BAYESIAN APPROACH (1)

- Θ is realization of r.v. $\vec{\theta} \in \Theta$, whose prior distribution is Π.
  Sample $X \in \mathcal{X}$ from $P_\theta = P_{X|\theta}$ (X given $\vec{\theta} = \theta$)
  Update to Posterior distribution $P_{\theta|X}$
- Joint distribution of X and $\vec{\theta}$,
  $$P(A \times B) = \int_B P_{X|\theta}(A)\, d\Pi(\theta), \quad A \in \mathcal{B}_{\mathcal{X}},\ B \in \mathcal{B}_\Theta$$

- **Theorem 4.1** (Bayes formula)
  $\mathcal{P} = \{P_{X|\theta} : \theta \in \Theta\}$ dominated by σ-finite ν, $f_\theta(x) = \frac{dP_{X|\theta}}{d\nu}$.
  Π is prior, $m(x) = \int_\Theta f_\theta(x)\, d\Pi > 0$.
  (i) Posterior $P_{\theta|X} \ll \Pi$ and $\frac{dP_{\theta|X}}{d\Pi} = \frac{f_\theta(x)}{m(x)}$
  (ii) If $\Pi \ll \lambda$ and $d\Pi/d\lambda = \pi(\theta)$ (λ σ-finite),
  then $$\frac{dP_{\theta|X}}{d\lambda} = \frac{f_\theta(x)\pi(\theta)}{m(x)}$$

- **Definition 4.1** (Bayes Action)
  $\mathcal{A}$ is action space; $L(\theta, a) \geq 0$ loss function.
  $\forall x \in \mathcal{X}$, a Bayes action wrt Π is any $\delta(x) \in \mathcal{A}$ s.t.
  $$E_{\theta|X}\left[L(\vec{\theta}, \delta(x)) \mid X = x\right] = \min_{a \in \mathcal{A}} E\left[L(\vec{\theta}, a) \mid X = x\right]$$
- If $\int_\Theta [g(\theta)]^2 d\Pi < \infty$, $\mathcal{A}$ = range $g(\theta)$, with Sq. Err. loss,
  $$\delta(x) = \int_\Theta g(\theta) f_{\theta|x}\, d\Pi = \frac{\int_\Theta g(\theta) f_\theta(x)\, d\Pi}{\int_\Theta f_\theta(x)\, d\Pi} = \frac{\int_\Theta g(\theta) f_\theta(x)\, d\Pi}{m(x)}$$
  (the posterior expectation of $g(\vec{\theta})$ given $X = x$.)

GENERALIZED, EMPIRICAL, AND HIERARCHICAL BAYES METHODS (2)

- Generalized Bayes:
  Action: $\delta(x)$ s.t. $\int_\Theta L(\theta, \delta(x)) f_\theta(x)\, d\Pi = \min_{a \in \mathcal{A}} \int_\Theta L(\theta, a) f_\theta(x)\, d\Pi$
  - Defined if Π is not probability measure
    $\Pi(\Theta) \neq 1$ is improper prior.
  - Can use noninformative prior.

- Empirical Bayes
  ξ is vector of hyperparameters.
  Empirical Bayes estimates ξ with historical data.
  - X is a sample from
    $P_{X|\xi}(A) = \int_\Theta P_{X|\theta}(A)\, d\Pi_{\theta|\xi}$, $A \in \mathcal{B}_{\mathcal{X}}$,
    $\Pi_{\theta|\xi}$ is a prior depending on ξ, as for $m(x) = \int_\Theta f_\theta(x)\, d\Pi$
    if $P_{X|\theta}$ has a pdf f_θ.
  - Estimate ξ w/ M.O.M.
  - E.g. $X_i|\theta \sim N(\mu, \sigma^2)$, $\Pi_{\mu|\xi} = N(\mu_0, \sigma_0^2)$
    $$\int_{\mathbb{R}^n} x_i\, m(x)\, dx = \int_{\mathbb{R}} \int_{\mathbb{R}^n} x_i f_\mu(x)\, dx\, d\Pi_{\mu|\xi} = \int_{\mathbb{R}} \mu\, d\Pi_{\mu|\xi} = \mu_0$$
    $$\int_{\mathbb{R}^n} x_i^2\, m(x)\, dx = \int_{\mathbb{R}} \int_{\mathbb{R}^n} x_i^2 f_\mu(x)\, dx\, d\Pi_{\mu|\xi} = \sigma^2 + \int_{\mathbb{R}} \mu^2\, d\Pi_{\mu|\xi} = \sigma^2 + \mu_0^2 + \sigma_0^2$$
    MOM $\Rightarrow \hat{\mu}_0 = \bar{x}$, $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 - \sigma^2$.

- Hierarchical Bayes
  - Put a prior on hyperparameters.
    $\Pi_{\theta|\xi}$ 1st-stage prior
    Λ prior on Ξ, the range of ξ.
    "marginal prior" for Θ: $\Pi(B) = \int_\Xi \Pi_{\theta|\xi}(B)\, d\Lambda(\xi)$ $B \in \mathcal{B}_\Theta$
  - Second stage prior can be noninformative.
  - Shortcut: X has pdf $f_\theta(x)$ wrt ν
    $\Pi_{\theta|\xi}$ has pdf $\pi_{\theta|\xi}(\theta)$ wrt κ
    Then prior Π has pdf wrt κ
    $$\pi(\theta) = \int_\Xi \pi_{\theta|\xi}(\theta)\, d\Lambda(\xi)$$
    and $m(x) = \int_\Theta \int_\Xi f_\theta(x)\, \pi_{\theta|\xi}(\theta)\, d\Lambda\, d\kappa$
    Let $P_{\theta|x,\xi}$ be post. of $\vec{\theta}$ given X and ξ, and
    $$m_{x|\xi}(x) = \int_\Theta f_\theta(x)\, \pi_{\theta|\xi}(\theta)\, d\kappa \quad \text{(marginal of } X \text{ given } \xi\text{)}.$$
    Posterior $P_{\theta|x}$ has pdf
    $$\frac{dP_{\theta|x}}{d\kappa} = \frac{f_\theta(x)\pi(\theta)}{m(x)} = \int_\Xi \frac{f_\theta(x)\, \pi_{\theta|\xi}(\theta)}{m(x)}\, d\Lambda(\xi)$$
    $$= \int_\Xi \frac{f_\theta(x)\, \pi_{\theta|\xi}(\theta)}{m_{x|\xi}(x)} \frac{m_{x|\xi}(x)}{m(x)}\, d\Lambda(\xi)$$
    $$= \int_\Xi \frac{dP_{\theta|x,\xi}}{d\kappa}\, dP_{\xi|x}$$
    $P_{\xi|x}$ is posterior of ξ given X. Hierarchical Bayes action is
    $$\delta(x) = \int_\Xi \delta(x, \xi)\, dP_{\xi|x} \quad \text{(weighted fn } \delta(x,\xi)\text{)}$$

BAYES RULES AND ESTIMATORS (3)

- $\delta(x)$ minimizes Bayes risk $r_T(\Pi) = \int_\Theta R_T(\theta)\, d\Pi$
  over all (randomized or nonrandomized) decision rules T.
  $R_T(\theta) = E_X\left[L(\theta, T(X))\right]$; $\delta(x)$ is Bayes rule.
- Admissibility
  Given $R_T(\theta) = E[L(T(x), \theta)]$, T is $\mathfrak{J}$-admissible iff
  there is no $T_0 \in \mathfrak{J}$ w/ $R_{T_0}(\theta) \leq R_T(\theta)$ $\forall \theta$
  and $R_{T_0}(\theta) < R_T(\theta)$ for some θ.
  - Bayes rules are typically admissible.
  - If T is better than Bayes rule δ, T is also Bayes rule.

- **Theorem 4.2** (Admissibility of Bayes rules)
  Let $\delta(x)$ be Bayes rule wrt prior Π.
  (i) If $\delta(x)$ is unique Bayes rule, $\delta(x)$ is admissible.
  (ii) If Θ is countable set, Bayes risk $r_\delta(\Pi) < \infty$,
  and Π gives prob to each $\theta \in \Theta$, then $\delta(x)$ is admissible.
  (iii) Let $\mathfrak{J}$ be rules w/ cts risk functions. If $\delta(x) \in \mathfrak{J}$,
  $r_\delta(\Pi) < \infty$, and Π gives prob to any open subset of Θ,
  then $\delta(x)$ is $\mathfrak{J}$-admissible.

- **Theorem 4.3**
  Suppose $\Theta \subseteq \mathbb{R}$ open. $\mathfrak{J}$ = decision rules w/ cts risk. $T \in \mathfrak{J}$ is
  $\mathfrak{J}$-admissible if $\exists \{\Pi_j\}$ priors s.t.
  (a) $r_T(\Pi_j) < \infty$ $\forall j$
  (b) $\forall \theta_0 \in \Theta, \forall \eta > 0$, $\lim_{j \to \infty} \frac{r_T(\Pi_j) - r_j^*(\Pi_j)}{\Pi_j(O_{\theta_0,\eta})} = 0$
  $r_j^*(\Pi_j) = \inf_{S \in \mathfrak{J}} r_S(\Pi_j)$, $O_{\theta_0,\eta} = \{\theta \in \Theta : \|\theta - \theta_0\| < \eta\}$.

- Bias
  $\forall T$ estimator of ϑ, $\text{bias}_T = E(T) - \vartheta$.
  Bayes estimator is usually biased.
- **Proposition 4.2**
  Let $\delta(x)$ be a Bayes estimator of $\vartheta = g(\theta)$ under
  Squared error loss. Then $\delta(x)$ is not unbiased unless
  risk $r_\delta(\Pi) = 0$. [use to check if estimator is Bayes].

- Consistency
  Bayes estimators are usually consistent and asymptotically unbiased.

- Minimaxity
  Theorem 4.11 (minimaxity of Bayes estimator)
  Let Π be a proper prior on Θ and δ be a Bayes estimator
  of ϑ wrt Π. Suppose δ has constant risk on Θ_Π.
  If $\Pi(\Theta_\Pi) = 1$, δ is minimax.
  If, in addition, δ is the unique Bayes estimator wrt Π,
  it is the unique minimax estimator [choose prior to constant risk]

MINIMAX ESTIMATORS (4)

- Estimation of $\vartheta = g(\theta) \in \mathbb{R}$, based on X from P_θ: $\theta \in \Theta$
  under loss L and risk $R_T(\theta) = E[L(T(x), \theta)]$.
- Minimax estimator:
  minimizes $\sup_{\theta \in \Theta} R_T(\theta)$ over all estimators T.
- Unique minimax estimators are admissible.
- **Theorem 4.12**
  Let Π_j, $j = 1, 2, \ldots$ Seq. of priors, r_j is Bayes
  risk of Bayes estimator of ϑ wrt Π_j.
  Let T be a constant risk estimator of ϑ.
  If $\liminf_j r_j \geq R_T$, T is minimax.

- **Lemma 4.3**
  $\Theta_0 \subseteq \Theta$. T minimax of ϑ on Θ_0.
  Then T minimax on Θ if $\sup_{\theta \in \Theta} R_T(\theta) = \sup_{\theta \in \Theta_0} R_T(\theta)$.

- **Theorem 4.14**
  Suppose X has pdf $c(\theta) e^{\theta T(x)}$ wrt ν, $T(x) \in \mathbb{R}$,
  and $\theta \in (\theta_-, \theta_+) \subseteq \mathbb{R}$. Estimate $\vartheta = E(T(x))$ under
  Sq. Err. loss. Let $\lambda \geq 0$, γ known, and let
  $$T_{\lambda,\gamma} = (T + \gamma\lambda)/(1 + \lambda)$$
  Sufficient condition for admissibility of $T_{\lambda,\gamma}$ is that
  $$\int_{\theta_0}^{\theta_+} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda}\, d\theta = \int_{\theta_-}^{\theta_0} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda}\, d\theta = \infty, \quad \theta_0 \in (\theta_-, \theta_+)$$

- If T has constant risk and is admissible, T is
  minimax. If L is strictly convex, T is unique [Thm 4.11]
- **Corollary 4.3**
  Assume X has pdf as in Thm 4.14 w/ $\theta_- = -\infty$, $\theta_+ = \infty$.
  (i) Estimate $\vartheta = E(T)$, $T(x)$ admissible under
  Sq. Err. loss and loss $(a - \vartheta)^2/V_\theta(T)$.
  (ii) T is the unique minimax estimator of ϑ under
  the loss $(a - \vartheta)^2/V_\theta(T)$.

