STATS 300A Notes

Kenneth Tay

Set-up (Lec 1)

- **Risk function** $R(g(\theta), \delta) = \mathbb{E}_\theta[L(g(\theta), \delta(X))]$. Expectation is taken over all X, NOT all θ.

Exponential families (Lec 1-2)

- Family of probability distributions $\{P_\theta\}$ is an s-**parameter exponential family** if all the P_θ's have densities of the form
$$p_\theta(x) = \exp\left[\sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta)\right] h(x)$$
w.r.t. some dominating measure μ.
- Canonical form uses $\eta_1, \dots, \eta_s$ as parameters:
$$p(x, \eta) = \exp\left[\sum_{i=1}^s \eta_i T_i(x) - A(\eta)\right] h(x).$$
- **Natural parameter space** is the set of values η for which the density makes sense, i.e.
$$\left\{\eta = (\eta_1, \dots, \eta_s) : \int \exp\left[\eta_i T_i(x)\right] h(x) \mu(dx) < \infty\right\}.$$
- Natural parameter space is convex.
- s-dimensional exponential family has **full rank** if the natural parameter space contains an s-dimensional rectangle.
- (300B Lec 5) $A(\eta)$ is known as the **cumulant function**. $A(\eta)$ is convex and infinitely differentiable.
- For any integrable function f and any η in the natural parameter space,
$$\int f(x) \exp\left[\sum \eta_i T_i(x) - A(\eta)\right] h(x) \mu(dx)$$
is infinitely differentiable w.r.t. η_i's, and the derivatives can be obtained by differentiating inside the integral. As a consequence of this,
$$\mathbb{E}_\eta T_j(x) = \frac{\partial}{\partial \eta_j} A(\eta),$$
$$\text{Cov}(T_i, T_j) = \frac{\partial}{\partial \eta_i} \frac{\partial}{\partial \eta_j} A(\eta).$$
In 300B notation (Lec 5), we write the above as $\mathbb{E}[T(X)] = \dfrac{\partial A}{\partial \theta^T}$, Var $[T(X)] = \dfrac{\partial^2 A}{\partial \theta \partial \theta^T}$.

- (Lec 9) Exponential families have continuous risk functions.
- (Lec 14) Exponential families have continuous power functions for any test.
- (Lec 13-14) 1-parameter exponential families have monotone likelihood ratio.

Sufficient, ancillary and complete statistics (Lec 2-4)

- Examples of sufficiency:
  - (Lec 2) $X_1, \ldots, X_n$ iid, $X_i \sim \text{Unif}[0, \theta]$. $\max(X_1, \ldots, X_n)$ is sufficient.
  - (Lec 2) $X_1, \ldots, X_n$ i.i.d., then the order statistics are sufficient.
- **Fisher-Neyman Factorization Theorem**: Main tool to determine sufficiency. $T = T(X)$ sufficient iff there exist non-negative functions g_θ and h such that $p_\theta(x) = g_\theta(T(x)) \cdot h(x)$ with probability 1. (Intuitively, the part which depends on θ only depends on X through $T(X)$.)
- **Minimal Sufficiency**: T minimal sufficient for X if for any other sufficient statistic T', T is a function of T'. Lecture 3 has a theorem to determine whether a statistic is minimal sufficient.
- A statistic V is **ancillary** if its distribution does not depend on θ. It is **first-order ancillary** if $\mathbb{E}_\theta V(X)$ does not depend on θ.
- Examples of ancillary statistics:
  - (Lec 4) When in a location family model, the statistics $X_i - \bar{X}$ are ancillary, so are the statistics $X_i - X_j$.
- A statistic T is **complete** if $\mathbb{E}_\theta f(T) = 0$ for all θ implies that $f = 0$ with probability 1.
- Let $\mathcal{P}_0 \subseteq \mathcal{P}_1$ such that every null set of $\mathcal{P}_0$ is a null set of $\mathcal{P}_1$. Then if T is complete sufficient for $\mathcal{P}_0$, it is also complete sufficient for $\mathcal{P}_1$.
- **Basu's Theorem**: If T is complete sufficient and V is ancillary, then T and V are independent.
- **Rao-Blackwell Theorem**: A theorem that tells us how we can, from an existing estimator δ, construct another estimator which has better risk than δ. (From existing estimator $\delta(X)$ and sufficient statistic T, use $\mathbb{E}[\delta(X) \mid T(X) = t]$).

UMVU estimation (Lec 4-6)

- $g(\theta)$ is **U-estimable** if it has an unbiased estimator.
- **Lehmann-Scheffé Theorem**: Suppose that there exists only one unbiased estimator of $g(\theta)$ based on sufficient statistic T. Then it must be UMVU.

  If we have an unbiased estimator of $g(\theta)$ based on a complete sufficient statistic, it must be UMVU.
- To show no UMVU for a family of distributions $\mathcal{F}$, try to find 2 different subfamilies of $\mathcal{F}$ with different UMVUs.
- UMVU can be inadmissible. E.g. $X \sim \text{Pois}(\lambda)$, estimating $e^{-a\lambda}$.
- Let $\mathcal{U}$ be the set of unbiased estimators for zero, δ_0 an unbiased estimator for $g(\theta)$. Then δ_0 is UMVU iff $\mathbb{E}[\delta_0 U] = \text{Cov}(\delta_0, U) = 0$ for all $U \in \mathcal{U}$.
- IF a UMVU exists, it is unique.

