_0}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right) - \log\left(\frac{1-\theta_1}{1-\theta_0}\right)}$$

is the $(1-\alpha)$-th quantile of the binomial(n, θ_0) (denoted $k_{1-\alpha}(\theta_0)$ and note that this is independent of θ_1

- therefore, φ given by k_0 and γ is UMP (uniformly most powerful) size α for the problem $H_0 = \{\theta_0\}$ versus $H_a = (\theta_0, 1]$

- also, the test φ is MP size $E_\theta(\varphi)$ for $H_0 = \{\theta\}$ versus $H_a = \{\theta_1\}$ when $\theta_1 > \theta_0$ and so by Corollary IV.3.1

$$E_\theta(\varphi) \;\le\; E_{\theta_1}(\varphi) = \gamma \binom{n}{k_{1-\alpha}(\theta_0)} \theta_1^{k_{1-\alpha}(\theta_0)} (1-\theta_1)^{n-k_{1-\alpha}(\theta_0)} + \sum_{i=k_{1-\alpha}(\theta_0)+1} \binom{n}{i} \theta_1^i (1-\theta_1)^{n-i}$$

which is continuous in θ_1 (it is a polynomial in θ_1) and so converges to $E_{\theta_0}(\varphi) = \alpha$ as $\theta_1 \downarrow \theta_0$ so $E_\theta(\varphi) \le \alpha$ when $\theta \le \theta_0$ which implies that φ is size α for $H_0 = [0, \theta_0]$ versus $H_a = (\theta_0, 1]$

- if φ_* is also size α for $H_0 = [0, \theta_0]$versus $H_a = (\theta_0, 1]$, then it is size α for $H_0 = \{\theta_0\}$ versus $H_a = \{\theta_1\}$ where $\theta_1 > \theta_0$ but then $E_{\theta_1}(\varphi) \ge E_{\theta_1}(\varphi_*)$ and so φ is UMP size α for $H_0 = [0, \theta_0]$ versus $H_a = (\theta_0, 1]$

- clearly in the binomial case we can find a UMP size α test, say φ_*, for $H_0 = [\theta_0, 1]$ versus $H_a = [0, \theta_0)$ based on the likelihood ratio $f_{\theta_0}(T(x))/f_{\theta_1}(T(x))$ for $\theta_1 < \theta_0$ and $\varphi_* \ne \varphi$

- this establishes that there is no UMP size α test for $H_0 = \{\theta_0\}$ versus $H_a = [0, \theta_0) \cup (\theta_0, 1]$ (two-sided problem) ■

- this argument will work for a number of problems to find UMP size α tests for one-sided problems

- it is also the case that to achieve exact size α, in some cases we will have $f_{\theta_1}(x)/f_{\theta_0}(x) > k_0$, and so $H_0 = \{\theta_0\}$ is rejected in favor of $H_a = \{\theta_1\}$ even though $f_{\theta_1}(x)/f_{\theta_0}(x) < 1$, i.e., optimal size α does not respect the likelihood ordering

Exercise IV.3.1 (A continuous problem that requires randomization.) Suppose $x = (x_1, \ldots, x_n)$ are iid $U(0, \theta)$ where $\Theta = (0, \infty)$. Determine the UMP size α test for $H_0 = (0, \theta_0]$ versus $H_a = (\theta_0, \infty)$

- now suppose the test function φ_{ψ_0} is size α for $H_0 = \Psi^{-1}\{\psi_0\}$ versus $H_a = \{\Psi^{-1}\{\psi_0\}\}^c$ for each $\psi_0 \in \Psi(\Theta)$ and let $u \sim U(0,1)$ independent of x and define

$$C_\Psi(u,x) = \{\psi_0 : \varphi_{\psi_0}(x) < u\} = \text{ set of } \psi_0 \text{ where } H_0 \text{ is accepted}$$