SIMULTANEOUS ESTIMATION AND SHRINKAGE ESTIMATORS (5)

- Simultaneous estimation
  $\vartheta \in \mathbb{R}^p$ (functions of θ)
  estimate w/ vector valued $T(x)$.
  Use single loss function $L(\vartheta, a)$
  e.g. Sq. Err. loss
  $$L(\theta, a) = \|a - \vartheta\|^2 = \sum_{i=1}^p (a_i - \vartheta_i)^2$$
- Most results extend straightforwardly, e.g.
  Unbiasedness, UMVUE, Bayes, minimaxity.
- Results for admissibility are quite different.
- **Theorem 4.15** (Risks of Shrinkage Estimators)
  $X \sim N_p(\theta, I_p)$ $p \geq 3$. Under Sq. Err. loss,
  with $\delta_{c,r} = X - \frac{r(p-2)}{\|X - c\|^2}(X - c)$,
  $c \in \mathbb{R}^p$, $r \in \mathbb{R}$ known, has risk
  $$R_{\delta_{c,r}}(\theta) = p - (2r - r^2)(p-2)^2\, E\left(\|X - c\|^{-2}\right).$$
  - risk smaller than p.
  - $\delta_{c,1}$ is best among $\delta_{c,r}$.
  - $\delta_{c,r}$ is minimax if $p \geq 3$, $0 \leq r \leq 2$,
  - $\delta_{c,r}$ dominated by $\delta_c^+ = X - \min\{1, \frac{p-2}{\|X - c\|^2}\}(X - c)$
  - if $\text{Var}(X) = \sigma^2 D$, σ^2 known,
    $$\tilde{\delta}_{c,r} = X - \frac{(p-2) r \sigma^2}{\|D^{-1}(x - c)\|^2} D^{-1}(x - c).$$
    $$R_{\tilde{\delta}_{c,r}} = \sigma^2\left[\text{tr}(D) - (2r - r^2)(p-2)^2 \sigma^2 E\left(\|D^{-1}(x - c)\|^{-2}\right)\right]$$
  - Shrinkage towards $\bar{x}$.
    $$X - \frac{p-3}{\|X - \bar{x}J_p\|^2}(X - \bar{x}J_p) \quad J_p \text{ is p-vector of 1's}$$
    $$X - \frac{(p-3)\hat{\sigma}^2}{\|D^{-1}(x - \bar{x}J_p)\|^2} D^{-1}(x - \bar{x}J_p)$$
    minimax when $p \geq 4$ under Sq. Err. loss.

LIKELIHOOD AND MLE (6)

- **Definition 4.3**
  $X \in \mathcal{X}$ Sample w/ pdf f_θ wrt ν, $\theta \in \Theta \subset \mathbb{R}^k$
  (i) $\forall x \in \mathcal{X}$, $f_\theta(x)$ is called likelihood function $\ell(\theta)$.
  (ii) $\bar{\Theta}$ is closure of Θ. If $\hat{\theta} \in \bar{\Theta}$ s.t.
  $$\ell(\hat{\theta}) = \max_{\theta \in \bar{\Theta}} \ell(\theta), \ \hat{\theta} \text{ is MLE of } \theta$$
  (iii) g Borel fn Θ to $\mathbb{R}^p$ $p \leq k$. If $\hat{\theta}$ is MLE of θ,
  $\hat{\vartheta} = g(\hat{\theta})$ is MLE of $\vartheta = g(\theta)$.
- Computation of MLE
  - Often more convenient to work w/ $\log \ell(\theta)$
  - Possible candidates solve $\frac{\partial \log \ell(\theta)}{\partial \theta} = 0$.
  - Max if $\frac{\partial^2 \log \ell(\theta)}{(\partial \theta)^2} < 0$

    if unique solution to likelihood eq. and $\ell(\theta) \to 0$ on boundary.
    If no solution, see if increasing or decreasing and look at boundary pts.

Methods of Computing MLE

- likelihood on integers:
  look for $\frac{\ell(\theta)}{\ell(\theta-1)} > 1$ or < 1 for particular θ.
  MLE θ at "tipping point".
- Mixed distribution.
  $f_\theta(x) = \theta f_1(x) + (1-\theta) f_2(x)$ $\theta \in (0,1)$.
  $S(\theta) = \frac{\partial \log \ell(\theta)}{\partial\theta} = \sum_{i=1}^n \frac{f_1(x_i) - f_2(x_i)}{f_2(x_i) + \theta[f_1(x_i) - f_2(x_i)]}$
  $S'(\theta) = -\sum_{i=1}^n \frac{[f_1(x_i) - f_2(x_i)]^2}{[f_2(x_i) + \theta[f_1(x_i) - f_2(x_i)]]^2} < 0$.
  $S(\theta) = 0$ has solution iff
  $\lim_{\theta \to 0} S(\theta) > 0$ AND $\lim_{\theta \to 1} S(\theta) < 0$ i.e.
  $\sum_{i=1}^n \frac{f_1(x_i)}{f_2(x_i)} > n$ AND $\sum_{i=1}^n \frac{f_2(x_i)}{f_1(x_i)} > n$.
  o/w, MLE is 0 or 1
- Newton-Raphson
  $$\hat\theta^{(t+1)} = \hat\theta^{(t)} - \left[\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^T}\Big|_{\theta=\theta^{(t)}}\right]^{-1} \frac{\partial \log \ell(\theta)}{\partial\theta}\Big|_{\theta=\theta^{(t)}}$$
  full rank.
- Fisher Scoring.
  Replace $\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta\partial\theta^T}\Big|_{\theta=\theta^{(t)}}\right]$ by $\left[E\left(\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^T}\right)\right]\Big|_{\theta=\theta^{(t)}}$

MLE in GLM and Quasi-MLE (8)

- MLE in Exponential Families
  - $\ell(\eta) = \exp\{\eta^T T(x) - \zeta(\eta)\} h(x)$, $\eta \in \Xi$.
  - likelihood eq: $\frac{\partial \log \ell(\eta)}{\partial\eta} = T(x) - \frac{\partial \zeta(\eta)}{\partial\eta} = 0$,
  - If $T(x)$ in range of $\frac{\partial \zeta(\eta)}{\partial\eta}$, it is unique MLE of $\mu(\eta) = \frac{\partial \zeta(\eta)}{\partial\eta}$
  - MLE of η is $\hat\eta = \mu^{-1}(T(x))$ (since $\mu(\eta)$ is one-to-one)
  - Similar results if not natural exp form.
- GLM Structure
  - $X_i \sim \exp\left\{\frac{\eta_i x_i - \zeta(\eta_i)}{\phi_i}\right\} h(x_i, \phi_i)$, $i = 1, 2, \ldots n$
    pdf wrt ν, $\eta_i \in \Xi = \{\eta : 0 < \int h(x,\phi) e^{\eta x/\phi} d\nu(x) < \infty\}$
    ζ, h known, $\zeta''(\eta) > 0$ $\forall \eta \in \Xi^0$
  - $E(X_i) = \zeta'(\eta_i)$ AND $Var(X_i) = \phi_i \zeta''(\eta_i)$ $i = 1, \ldots, n$.
  - $\mu(\eta) = \zeta'(\eta)$ $g(\mu(\eta_i)) = \beta^T Z_i$, g unknown
  - Assume g is C^3, and $\frac{d}{d\eta}(g \circ \mu)(\eta) \neq 0$ $\forall \eta$.
  - Range of β: $B = \{\beta : (g \circ \mu)^{-1}(\beta^T z) \in \Xi^0 \ \forall z \in \mathcal{Z}\}$
  - ϕ_i's nuisance parameters that scale things.
- MLE in GLM
  - Assume $\phi_i = \phi/t_i$ $\forall i$, unknown ϕ, known t_i.
  - $\theta = (\beta, \phi)$, $\psi = (g \circ \mu)^{-1}$ $(\eta_i = \psi(\beta Z_i))$
  - $\log \ell(\theta) = \sum_{i=1}^n \left[\log h(x_i, \phi/t_i) + \frac{\psi(\beta^T Z_i) x_i - \zeta(\psi(\beta^T Z_i))}{\phi/t_i}\right]$
  - $\frac{\partial \log \ell(\theta)}{\partial\beta} = \frac{1}{\phi}\sum_{i=1}^n \{[x_i - \mu(\psi(\beta^T Z_i))]\psi'(\beta^T Z_i) t_i Z_i\} = 0$
    $\frac{\partial \log \ell(\theta)}{\partial\phi} = \sum_{i=1}^n \left\{\frac{\partial \log h(x_i, \phi/t_i)}{\partial\phi} - \frac{t_i[\psi(\beta^T Z_i) x_i - \zeta(\psi(\beta^T Z_i))]}{\phi^2}\right\} = 0$

- Can usually solve for β w/out ϕ. If a solution for β exists,
  $Var\left(\frac{\partial \log \ell(\theta)}{\partial\beta}\right) = \frac{M_n(\beta)}{\phi}$, $\frac{\partial^2 \log \ell(\theta)}{\partial\beta\partial\beta^T} = [R_n(\beta) - M_n(\beta)]/\phi$
  $M_n(\beta) = \sum_{i=1}^n [\psi'(\beta^T Z_i)]^2 \zeta''(\psi(\beta^T Z_i)) t_i Z_i Z_i^T$
  $R_n(\beta) = \sum_{i=1}^n [x_i - \mu(\psi(\beta^T Z_i))]\psi''(\beta^T Z_i) t_i Z_i Z_i^T$
- If g is canonical, $\psi'' \equiv 0$ and $R_n = 0$, easy to solve, $M_n(\beta)$ pos def $\Rightarrow \hat\beta$ is unique solution.
- If g noncanonical, if $[M_n(\beta)]^{-1/2} R_n(\beta) [M_n(\beta)]^{-1/2} \to 0$, then $\hat\beta$ is MLE for large n.
- Computation of MLE in GLM
  - $\hat\beta^{(t+1)} = \hat\beta^{(t)} - [R_n(\hat\beta^{(t)}) - M_n(\hat\beta^{(t)})]^{-1} S_n(\hat\beta^{(t)})$ $t = 0, 1, \ldots$
    where $S_n(\beta) = \phi \frac{\partial \log \ell(\theta)}{\partial\beta}$. (Newton-Raphson)
  - $\hat\beta^{(t+1)} = \hat\beta^{(t)} + [M_n(\hat\beta^{(t)})]^{-1} S_n(\hat\beta^{(t)})$ (Fisher scoring).
- Quasi-MLE
  If assumption ϕ_i is arbitrary, or distribution assumption on X_i does not hold, but $E(X_i) = \zeta'(\eta_i)$ AND $Var(X_i) = \phi_i \zeta''(\eta_i)$ AND $g(\mu(Z_i)) = \beta^T Z_i$ $\forall i$,
  can estimate β by solving
  $$\sum_{i=1}^n \{[x_i - \mu(\psi(\beta^T Z_i))]\psi'(\beta^T Z_i) t_i Z_i\} = 0$$
  - This estimate is quasi-MLE.

Useful Results

Empirical Bayes: $X|\theta \sim N(\mu, \sigma^2)$
$\Pi_{\mu|\xi} = N(\mu_0, \sigma_0^2)$ $\xi = (\mu_0, \sigma_0^2)$
$\int_{\mathbb{R}^n} x_i m_\xi(x) dx = \int_\Theta \int_{\mathbb{R}^n} x_i f_\mu(x) dx d\Pi_{\mu|\xi} = \int_{\mathbb{R}} \mu d\Pi_{\mu|\xi} = \mu_0 = \bar{x}$
$\int_{\mathbb{R}^n} x_i^2 m_\xi(x) dx = \int_\Theta \int_{\mathbb{R}^n} x_i^2 f_\mu(x) dx d\Pi_{\mu|\xi} = \sigma^2 + \int_{\mathbb{R}} \mu^2 d\Pi_{\mu|\xi}$
$= \sigma^2 + \mu_0^2 + \sigma_0^2 = \frac{1}{n}\sum x_i^2$

Conjugate Priors

Likelihood	Prior	Posterior
$Bern(p)$	$Beta(\alpha, \beta)$	$Beta(\alpha + \sum x_i, \beta + n - \sum x_i)$
$Bin(N_i, p)$	$Beta(\alpha, \beta)$	$Beta(\alpha + \sum x_i, \beta + \sum N_i - \sum x_i)$
$NB(p, r)$	$Beta(\alpha, \beta)$	$Beta(\alpha + \sum x_i, \beta + rn)$
$Pois(\lambda)$	$\Gamma(\alpha, \gamma)$	$\Gamma(\alpha + \sum x_i, \frac{\gamma}{n\gamma + 1})$
$G(p)$	$Beta(\alpha, \beta)$	$Beta(\alpha + n, \beta + \sum x_i - n)$
$N(\mu, \sigma^2)$	$N(\mu_0, \sigma_0^2)$	$N\left(\frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}\right), \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$
$U(0, \theta)$	$Pa(a, \theta)$	$Pa(\max\{x_1, \ldots, x_n, a\}, \theta + n)$
$E(0, \theta)$	$\Gamma(\alpha, \gamma)$	$\Gamma(\alpha + n, \frac{\gamma}{1 + \gamma \sum x_i})$
$\Gamma(\alpha, \gamma)$	$\Gamma(\alpha_0, \gamma_0)$	$\Gamma(\alpha_0 + n\alpha, \frac{\gamma_0}{1 + \gamma_0 \sum x_i})$

Common Distributions

Poisson
$P(\theta)$ $\theta^x e^{-\theta}/x!$ $x = 0, 1, \ldots$
$E(x) = \theta$ $Var(x) = \theta$

Geometric
$G(p)$ $(1-p)^{x-1} p$ $x = 1, 2, \ldots$
$E(x) = 1/p$ $Var(x) = (1-p)/p^2$

Negative Binomial
$NB(p, r)$ $\binom{x-1}{r-1} p^r (1-p)^{x-r}$ $x = r, r+1, \ldots$
$E(x) = r/p$ $Var(x) = r(1-p)/p^2$

Exponential
$E(a, \theta)$ $\theta^{-1} e^{-(x-a)/\theta}$
$E(x) = \theta + a$ $Var(x) = \theta^2$

Gamma
$\Gamma(\alpha, \gamma)$ $\frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x)$
$E(x) = \alpha\gamma$ $Var(x) = \alpha\gamma^2$

Beta
$B(\alpha, \beta)$ $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x)$
$E(x) = \alpha/(\alpha+\beta)$ $Var(x) = \alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$

Pareto
$Pa(a, \theta)$ $\theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x)$
$E(x) = \theta a/(\theta - 1)$ $Var(x) = \theta a^2/[(\theta-1)^2(\theta-2)]$.