Estimation for location families (Lec 6-7)

- The bias, variance and risk functions of any location equivariant estimator are constant (do not depend on θ).
- Any location equivariant estimator is the sum of some (fixed) location equivariant estimator and a location invariant estimator.
- For $n > 1$, u is location invariant iff it is a function of the differences $Y_i = X_i - X_n$, $i = 1, \dots, n-1$. For $n = 1$, u is location invariant iff it is a constant.
- In location families, the differences $X_i - X_j$ $(i \neq j)$ are ancillary.
- (TPE Thm 3.1.10, Cor 3.1.11 p151) Let $Y = (X_1 - X_n, \dots, X_{n-1} - X_n)$. Suppose there exists an equivariant δ_0 with finite risk. Assume that for each $y = (y_1, \dots, y_{n-1})$, there exists a number $v(y) = v^*(y)$ which minimizes $\mathbb{E}_0 \{\rho[\delta_0(X) - v(y)] \mid Y = y\}$. Then the estimator $\delta_0(X) - v^*(Y)$ is minimium risk equivariant (MRE).
  - If ρ convex and not monotone, then the MRE exists.
  - If ρ strictly convex, the MRE is unique.
  - Under squared error loss, $v^*(Y) = \mathbb{E}_0[\delta_0(X) \mid Y]$.
  - Under absolute error loss, $v^*(Y) =$ any median of the conditional distribution $\delta_0(X) \mid Y$.

  Often, we will have X being complete sufficient, so by Basu's Theorem the conditional expectation is the same as the unconditional expectation.
- **Pitman estimator**: Under squared error loss, MRE estimator is given by

$$\delta^* = \frac{\int_{-\infty}^{\infty} u f(x_1 - u, \dots, x_n - u) du}{\int_{-\infty}^{\infty} f(x_1 - u, \dots, x_n - u) du}.$$

- For squared error loss:
  - If δ is location equivariant with constant bias $b = \mathbb{E}_\theta \delta - \theta$, then $\delta - b$ is unbiased, location equivariant and has smaller risk than δ (unless $b = 0$).
  - The unique MRE is unbiased.
  - If UMVU exists and is location equivariant, it must be MRE as well.
  - If UMVU exists, then it is the Pitman estimator (except in exceptional settings).

Bayes estimators (Lec 8-9)

- In this setting, the quantity to be minimized is

$$\int_\theta \int_x L(\theta, \delta(x)) dP_\theta(x) d\Lambda(\theta) = \mathbb{E}[L(\Theta, \delta(X))],$$

  where the expectation is taken over the joint distribution of (x, θ), i.e. $\Theta \sim \Lambda$, and given $\Theta = \theta$, $X \sim P_\theta$.
- Under **squared error loss**, $\delta_\Lambda(X) = \mathbb{E}[\Theta \mid X]$, i.e. the mean of the posterior distribution. Under **absolute error loss**, the Bayes estimator would be the median of the posterior distribution.

- (TPE Cor 4.1.2 p228) Under **0-1 loss**

$$L(\theta, d) = \begin{cases} 0 \text{if } |d-\theta| \leq c, \\ 1 \text{if } |d-\theta| > c, \end{cases}$$

  then $\delta_\Lambda(X)$ is the midpoint of the interval I of length $2c$ which maximizes $\mathbb{P}(\Theta \in I \mid x)$.

- **Weighted squared loss:** $L(\theta, d) = w(\theta)(d-\theta)^2$ for some weight function w. Under weighted squared loss, we have Bayes estimator $d^* = \dfrac{\mathbb{E}[\Theta w(\Theta)]}{\mathbb{E}[w(\Theta)]}$, where $\Theta \sim \Lambda$. (If data is observed, use the posterior distribution of Θ instead.)

- **Uniqueness of Bayes estimators**: Let Q be the marginal distribution of X. δ_Λ is unique if (i) its average risk w.r.t. Λ is finite, and (ii) $Q(N) = 0 \implies P_\theta(N) = 0$ for all θ. (Condition (ii) is satisfied if the parameter space Ω is an open set and equal to the support of Λ, and $P_\theta(E)$ is continuous in θ for every E.)

- If $\delta_\Lambda(X)$ is a unique Bayes estimator, then it is admissible.

- **Bayes estimators are biased**: Under squared error loss, no unbiased estimator $\delta(X)$ is Bayes unless its average risk is 0, i.e. $\mathbb{E}[(\delta(X) - g(\Theta))^2] = 0$.

Minimax estimation (Lec 9-11)

- If a Bayes estimator has constant risk, it is minimax.

- For a prior distribution Λ, let r_Λ be the Bayes risk of the Bayes estimator δ_Λ. If $r_\Lambda = \sup_\theta R(\theta, \delta_\Lambda)$, then

  - δ_Λ is minimax.
  - If δ_Λ is uniquely Bayes, it is uniquely minimax as well.
  - r_Λ is least favorable, i.e. $r_\Lambda \geq r_{\Lambda'}$ for any Λ'.

