Large Sample Theory

Basic Definition:

1. consistent estimator δ_n for $g(\theta)$, if $\delta_n \to g(\theta)$ in probability.
2. asymptotic relative efficiency of $\hat{\theta}_n$ w.r.t. $\tilde{\theta}_n$ is $\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}$.

τ-percentile: $\tilde{\theta}_n$ be the $\lfloor \tau n \rfloor$-th order statistic, $F(\theta) = \tau$, then

$$\sqrt{n}(\tilde{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \frac{\tau(1-\tau)}{(F'(\theta))^2}).$$

(Hint: consider $S_n = \#\{i \le n : X_i \le \theta + a/\sqrt{n}\}$.)

Delta method: $\sqrt{n}(f(\bar{X}_n) - f(\mu)) \Rightarrow \mathcal{N}(0, (f'(\mu))^2\sigma^2)$.

MLE: $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, I^{-1}(\theta))$.

Data Reduction

Sufficient:

1. def. conditional distribution $[X \mid T = t]$ doesn't depend on θ.
2. B&C$^{6.2.2}$. if $p(x \mid \theta)/q(T(x) \mid \theta)$ is free of θ, then $T(X)$ is suff.
3. NFFC. $T(X)$ is sufficient if.f. $p_\theta(x) = g_\theta(T(x))h(x)$.

Minimal sufficient:

1. def. sufficient T is min.suff. if T is function of any other suff. T'.
2. if $p(x;\theta) = c_{x,y}p(y;\theta) \Leftrightarrow T(x) = T(y)$, then T is min.suff.

Complete:

1. def. V is ancillary if the distribution of V is free of θ.
2. def. V is first-order ancillary if $\mathbb{E}_\theta[V]$ is free of θ.
3. def. T is complete if $\mathbb{E}_\theta[f(T)] = 0$ for all θ implies $f(T) = 0$ a.e.
4. (Basu) complete and sufficient $U \perp\!\!\!\perp$ ancillary V.
5. (Rao-Blackwell) for conv.loss and suff T, $R(\theta, \mathbb{E}[\delta \mid T]) \le R(\theta, \delta)$.
6. common steps:
   (a) suppose $\int f(x)h(x)e^{\theta x}\,dx = 0$ for all $\theta \in \Omega$,
   (b) decompose $f = f_+ - f_-$ with $f_+, f_- \ge 0$,
   (c) view f_+ and f_- as un-normalised densities p_+ and p_-,
   (d) argue that MGF of p_+ and p_- are equal, then $f_+ = f_-$ a.e.

UMRUE:

1. def. $R(\theta, \delta) \le R(\theta, \delta')$ for $\forall\, \theta \in \Omega$ and $\forall$ unbiased δ'.
2. (Lehmann-Scheffe) if T is comp.suff. and $\mathbb{E}_\theta[h(T)] = g(\theta)$, then $h(T)$ is i) only unbiased fun. of T, ii) UMRUE under conv.loss.
3. δ_0 is UMVUE of $g(\theta)$ if.f. $\mathbb{E}[\delta_0(X)U] = 0$ for all U with mean 0. (Hint: consider $\delta_\lambda = \delta_0 + \lambda U$.)
4. strategies to find UMRUE
   (a) Rao-B. condition of comp.suff. on unbiased estimator. cond.probability for discrete, ancillary for continuous.
   (b) Solve the unique δ satisfying $\mathbb{E}_\theta[\delta(T)] = g(\theta)$.
   (c) Guess.

Fisher Information:

1. $I(\theta) = \mathbb{E}[(\frac{\partial}{\partial\theta}\log f_\theta(x))^2] = -\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\log f_\theta(x)]$.
2. Cramer-Rao lower bound: $\mathrm{Var}(\delta) \ge [g'(\theta)]^2/I(\theta)$.
   $\varphi(x) = \frac{\partial}{\partial\theta}\log f_\theta(x), \mathbb{E}_\theta[\varphi(X)] = 0, E_\theta[\delta^2] < \infty, g'(\theta) = \mathbb{E}_\theta[\delta\varphi]$.

Exponential Family

general form:

$$p(x;\theta) = \exp\{\sum_{i=1}^n \eta_i(\theta)T_i(x) - B(\theta)\}h(x).$$

1. standardiser: $B(\theta) = \log\int \exp\{\sum_{i=1}^n \eta_i(\theta)T_i(x)\}h(x)\,dx$.
2. parameter space: $\Theta = \{\theta : B(\theta) < \infty\}$.

canonical form:

$$p(x;\eta) = \exp\{\sum_{i=1}^n \eta_i T_i(x) - B(\theta)\}h(x).$$

1. natural parameter η_i, and nature parameter space.
2. def. canonical exp.fam. is minimal, if no affine T_i's and η_i's. ($\sum_i \lambda_i T_i(x) = \lambda_0$ implies $\lambda_i = 0$, similar for η_i's.)
3. def. min.exp.fam. is full-rank, if nat.par.space contain open rect.
4. if exp.fam. is full-rank, then T is minimal sufficient and complete.