5

Asymptotically Efficient Estimation

- $\{\hat\theta_n\}$ estimating θ is asymptotically normal if
  $$[V_n(\theta)]^{-\frac{1}{2}}(\hat\theta_n - \theta) \xrightarrow{d} N_k(0, I_k),$$
  with $V_n(\theta)_{k\times k}$ positive definite.
  - $k=1$, $V_n(\theta)$ is variance and amse of $\hat\theta_n$
  - $k>1$, $V_n(\theta)$ is asymptotic covariance matrix.
- For $\hat\theta_{jn}$ AN with asy covariance $V_{jn}(\theta)$, $j=1,2$, say $\hat\theta_{1n}$ asymptotically more efficient than $\hat\theta_{2n}$ if
  $$V_{1n}(\theta) \le V_{2n}(\theta).$$
- If $V_n(\theta) = Var(\hat\theta_n)$, then under some regularity conditions from Thm 3.3,
  $$V_n(\theta) \ge [I_n(\theta)]^{-1}$$
  doesn't always work if V_n is cov-matrix.
- Theorem 4.16

  $X_1, \dots, X_n$ iid from p.d.f. wrt σ-finite ν, $\theta \in \Theta \underset{open}{\subset} \mathbb{R}^k$.
  $\forall x$ in range of X, $f_\theta(x)$ is C^2 in θ,
  $$\frac{\partial}{\partial\theta}\int \psi_\theta(x)d\nu = \int \frac{\partial}{\partial\theta}\psi_\theta(x)d\nu, \quad \psi_\theta(x) = f_\theta(x), \frac{\partial}{\partial\theta}f_\theta(x),$$
  $$I_1(\theta) = E\left[\frac{\partial}{\partial\theta}\log f_\theta(x)\left(\frac{\partial}{\partial\theta}\log f_\theta(x)\right)^T\right] \text{ positive definite},$$
  $\forall \theta \in \Theta$, $\exists c_\theta > 0$, $h_\theta \ge 0$ s.t. $E[h_\theta(X_1)] < \infty$ and
  $$\sup_{\gamma: \|\gamma - \theta\| < c_\theta} \left\|\frac{\partial^2 \log f_\gamma(x)}{\partial\gamma\partial\gamma^T}\right\| \le h_\theta(x) \quad \forall x,$$
  where $\|A\| = \sqrt{Tr(A^T A)}$ for matrix A.

  If $\hat\theta_n$ is estimator of θ, and is AN with $V_n(\theta) = \frac{V(\theta)}{n}$, then the information inequality holds except on a set of measure 0.
- Definition 4.4 (Asymptotic Efficiency)

  Assume Fisher information $I_n(\theta)$ is well defined. $\{\hat\theta_n\}$ that is AN is asymptotically efficient iff
  $$V_n(\theta) = [I_n(\theta)]^{-1}.$$
  $\hat\vartheta_n = g(\hat\theta_n)$ is asymptotically efficient for $\vartheta = g(\theta)$ if $\hat\theta_n$ is asymptotically efficient for θ.
- Theorem 4.17

  Assume the conditions of Thm 4.16.
  (i) $\exists \{\hat\theta_n\}$ s.t. $P(s_n(\hat\theta_n) = 0) \to 1$ AND $\hat\theta_n \xrightarrow{P} \theta$ where $s_n(\gamma) = \frac{\partial \log \ell(\gamma)}{\partial\gamma}$

  (ii) Any consistent sequence $\hat\theta_n$ of RLE's is asymptotically efficient.
- For Exponential Families:
  $$f_\eta(x_i) = \exp\{\eta^T T(x_i) - \zeta(\eta)\}h(x_i)$$
  $$-\frac{\partial^2 \log f_\eta(x_i)}{\partial\eta\partial\eta^T} = \frac{-\partial^2 \zeta(\eta)}{\partial\eta\partial\eta^T} \Rightarrow \sup_{\gamma: \|\gamma - \eta\| < c_\eta}\left\|\frac{\partial^2 \log f_\gamma(x)}{\partial\gamma\partial\gamma^T}\right\| \le h_\eta(x)$$
  Prop 3.2 $\Rightarrow$ other conditions in Thm 4.16 are satisfied.
  $$s(\eta) = \sum_{i=1}^n \left[T(X_i) - \frac{\partial\zeta(\eta)}{\partial\eta}\right] \quad \left(\frac{\partial^2\zeta(\eta)}{\partial\eta\partial\eta^T} = Var(T(X_i))\right)$$
  $\hat\theta_n = \frac{1}{n}\sum T(X_i) \in \Theta \Rightarrow \hat\theta_n$ is unique RLE $\Rightarrow$ unique MLE

RLE of η is $\hat\eta_n = g^{-1}(\hat\theta_n)$ $\quad (\eta = g^{-1}(\theta))$.

- Theorem 4.18

  Consider GLM with $\phi_i = \phi/t_i$, $t_i \in (t_0, t_\infty)$, $0 < t_0 \le t_\infty < \infty$.
  Assume the range of β is open subset of $\mathbb{R}^p$; $0 < \inf \varphi(\beta^T Z_i) \le \sup \varphi(\beta^T Z_i) < \infty$
  $\varphi(t) = [\psi'(t)]^2 \zeta''(\psi(t))$; $\max_{i \le n} Z_i^T (Z^T Z)^{-1} Z_i \xrightarrow{n\to\infty} 0$ and $\lambda_-[Z^T Z] \to \infty$.

  (i) $\exists! \{\hat\beta_n\}$ s.t. $P(s_n(\hat\beta_n) = 0) \to 1$ AND $\hat\beta_n \xrightarrow{P} \beta$
  $s(\beta) = \frac{\partial \log \ell(\beta, \phi)}{\partial\beta}$ is the score

  (ii) Let $I_n(\beta) = Var(s_n(\beta))$. Then
  $$[I_n(\beta)]^{\frac{1}{2}}(\hat\beta_n - \beta) \xrightarrow{d} N(0, I_p).$$

  (iii) If ϕ is known or the p.d.f. indexed by $\theta = (\beta, \phi)$ satisfies the conditions for f_θ in Thm 4.16, $\hat\beta_n$ is asymptotically efficient.
- One Step MLEs
  $$\hat\theta_n^{(1)} = \hat\theta_n^{(0)} - \left[\nabla s_n(\hat\theta_n^{(0)})\right]^{-1} s_n(\hat\theta_n^{(0)})$$
  [think NR with some initial $\hat\theta_n^{(0)}$]
- Theorem 4.19

  Assume that the conditions of Thm 4.16 hold and $\hat\theta_n^{(0)}$ is $\sqrt{n}$-consistent for θ.
  (i) The one-step MLE $\hat\theta_n^{(1)}$ is asymptotically efficient.
  (ii) The one-step MLE that replaces $\nabla s_n(\gamma)$ with its expected value, $-I_n(\gamma)$, is asymptotically efficient (F-S).
- Theorem 4.20

  "Under some nice conditions Bayes estimators are asymptotically efficient"

Pseudo-Likelihood Approach

[omitted]

Empirical CDF and Nonparametric MLE

- Empirical c.d.f.:
  $$F_n(t) = \frac{1}{n}\sum_{i=1}^n I_{(-\infty, t]}(X_i) \quad t \in \mathbb{R}^d$$
  $(-\infty, a] = (-\infty, a_1] \times \dots \times (-\infty, a_d]$ for $a = (a_1, \dots, a_d) \in \mathbb{R}^d$.
- Properties:
  - $\forall t \in \mathbb{R}^d$, $nF_n(t) \sim Bin(F(t), n)$
  - $E(F_n(t)) = F(t)$; $Var(F_n(t)) = F(t)[1 - F(t)]/n$.
  - $F_n(t)$ is $\sqrt{n}$-consistent for $F(t)$
  - $\sqrt{n}[(F_n(t_1), \dots, F_n(t_m)) - (F(t_1), \dots, F(t_m))] \xrightarrow{d} N(0, \Sigma)$
    $\Sigma_{m\times m}$ has $(i,j)^{th}$ entry $P(X_1 \in (-\infty, t_i] \cap (-\infty, t_j]) - F(t_i)F(t_j)$
  - $\rho_\infty(G_1, G_2) = \|G_1 - G_2\|_\infty = \sup_{t \in \mathbb{R}^d}|G_1(t) - G_2(t)|$
- Lemma 5.1 (DKW inequality)

  (i) When $d=1$, $\exists C > 0$ (not depending on F) st.
  $$P(\rho_\infty(F_n, F) > z) \le Ce^{-2nz^2}, \quad z > 0.$$
  (ii) When $d \ge 2$, $\forall \varepsilon > 0$ $\exists C_{\varepsilon, d} > 0$ s.t.
  $$P(\rho_\infty(F_n, F) > z) \le C_{\varepsilon, d}\, e^{-(2-\varepsilon)nz^2} \quad z > 0.$$

- Theorem 5.1

  Let F_n be e.c.d.f. on iid $X_1, \dots, X_n$ from F on $\mathbb{R}^d$
  (i) $\rho_\infty(F_n, F) \xrightarrow{a.s.} 0$ as $n \to \infty$
  (ii) $E[\sqrt{n}\rho_\infty(F_n, F)]^s = O(1)$ $\forall s > 0$.
  ($n^{-\frac{1}{2}}$ rate of convg)
- $$\rho_{L_p}(G_1, G_2) = \|G_1 - G_2\|_{L_p} = \left[\int |G_1(t) - G_2(t)|^p dt\right]^{1/p}$$
  for $G_i \in \mathscr{F}_1 = \{G \in \mathscr{F} : \int |t| dG(t) < \infty\}$.
- Theorem 5.2

  Let F_n be e.c.d.f. based on $X_1, \dots, X_n \overset{iid}{\sim} F \in \mathscr{F}_1$
  (i) $\rho_{L_p}(F_n, F) \xrightarrow{a.s.} 0$
  (ii) $E[\sqrt{n}\rho_{L_p}(F_n, F)] = O(1)$ if $1 \le p < 2$ AND
  $\int \{F(t)[1 - F(t)]\}^{p/2} dt < \infty$ OR $p \ge 2$.
- Nonparametric MLE

  P_G is prob. measure corresponding to $G \in \mathscr{F}$. Given $X_1 = x_1, \dots, X_n = x_n$, the nonparametric likelihood is function from $\mathscr{F}$ to $[0, \infty)$.
  $$\ell(G) = \prod_{i=1}^n P_G(\{x_i\})$$
- Theorem 5.3

  $X_1, \dots, X_n \overset{iid}{\sim} F \in \mathscr{F}$.
  The e.c.d.f. F_n maximizes $\ell(G)$ over $G \in \mathscr{F}$.