  (The assumption in the theorem holds if the risk of the estimator is constant. It can also hold if Λ puts all its mass on θ values where the risk function is worst.)

- Suppose we have an estimator δ and a sequence of priors $\{\Lambda_m\}$ such that $\sup_\theta R(\theta, \delta) = r$ and $r_{\Lambda_m} \to r$. Then δ is minimax and $\{\Lambda_m\}$ is least favorable.

- Suppose X has unknown distribution F from family $\mathcal{F}$. Suppose δ is minimax when estimating $g(F)$ for $F \in \mathcal{F}_0$, where $\mathcal{F}_0 \subseteq \mathcal{F}$. If $\sup_{F \in \mathcal{F}_0} R(\delta, F) = \sup_{F \in \mathcal{F}} R(\delta, F)$, then δ is minimax for $\mathcal{F}$ as well.

- If an estimator has constant risk and is admissible, then it is minimax. (A minimax estimator with constant risk need not be admissible.)

- (TPE Prob 5.1.17 p392, HW5 Qn4) Let r_Λ denote the Bayes risk w.r.t. prior Λ. If $r_\Lambda = \infty$ for some Λ, then any estimator δ has unbounded risk.

UMP tests (Lec 12-14)

- The **size** of a test φ is $\sup_{\theta \in \Omega_0} \mathbb{E}_\theta \varphi(X)$.

- We can always restrict attention to tests based on sufficient statistics.

- **Monotone Likelihood Ratio (MLR)**: A family of densities p_θ with $\theta \in \mathbb{R}$ has MLR in $T(x)$ if for all $\theta < \theta'$, $\dfrac{p_{\theta'}(x)}{p_\theta(x)}$ is a non-decreasing function of $T(x)$.

- (TSH Thm 3.4.1 p65) For families with MLR, rejecting for large values of likelihood ratio is equivalent to rejecting for large values of T. Hence, for testing $\theta = \theta_0$ vs. $\theta > \theta_0$, there exists a UMP level α test of the form

$$\varphi(X) = \begin{cases} 1 & \text{if } T(X) > c, \\ \gamma & \text{if } T(X) = c, \\ 0 & \text{if } T(X) < c, \end{cases}$$

  where c and γ are determined by $\mathbb{E}_{\theta_0}\varphi(X) = \alpha$.

- (TSH Thm 3.4.1 p65) Families with MLR have a power function $\beta(\theta) = \mathbb{E}_\theta \varphi(X)$ that is strictly increasing for all points θ such that $0 < \beta(\theta) < 1$. This allows us to extend UMP tests for $\theta = \theta_0$ vs. $\theta > \theta_0$ to $\theta \leq \theta_0$ vs. $\theta > \theta_0$.

  Also, for any $\theta < \theta_0$, this test minimizes $\beta(\theta)$ (the probability of Type 1 error) among all tests satisfying $\mathbb{E}_{\theta_0}\varphi(X) = \alpha$.

- Examples of MLR families:
  - (TSH Cor 3.4.1 p67) 1-parameter exponential families.
  - Double exponential distribution with $\theta \geq 0$.
  - (Lec 16) Non-central chi-squared $\chi^2_n(\theta)$ for $\theta \geq 0$.
  - (TSH Eg 3.4.1 p66) Hypergeometric distribution.

- (TSH Thm 3.2.1 p60) **Neyman-Pearson Lemma** for Simple vs. Simple:
  (i) For testing P_0 vs. P_1, there exists a (possibly randomized) test φ and a constant k such that:
    (a) $\mathbb{E}_0\varphi(X) = \alpha$, and
    (b) $\varphi(X) = 1$ if $p_1(X) > kp_0(X)$, and $\varphi(X) = 0$ if $p_1(X) < kp_0(X)$.
  (ii) (Sufficiency to get a MP level α test) A sufficient condition for a test φ to be MP level α is it satisfies (a) and (b) in (i).
  (iii) (Necessity) If φ is MP level α, then for some k, it satisfies (b), and it also satisfies (a) unless there exists a test φ' whose size is $< \alpha$ with power $= 1$.

- For Simple vs. Simple, the power of a MP level α test is $> \alpha$.

- **How to find a UMP test**:
  - **Simple H_0 vs. simple H_1.**
    This case is completely solved using the Neyman-Pearson Lemma.

- **Simple H_0 vs. composite H_1.**

  Say we have $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \in \omega$, where ω is some subset of the parameter space Ω not containing θ_0.

  Fix $\theta' \in \omega$ and use the Neyman-Pearson Lemma to determine what a MP level α test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta'$ looks like. If there exists such a test that does not depend on θ', then it is UMP.

- **Composite H_0 vs. simple H_1.** Say $H_0 : X \sim f_\theta, \theta \in \omega$, $H_1 : X \sim g$.

  Reduce the problem to Simple vs. Simple by introducing a mixture density: $h_\Lambda(x) = \displaystyle\int_{\theta\in\omega} f_\theta(x) d\Lambda(\theta)$ for some prior Λ of θ.