Hypothesis Testing

Basic definition:

1. test function $\phi(x)$: the prob. rejects H_0 given $X = x$,
2. power function: $\beta(\theta) = \mathbb{E}_\theta[\phi(X)] = P_\theta(\text{rejects } H_0)$,
3. significant level α: $\sup_{\theta_0 \in \Omega_0} \mathbb{E}_\theta[\phi(X)] \le \alpha$,
4. level-α uniformly most powerful test ϕ: if
   $$\mathbb{E}_\theta[\phi(X)] \ge \mathbb{E}_\theta[\phi^*(X)] \quad \text{for all } \theta \in \Omega_1,$$
   for any other level-α test ϕ^*.
5. families with monotone likelihood ratio in $T(X)$:
   (a) $\theta \ne \theta'$ implies $p_\theta \ne p_{\theta'}$, (identifiability)
   (b) $\theta < \theta'$ implies the ratio $p_\theta(x)/p_{\theta'}(x)$ is a non-decreasing function of $T(X)$. (monotonicity)

Find UMP test:

1. Neyman-Pearson Lemma for simple vs simple:
   Existence there exist $\phi(x)$ and constant k
   (a) $\mathbb{E}_{p_0}[\phi(X)] = \alpha$, (size = level)
   (b) $\phi(x) = \begin{cases} 1 & \text{if } \frac{p_1(x)}{p_0(x)} > k_\alpha \\ 0 & \text{if } \frac{p_1(x)}{p_0(x)} < k_\alpha \end{cases}$

   Sufficiency if a test holds (a) and (b) for some k, then it is MP.

   Necessity if a test ϕ is MP at level α, then it holds (b) for some k, and also holds (a) unless $\exists$ a test of size $< \alpha$ and power 1.
2. For MLR family, test $H_0 : \theta \le \theta_0$ vs $H_1 : \theta \ge \theta_1$:
   (a) there exist a UMP test at level α of the form
   $$\phi(x) = \begin{cases} 1 & \text{if } T(x) > k \\ \gamma & \text{if } T(x) = k \\ 0 & \text{if } T(x) < k \end{cases},$$
   (b) the power function $\beta(\theta) = \mathbb{E}_\theta[\phi(X)]$ is strictly increasing for $0 < \beta(\theta) < 1$.

Optimal Tests for Composite Nulls:

Hypothesis: $H_0 : X \sim f_\theta,\ \theta \in \Omega$ vs $H_1 : X \sim g$.

Consider new hypothesis $H_\Lambda :\ X \sim h_\Lambda(x) = \int_{\Omega_0} f_\theta(x)\,d\Lambda(\theta)$.

Let β_Λ be the power of MP level-α test ϕ_Λ for H_Λ vs H_1.

Prior Λ is a least favorable dist if $\beta_\Lambda \le \beta_{\Lambda'}$ for any prior Λ'.

Suppose ϕ_Λ is an MP level α-test for testing H_Λ against H_1. If ϕ_Λ is level-α for the original hypothesis H_0, then

1. The test ϕ_Λ is MP for original test vs alternative,
2. The prior distribution Λ is least favorable.

Bayes Esti. and Average Risk Optimality

Basic definition:

1. loss function: $L(\theta, d)$,
2. risk: $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$,
3. average/Bayes risk: $r(\Lambda, \delta) = \int_\Omega R(\theta, \delta)\, d\Lambda(\theta)$,
4. posterior risk: $\mathbb{E}[L(\Theta, \delta(X)) \,|\, X = x]$,
5. Bayes estimator δ_Λ: δ that minimizes $r(\Lambda, \delta)$,
6. Bayes risk: $r(\Lambda, \delta_\Lambda)$ for prior Λ,
7. inadmissible (esimator δ): if there exists another estimator δ' which dominates it (that is, such that $R(\theta, \delta') \le R(\theta, \delta)$ for all θ, with strict inequality for some θ),
8. admissible: no such dominating estimator δ' exists.

Find Bayes estimator:

1. minimizing posterior loss (take derivative w.r.t. d).
   - there exists δ_0 with finite risk for all θ.
2. posterior mean, for the squared loss function.