Empirical Likelihoods

- Adding info to nonparametric likelihood is called empirical likelihood, leads to MELE.
- Suppose $\int u(x)dF = 0$ for $u: \mathbb{R}^d \to \mathbb{R}^s$, Borel. maximize $\ell(G)$ subject to $p_i > 0$, $\sum_{i=1}^n p_i = 1$, and $\sum_{i=1}^n p_i u(x_i) = 0$ for $p_i = P_G(\{x_i\})$.
  An MELE of F is $\hat F(t) = \sum_{i=1}^n \hat p_i I_{(-\infty, t]}(X_i)$
  where $\hat p_i = n^{-1}[1 + \lambda_n^T u(X_i)]^{-1}$ and λ_n is such that
  $$\sum_{i=1}^n \hat p_i u(X_i) = \frac{1}{n}\sum_{i=1}^n \frac{u(X_i)}{1 + \lambda_n^T u(X_i)} = 0$$
  Can show this holds asymptotically (Similar to Thm 4.17).
- Theorem 5.4

  Let u Borel st. $\int u(x)dF = 0$ and $\hat F$ is MELE of F.
  Suppose $U = Var(u(X_1))$ pos. definite.
  Then $\sqrt{n}[(\hat F(t_1), \dots, \hat F(t_m)) - (F(t_1), \dots, F(t_m))] \xrightarrow{d} N(0, \Sigma_u)$
  where $\Sigma_u = \Sigma - W^T U^{-1} W$
  and $W = (W(t_1), \dots, W(t_m))$, $W(t_j) = E[u(X_1)I_{(-\infty, t_j]}(X_1)]$
  [i.e. $\hat F$ is asymptotically more efficient than F_n]

Density Estimation

- $$f_n(t) = \frac{F_n(t + \lambda_n) - F_n(t - \lambda_n)}{2\lambda_n}, \quad t \in \mathbb{R}$$
  where F_n is e.c.d.f. and $\{\lambda_n\}$ is a sequence of positive numbers.
- $2n\lambda_n f_n(t) \sim Bin(F(t + \lambda_n) - F(t - \lambda_n), n)$.
- $E[f_n(t)] \to f(t)$ if $\lambda_n \to 0$ as $n \to \infty$
- $Var(f_n(t)) \to 0$ if $\lambda_n \to 0$ and $n\lambda_n \to \infty$
- $\lambda_n \to 0$ and $n\lambda_n \to \infty$, f C^1 $\Rightarrow$
  $$mse_{f_n(t)}(F) = \frac{f(t)}{2n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right) + O(\lambda_n^2).$$
- if, additionally, $n\lambda_n^3 \to 0$,
  $$\sqrt{n\lambda_n}[f_n(t) - f(t)] \xrightarrow{d} N\left(0, \tfrac{1}{2}f(t)\right)$$
- Kernel density estimators
  $$\hat f(t) = \frac{1}{n\lambda_n}\sum_{i=1}^n w\left(\frac{t - x_i}{\lambda_n}\right)$$
  where w is known Leb. p.d.f.
- Properties of kernel density estimator
  - $\hat f$ is a Lebesgue density
  - $bias\,\hat f(t) = \int w(y)[f(t - \lambda_n y) - f(t)]dy$
    - if f Bounded + Cts, DCT $\Rightarrow$ $bias\,\hat f(t) \to 0$ as $\lambda_n \to 0$
    - If f Bounded + Cts, and $w_0 = \int [w(t)]^2 dt < \infty$, then
      $$Var(\hat f(t)) = \frac{1}{n\lambda_n^2}Var\left(w\left(\frac{t - X_1}{\lambda_n}\right)\right) = \dots$$
      $$\dots = \frac{w_0 f(t)}{n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right)$$
    - If $\lambda_n \to 0$, $n\lambda_n \to \infty$, f' is Bounded + Cts, then
      $$mse_{\hat f(t)}(F) = \frac{w_0 f(t)}{n\lambda_n} + O(\lambda_n^2)$$
    - If, additionally, $w_0 < \infty$,
      $$\sqrt{n\lambda_n}\{\hat f(t) - E[\hat f(t)]\} \xrightarrow{d} N(0, w_0 f(t)).$$
    - If, additionally, f' is B+C and $n\lambda_n^3 \to 0$, then
      $$\sqrt{n\lambda_n}\{\hat f(t) - f(t)\} \xrightarrow{d} N(0, w_0 f(t))$$
      "slower than $\sqrt{n}$"

Sample Quantiles and Their Asymptotic Properties

- $p \in (0,1)$. $G^{-1}(p) = \inf\{x : G(x) \ge p\}$ is the p^{th} quantile of G. (also called θ_p).
- $\hat\theta_p = F_n^{-1}(p) = c_{np}X_{(m_p)} + (1 - c_{np})X_{(m_p + 1)}$
  where $m_p = \lfloor np \rfloor$
- $\lim_{x \uparrow \theta_p} F(x) = F(\theta_p^-) \le p \le F(\theta_p^+) = \lim_{x \downarrow \theta_p} F(x)$.

Theorem 5.9

Let $X_1, \dots, X_n \overset{iid}{\sim} F$ st. $p < F(\theta_p + \varepsilon)$ $\forall \varepsilon > 0$.
Then, $\forall \varepsilon > 0$, $P(|\hat\theta_p - \theta_p| > \varepsilon) \le 2Ce^{-2n\delta_\varepsilon^2}$,
where $\delta_\varepsilon = \min(F(\theta_p + \varepsilon) - p,\ p - F(\theta_p - \varepsilon))$.
[$\hat\theta_p$ strongly consistent for θ_p]

- $P(\hat\theta_p \le t) = \sum_{i = l_p}^n \binom{n}{i}[F(t)]^i[1 - F(t)]^{n-i}$, $l_p = \lceil np \rceil$.
- $\hat\theta_p$ has p.d.f. $\varphi_n(t) = n\binom{n-1}{l_p - 1}[F(t)]^{l_p - 1}[1 - F(t)]^{n - l_p}f(t)$

Theorem 5.10

Let $X_1, \dots, X_n \overset{iid}{\sim} F$, Φ is c.d.f. of $N(0,1)$.
(i) If $F(\theta_p) = p$, then $P(\sqrt{n}(\hat\theta_p - \theta_p) \le 0) \to \Phi(0) = \frac{1}{2}$
(ii) If F cts at θ_p and $\exists F'(\theta_p^-) > 0$,
$$P(\sqrt{n}(\hat\theta_p - \theta_p) \le t) \to \Phi(t/\sigma_F^-), \quad t < 0, \quad \sigma_F^- = \frac{\sqrt{p(1-p)}}{F'(\theta_p^-)}$$
(iv) If $F'(\theta_p) > 0$, $\sqrt{n}(\hat\theta_p - \theta_p) \xrightarrow{d} N(0, \sigma_F^2)$, $\sigma_F = \frac{\sqrt{p(1-p)}}{F'(\theta_p)}$

• Theorem 5.11

$X_1, \dots, X_n \overset{iid}{\sim} F$.

If $F'(\theta_p) > 0$ exists, then

$$\sqrt{n}(\hat{\theta}_p - \theta_p) = \sqrt{n}\frac{[F_n(\theta_p) - F(\theta_p)]}{F'(\theta_p)} + o_p(1).$$

• Corollary 5.1

Let $X_1, \dots, X_n \overset{iid}{\sim} F$ with positive derivative at θ_{p_i}, $0 < p_i < p_{i+1}$

Then $\sqrt{n}[(\hat{\theta}_{p_1}, \dots, \hat{\theta}_{p_m}) - (\theta_{p_1}, \dots, \theta_{p_m})] \xrightarrow{d} N_m(0, D)$,

where $D_{m \times m}$ has $(i,j)^{th}$ element $\dfrac{p_i(1-p_j)}{F'(\theta_{p_i})F'(\theta_{p_j})}$ $\quad i \le j$

ROBUSTNESS AND EFFICIENCY

• If $Var(X_1) < \infty$,

$\sqrt{n}(\bar{X} - \theta) \xrightarrow{d} N(0, Var(X_1))$ (CLT)

$\sqrt{n}(\hat{\theta}_{0.5} - \theta) \xrightarrow{d} N(0, [2F'(\theta)]^{-2})$ (Thm 5.10)

asymptotic relative efficiency of $\hat{\theta}_{0.5}$ w.r.t. $\bar{X}$ is

$e(F) = 4[F'(\theta)]^2 Var(X_1)$

[generally $e(F) > 1$ if F has fat tails]

• The α-trimmed sample mean is defined as

$$\bar{X}_\alpha = \frac{1}{(1-2\alpha)n}\sum_{j=m_\alpha+1}^{n-m_\alpha} X_{(j)}$$

where $m_\alpha = \lfloor n\alpha \rfloor$ and $\alpha \in (0, \frac{1}{2})$.

• If $F(x) = F_0(x - \theta)$, with F_0 symmetric around 0 and has Lb p.d.f. positive in range of X_1, then

$\sqrt{n}(\bar{X}_\alpha - \theta) \xrightarrow{d} N(0, \sigma_\alpha^2)$,

where $\sigma_\alpha^2 = \dfrac{2}{(1-2\alpha)^2}\left\{\int_0^{F_0^{-1}(1-\alpha)} x^2 dF_0(x) + \alpha[F_0^{-1}(1-\alpha)]^2\right\}$

• M-estimators

[omitted]

L-ESTIMATORS AND TRIMMED SAMPLE MEAN

[omitted]

PROFILE LIKELIHOODS, GEE, and GMM

• $\ell(\theta, \xi)$ a likelihood,

For each fixed θ, let $\xi(\theta)$ satisfy

$\ell(\theta, \xi(\theta)) = \sup_{\xi} \ell(\theta, \xi)$.

The function $\ell_p(\theta) = \ell(\theta, \xi(\theta))$ is profile likelihood

If $\hat{\theta}_p$ maximizes $\ell_p(\theta)$, it is MPLE of θ

PREVIOUS RESULTS

• $\hat{\theta}$ is MLE if $\ell(\hat{\theta}) = \max_{\theta \in \Theta} \ell(\theta)$.

• Possible candidates:

$\dfrac{\partial \log \ell(\theta)}{\partial \theta} = 0$ or boundary pts.

• Check $\dfrac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^T} < 0$ for max.

• Other methods of computing:

- integers $\dfrac{\ell(\theta)}{\ell(\theta-1)} > 1$ or < 1 for particular θ.

- Mixed distribution:

$f_\theta(x) = \theta f_1(x) + (1-\theta)f_2(x)$ $\quad \theta \in (0,1)$.

$s(\theta) = \dfrac{\partial \log \ell(\theta)}{\partial\theta} = \sum_{i=1}^{n} \dfrac{f_1(x_i) - f_2(x_i)}{f_2(x_i) + \theta[f_1(x_i) - f_2(x_i)]}$

$s'(\theta) < 0$.

$s(\theta)$ has solution iff

$\lim_{\theta \to 0} s(\theta) > 0$ AND $\lim_{\theta \to 1} s(\theta) < 0$

i.e. $\sum_{i=1}^{n} \dfrac{f_1(x_i)}{f_2(x_i)} > n$ AND $\sum_{i=1}^{n} \dfrac{f_2(x_i)}{f_1(x_i)} > n$

o/w MLE is 0 or 1.

• Newton-Raphson:

$$\theta^{(t+1)} = \theta^{(t)} - \left[\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^T}\Big|_{\theta=\theta^{(t)}}\right]^{-1} \frac{\partial \log \ell(\theta)}{\partial\theta}\Big|_{\theta=\theta^{(t)}}$$

• Fisher-Scoring:

Replace $\left[\dfrac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^T}\Big|_{\theta=\theta^{(t)}}\right]$ by $\left[E\left(\dfrac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^T}\right)\right]\Big|_{\theta=\theta^{(t)}}$

• Exp. families:

$\ell(\eta) = \exp\{\eta^T T(x) - \zeta(\eta)\} h(x)$, $\quad \eta \in \Xi$.

$s(\eta) = \dfrac{\partial \log \ell(\eta)}{\partial\eta} = T(x) - \dfrac{\partial \zeta(\eta)}{\partial\eta} = 0$

If $T(x)$ in range of $\dfrac{\partial\zeta(\eta)}{\partial\eta}$, it is unique MLE of $\mu(\eta) = \dfrac{\partial\zeta(\eta)}{\partial\eta}$.

MLE of η is $\hat{\eta} = \mu^{-1}(T(x))$

• GLM Structure

- $X_i \sim \exp\left\{\dfrac{\eta_i x_i - \zeta(\eta_i)}{\phi_i}\right\} h(x_i, \phi_i)$ $\quad i = 1, \dots, n$,

$\eta_i \in \Xi = \{\eta : 0 < \int h(x,\phi) e^{\eta x/\phi} d\nu(x) < \infty\}$,

$\zeta''(\eta) > 0 \ \forall \eta \in \Xi^\circ$

- $E(X_i) = \zeta'(\eta_i)$ and $Var(X_i) = \phi_i \zeta''(\eta_i)$

- $\mu(\eta) = \zeta'(\eta)$

- Assume g is C^3 and $g(\mu(\eta_i)) = \beta^T z_i$ $\quad g$ = link function, and $\dfrac{d}{d\eta}(g \circ \mu)(\eta) \ne 0 \ \forall \eta$

- $B = \{\beta : (g \circ \mu)^{-1}(\beta^T z) \in \Xi^\circ \ \forall z \in \mathcal{Z}\}$.

