S620 - Introduction To Statistical Theory - Spring 2014 - Information Sheet Exam I - Enrique Areyan

6 elements of statistical inference (1) Parameter Space Θ (2) Sample Space $\mathfrak{X}$ (3) Family of prob. distributions on $\mathfrak{X}$ (4) Action Space (5) Loss function $L: \Theta \times \mathcal{A} \to \mathbb{R}$ (6) a set $\mathcal{D}$ of decision rules. $d: \mathfrak{X} \to \mathcal{A}$

Risk Function: if state of nature $\theta \in \Theta$ obtains, then the risk associated with non-randomized decision rule d is $R(\theta, d) = E_\theta L(\theta, d(X)) = \begin{cases} \int_{\mathfrak{X}} L(\theta, d(x)) f(x;\theta)dx \\ \sum_{x \in \mathfrak{X}} L(\theta, d(x)) f(x;\theta) \end{cases}$. Associated with each d is a risk function: $R(\cdot, d): \Theta \to \mathbb{R}$.

The randomized decision rule $d^* = \lambda d_i + (1-\lambda) d_j$ is the rule that takes action $d_i(x)$ with prob. λ and action $d_j(x)$ with prob. $(1-\lambda)$. The risk function of d^* is: $R(\theta, d^*) = \lambda R(\theta, d_i) + (1-\lambda) R(\theta, d_j)$.

KEY NOTION: different decision rules should be compared by comparing their risk function (as function of θ)

Common loss function: abs. error $L(\theta, a) = |\theta - a|$, squared error loss $L(\theta, a) = (\theta - a)^2$

Admissibility: Let d_1, d_2 be decision rules s.t.: $R(\theta, d_1) \le R(\theta, d_2)$, $\forall \theta \in \Theta$, with $R(\theta, d_1) < R(\theta, d_2)$ for at least one $\theta \in \Theta$, then, d_1 strictly dominates d_2 $[d_1 \succ d_2]$ / If d is strictly dominated by some other rule, then d is inadmissible, otherwise it is admissible.

UMR: Let $\mathcal{D}_0$ be a collection of decision rules. The rule $d^* \in \mathcal{D}_0$ has uniformly min. risk in $\mathcal{D}_0$ iff $R(\theta, d^*) \le R(\theta, d)$; $\forall \theta \in \Theta$ and $\forall d \in \mathcal{D}_0$. [UMR may be impossible to obtain. Two strategies then:]

I. impartiality principles: restrict attention to "reasonable rules" that have constant risk so minimizing them is equivalent to minimizing the single risk value. A. Unbiased rules. d is L-unbiased iff: $E_\theta L(\theta', d(X)) \ge E_\theta L(\theta, d(X)) = R(\theta, d)$ $\forall \theta, \theta' \in \Theta$, where θ = true state of nature, θ' = some other state. B. Equivariant rules. II Relax the optimality criterion: A Minimax Principle: a decision rule $\bar{d}$ is minimax iff $\sup_\theta R(\theta, \bar{d}) \le \sup_\theta R(\theta, d)$, $\forall d \in \mathcal{D}$. Minmax principle: use minimax rule. this is a very conservative principle.

B. BAYES principle: Let π be a prob. dist. on Θ. Bayes risk of d: $r(d, \pi) = \int_\Theta R(\theta, d)\pi(\theta)d\theta$. Bayes principle = choose d_π, the rule that minimizes the Bayes risk for a given $\pi(\theta)$: $\int_\Theta R(\theta, d_\pi)\pi(\theta)d\theta \le \int_\Theta R(\theta, d)\pi(\theta)d\theta$ $\forall d \in \mathcal{D}$

THEOREM: d_π Bayes rule with constant risk $\Rightarrow$ d_π is MINIMAX.

Admissibility of Bayes Rules: THM 2.3: Assume $\Theta = \{\theta_1, \dots, \theta_k\}$ and $\pi(\cdot)$ is a prob. dist. on Θ. then, a Bayes rule w.r.t. π is admissible. THM 2.4: If a Bayes rule is unique then it is admissible. THM 2.5: d_π is admissible (continuous case).

BAYESIAN INFERENCE: Treat θ as a r.v. (1) prior distribution on θ (2) inference based on posterior dist. POSTERIOR $\pi(\theta|x) = \theta|x \propto$ prior $\times$ likelihood $= \pi(\theta) f(x;\theta) \Rightarrow \theta|x = \pi(\theta)f(x;\theta) / \int_\Theta \pi(\theta')f(x;\theta')d\theta'$ (joint of (θ, x) / marginal of x). To find Bayes rule, d_π, define $d_\pi(x)$ for each $x \in \mathfrak{X}$ to minimize expected posterior loss $\int_\Theta L(\theta, d(x))\pi(\theta|x)d\theta$ (as a function of d).