  Let φ_Λ be the MP level α test for H_Λ vs. H_1. Suppose φ_Λ is level α for the composite H_0, i.e. $\sup_{\theta\in\omega} \mathbb{E}_\theta \varphi_\Lambda(X) \leq \alpha$. Then ϕ_Λ is MP for testing H_0 vs. H_1, and Λ is least favorable.

- **Composite H_0 vs. composite H_1.**

  Say we have $H_0 : \theta \in \omega_0$ vs. $H_1 : \theta \in \omega_1$.

  Fix $\theta' \in \omega_1$ and determine the UMP test for $H_0 : \theta \in \omega_0$ vs. $H_1 : \theta = \theta'$. If this test does not depend on θ', then it is UMP for the original setting.

- (TSH Thm 3.7.1 p81) **UMP 2-sided test for 1-parameter exponential family:** For testing $H : \theta \leq \theta_1$ or $\theta \geq \theta_2$ vs. $K : \theta_1 < \theta < \theta_2$ in the one-parameter exponential family, there exists a UMP test given by

$$\varphi(x) = \begin{cases} 1 & \text{if } C_1 < T(x) < C_2, \\ \gamma_i & \text{if } T(x) = C_i, i = 1, 2 \\ 0 & \text{if } T(x) < C_1 \text{ or } T(x) > C_2, \end{cases}$$

  where the C's and γ's are determined by $\mathbb{E}_{\theta_1}\phi(X) = \mathbb{E}_{\theta_2}\phi(X) = \alpha$.

  For $0 < \alpha < 1$, the power function of this test has a maximum at a point $\theta_0 \in (\theta_1, \theta_2)$, and decreases strictly as θ tends away from θ in either direction, unless there exist t_1 and t_2 such that $P_\theta\{T(X) = t_1\} + P_\theta\{T(X) = t_2\} = 1$ for all θ.

Uniformly most accurate (UMA) bounds (TSH Sec 3.5.1)

- The theory of UMP 1-sided tests can be applied to the problem of obtaining a lower or upper bound for a real-valued parameter θ.
- $\underline{\theta} = \underline{\theta}(X)$ is a **lower confidence bound** for θ at confidence level $1-\alpha$ if $\mathbb{P}_\theta\{\underline{\theta}(X) \leq \theta\} \geq 1 - \alpha$ for all θ.
- We want $\underline{\theta}$ to underestimate θ by as little as possible. A lower confidence bound $\underline{\theta}$ for which $\mathbb{P}_\theta\{\underline{\theta}(X) \leq \theta'\} =$ minimum for all $\theta' < \theta$ is called a **uniformly most accurate lower confidence bound** for θ and confidence level $1 - \alpha$.
- (TSH Cor 3.5.1 p73) Let family of densities $\{p_\theta(x)\}$ have MLR in $T(x)$, and suppose that the CDF $F_\theta(t)$ of $T = T(X)$ is a continuous function in both t and θ when the other is fixed.
  1. There exists a UMA confidence bound $\underline{\theta}$ for θ at each confidence level $1 - \alpha$.
  2. If x denotes the observed values of X and $t = T(x)$, and if the equation $F_\theta(t) = 1 - \alpha$ has a solution in $\theta = \hat{\theta}$, then this solution is unique and $\underline{\theta}(x) = \hat{\theta}$.
- Finding a UMA upper bound for θ corresponds to inverting the UMP test for $\theta = \theta_0$ vs. $\theta < \theta_0$.

  Finding a UMA lower bound for θ corresponds to inverting the UMP test for $\theta = \theta_0$ vs. $\theta > \theta_0$.

UMPU tests (Lec 14-16)

- **Definition of unbiased test**: Say we are testing $\theta \in \Omega_0$ vs. $\theta \in \Omega_1$. A test φ is unbiased at level α if (i) $\sup_{\theta \in \Omega_0} \mathbb{E}_\theta \varphi \leq \alpha$, and (ii) $\inf_{\theta \in \Omega_1} \mathbb{E}_\theta \varphi \geq \alpha$.
- **Definition of similar test**: Say we are testing $\theta \in \Omega_0$ vs. $\theta \in \Omega_1$. Let $\omega := \bar{\Omega}_0 \cap \bar{\Omega}_1$. A test φ is similar if it satisfies $\mathbb{E}_\theta \varphi = \alpha$ for all $\theta \in \omega$.
- **Neyman structure**: A test has Neyman structure w.r.t. T if $\mathbb{E}_{\theta_0}[\varphi(X) \mid T(X) = t] = \alpha$ for all t.
- If T is complete sufficient for ω, then every similar test has Neyman structure.
- To find UMPU test, one approach is to find UMP tests among all similar tests. If the UMP test we found is unbiased as well, then it will be UMPU.
- Assume that we are in an exponential family model, i.e.

$$p_{\theta,\vartheta}(x) \propto \exp\left[\theta U(x) + \sum_{i=1}^d \vartheta_i T_i(x)\right] h(x).$$

  Then:

  1. With $\theta = \theta_0$ fixed, $(T_1, \ldots, T_d)$ is sufficient, and also complete if $\{\vartheta_1, \ldots, \vartheta_d\}$ contains a d-dimensional rectangle.
  2. $T = (T_1, \ldots, T_d)$ has an exponential family of distributions.
  3. The conditional distribution of $U \mid T = t$ is a 1-parameter exponential family. (By sufficiency, this exponential family does not depend on ϑ.)