- ϕ_i are nuisance parameters.

• MLE in GLM

- Assume $\phi_i = \phi/t_i \ \forall i$, where t_i known.

- $\theta = (\beta, \phi)$, $\quad \psi = (g \circ \mu)^{-1}$ $\quad (\eta_i = \psi(\beta^T z_i))$.

- $\log \ell(\theta) = \sum_{i=1}^{n}\left[\log h(x_i, \phi/t_i) + \dfrac{\psi(\beta^T z_i)x_i - \zeta(\psi(\beta^T z_i))}{\phi/t_i}\right]$

- $\dfrac{\partial \log \ell(\theta)}{\partial\beta} = \dfrac{1}{\phi}\sum_{i=1}^{n}\{[x_i - \mu(\psi(\beta^T z_i))]\psi'(\beta^T z_i) t_i z_i\} = 0$

$\dfrac{\partial \log \ell(\theta)}{\partial\phi} = \sum_{i=1}^{n}\left\{\dfrac{\partial \log h(x_i, \phi/t_i)}{\partial\phi} - \dfrac{t_i[\psi(\beta^T z_i)x - \zeta(\psi(\beta^T z_i))]}{\phi^2}\right\}$

- can usually solve for β w/out ϕ. If a soln for β exists,

$Var\left(\dfrac{\partial \log \ell(\theta)}{\partial\beta}\right) = \dfrac{M_n(\beta)}{\phi}$, $\quad \dfrac{\partial^2 \log \ell(\theta)}{\partial\beta\partial\beta^T} = \dfrac{R_n(\beta) - M_n(\beta)}{\phi}$

$M_n(\beta) = \sum_{i=1}^{n}[\psi'(\beta^T z_i)]^2 \zeta''(\psi(\beta^T z_i)) t_i z_i z_i^T$

$R_n(\beta) = \sum_{i=1}^{n}[x_i - \mu(\psi(\beta^T z_i))]\psi''(\beta^T z_i) t_i z_i z_i^T$

- If g is canonical, $\psi'' \equiv 0$ and $R_n = 0$, easy to solve for β. $M_n(\beta)$ pos. def. $\Rightarrow \hat{\beta}$ is unique soln

- If g non-canonical, and if $[M_n(\beta)]^{-1/2} R_n(\beta) [M_n(\beta)]^{-1/2} \to 0$ then $\hat{\beta}$ is MLE for large n

• Computation of MLE in GLM:

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - [R_n(\hat{\beta}^{(t)}) - M_n(\hat{\beta}^{(t)})]^{-1} S_n(\hat{\beta}^{(t)})$,

$S_n(\beta^{(t)}) = \phi \dfrac{\partial \log \ell(\theta)}{\partial\beta}$ $\quad$ (NR)

- $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + [M_n(\hat{\beta}^{(t)})]^{-1} S_n(\hat{\beta}^{(t)})$

• Convergence Theorems

- CTS Mapping: for a cts g (a.e.)

$X_n \to X \Rightarrow g(X_n) \to g(X)$

for $\xrightarrow{a.s.}$, $\xrightarrow{P}$, $\xrightarrow{d}$.

- Slutsky:

$X_n \xrightarrow{d} X$ AND $Y_n \xrightarrow{P} c \in \mathbb{R}^k$.

Then (i) $X_n + Y_n \xrightarrow{d} X + c$

(ii) $X_n Y_n \xrightarrow{d} cX$

(iii) $X_n / Y_n \xrightarrow{d} X/c$ if $c \ne 0$.

• O, o, O_p, and o_p

- $X_n = O(Y_n) \iff \dfrac{X_n}{Y_n} \xrightarrow{n\to\infty} c \in \mathbb{R}$. $\quad$ ($a_0 n^2 - a_1 n + a_0$ is $O(n^2)$)

- $X_n = o(Y_n) \iff \dfrac{X_n}{Y_n} = o(1) \iff \dfrac{X_n}{Y_n} \to 0$

- $X_n = O_p(Y_n) \iff \forall \varepsilon > 0 \ \exists C > 0$ s.t. $\sup_n P(|X_n| \ge C|Y_n|) < \varepsilon$

- $X_n = o_p(Y_n) \iff \dfrac{X_n}{Y_n} \xrightarrow{P} 0$.

- $X_n = o_p(Y_n) \Rightarrow X_n = O_p(Y_n)$

- $X_n \xrightarrow{d} X \Rightarrow X_n = O_p(1)$

- $E|X_n| = O(a_n) \Rightarrow X_n = O_p(a_n)$

• CRLB + Fisher Information

- under some conditions (in Thm 4.16) if $E[T(x)] = g(\theta)$,

Then $Var(T(x)) = \left[\dfrac{\partial}{\partial\theta} g(\theta)\right]^T [I(\theta)]^{-1} \dfrac{\partial}{\partial\theta} g(\theta)$

where $I(\theta) = E\left[\dfrac{\partial}{\partial\theta}\log f_\theta(x)\left[\dfrac{\partial}{\partial\theta}\log f_\theta(x)\right]^T\right] = -E\left[\dfrac{\partial^2}{\partial\theta\partial\theta^T}\log f_\theta(x)\right]$

• WLLN

If $\exists p \in [1, 2]$ s.t. $\lim_{n\to\infty} \dfrac{1}{n^p}\sum_{i=1}^{n} E|X_i|^p = 0$, then

$\dfrac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \xrightarrow{P} 0$

• δ-method

$a_n(X_n - c) \xrightarrow{d} Y = N_k(0, \Sigma)$

$\Rightarrow a_n(g(X_n) - g(c)) \xrightarrow{d} N(0, \nabla g(c)^T \Sigma \nabla g(c))$

If $g'(c) = 0$, use cts mapping.

• $V_n^{-1/2}(\theta)(\hat{\theta} - \theta) \xrightarrow{d} N(0, I_k)$ $\quad \theta \in \mathbb{R}^k$

asymptotically efficient $\Rightarrow V_n(\theta) = I_n(\theta)^{-1}$

• e of A w.r.t. $B = \dfrac{\text{amse } B}{\text{amse } A}$

BOOTSTRAP

- Heuristic description
  - $\mathcal{P}$: population producing X
  - $\hat{\mathcal{P}}$: estimated population.
  - $\mathcal{P} \Rightarrow X \Rightarrow \hat{\theta} = \hat{\theta}(X)$
  - $\hat{\mathcal{P}} \Rightarrow X^* \Rightarrow \hat{\theta}^* = \hat{\theta}(X^*)$ ← bootstrap
  - $V_{ar}(\hat{\theta})$ approximated by $V_{ar}^*(\hat{\theta}^*)$
  - or by $\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \frac{1}{B}\sum_{l=1}^{B}\hat{\theta}^{*l}\right)^2$, $\hat{\theta}^{*b} = \hat{\theta}(X^{*b})$

- Parametric Bootstrap
  - $X_i \sim F_\theta$ iid, $\hat{\theta} = \hat{\theta}(X)$, $X = (X_1, \dots, X_n)$
  - $X^* = (X_1^*, \dots, X_n^*)$, $X_i^* \overset{iid}{\sim} F_{\hat{\theta}}$.

- e.g. $F_\theta(x) = F_0\left(\frac{x-\mu}{\sigma}\right)$ $\quad \theta = (\mu, \sigma^2)$
  - $T = \frac{\sqrt{n}(\bar{X}-\mu)}{S}$ parameter free.
  - $\hat{\theta} = (\bar{x}, s^2)$, $X_i^* \overset{iid}{\sim} F_{\hat{\theta}}$, $\frac{X^* - \bar{x}}{s} \sim F_0$
  - $T^* = \sqrt{n}\sum_{i=1}^{n}\frac{X_i^{*} - \bar{x}}{s} \sim T$
  - $V_{ar}^*(T^*) = V_{ar}(T)$.

- If $\exists\, \tau$ st. $V_{ar_\theta}(\hat{\theta}) = \tau(\theta)$, $X_i \overset{iid}{\sim} F_\theta$
  - then $V_{ar_{\hat{\theta}}}^*(\hat{\theta}^*) = \tau(\hat{\theta})$; $X_i^* \overset{iid}{\sim} F_{\hat{\theta}}$.
  - "substitution approach"

- Nonparametric bootstrap
  - X^* is SRS w/ replacement of size n from X.

- Properties of $V_{ar}^*(\hat{\theta}^*)$.
  - $\hat{\theta} = \bar{x}$, $\hat{\theta}^* = \bar{X}^*$.
  - $E^*(\bar{X}^*) = \frac{1}{n}\sum_{i=1}^{n}E^*(X_i^*) = \frac{1}{n}\sum_{i=1}^{n}\bar{x} = \bar{x}$
  - $V_{ar}^*(\bar{X}^*) = \frac{1}{n^2}\sum V_{ar}^*(X_i^*) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$
  - $= \frac{1}{n^2}\sum_{j=1}^{n}(x_j-\bar{x})^2 = \frac{n-1}{n^2}S^2 \approx S^2/n$.
  - $\hat{\theta} = g(\bar{x})$, $\hat{\theta}^* = g(\bar{X}^*)$
  - $g(\bar{X}^*) \approx g(\bar{x}) + g'(\bar{x})(\bar{X}^* - \bar{x})$
  - $\Rightarrow V_{ar}^*(\hat{\theta}^*) = V_{ar}^*[g(\bar{X}^*)] \approx [g'(\bar{x})]^2 V_{ar}^*(\bar{X}^* - \bar{x})$
  - $= [g'(\bar{x})]^2 V_{ar}^*(\bar{X}^*) \approx \frac{n-1}{n}\frac{[g'(\bar{x})]^2 S^2}{n}$.

- Sample Median
  - $\hat{Q}_{1/2} = \hat{F}_n^{-1}(\tfrac{1}{2})$, $\hat{F}_n$ is empirical distr.
  - $n = 2m-1$, $m \in \mathbb{N}$. $\Rightarrow \hat{Q}_{1/2} = X_{(m)}$.
  - $V_{ar}^*(X_{(m)}^*) = \sum_{k=1}^{n} p_k\left(X_{(k)} - \sum_{j=1}^{n} p_j(X_{(j)})\right)^2$, where
  - $p_k = P^*\{X_{(m)}^* = X_{(k)} \mid \text{data}\} = \sum_{j=0}^{m-1}\binom{n}{j}\frac{(k-1)^j(n-k+1)^{n-j} - k^j(n-k)^{n-j}}{n^n}$

CHAPTER 6: HYPOTHESIS TESTS

UMP TESTS AND NEYMAN-PEARSON LEMMA

- Power function
  - $\beta_T(P) = E[T(X)]$, $P \in \mathcal{P}$
  - $= P(\text{Type I error})$ if $P \in \mathcal{P}_0$
  - $= 1 - P(\text{Type II error})$ if $P \in \mathcal{P}_1$

- Significance Tests
  - Maximize $\beta_T(P)$ over $P \in \mathcal{P}_1$ subject to
  - $\underbrace{\sup_{P \in \mathcal{P}_0}\beta_T(P)}_{\text{size}} \leq \underbrace{\alpha}_{\text{level of significance}}$

- Definition 6.1
  - A test T_* of size α is UMP
  - $\iff \beta_{T_*}(P) \geq \beta_T(P)$ $\forall P \in \mathcal{P}_1$ and $\forall T$ of level α.

- If $U(X)$ is a sufficient statistic for $P \in \mathcal{P}$, then UMP tests need only be a function of U.

- Theorem 6.1
  - $\mathcal{P}_0 = \{P_0\}$, $\mathcal{P}_1 = \{P_1\}$,
  - (i) $\forall \alpha$ $\exists$ UMP test of size α:
  - $$T_*(X) = \begin{cases} 1 & f_1(x) > c f_0(x) \\ \gamma & f_1(x) = c f_0(x) \\ 0 & f_1(x) < c f_0(x) \end{cases}$$
  - for $\gamma \in [0,1]$, $c \geq 0$ st. $E[T_*(X)] = \alpha$ when $P = P_0$.
  - (ii) If T_{**} is a UMP test of size α, then
  - $$T_{**}(X) = \begin{cases} 1 & f_1(x) > c f_0(x) \\ 0 & f_1(x) < c f_0(x) \end{cases} \quad \text{a.s. } \mathcal{P}$$

- Lemma 6.1
  - Suppose $\exists\, T_*$ of size α st. $\forall P_1 \in \mathcal{P}_1$,
  - T_* is UMP for testing H_0 vs $H_1: P = P_1$
  - Then T_* is UMP for H_0 vs $H_1: P \in \mathcal{P}_1$.