CASE 1: Bayesian approach to hypothesis testing: $L(\theta, a_i)$: $\theta \in \Theta_0$, $\theta \in \Theta_1$, type II error, type I error. The expected posterior loss is: $\int_\Theta L(\theta, d(x))\pi(\theta|x)d\theta = \begin{cases} \int_{\Theta_1} \pi(\theta|x)d\theta \to \text{posterior prob. of } \Theta_1 & \text{if } d(x) = a_0 \\ \int_{\Theta_0} \pi(\theta|x)d\theta \to \text{posterior prob of } \Theta_0 & \text{if } d(x) = a_1 \end{cases}$ Bayes rule: choose Θ_i with the larger posterior prob.

CASE 2: Point estimation (a) For squared error loss $L(\theta, a) = (\theta - a)^2$, the expected posterior loss is minimized [Bayes rule] by choosing $d(x) = \int_\Theta \theta\pi(\theta|x)d\theta$, the posterior mean [mean of $\pi(\theta|x)$]. (b) For absolute error loss $L(\theta, a) = |\theta - a|$, the Bayes rule is the posterior median of θ. $\int_{-\infty}^{d} \pi(\theta|x)d\theta = \int_{d}^{\infty} \pi(\theta|x)d\theta = \frac{1}{2}$.

HYPOTHESIS TESTING: Partition Θ into Θ_0, Θ_1. Test hypothesis: $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$. If Θ_i contains a single element then H_i is simple, o/w H_i is composite.

Neyman-Pearson Formulation of Hypothesis testing: Fix $\alpha \in (0,1)$, the significance level. Require that $P_\theta(\text{reject } H_0) \le \alpha$, $\forall \theta \in \Theta_0$ (α is upper bound on prob. of type I error). Test that satisfy this are Level-α Tests.

A Test is a function $\phi: \mathfrak{X} \to [0,1]$ interpreted as: if x is observed, then H_0 is rejected with prob $\phi(x)$. Now, $P_\theta(\text{reject } H_0) = E_\theta \phi(X)$. the size of a test ϕ is: $\sup_{\theta \in \Theta_0} E_\theta \phi(X)$. A Test is level $\alpha \iff$ size $\le \alpha$.

The power function of a test ϕ is the function $W: \Theta \to [0,1]$, defined by $W(\theta) = P_\theta(\text{reject } H_0) = E_\theta \phi(X)$. The power of a test is $W(\theta)$, where $\theta \in \Theta_1$. [A good test is one wich makes $W(\theta)$ as large as possible on Θ_1, while satisfying $W(\theta) \le \alpha$, $\forall \theta \in \Theta_0$].

Simple Hypotheses: $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$. Assume densities $f_i(x) = f(x; \theta_i)$ density when θ_i is the true state of nature. Define the likelihood ratio $\Lambda(x) = \frac{f_1(x)}{f_0(x)}$, the larger $\Lambda(x)$, the stronger the evidence against H_0.

A likelihood ratio test (LRT) of H_0 vs H_1 is a test $\phi(x)$ of the form:

$$\phi_0(x) = \begin{cases} 1 & f_1(x) > k f_0(x) \iff \Lambda(x) > k \ge 0 \\ \gamma(x) & f_1(x) = k f_0(x), \text{ where } \gamma(x): \mathfrak{X} \to [0,1] \\ 0 & f_1(x) < k f_0(x) \end{cases}$$

The NPL says the test ϕ_0 is the best test of size α. Also, $\gamma(x) \equiv \gamma_0$. NPL: (a) Optimality condition (b) Existence Condition (c) Uniqueness condition.

the critical region of a test is the set of $x \in \mathcal{X}$ for which $\phi(x) = 1$. [Reject H_0] $\Leftrightarrow \phi'(1)$.

Composite Hypotheses: Def: A test ϕ_0 is Uniformly most Powerful (UMP) iff (1) ϕ_0 is level-α test: $E_\theta \phi_0(X) \le \alpha \ \forall \theta \in \Theta_0$ AND (2) if ϕ is another level-α test and $\theta \in \Theta_1$: $E_\theta \phi_0(x) \ge E_\theta \phi(x)$

Def: the family of densities $\mathcal{F} = \{f(x;\theta) : \theta \in \Theta \subseteq \mathbb{R}\}$ is of monotone likelihood ratio (MLR) iff $\exists$ a function $t : \mathcal{X} \to \mathbb{R}$ s.t. $\Lambda(x) = f(x;\theta_2)/f(x;\theta_1)$ is a nondecreasing function of $t(x)$ for any $\theta_1 \le \theta_2$. THEOREM: Suppose that the density of x belongs to a family that has MLR with respect to $t(x)$. For testing $H_0: \theta \le \theta_0$ vs. $H_1: \theta > \theta_0$, the test: $\phi_0(x) = \begin{cases} 1 & t(x) > t_0 \\ 0 & t(x) \le t_0 \end{cases}$ [need to randomize in discrete case], is UMP among all test having same or smaller size.