- then denoting joint distribution of (u,x) by P_θ^*

$$\begin{aligned} & P_\theta^*(\{x : \Psi(\theta) \in C_\Psi(u,x)\}) = E_\theta(P(\{x : \Psi(\theta) \in C_\Psi(u,x)\} \,|\, x)) \\ = \;& E_\theta(P(\{x : \varphi_{\Psi(\theta)}(x) < u \,|\, x)) = 1 - E_\theta(\varphi_{\Psi(\theta)}) \geq 1 - \alpha \end{aligned}$$

and so is a $(1-\alpha)$-confidence region for Ψ

- note - in general, $C_\Psi(u,x)$ will be a randomized region, namely, for observed x, when $\varphi_{\psi_0}(x) \neq 0, 1$ then $\psi_0 \in C_\Psi(u,x)$ with probability $1 - \varphi_{\psi_0}(x)$

- also when $\Psi(\theta) \neq \Psi(\theta')$, then

$$
\begin{aligned}
P_\theta^*(\{x : \Psi(\theta') \in C_\Psi(u,x)\}) &= 1 - E_\theta(\varphi_{\Psi(\theta')}) \\
&= \text{probability of } C_\Psi \text{ covering the false value } \Psi(\theta')
\end{aligned}
$$

- so a size α test that maximizes the power, produces a $(1-\alpha)$-confidence region for Ψ that minimizes the probability of covering a false value

Definition IV.3.2 A $(1-\alpha)$-confidence region C_Ψ for $\psi = \Psi(\theta)$ is *uniformly most accurate* (UMA) if for all $\theta, \theta' \in \Theta$ with $\Psi(\theta) \neq \Psi(\theta')$, we have that $P_\theta^*(\{x : \Psi(\theta') \in C_\Psi(u,x)\})$ is minimized among all $(1-\alpha)$-confidence regions for $\psi = \Psi(\theta)$. ■

Exercise IV.3.2 If C_Ψ is a uniformly most accurate $(1-\alpha)$-confidence region C_Ψ for $\psi = \Psi(\theta)$, then show that

$$
\varphi_{\Psi(\theta)}(x) = P_\theta^*(\{x : \Psi(\theta) \notin C_\Psi(u,x)\})
$$

is UMP size α for $H_0 = \{\Psi(\theta)\}$ versus $H_a = \{\Psi(\theta)\}^c$.

- note - define $\varphi_\psi(x) \equiv \alpha$ for every ψ and then

$$C_\Psi(u,x) = \begin{cases} \Psi(\Theta), & \alpha < u \\ \phi, & \text{otherwise} \end{cases}$$

is a $(1-\alpha)$-confidence region C_Ψ for $\psi = \Psi(\theta)$ with exact coverage so with randomized regions we can always obtain exact coverage and improve accuracy by doing so

- note often an alternative criterion for optimality is considered, namely, seek a $(1-\alpha)$-confidence region C_Ψ for $\psi = \Psi(\theta)$ that minimizes $E_\theta(\nu(C_\Psi(u,x)))$ uniformly in θ where $\nu(C_\Psi(u,x))$ is a measure of the size of $C_\Psi(u,x)$ (like volume)

- we have

$$E_\theta(\nu(C_\Psi(u,x))) = \int_{\mathcal{X}}\int_0^1\int_{\Psi(\Theta)} I_{C_\Psi(u,x)}(\psi)\,\nu(d\psi)\,du\,P_\theta(dx)$$

$$\begin{aligned}
&= \int_{\Psi(\Theta)}\int_{\mathcal{X}}\int_0^1 I_{C_\Psi(u,x)}(\psi)\,du\,P_\theta(dx)\,\nu(d\psi) \\
&= \int_{\Psi(\Theta)} P_\theta^*(\{x:\psi\in C_\Psi(u,x)\}\,\nu(d\psi) \\
&= \int_{\Psi(\Theta)\setminus\Psi(\theta)} P_\theta^*(\{x:\psi\in C_\Psi(u,x)\}\,\nu(d\psi) + \\
&\quad \nu(\{\Psi(\theta)\})P_\theta^*(\{x:\Psi(\theta)\in C_\Psi(u,x)\} \\
&= \int_{\Psi(\Theta)\setminus\Psi(\theta)} P_\theta^*(\{x:\psi\in C_\Psi(u,x)\}\,\nu(d\psi) + \nu(\{\Psi(\theta)\})(1-\alpha)
\end{aligned}$$

- so, if $\nu(\{\Psi(\theta)\})$ is constant in θ, minimizing size is equivalent to maximizing accuracy