MONOTONE LIKELIHOOD RATIO AND UMP TESTS

- Definition 6.2
  - Suppose $X \sim P_\theta$, $P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$.
  - Let $f_\theta = dP_\theta/d\nu$
  - $\mathcal{P}$ is said to have Monotone likelihood ratio in $Y(X)$ (real-valued statistic) $\iff \forall \theta_1 < \theta_2$,
  - $f_{\theta_2}(x)/f_{\theta_1}(x)$ is nondecreasing in $Y(x)$ for values at which at least one of $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ is positive.

- Lemma 6.3
  - Suppose X in $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $\theta \in \mathbb{R}$, and $\mathcal{P}$ has MLR in $Y(X)$.
  - If ψ is a nondecreasing function of Y, then $g(\theta) = E[\psi(Y)]$ is a nondecreasing function of θ.

- $\theta \in \mathbb{R}$, $\eta(\theta)$ nondecreasing function of θ $\Rightarrow$
  - $f_\theta(x) = \exp\{\eta(\theta)Y(x) - \xi(\theta)\}h(x)$
  - has MLR in $Y(X)$.

- Theorem 6.2
  - Suppose X in $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ that has MLR in $Y(X)$. $\quad H_0: \theta \leq \theta_0$, $H_1: \theta > \theta_0$
  - (i) $\exists$ UMP test of size α:
  - $$T_*(X) = \begin{cases} 1 & Y(X) > c \\ \gamma & Y(X) = c \\ 0 & Y(X) < c \end{cases}$$
  - where $\gamma \in [0,1]$, $c \geq 0$ st. $\beta_{T_*}(\theta_0) = E_{\theta_0}[T_*(X)] = \alpha$.
  - (ii) $\beta_{T_*}(\theta)$ is strictly increasing $\forall \theta$ st. $0 < \beta_{T_*}(\theta) < 1$.
  - (iii) $\forall \theta < \theta_0$, T_* minimizes $\beta_T(\theta)$ over all T st. $\beta_T(\theta_0) = \alpha$.
  - (iv) Assume $P_\theta(f_\theta(x) = c f_{\theta_0}(x)) = 0$ $\forall \theta > \theta_0$ and $c \geq 0$.
    - If T st. $\beta_T(\theta_0) = \beta_{T_*}(\theta_0)$, then
    - $\forall \theta > \theta_0$, either $\beta_T(\theta) < \beta_{T_*}(\theta)$ or $T = T_*$ a.s. P_θ.
  - (v) $\forall \theta_1$ fixed, T_* is UMP for testing $H_0: \theta \leq \theta_1$ vs $H_1: \theta > \theta_1$ with size $\beta_{T_*}(\theta_1)$.

- Corollary 6.1
  - Let X has p.d.f. in one-parameter exponential family with η strictly monotone in θ.
  - If η is increasing in θ, T_* in Theorem 6.2 is UMP for $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$, c, γ st. $\beta_{T_*}(\theta_0) = \alpha$
  - If η is decreasing in θ, the result holds for $H_0: \theta \geq \theta_0$ vs $H_1: \theta < \theta_0$.

UMP TESTS FOR TWO-SIDED HYPOTHESES AND UNBIASED TESTS

- Proposition 6.1 (Generalized N-P Lemma)
  - Let $f_1, \dots, f_{m+1}$ Borel in $\mathbb{R}^p$, integrable w.r.t. ν.
  - For $t_1, \dots, t_m \in \mathbb{R}$, let $\mathcal{J} = \{\phi : \mathbb{R}^p \to [0,1] \mid \int \phi f_i\, d\nu \leq t_i\ \forall i\}$
  - and $\mathcal{J}_0$ be $\mathcal{J}$ w/ equality.
  - If $\exists\, c_1, \dots, c_m \in \mathbb{R}$ st.
  - $$\phi_*(x) = \begin{cases} 1 & f_{m+1}(x) > c_1 f_1(x) + \dots + c_m f_m(x) \\ 0 & f_{m+1}(x) < c_1 f_1(x) + \dots + c_m f_m(x) \end{cases}$$
  - where $\phi_*(x) \in \mathcal{J}_0$, then ϕ_* maximizes $\int \phi f_{m+1}\, d\nu$ over $\phi \in \mathcal{J}_0$.
  - If $c_i \geq 0\ \forall i$, ϕ_* maximizes $\int \phi f_{m+1}\, d\nu$ over $\phi \in \mathcal{J}$.

- Lemma 6.2
  - Let $f_1, \dots, f_m$ and ν be given in Proposition 6.1.
  - Then $M = \{(\int \phi f_1\, d\nu, \dots, \int \phi f_m\, d\nu) : \phi : \mathbb{R}^p \to [0,1]\}$ is convex and closed.
  - If $(t_1, \dots, t_m) \in M^\circ_{\text{interior}}$, then $\exists\, c_1, \dots, c_m$ st. $\phi_* \in \mathcal{J}_0$.

- Two-sided hypotheses
  - $\begin{cases} H_0: \theta \leq \theta_1 \text{ or } \theta \geq \theta_2 \\ H_1: \theta_1 < \theta < \theta_2 \end{cases}$ (2)
  - $\begin{cases} H_0: \theta_1 \leq \theta \leq \theta_2 \\ H_1: \theta < \theta_1 \text{ or } \theta > \theta_2 \end{cases}$ (3)
  - $\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$ (4)

- Theorem 6.3 (UMP tests for two-sided hypotheses)
  - Suppose X has p.d.f. in one-parameter exponential family:
  - $f_\theta(x) = \exp\{\eta(\theta)Y(x) - \xi(\theta)\}h(x)$
  - w.r.t. σ-finite measure. η strictly increasing function of θ.
  - (i) For testing (2), a UMP test of size α is:
  - $$T_*(X) = \begin{cases} 1 & c_1 < Y(X) < c_2 \\ \gamma_i & Y(X) = c_i,\ i = 1,2 \\ 0 & Y(X) < c_1 \text{ or } Y(X) > c_2 \end{cases} \quad (5)$$
  - where c_i's and γ_i's determined by
  - $\beta_{T_*}(\theta_1) = \beta_{T_*}(\theta_2) = \alpha$. (6)
  - (ii) T_* minimizes $\beta_T(\theta)$ over all $\theta < \theta_1$, $\theta > \theta_2$, and T satisfying (6).
  - (iii) If T_*, T_{**} both satisfy (5), and $\beta_{T_*}(\theta_1) = \beta_{T_{**}}(\theta_1)$, and if the rejection region $\{T_{**} = 1\}$ is to the right of $\{T_* = 1\}$, then $\beta_{T_*}(\theta) < \beta_{T_{**}}(\theta)$ for $\theta > \theta_1$ and $\beta_{T_*}(\theta) > \beta_{T_{**}}(\theta)$ for $\theta < \theta_1$.
    - If both T_* and T_{**} satisfy (5)+(6), $T_* = T_{**}$ a.s. $\mathcal{P}$.

- Definition 6.3
  - Let α be the given level of significance.
  - A test T for $H_0: P \in \mathcal{P}_0$ vs $H_1: P \in \mathcal{P}_1$ is unbiased at level α $\iff$ $\beta_T(P) \leq \alpha$ $\quad P \in \mathcal{P}_0$; $\beta_T(P) \geq \alpha$ $\quad P \in \mathcal{P}_1$
  - A test is UMPU $\iff$ it is UMP w/in unbiased tests of level α

- Definition 6.1
  - $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$, level of sig α.
  - Let $\bar{\Theta}_{01}$ be the common boundary of Θ_0, Θ_1.
  - A test T is similar on $\bar{\Theta}_{01}$ $\iff \beta_T(P_\theta) = \alpha$ $\forall \theta \in \bar{\Theta}_{01}$

UMPU TESTS IN EXPONENTIAL FAMILIES

- Continuity of power function
  - $\beta_T(P)$ is cts in θ $\iff \forall \{\theta_j\} \subset \Theta$, $\theta_j \to \theta_0$
  - $\Rightarrow \beta_T(P_j) \to \beta_T(P_0)$.

- Lemma 6.5
  - $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$.
  - Suppose $\forall T$, $\beta_T(P)$ is cts in θ.
  - If T_* is UMP among similar tests, and has size α,
  - then T_* is a UMPU test

- **Neyman Structure**

Let $U(x)$ be sufficient statistic for $P \in \bar{\mathcal{P}}$, let $\bar{\mathcal{P}}_U$ be family of distns of U as P ranges over $\bar{\mathcal{P}}$.
A test has Neyman Structure wrt U if
$$E[T(X)\mid U] = \alpha \quad \text{a.s. } \bar{\mathcal{P}}_U$$
Then $E[T(X)] = E[E[T(X)\mid U]] = \alpha \quad P \in \bar{\mathcal{P}}$,
so T is similar on $\bar{\Theta}_{01}$.

- **Lemma 6.6**

Let $U(x)$ be sufficient for $P \in \bar{\mathcal{P}}$.
All tests similar on $\bar{\Theta}_{01}$ have Neyman structure wrt U
$\iff$ U is boundedly complete for $P \in \bar{\mathcal{P}}$.

- **Theorem 6.4** (UMPU tests in Multiparameter exp. fam.)

Suppose that X has the following pdf wrt σ-finite measure
$$f_{\theta,\varphi}(x) = \exp\{\theta Y(x) + \varphi^T U(x) - \zeta(\theta,\varphi)\}$$
where θ is a real-valued param, φ is vector-valued and Y, U are statistics.

(i) $H_0: \theta \le \theta_0$ vs $H_1: \theta > \theta_0$, UMPU test of size α is
$$T_*(Y,U) = \begin{cases} 1 & Y > c(U) \\ \gamma(U) & Y = c(U) \\ 0 & Y < c(U) \end{cases}$$
$c(u), \gamma(u)$ Borel st $E_{\theta_0}[T_*(Y,U) \mid U=u] = \alpha \quad \forall u$.

(ii) $H_0: \theta \le \theta_1$ or $\theta \ge \theta_2$ vs. $H_1: \theta_1 < \theta < \theta_2$
UMPU test of size α is:
$$T_*(Y,U) = \begin{cases} 1 & c_1(u) < Y < c_2(u) \\ \gamma_i(u) & Y = c_i(u),\ i=1,2 \\ 0 & Y < c_1(u) \text{ or } Y > c_2(u) \end{cases}$$
$c_i(u)$'s, $\gamma_i(u)$'s Borel funcs st
$$E_{\theta_1}[T_*(Y,U) \mid U=u] = E_{\theta_2}[T_*(Y,U) \mid U=u] = \alpha \quad \forall u.$$

(iii) $H_0: \theta_1 \le \theta \le \theta_2$ vs $H_1: \theta < \theta_1$ or $\theta > \theta_2$ UMPU$_\alpha$
$$T_*(Y,U) = \begin{cases} 1 & Y > c_1(u) \text{ or } Y < c_2(u) \\ \gamma_i(u) & Y = c_i(u),\ i=1,2 \\ 0 & c_1(u) < Y < c_2(u) \end{cases}$$
st $E_{\theta_1}[T_*(Y,U) \mid U=u] = E_{\theta_2}[T_*(Y,U) \mid U=u] = \alpha \quad \forall u$.

(iv) $H_0: \theta = \theta_0$ vs $H_1: \theta \ne \theta_0$, UMPU$_\alpha$ is (iii)
st $E_{\theta_0}[T_*(Y,U) \mid U=u] = \alpha \quad \forall u$
$E_{\theta_0}[T_*(Y,U)\, Y \mid U=u] = \alpha E_{\theta_0}(Y \mid U=u) \quad \forall u$

- **Lemma 6.7**

$X \sim f_{\theta,\varphi}(x) = \exp\{\theta Y(x) + \varphi^T U(x) - \zeta(\theta,\varphi)\}$.
Suppose $V(Y,U) \perp\!\!\!\perp U$ when $\theta = \theta_j$ (θ_j given in Thm 6.4)

(i) If $V(y,u)$ is increasing in y for each u, then UMPU tests in (i)–(iii) of Thm 6.4 are equivalent to those but replace $Y, (Y,U)$ by V, $c_i(u), \gamma_i(u)$ by c_i, γ_i.

(ii) If $\exists\, a(u) > 0$, $b(u)$ Borel st. $V(y,u) = a(u)y + b(u)$, then UMPU test in Thm 6.4 (iv) equivalent to replace $Y, (Y,U)$ by V and $c_i(u) = c_i$, $\gamma_i(u) = \gamma_i$.