- In the above $(d+1)$-parameter exponential family, assume that for fixed $\theta = \theta_0$, the family is of full-rank (i.e. $(T_1, \ldots, T_d)$ complete for ω).

  For testing $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$, there exists UMPU level α test of the form

$$\varphi(u,t) = \begin{cases} 1 & \text{if } u > c(t), \\ \gamma(t) & \text{if } u = c(t), \\ 0 & \text{if } u < c(t), \end{cases}$$

  where $u = U(x)$, $t = T(x)$, $c(t)$ and $\gamma(t)$ are determined such that $\mathbb{E}_{\theta_0}[\varphi(U,T) \mid T] = \alpha$.
- To find UMPU test in multi-parameter exponential family, condition on the nuisance parameters.
- (TSH Prob 4.1 p139, HW7 Qn1) Any UMPU test is admissible (i.e. no other test is at least as powerful against all alternatives and more powerful against some).
- (TSH Thm 4.4.1 p121) Let X be distributed according to $dP^X_{\theta,\vartheta}(x) = C(\theta,\vartheta) \exp\left[\theta U(X) + \sum_{i=1}^k \vartheta_i T_i(x)\right] d\mu(x)$.

  Let $\vartheta = (\vartheta_1, \ldots, \vartheta_k)$ and $T = (T_1, \ldots, T_k)$. Write the above as a function of the sufficient statistics (U,T):

$$dP^{U,T}_{\theta,\vartheta}(u,t) = C(\theta,\vartheta) \exp\left[\theta u + \sum_{i=1}^k \vartheta_i t_i\right] d\nu(u,t).$$

  When $T = t$ is given, U is the only remaining variable, and the conditional distribution of U given t is an exponential family:

$$dP^{U|t}_\theta(u) = C_t(\theta) e^{\theta u} d\nu_t(u).$$

(i) For testing $\theta \le \theta_0$ vs. $\theta > \theta_0$, there is a UMPU level α test ϕ_1 defined by

$$\phi_1(u,t) = \begin{cases} 1 & \text{if } u > C_0(t), \\ \gamma_0(t) & \text{if } u = C_0(t), \\ 0 & \text{if } u < C_0(t), \end{cases}$$

where functions C_0 and γ_0 are determined by $\mathbb{E}_{\theta_0}[\phi_1(U,T) \mid t] = \alpha$ for all t.

(ii) For testing $\theta \le \theta_1$ or $\theta \ge \theta_2$ vs. $\theta_1 < \theta < \theta_2$, there is a UMPU level α test ϕ_2 defined by

$$\phi_2(u,t) = \begin{cases} 1 & \text{if } C_1(t) < u < C_2(t), \\ \gamma_i(t) & \text{if } u = C_i(t), \quad i = 1,2, \\ 0 & \text{if } u < C_1(t) \text{ or } u > C_2(t), \end{cases}$$

where functions C's and γ's are determined by $\mathbb{E}_{\theta_1}[\phi_2(U,T) \mid t] = \mathbb{E}_{\theta_2}[\phi_2(U,T) \mid t] = \alpha$ for all t.

(iii) For testing $\theta_1 \le \theta \le \theta_2$ vs. $\theta < \theta_1$ or $\theta > \theta_2$, there is a UMPU level α test ϕ_3 defined by

$$\phi_3(u,t) = \begin{cases} 1 & \text{if } u < C_1(t) \text{ or } u > C_2(t), \\ \gamma_i(t) & \text{if } u = C_i(t), \quad i = 1,2, \\ 0 & \text{if } C_1(t) < u < C_2(t), \end{cases}$$

where functions C's and γ's are determined by $\mathbb{E}_{\theta_1}[\phi_3(U,T) \mid t] = \mathbb{E}_{\theta_2}[\phi_3(U,T) \mid t] = \alpha$ for all t.

(iv) For testing $\theta = \theta_0$ vs. $\theta \ne \theta_0$, there is a UMPU level α test ϕ_4 defined by $\mathbb{E}_{\theta_0}[\phi_4(U,T) \mid t] = \alpha$ and $\mathbb{E}_{\theta_0}[U\phi_4(U,T) \mid t] = \alpha\mathbb{E}_{\theta_0}[U \mid t]$.

Invariant and UMPI tests (Lec 16-18)

Set-up: We have data $X \sim P_\theta$, $\theta \in \Omega$. Let $\mathcal{S}$ be the sample space for X. Testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$.

Assume that there is a group of 1-to-1 transformations G which act on the data such that for any $g \in G$,

$$X \sim P_\theta \quad \Rightarrow \quad gX \sim P_{\theta'}$$

for some $\theta' \in \Omega$. If this is the case, we write $\theta' = \bar{g}\theta$, and we say that g induces a transformation on the parameter space.