BAYES FACTORS: Posterior Odds = Prior odds $\times$ Bayes FACTOR $\Leftrightarrow \dfrac{P(\theta_0 \mid x)}{P(\theta_1 \mid x)} = \dfrac{\pi_0}{\pi_1} \dfrac{f_0(x)}{f_1(x)}$, where $f_0(x) = f(x \mid \theta_0)$; $f_1(x) = f(x;\theta_1)$. BAYES FACTOR $B = f_0(x)/f_1(x)$ is sufficiently smaller than 1.

USE B to test $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$, reject H_0 iff B is sufficiently smaller than 1. Tests are of the form: reject H_0 iff $B < K \Leftrightarrow$ reject H_0 iff $f_1(x)/f_0(x) > \frac{1}{K} = C$; form of MP test by NPL.

Examples: Test simple hypothesis: $X \sim Exp(\theta)$, so X has pdf $f(x) = f(x \mid \theta) = \begin{cases} 0 & \text{if } x < 0 \\ \theta e^{-\theta x} & \text{if } x \ge 0 \end{cases}$

Test $H_0: \theta = 1$ vs $H_1: \theta = 2 \Rightarrow \lambda(x) = f_1(x)/f_0(x) = 2e^{-2x}/1e^{-1x} = 2e^{-x}$. By NPL, the MP Test is

$\phi(x) = \begin{cases} 1 & 2e^{-x} > K \\ \gamma_0 & 2e^{-x} = K \\ 0 & 2e^{-x} < K \end{cases}$ ← don't worry about this b/c continuous. To determine K, first fix $\alpha \in (0,1)$: α = size of $\phi(x) = \sup_{\theta \in \Theta_0} E_\theta \phi(x) = \sup_{\theta \in \{1\}} E_\theta \phi(x) = E_{\theta=1} \phi(x) = \int_0 \phi(x) f_0(x)\,dx = \int 1 \cdot e^{-x}\,dx$, where $2e^{-x} > K \Rightarrow e^{-x} > K/2 \Rightarrow -x > \log(K/2) \Rightarrow x < -\log(\frac{K}{2})$, So $= \int_0^{-\log(K/2)} e^{-x}\,dx = \dots = 1 - K/2 \Rightarrow \frac{K}{2} = 1 - \alpha$ $\Rightarrow \boxed{K = 2(1-\alpha)}$.

Example 4a: $H_0: x \sim P_0$ vs $H_1: x \sim P_1$. $\mathcal{X} = \{1,2,3\}$. Distributions are given by:

	$x=1$	$x=2$	$x=3$
$P_0(x)$	.009	.001	0.99
$P_1(x)$	.001	.989	0.01
$\frac{P_1(x)}{P_0(x)}$	$\frac{1}{9}$	989	$\frac{1}{99}$

Construct a MP test with $\alpha = 0.01$

NPL: $x = 2$, if we reject H_0 when $x = 2$, Then $P(\text{Type I error}) = P_{P_0}(x=2) = 0.001$

$x = 1$, if we reject H_0 when $x = 1$, $+ 0.01$ Then $P(\text{Type I error}) = P_{P_0}(x=1) = 0.009$

So, The MP level-α Test is:

$\phi(x) = \begin{cases} 1 & \text{if } x = 1,2 \\ 0 & \text{if } x = 3 \end{cases}$

If we observe $x = 1, 2$, then we reject H_0 at $\alpha = 0.01$

Exercise 2.4. (1) $\Theta = \{0,1\}$, $\theta = 0$ component not functioning, $\theta = 1$ component functioning (2) $\mathcal{X} = \{0,1\}$ $x = 0$, warning light off, $x = 1$ warning light on (3) $\mathcal{F} = \{P_0, P_1\}$ $P_0 = \text{Bernoulli}(2/3)$; $P_1 = \text{Bernoulli}(1/4)$ [i.e. $P_0(0) = \frac{1}{3}$; $P_0(1) = \frac{2}{3}$]. (4) $\mathcal{A} = \{0,1\}$; $a = 0$ don't launch; $a = 1$ launch (5) Loss function $L : \Theta \times \mathcal{A} \to \mathbb{R}$. $L(0,0) = 0$, $L(0,1) = 10$, $L(1,0) = 5$; $L(1,1) = 0$ (6) Nonrandomized decision rules: $d : \mathcal{X} \to \mathcal{A}$ there are $2^2 = 4$ nonrandomized d. rules: $d_1(x) = 1$; $d_2(x) = 0$; $d_3(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x = 0 \end{cases}$; $d_4(x) = \begin{cases} 0 & \text{if } x = 1 \\ 1 & \text{if } x = 0 \end{cases}$