- If c.d.f. of V is cts, or c.d.f. of $Y \mid U=u$ cts, don't need γ_i's.

LIKELIHOOD RATIO TESTS

- **Definition 6.2**

Let $\ell(\theta) = f_\theta(x)$ be the likelihood function.
$H_0: \theta \in \Theta_0$, $H_1: \theta \in \Theta_1$. A likelihood ratio (LR) test is any test that rejects $H_0 \iff \lambda(x) < c$, where $c \in [0,1]$ and
$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} \ell(\theta)}{\sup_{\theta \in \Theta} \ell(\theta)}$$

- If there is a sufficient statistic, $\lambda(x)$ depends only on sufficient statistic.
- $\hat\theta$ MLE of θ, $\hat\theta_0$ MLE of θ under $\theta \in \Theta_0$, then
$$\lambda(x) = \ell(\hat\theta_0)/\ell(\hat\theta).$$

- If $\exists\, c_\alpha \in [0,1]$ st. $\sup_{\theta \in \Theta_0} P_\theta(\lambda(X) < c_\alpha) = \alpha$, a LR test of size α can be obtained.

- **Proposition 6.5**

$X \sim f_\theta(x) = \exp\{\eta(\theta) Y(x) - \xi(\theta)\} h(x)$
where η is a strictly increasing function and differentiable function of θ

(i) $H_0: \theta \le \theta_0$ vs $H_1: \theta > \theta_0$. $\exists$ LR test with RR same as UMP test in Thm 6.2

(ii) $H_0: \theta \le \theta_1$ or $\theta \ge \theta_2$ vs $H_1: \theta_1 < \theta < \theta_2$. $\exists$ LR test with RR same as UMP test in Thm 6.3

(iii) Other two sided hypotheses: $\exists$ LR test with RR equivalent to $Y(x) < c_1$ or $Y(x) > c_2$ for $c_1, c_2 \in \mathbb{R}$.
[May not have same size].

- **Definition 2.13** (asymptotic tests)

Let $X = (X_1, \dots, X_n)$ sample from $P \in \mathcal{P}$ and $T_n(x)$ be test for $H_0: P \in \mathcal{P}_0$ vs $H_1: P \in \mathcal{P}_1$.

(i) If $\limsup_n \alpha_{T_n}(P) \le \alpha \quad \forall P \in \mathcal{P}_0$, then α is asymptotic significance level of T_n

(ii) If $\lim_{n\to\infty} \sup_{P \in \mathcal{P}_0} \alpha_{T_n}(P)$ exists, it is called the limiting size of T_n

(iii) T_n is consistent $\iff$ Type II error prob $\to 0$.

- **Theorem 6.5**

Assume the conditions in Thm 4.16
$H_0: \theta = g(\vartheta)$, $\vartheta \in \mathbb{R}^{k-r}$ unknown, $g: \mathbb{R}^{k-r} \to \mathbb{R}^k$, $\mathcal{C}^1$; with full rank $\frac{\partial g(\vartheta)}{\partial \vartheta}$.
Under H_0: $-2\log\lambda_n \xrightarrow{d} \chi^2_r$
LR test w/ RR $\lambda_n < e^{-\chi^2_{r,\alpha}/2}$ has asymptotic signif level α.

ASYMPTOTIC TESTS

- $H_0: \theta = g(\vartheta)$, $\vartheta \in \mathbb{R}^r$ $\iff$ $H_0: R(\theta) = 0$ for $R: \mathbb{R}^k \to \mathbb{R}^r$, $R \in \mathcal{C}^1$.

- Wald test, Reject H_0 when
$$W_n = [R(\hat\theta)]^T \left(C(\hat\theta)^T I_n(\hat\theta)^{-1} C(\hat\theta)\right)^{-1} [R(\hat\theta)]$$
is large, $C(\theta) = \frac{\partial R(\theta)}{\partial\theta}$, $I_n(\theta) = -E\left[\frac{\partial^2 \log \ell(\theta)}{\partial\theta\,\partial\theta^T}\right]$
$\hat\theta$ is MLE or RLE of θ
If $H_0: \theta = \theta_0$, $W_n = (\hat\theta - \theta_0)^T I_n(\hat\theta)(\hat\theta - \theta_0)$.

- Rao test: reject H_0 when
$$R_n = [S_n(\tilde\theta)]^T [I_n(\tilde\theta)]^{-1} S_n(\tilde\theta)$$
is large. $S_n(\theta) = \frac{\partial \log \ell(\theta)}{\partial\theta}$, $\tilde\theta$ is MLE or RLE of θ under H_0

- **Theorem 6.6**

Assume the conditions of Theorem 4.16,

(i) Under $H_0: R(\theta) = 0$, $W_n \xrightarrow{d} \chi^2_r$.
Wald test rejects $H_0 \iff W_n > \chi^2_{r,\alpha}$ has asymptotic significance level α.

(ii) (i) holds when W_n is replaced by R_n.

- **Theorem 6.12**

Assume that $V_n^{-1/2}(\hat\theta_n - \theta) \xrightarrow{d} N_k(0, I_k) \quad \forall P$
Also assume $\lambda_+[V_n] \to 0$.

(i) The test with rejection region
$$(\hat\theta_n - \theta_0)^T V_n^{-1} (\hat\theta_n - \theta_0) > \chi^2_{k,\alpha}$$
is consistent (also if $\hat V_n^{-1}$ is consistent ($\hat V_n^{-1} V_n \to 1$))

- **Multinomial distributions**

$X = (X_1, \dots, X_k) \sim \text{Multinomial}(p)$, $p = (p_1, \dots, p_k)$.
$$Z_n(p) = \sqrt{n}\left(\tfrac{X}{n} - p\right) \xrightarrow{d} N_k(0, \Sigma), \quad \Sigma = \mathrm{Var}\left(\tfrac{X}{\sqrt{n}}\right)$$
$\Sigma_{ii} = p_i(1-p_i)$, $\Sigma_{ij} = -p_i p_j$, $i \ne j$.
$H_0: p = p_0$ vs $H_1: p \ne p_0$.
$$\chi^2 = \sum_{j=1}^k \frac{(X_j - np_{0j})^2}{np_{0j}} = \|D(p_0) Z_n(p_0)\|^2$$
$D(c)$ is block diag with j^{th} element $c_j^{-1/2}$.
$$\chi^2 = \sum_{j=1}^k \frac{(X_j - np_{0j})^2}{X_j} = \|D(X/n) Z_n(p_0)\|^2$$
asymptotic sig level α, reject if $\chi^2 > \chi^2_{k-1,\alpha}$.

- **Theorem 6.8**

Let $\Phi = (\sqrt{p_1}, \dots, \sqrt{p_k})$, Λ be $k \times k$ projection matrix.
(i) If $\Lambda\Phi = a\Phi$, then
$$Z_n(p)^T D(p) \Lambda D(p) Z_n(p) \xrightarrow{d} \chi^2_r$$
$r = \mathrm{tr}(\Lambda) - a$.
(ii) The result holds if $D(p) \to D(X/n)$.

- **Sketch of Thm 6.5**

$$\lambda_n = \ell(g(\hat\vartheta))/\ell(\hat\theta).$$
$$\sqrt{n}\, I_1(\theta)(\hat\theta - \theta) = n^{-1/2} S_n(\theta) + o_p(1).$$
$$2[\log \ell(\hat\theta) - \log \ell(\theta)] = n(\hat\theta - \theta)^T I_1(\theta)(\hat\theta - \theta) + o_p(1)$$
$$= n^{-1}[S_n(\theta)]^T [I_1(\theta)]^{-1} S_n(\theta) + o_p(1)$$
Under H_0:
$$2[\log \ell(g(\hat\vartheta)) - \log \ell(g(\vartheta))] = n^{-1} \tilde S_n(\vartheta)^T \tilde I_1(\vartheta)^{-1} \tilde S_n(\vartheta) + o_p(1)$$
$$\tilde S_n(\vartheta) = \frac{\partial \log \ell(g(\vartheta))}{\partial\vartheta} = D(\vartheta) S_n(g(\vartheta)), \quad D(\vartheta) = \frac{\partial g(\vartheta)}{\partial\vartheta}.$$
$$\Rightarrow -2\log\lambda_n = n^{-1} S_n(g(\vartheta))^T B(\vartheta) S_n(g(\vartheta)) + o_p(1)$$
$$B(\vartheta) = [I_1(g(\vartheta))]^{-1} - D(\vartheta)^T \tilde I_1(\vartheta)^{-1} D(\vartheta)$$
By CLT: $n^{-1/2}[I_1(\theta)]^{-1/2} S_n(\theta) \xrightarrow{d} Z = N_k(0, I_k)$
$$Z^T [I_1(g(\vartheta))]^{1/2} B(\vartheta) [I_1(g(\vartheta))]^{1/2} Z = \chi^2_r.$$
$-2\log\lambda_n \to$ [illegible].

- **Order statistics**

$X_{(n)}$: p.d.f. $n f(x) F(x)^{n-1}$
c.d.f. $P(X_{(n)} < x) = [F(x)]^n$
$X_{(1)}$: p.d.f. $n f(x)(1 - F(x))^{n-1}$
c.d.f. $P(X_{(1)} < x) = 1 - (1 - F(x))^n$

- **Gamma / Chi square**

- $X \sim \Gamma(\alpha, \gamma)$ $\quad f_X(x) = \frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-x/\gamma} I_{(0,\infty)}(x)$
$\Rightarrow Y = cX \sim \Gamma(\alpha, c\gamma)$
- $X \sim \Gamma\left(\frac{\nu}{2}, 2\right) \Rightarrow X \sim \chi_\nu$
- mgf of $E(a, \theta)$ is $e^{at}(1 - \theta t)^{-1} = \phi(t)$
- mgf of $\Gamma(\alpha, \gamma)$ is $(1 - \gamma t)^{-\alpha} = \phi(t)$
- $\phi_{a_1X_1 + \dots + a_nX_n}(t) = \phi_{X_1}(a_1 t) \cdots \phi_{X_n}(a_n t)$

- **Wald / Rao**

- Thm 1.12 + Thm 4.17 $\Rightarrow \sqrt{n}(R(\hat\theta) - R(\theta)) \xrightarrow{d} N_r(0, C(\theta)^T I_1(\theta)^{-1} C(\theta))$.
$H_0 \Rightarrow R(\theta) = 0 \Rightarrow n R(\hat\theta)^T [C(\theta)^T I_1(\theta)^{-1} C(\theta)]^{-1} R(\hat\theta) \xrightarrow{d} \chi^2_r$.
Use Slutsky's on $[C(\theta)^T I_1(\theta)^{-1} C(\theta)]^{-1}$ since $\hat\theta \xrightarrow{P} \theta$, $I(\theta)$, $C(\theta)$ cts
- Lagrange mult $\Rightarrow S_n(\tilde\theta) + C(\tilde\theta)\lambda_n = 0$ and $R(\tilde\theta) = 0$.
Taylor's expansion $\Rightarrow C(\theta)^T(\tilde\theta - \theta) = o_p(n^{-1/2})$
$S_n(\theta) - I_n(\theta)(\tilde\theta - \theta) + C(\theta)\lambda_n = o_p(n^{1/2})$
$I_n(\theta) = n I_1(\theta)$.
Then $C(\theta)^T I_n(\theta)^{-1} C(\theta)\lambda_n = -C(\theta)^T I_n(\theta)^{-1} S_n(\theta) + o_p(n^{-1/2})$
$\Rightarrow \lambda_n^T C(\theta)^T I_n(\theta)^{-1} C(\theta)\lambda_n \xrightarrow{d} \chi^2_r$.
But $C(\tilde\theta)\lambda = -S_n(\tilde\theta)$, $I_n(\theta) = n I_1(\theta)$, so done.

CHAPTER 7: CONFIDENCE SETS

PIVOTAL QUANTITIES AND CONFIDENCE SETS

- Confidence sets
  X: a sample from population $P \in \mathcal{P}$
  $\theta = \theta(P)$, a functional from $\mathcal{P}$ to $\Theta \subset \mathbb{R}^k$
  $C(X)$ a confidence set for θ, $C(X) \in \mathcal{B}_\Theta$.
  $\inf_{P\in\mathcal{P}} P(\theta \in C(X))$: confidence coefficient of $C(X)$
  if $\inf_{P\in\mathcal{P}} P(\theta \in C(X)) \geq 1-\alpha$, $C(X)$ has confidence level $1-\alpha$.

- Definition 7.1
  A known Borel function R of (X, θ) is a pivotal quantity $\iff$ distn of $R(X,\theta)$ doesn't depend on P.