We also assume that Ω_0 and Ω_1 are *preserved* in the sense that $\bar{g}\theta \in \Omega_i$ iff $\theta \in \Omega_i$ for $i = 0, 1$.

- A test φ is **invariant** if $\varphi(x) = \varphi(gx)$ for all $g \in G$.
- A transformation group G is **transitive** if for any $x_1, x_2 \in X$, there is a $g \in G$ such that $g(x_1) = x_2$.
- (TSH Thm 6.3.1 p219) **UMPI test for finite G:** Suppose testing Ω_0 vs. Ω_1 remains invariant under a finite group $G = \{g_1, \ldots, g_N\}$ and that $\bar{G}$ is transitive over Ω_0 and over Ω_1. Then there exists a UMPI test, and it rejects Ω_0 when
$$\frac{\sum_{i=1}^N p_{\bar{g}_i\theta_1}(x)/N}{\sum_{i=1}^N p_{\bar{g}_i\theta_0}(x)/N}$$
is large, where θ_0 and θ_1 are any elements of Ω_0 and Ω_1 respectively.

- A statistic $T = T(X)$ is **maximal invariant** if (i) T is invariant (i.e. $T(x) = T(gx)$ for all x, g), and (ii) if $T(x_1) = T(x_2)$, then there exists some $g \in G$ such that $x_2 = gx_1$.
- A test is invariant iff it is a function of a maximal invariant statistic.
- Examples of maximal invariant statistics:
  - Family of transformations $X_i' = X_i + a$, where a is a real constant: $(X_1 - X_n, \dots, X_{n-1} - X_n)$ is maximal invariant.
  - Family of transformations $X_i' = cX_i$, where c is a non-zero real constant, and X_i's assumed to be non-zero: $\left(\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}\right)$ is maximal invariant.
  - $G = n!$ permutations of $X_1, \dots, X_n$: $(X_{(1)}, \dots, X_{(n)})$ order statistics is maximal invariant.
  - Family of transformations $X_i' = f(X_i)$, where f is a continuous, strictly increasing function: The ranks of the data are maximal invariant.
- A test φ is **almost invariant** w.r.t. group G if for all $g \in G$, $\varphi(x) = \varphi(g(x))$ a.e. Here, the null set N_g can depend on g.
- Assume that there exists a unique (a.e.) UMPU test φ^* and also a UMPaI test w.r.t. group G. Then the latter test is unique (a.e.) and equal to φ^* (a.e.).

Examples of complete sufficient statistics

- (Lec 4) $X_1, \dots, X_n$ iid, $X_i \sim \text{Bernoulli}(p)$. $\sum X_i$ is complete sufficient.
- (Lec 7) $X_1, \dots, X_n \overset{iid}{\sim} E(\theta, b)$, with θ unknown b known. $X_{(1)}$ is complete sufficient.
- (TPE Eg 1.6.24 p43) $X_1, \dots, X_n \overset{iid}{\sim} E(\theta, b)$, with both θ and b unknown. $\left(X_{(1)}, \sum[X_i - X_{(1)}]\right)$ are complete sufficient. See Eg 2.2.5 p98 for more details on distribution.
- (Lec 3) $X_1, \dots, X_n$ iid, $X_i \sim \text{Unif}(0, \theta)$. $T = \max(X_1, \dots, X_n)$ is complete sufficient.
- (Lec 5) $X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}(\theta - \sigma, \theta + \sigma)$, both θ and σ unknown. $(X_{(1)}, X_{(n)})$ is complete sufficient.
- (Lec 3) For exponential family model of full rank, the T_i's (all together as a vector) are complete sufficient.
- (TPE Prob 1.6.33 p72, HW2 Qn 2) $X_1, \dots, X_n$ i.i.d. with unknown density f (w.r.t. Lebesgue measure). The order statistics are complete.

Examples of UMVU estimators

- (Lec 4) $X_1, \dots X_n$ iid, $X_i \sim \text{Bernoulli}(p)$. Let $T = \sum X_i$. $\frac{T}{n}$ is UMVU for p. Can also get UMVU for $p(1-p)$.
- (TPE Eg 2.1.13 p88) $T \sim \text{Binom}(n, p)$. UMVU estimator for $p(1-p)$ is $\frac{T(n-T)}{n(n-1)}$.
- (TPE Eg 2.2.5 p98) $X_1, \dots, X_n \sim E(\xi, b)$, b known. UMVU estimator for ξ is $X_{(1)} - \frac{b}{n}$.