Risk functions d_1: $R(0, d_1) = \sum_{x \in \mathcal{X}} L(0, d_1(x)) P_0(x) = L(0, d_1(0)) P_0(0) + L(0, d_1(1)) P_0(1) = L(0,1) P_0(0) + L(0,1) P_0(1) = 10 \cdot \frac{1}{3} + 10 \cdot \frac{2}{3} = 10$

$R(1, d_1) = \sum_{x \in \mathcal{X}} L(1, d_1(x)) P_1(x) = L(1, d_1(0)) P_1(0) + L(1, d_1(1)) P_1(1) = L(1,1) P_1(0) + L(1,1) P_1(1) = 0 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = 0$.. And so on for d_i.

the risk set of randomized rules is the parallelogram (convex hull) whose vertices are the above risk points.

For Bayes Rule: Suppose $\pi(0) = \psi = 2/5$ and $\pi(1) = 1 - \psi = 3/5$. The Bayes risk of d is: $\frac{2}{5} R(0,d) + \frac{3}{5} R(1,d)$. So lines of the form $\frac{2}{5} r_0 + \frac{3}{5} r_1 = C$ have constant Bayes risk. The risk point of Bayes is the (r_1, r_2) for which c is smallest. Or, compute all $\frac{2}{5} R(0,d) + \frac{3}{5} R(1,d)$ and choose d that minimizes this.

HW5. Suppose that $x_1, \dots, x_5 \sim \text{Pois}(\lambda)$: $P_r\{X_i = r\} = (e^{-\lambda} \lambda^r)/r!$; for $r = 0, 1, 2, \dots$ Compare the two tests: $H_0: \lambda = \frac{1}{2}$ vs $H_1: \lambda = 1$

(a) $Y = X_1 + \dots + X_5$; Reject H_0 iff $Y \ge 4$. Note $Y \sim \text{Pois}(5\lambda)$. (b) Reject H_0 iff at least two X_i are non zero.

Size of (a): $P_{\lambda=\frac{1}{2}}(\text{reject } H_0) = P_{\lambda=\frac{1}{2}}(Y \ge 4) = 1 - P_{\lambda=\frac{1}{2}} P(Y \le 3) = 1 - \text{ppois}(3, 2.5)$.

Power of (a): $P_{\lambda=1}(\text{reject } H_0) = P_{\lambda=1}(Y \ge 4) = 1 - P_{\lambda=1} P(Y \le 3) = 1 - \text{ppois}(3, 5)$

Size of (b): $P_{\lambda=\frac{1}{2}}(X_i > 0) = 1 - P_{\lambda=\frac{1}{2}}(X_i = 0) = P_{\lambda=\frac{1}{2}}$ Let $Z \sim Bin(5; P_{\lambda=\frac{1}{2}})$; Size $= P(Z > 3) = 1 - P(Z \le 2) = 1 - \text{pbinom}(2, 5, P_{\lambda=\frac{1}{2}})$

Power of (b): $P_{\lambda=1}(X_i > 0) = 1 - P_{\lambda=1}(X_i = 0) = P_{\lambda=1}$ Then $P(Z > 3) = 1 - P(Z \le 2) = 1 - \text{pbinom}(2, 5, P_{\lambda=1})$.

If $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ with μ unknown and σ^2 known. Then: $\bar{X} \sim N(\mu, \sigma^2/n)$ & $\sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0,1)$

$X \sim N(\mu, \sigma^2)$. p.d.f $f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$; $T \sim Exp(\lambda)$ p.d.f. $f(x;\lambda) = \lambda e^{-\lambda x}$ mean $= \frac{1}{\lambda}$

Example 4.1: $X \sim \text{Binomial}(10, \theta)$. $H_0: \theta \le \frac{1}{2}$ vs $H_1: \theta > \frac{1}{2}$. Reject H_0 iff $x \ge K$. Given α choose K: $P_\theta(\text{reject } H_0) = P_\theta(X \ge K)$ is an increasing function of θ, hence, the size of the test is $P_{\theta=\frac{1}{2}}(X \ge K) = 1 - P_{\theta=\frac{1}{2}}(X \le K-1) = 1 - \text{pbinom}(K-1, 10, .5)$. Define: $\phi(x) = \begin{cases} 1 & x \ge 9 \\ \lambda & x = 8 \\ 0 & x \le 7 \end{cases}$ To find γ_0: we want size $= 0.05$. $\Rightarrow$ size $= \max\{E_\theta \phi(x) : \theta \le 0.05\} = 0.05$

$E_\theta \phi(x) = P(X \ge 9) + \lambda P(X = 8) = 0.05$, solve for $\lambda = 89/45$. Test has size 0.05.