Properties:

1. unbiased estimator δ for $g(\theta)$ is not Bayes est. under the squared loss function unless its average risk is zero.
   - $r(\Lambda, \delta) < \infty$ and $\mathbb{E}[g(\Theta)^2] < \infty$.
   - (Hint: $\mathbb{E}_{(X,\Theta)}[\delta(X)g(\Theta)] = \mathbb{E}_X[\delta^2(X)] = \mathbb{E}_\Theta[g^2(\Theta)]$.)
2. unique Bayes estimator is admissible.
3. Bayes estimator is unique when
   (a) under strictly convex loss function,
   (b) $r(\Lambda, \delta) < \infty$, finiteness for comparison,
   (c) $P_\theta << Q$, where Q is the marginal dist. of X.
   (open support of Λ, and $P_\theta(A)$ cont. w.r.t. θ)
   (fitness for comparison, same support)
4. all admissible estimators are limits of Bayes estimators.

Empirical Bayes estimator:

1. calculate marginal distribution of X:
$$m(x \,|\, r) = \int f(x \,|\, \theta)\pi(\theta \,|\, r)\, d\theta$$
2. estimate the hyperparameter based on max $m(x \,|\, r)$.
3. minimize the empirical posterior loss:
$$\min_\delta \int L(\theta, \delta(x))\pi(\theta \,|\, x, \hat{r}(x))d\theta.$$

Minimaxity & Worst-Case Optimality

Basic Definition:

1. minimax estimator: δ that minimize $\sup_{\theta\in\Omega} R(\theta, \delta)$.
2. least favourable prior Λ: $r_\Lambda \ge r_{\Lambda'}$ for any prior Λ'.
3. least favourable sequence of priors $\{r_{\Lambda_m}\}$:
   (a) $r_{\Lambda_m} = r(\Lambda_m, \delta_{\Lambda_m}) \to r < \infty$,
   (b) $r \ge r_{\Lambda'}$ for any prior Λ'.

Find minimax estimator:

1. If Bayes risk = minimax risk, i.e. $r_\Lambda = \sup_\theta R(\theta, \delta_\Lambda)$,
   (a) Bayes estimator δ_Λ is minimax,
   (b) Λ is a least favourable prior,
   (c) unique Bayes esti. implies unique minimax esti.
2. If a Bayes estimator has constant risk, it's minimax.
3. $\omega_\Lambda = \{\theta\colon R(\theta, \delta_\Lambda) = \sup_{\theta'} R(\theta', \delta_\Lambda)\}$, δ_Λ is minimax if $\Lambda(\omega_\Lambda) = 1$.
4. If a sequence of priors $\{r_{\Lambda_m}\}$ with $r_{\Lambda_m} \to r < \infty$, and there exists estimator δ with $\sup_\theta R(\theta, \delta) = r$, then
   (a) δ is minimax,
   (b) $\{r_{\Lambda_m}\}$ is least-favourable.

Property:

1. minimax esti. may not necessarily be Bayes esti.
2. admissible with constant risk, implies minimax.
3. minimaxity may not guarantee admissibility.

Randomized minimax estimator for non-convex losses.

Prove (in)admissibility:

1. support of parameter,
2. risk equal to 0 at some par. point,
3. unique Bayes estimator, or convex combination,
4. limiting Bayes method:
   (a) assume minimax esti. is inadmissible,
   (b) construct strictly dominating esti.,
   (c) calculate average risk of Bayes esti. and dominating esti. under same conjucate prior,
   (d) calculate the ratio of diff. of minimax risk and average risk, take hyperpar. to infinity.
5. (def) find dominating estimator.

Common Distributions

1. Gamma distribution:
$$f_X(x \,|\, \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$
   with mean $\alpha\beta$, variance $\alpha\beta^2$, k-th moment $\dfrac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\beta^n$.
   Gamma function:
   (a) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \alpha > 0$,
   (b) $\Gamma(n) = (n-1)!, \quad n$ is integer,
   (c) $\Gamma(1) = 1,\ \Gamma(\frac{1}{2}) = \sqrt{\pi}$.
2. Beta distribution:
$$f(x \,|\, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$
   with moment $(\dfrac{\alpha}{\alpha+\beta})\ \dfrac{B(\alpha+n, \beta)}{B(\alpha, \beta)}$, where $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
3. Exponential distribution:
$$f_X(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \quad 0 \le x < \infty,\ \lambda > 0.$$
   with $\mathbb{E}[X] = \lambda$, $\mathrm{Var}X = \lambda^2$.
4. Poisson distribution:
$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots.$$
   with $\mathbb{E}[X] = \lambda$, $\mathrm{Var}X = \lambda$.