- (Lec 4) $X_1, \dots X_n$ iid, $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Sample mean is UMVU for μ, whether σ^2 is known or unknown. Sample variance is UMVU for σ^2.
- (TPE Prob 2.2.1 p132, HW2 Qn6) $X_1, \dots X_n$ iid, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, with σ^2 known. Can find UMVU estimators for μ^2, μ^3 and μ^4.
- (Lec 4) $X_1, \dots X_n$ iid, $X_i \sim \text{Pois}(\lambda)$. Can get UMVU estimator for $e^{-\lambda}$.
- (TPE Prob 2.1.15 p131, HW2 Qn5) $X_1, \dots X_n$ iid, $X_i \sim \text{Pois}(\lambda)$. $\bar{X}$ is UMVU for λ.
- (Lec 4) $X_1, \dots X_n$ iid, $X_i \sim \text{Unif}(0, \theta)$. $\dfrac{n+1}{n}\max(X_1, \dots, X_n)$ is UMVU for θ.
- (TPE Prob 2.2.24 p134, HW3 Qn4) $X_1, \dots X_n$ iid, $X_i \sim \text{Unif}(\xi - b, \xi + b)$, both unknown. Can find UMVU estimators for ξ, b and ξ/b.
- (TPE Prob 2.4.6 p137, HW3 Qn7) $X_1, \dots X_m$ i.i.d. according to F, $Y_1, \dots, Y_n$ i.i.d. according to G. Can find UMVU estimator for $\mathbb{P}(X_i < Y_j)$.

Examples of MRE estimators

- (Lec 7, TPE Prob 3.1.5 p207, HW4 Qn3) $X_1, \dots, X_n \overset{iid}{\sim} E(\theta, b)$, with θ unknown b known. Under squared error loss, MRE is $X_{(1)} - \dfrac{b}{n}$. Under absolute error loss, it is $X_{(1)} - \dfrac{b \log 2}{n}$. Under 0-1 loss (1 if our guess is distance k away from the actual), MRE is $X_{(1)} - k$.

Examples of Bayes estimators

- (Lec 8) $X \sim \text{Binom}(n, \theta)$, $\Lambda = \text{Beta}(a, b)$, squared error loss. Posterior distribution of Θ is $\text{Beta}(x + a, n - x + b)$, so Bayes estimator of θ is $\dfrac{x+a}{n+a+b}$.
- (Lec 8) $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$, σ^2 fixed and known. $\Lambda = \mathcal{N}(\mu, b^2)$. Posterior distribution of θ is

$$\mathcal{N}\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{b^2}}{\frac{n}{\sigma^2} + \frac{1}{b^2}}, \left(\frac{n}{\sigma^2} + \frac{1}{b^2} \right)^{-1} \right).$$

Examples of minimax estimators

- (Lec 10) $X \sim \text{Binom}(n, p)$, squared error loss. $\dfrac{X + \sqrt{n}/2}{n + \sqrt{n}}$ is minimax.
- (Lec 10) $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, 1)$. Under squared error loss, $\bar{X}$ is minimax.
- (Lec 11) $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$, squared error loss. If σ^2 known, $\bar{X}$ is minimax. If σ^2 unknown, every estimator has sup risk $= \infty$. If σ^2 unknown but bounded by known M, then $\bar{X}$ is minimax.
- (TPE Prob 5.1.26 p392, HW5 Qn7) Let $X_1, \dots, X_m \overset{iid}{\sim} \mathcal{N}(\xi, \sigma^2)$, $Y_1, \dots, Y_n \overset{iid}{\sim} \mathcal{N}(\eta, \tau^2)$, with variances known. Under squared error loss, $\bar{Y} - \bar{X}$ is minimax for $\eta - \xi$.

- (Stated in Lec 11) If X_i's iid supported on $\{0,1\}$, then the minimax estimator for the mean under squared error loss is $\dfrac{\sqrt{n}}{1+\sqrt{n}}\bar{X} + \dfrac{1}{2(1+\sqrt{n})}$.
- (TPE Prob 5.1.25 p392, HW5 Qn6) If $X_1, \ldots, X_n \overset{iid}{\sim} F$, F unknown, then a minimax estimator for $F(0) = \mathbb{P}(X_i \le 0)$ under squared error loss is $\dfrac{\text{No. of } X_i \le 0}{\sqrt{n}} \dfrac{1}{1+\sqrt{n}}\bar{X} + \dfrac{1}{2(1+\sqrt{n})}$.

Examples of UMP tests

- (TSH Prob 3.32 p99, HW6 Qn4) Let $X \sim$ Cauchy with scale parameter 1 and location parameter θ. Then there is no UMP test for $\theta = 0$ vs. $\theta > 0$.
- (TSH Prob 3.34 p99, HW6 Qn5) Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Gamma}(g, b)$ (shape-scale).
  - When g is known, there is a UMP test for $b \le b_0$ vs. $b > b_0$.
  - When b is known, there is a UMP test for $g \le g_0$ vs. $g > g_0$.
- (Lec 12) Testing mean in normal model with σ^2 known. There is a UMP test for $\theta = 0$ vs. $\theta > 0$ or $\theta < 0$. Testing $\theta = 0$ vs. $\theta \ne 0$ has no UMP test.
- (Lec 13) 1-sided normal variance, where both mean and variance unknown.
- (Lec 14) $X \sim \mathcal{N}(\mu_x, \sigma^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma^2)$ with σ^2 known. Testing $\mu_y \le \mu_x$ vs. $\mu_y > \mu_x$.
- (TSH Eg 3.9.2 p90) Multivariate normal, testing $\sum_{i=1}^k a_i\mu_i \le \delta$ vs. $\sum_{i=1}^k a_i\mu_i > \delta$
- (TSH Prob 3.2 p92, HW6 Qn1) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$. There are UMP tests for testing $\theta \le \theta_0$ vs. $\theta > \theta_0$. There is a unique UMP test for testing $\theta = \theta_0$ vs. $\theta \ne \theta_0$.
- (Lec 14) Testing $\mathbb{P}(X_i \le u) \ge p_0$ vs. $\mathbb{P}(X_i \le u) < p_0$, where u and p_0 fixed. Results in sign test.

Examples of UMPU tests

- (TSH Prob 4.25 p144, HW7 Qn5) $X \sim \text{NegBin}(p_1, m)$ and $Y \sim \text{NegBin}(p_2, n)$ independent. Let $q_i = 1 - p_i$. There is a UMPU test for $q_2/q_1 \le \theta_0$, hence there is a UMPU test for $p_1 \le p_2$.
- (Lec 15) Normal setting, testing mean & variance.
- (Lec 16) Testing independence in bivariate normal family.
- (TSH Prob 5.21 p197, HW7 Qn6) $X \sim \mathcal{N}(\xi, 1)$ and $Y \sim \mathcal{N}(\eta, 1)$ independent. There is no UMPU test for $\xi = \eta = 0$ vs. $\xi > 0, \eta > 0$.
- (Lec 15) Testing 2 Poisson means, testing 2 Binomial probabilities.
- (Lec 14) 1-parameter exponential family, testing $\theta = \theta_0$ vs. $\theta \ne \theta_0$.
- (Lec 16) One-sample randomization tests.

Examples of UMPI tests

- (Lec 16) $X_i \sim \mathcal{N}(\theta_i, 1)$, X_i's mutually independent. Testing $H_0 : \theta_1 = \cdots = \theta_n = 0$ vs. H_1 : not all 0.
- (Lec 17) $X_1, \dots, X_n$ iid, $X_i \sim \mathcal{N}(\xi, \sigma^2)$ with both parameters unknown. Testing $H_0 : \sigma = \sigma_0$ vs. $H_1 : \sigma < \sigma_0$.
- (Lec 17) $X_1, \dots, X_n$ iid, $X_i \sim \mathcal{N}(\xi, \sigma^2)$ with both parameters unknown. Testing $H_0 : \xi = 0$ vs. $H_1 : \xi > 0$.
- (Lec 17): $X_1, \dots, X_m$ iid, $X_i \sim \mathcal{N}(\xi, \sigma^2)$, $Y_1, \dots, Y_n$ iid, $Y_j \sim \mathcal{N}(\eta, \tau^2)$, all 4 parameters unknown. Testing $H_0 : \sigma^2 = \tau^2$ vs. $H_1 : \sigma^2 < \tau^2$.
- (Lec 16) $X_1, \dots, X_n$ i.i.d. on $(0, 1)$. Testing $H_0 : X_i \sim \text{Unif}(0, 1)$, vs. $H_1 : X_i$'s have density $f(x)$ or $f(1-x)$, where f is fixed.
- (Lec 17) $X = (X_1, \dots, X_n)$ real-valued with joint density $f_i(x_1 - \theta, \dots, x_n - \theta)$, $i = 0$ or 1 with θ unknown. Testing $H_0 : f_0$ is true with unknown θ vs. $H_1 : f_1$ is true with unknown θ.

Other Facts

- **Generalized Likelihood Ratio (GLR) test:** Say we have $X \sim P_\theta$, and we are testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$. The GLR test statistic is
$$T' = \frac{\sup\limits_{\theta \in \Omega_0 \cup \Omega_1} p_\theta(x)}{\sup\limits_{\theta \in \Omega_0} p_\theta(x)}.$$
We reject for large values of T'.
- Order statistics $\left(X_{(1)}, \dots, X_{(n)}\right)$ are equivalent to $\left(\sum X_i, \sum X_i^2, \dots, \sum X_i^n\right)$.
- (TPE Eg 1.6.10 p36, HW1 Qn7) Let $U_1 = \sum X_i, U_2 = \sum X_i X_j \ (i \neq j), \dots, U_n = X_1 \dots X_n$, and let $V_k = \sum X_i^k$. Then the order statistics are equivalent to $(U_1, \dots, U_n)$, which in turn (by Newton's identities) is equivalent to $(V_1, \dots, V_n)$.
- (Lec 9) $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$, σ^2 known, squared error loss. Then risk of $a\bar{X} + b$ is $\dfrac{a^2\sigma^2}{n} + (a\theta + b - \theta)^2$.
- (Lec 13) $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, μ and σ unknown. Then $\bar{X}$ and $S^2 = \dfrac{1}{n-1} \sum (X_i - \bar{X})^2$ are independent, and $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

  (TPE p92) If μ is known, then $\dfrac{\sum (X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$.
- (TSH Thm 11.28) **Normality of sample median:** Suppose $X_1, \dots, X_n$ i.i.d. with cdf F. Assume $F(\theta) = 1/2$, and that F is differentiable at θ with $F' = f$ and $f(\theta) > 0$. Let $\tilde{X}_n$ be the sample median. Then
$$\sqrt{n}(\tilde{X}_n - \theta) \to \mathcal{N}\left(0, \frac{1}{4f^2(\theta)}\right).$$