- CS from Pivotal Quantity
  1) Find c_1, c_2 st. $P(c_1 \leq R(X,\theta) \leq c_2) \geq 1-\alpha$
  2) $C(x) = \{\theta \in \Theta : c_1 \leq R(X,\theta) \leq c_2\}$
- If $R(x,\theta)$ has cts c.d.f., $C(X)$ has conf. coef. $1-\alpha$.
  3) Invert for θ: if $R(X,\theta)$ monotonic in θ, then $C(X) = [\underline{\theta}(X), \bar{\theta}(X)]$. Also can be union of intervals.

- Proposition 7.1 (Existence of pivotal quantities)
  Let $T(X) = (T_1(X), \dots, T_s(X))$ and $T_1, \dots, T_s$ independent statistics. Suppose each T_i has a continuous c.d.f. $F_{T_i,\theta}$ indexed by θ. Then $R(X,\theta) = \prod_{i=1}^{s} F_{T_i,\theta}(T_i(X))$ is pivotal.

- Theorem 7.1
  Suppose $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $T(X)$ real-valued statistic with c.d.f. $F_{T,\theta}(t)$, $\alpha_1, \alpha_2 > 0$ st. $\alpha_1 + \alpha_2 < \frac{1}{2}$
  (i) Suppose that $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are nonincreasing in θ, $\forall t$. Define $\bar{\theta} = \sup\{\theta : F_{T,\theta}(T) \geq \alpha_1\}$ AND $\underline{\theta} = \inf\{\theta : F_{T,\theta}(T-) \leq 1-\alpha_2\}$
  Then, $[\underline{\theta}(T), \bar{\theta}(T)]$ is a level $1-\alpha$ confidence interval for θ.
  (ii) If $F_{T,\theta}(t)$ and $F_{T,\theta}(t-)$ are nondecreasing, the result holds with
  $\bar{\theta} = \inf\{\theta : F_{T,\theta}(T) \geq \alpha_1\}$, $\underline{\theta} = \sup\{\theta : F_{T,\theta}(T-) \leq 1-\alpha_2\}$
  (iii) If $F_{T,\theta}$ is a continuous c.d.f. $\forall\theta$, then $F_{T,\theta}(T)$ is a pivotal quantity and (i)(ii) CI actually have confidence coefficient $1-\alpha$.

- When $\mathcal{P}$ in Thm 7.1 has MLR in $T(X)$, Lemma 6.3 $\Rightarrow$ conditions of Thm 7.1 (i) hold.
- If $F_{T,\theta}(t)$ cts in θ, $\lim_{\theta \to \theta_-} F_{T,\theta}(t) > \alpha$, $\lim_{\theta \to \theta_+} F_{T,\theta}(t) < \alpha$, then $\hat{\theta}$ is soln of $F_{T,\theta}(t) = \alpha$.

INVERTING ACCEPTANCE REGIONS OF TESTS

- For a test T, the set $\{x : T(x) \neq 1\}$ is called the acceptance region (careful when T is randomized).

- Theorem 7.2
  For each $\theta_0 \in \Theta$, let T_{θ_0} be a test for $H_0: \theta = \theta_0$ with significance level α and acceptance region $A(\theta_0)$.
  For each x in the range of X, define
  $$C(x) = \{\theta : x \in A(\theta)\}.$$
  Then $C(X)$ is a level $1-\alpha$ confidence set for θ.
  If T_{θ_0} is nonrandomized and of size α $\forall \theta_0$, then $C(X)$ has confidence coeff $1-\alpha$.

- Proposition 7.2
  Let $C(X)$ be a confidence set for θ with confidence level $1-\alpha$.
  $\forall \theta_0 \in \Theta$, define $A(\theta_0) = \{x : \theta_0 \in C(x)\}$.
  Then the test $T(X) = 1 - I_{A(\theta_0)}(X)$ has significance level α for testing $H_0: \theta = \theta_0$ vs. some H_1.

- Exp Family: $f_\theta(x) = \exp\{\eta(\theta) Y(x) - \xi(\theta)\} h(x)$, $\theta \in \mathbb{R}$, $\eta(\theta)$ nondecreasing in θ.
  Thm 2.2 ($H_0: \theta = \theta_0$, $H_1: \theta > \theta_0$)
  $A(\theta_0) = \{x : Y(x) \leq c(\theta_0)\}$, $c(\theta)$ nondecreasing in θ
  $C(X) = [\underline{\theta}(X), \infty)$, $\underline{\theta}$ is lower conf. bound

LENGTHS OF CONFIDENCE INTERVALS

- Sometimes can compare rates of convergence to 0.
- Theorem 7.3
  Let θ be real-valued param, $T(X)$ - real-valued statistic.
  (i) Let $U(X)$ be positive statistic.
  $\frac{T-\theta}{U}$ pivotal with Leb pdf f unimodal at $x_0 \in \mathbb{R}$
  $$\mathcal{C} = \left\{[T-bU, T-aU] : a \in \mathbb{R}, b \in \mathbb{R}, \int_a^b f(x)dx = 1-\alpha\right\}$$
  If $[T-b_*U, T-a_*U] \in \mathcal{C}$, $f(a_*) = f(b_*) > 0$, and $a_* \leq x_0 \leq b_*$, then the interval has shortest length within $\mathcal{C}$.
  (ii) $T > 0$, $\theta > 0$, T/θ pivotal with Leb pdf f.
  $x^2 f(x)$ unimodal at x_0.
  $$\mathcal{C} = \left\{[b^{-1}T, a^{-1}T] : a > b > 0, \int_a^b f(x)dx = 1-\alpha\right\}$$
  If $[b_*^{-1}T, a_*^{-1}T] \in \mathcal{C}$, $a_*^2 f(a_*) = b_*^2 f(b_*) > 0$ and $a_* \leq x_0 \leq b_*$, the interval has shortest length in $\mathcal{C}$.

- Note, when pdf is strictly increasing or decreasing (not unimodal), use endpoint as CI endpoint.

UMA AND UMAU CONFIDENCE SETS

- Definition 7.2
  Let $\theta \in \Theta$ be an unknown parameter and Θ' be a subset of Θ that does not contain the true parameter value of θ.
  A confidence set $C(X)$ for θ with confidence coef $1-\alpha$ is Θ'-Uniformly most accurate (UMA) $\iff$ $\forall C_1(X)$ with conf level $1-\alpha$, $P(\theta' \in C(X)) \leq P(\theta' \in C_1(X))$ $\forall \theta' \in \Theta'$.
  UMA means $\{\theta\}^c$-UMA.

- Seek lower conf. bound, $\Theta' = \{\underline{\theta} \in \Theta : \underline{\theta} < \theta\}$

- Theorem 7.4
  Let $C(X)$ be a confidence set for θ obtained by inverting the acceptance regions of nonrandomized tests T_{θ_0} testing $H_0: \theta = \theta_0$ vs $H_1: \theta \in \Theta_{\theta_0}$.
  Suppose $\forall \theta_0$, T_{θ_0} is UMP of size α.
  Then, $C(X)$ is Θ'-UMA, conf coef $1-\alpha$, when $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$.

  - Combine w/ Thm 6.2 $\Rightarrow$ UMA conf bounds for one param exponential family w/ MLR.

- Definition 7.3
  $\theta \in \Theta$ unknown, $\Theta' \subset \Theta \setminus \{\theta\}$
  (i) A level $1-\alpha$ confidence set $C(X)$ is Θ'-unbiased $\iff$ $P(\theta' \in C(X)) \leq 1-\alpha$ $\forall \theta' \in \Theta'$.
  (ii) Let $C(X)$ be a Θ'-unbiased confidence set with conf coef $1-\alpha$.
  If $P(\theta' \in C(X)) \leq P(\theta' \in C_1(X))$ $\forall \theta' \in \Theta'$ holds $\forall C_1(X)$ that is Θ'-unbiased,
  Then $C(X)$ is Θ'-uniformly most accurate unbiased (UMAU).

- Theorem 7.5
  Let $C(X)$ be conf set for θ obtained by inverting acceptance regions of nonrandomized tests T_{θ_0} for $H_0: \theta = \theta_0$ vs $H_1: \theta \in \Theta_{\theta_0}$.
  If T_{θ_0} is unbiased of size α $\forall \theta_0$,
  Then $C(X)$ is Θ'-unbiased w/ conf coef $1-\alpha$
  $\Theta' = \{\theta' : \theta \in \Theta_{\theta'}\}$.
  If T_{θ_0} is also UMPU $\forall \theta_0$, $C(X)$ is Θ'-UMAU.

- Volume
  $$\operatorname{vol}(C(X)) = \int_{C(X)} d\theta' \quad \text{for } C(X) \text{ a Borel set.}$$
- Theorem 7.6 (Pratt's theorem)
  Suppose $\operatorname{vol}(C(X)) = \int_{C(X)} d\theta'$ is finite a.s. $\mathcal{P}$.
  Then the expected volume of $C(X)$ is
  $$E[\operatorname{vol}(C(X))] = \int_{\theta \neq \theta'} P(\theta' \in C(X))\, d\theta'$$
  ($C(X)$ UMA $\Rightarrow$ smallest expected vol among all sets.)

ASYMPTOTIC CONFIDENCE SETS AND LIKELIHOODS

- Asymptotic Criterion
  - Asymptotic level $1-\alpha$ if $\liminf_n P(\theta \in C(X)) \geq 1-\alpha$ $\forall P \in \mathcal{P}$
  - If $\lim_{n\to\infty} P(\theta \in C(X)) = 1-\alpha$ $\forall P \in \mathcal{P}$, $C(X)$ is asymptotically correct (not same as limiting conf coef $1-\alpha$).

- Asymptotic pivotal quantity
  - limiting distribution of $R(X,\theta)$ does not depend on P.
  - Usually $\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta)$ when $V_n^{-1/2}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I_k)$ and $\hat{V}_n$ consistent for V_n.
  Then $C(X) = \{\theta : \|\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta)\|^2 \leq \chi^2_{k,\alpha}\}$.

- Example: Functions of means
  $\theta = g(\mu)$, $g: \mathbb{R}^d \to \mathbb{R}^k$, $k \leq d$.
  $V_n = \nabla g(\mu)^T \operatorname{Var}(X_1) \nabla g(\mu) / n$.
  $\hat{V}_n = \nabla g(\bar{X})^T S^2 \nabla g(\bar{X}) / n$.
  $C(X) = \{\theta : \|\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta)\|^2 \leq \chi^2_{k,\alpha}\}$.

- Proposition 7.4
  Let $C_j(X) = \{\theta : \|\hat{V}_{jn}^{-1/2}(\hat{\theta}_{jn} - \theta)\|^2 \leq \chi^2_{k,\alpha}\}$, $j = 1,2$
  satisfying $V_{jn}^{-1/2}(\hat{\theta}_{jn} - \theta) \xrightarrow{d} N_k(0, I_k)$,
  $\hat{V}_{jn}$ consistent for V_{jn}, $j = 1,2$.
  If $\operatorname{Det}(V_{1n}) < \operatorname{Det}(V_{2n})$ (for n sufficiently large) then
  Then $P(\operatorname{vol}(C_1(X)) < \operatorname{vol}(C_2(X))) \to 1$.
  - If $\hat{\theta}_{1n}$ is asymptotically more efficient than $\hat{\theta}_{2n}$, $\operatorname{Det}(V_{1n}) \leq \operatorname{Det}(V_{2n})$.

- Parametric likelihoods
  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$. $\theta = (\vartheta, \varphi)$
  Consider confidence sets for $\vartheta \in \mathbb{R}^r$.
  LR Test:
  $\Theta_0 = \{\theta : \vartheta = \vartheta_0\}$. (constraint under H_0)
  $A(\vartheta_0) = \{x : \ell(\vartheta_0, \hat{\varphi}_{\vartheta_0}) \geq e^{-c_\alpha/2} \ell(\hat{\theta})\}$
  $\hat{\theta}$ is MLE of $\theta \in \Theta$. $\ell(\vartheta_0, \hat{\varphi}_{\vartheta_0}) = \sup_\varphi \ell(\vartheta_0, \varphi)$.
  Thm 6.5 $\Rightarrow$ $c_\alpha = \chi^2_{r,\alpha}$
  $\Rightarrow C(X) = \{\vartheta : \ell(\vartheta, \hat{\varphi}_\vartheta) \geq e^{-c_\alpha/2} \ell(\hat{\theta})\}$ asymptotically correct ($1-\alpha$ level)
  [generally different than $\{\vartheta : \|\hat{V}_n^{-1/2}(\hat{\vartheta}_n - \vartheta)\|^2 \leq \chi^2_{r,\alpha}\}$.
  If $-\ell(\vartheta, \varphi)$ is convex in ϑ, $C(X)$ is bounded set.
  Wald test:
  $A(\vartheta_0) = \{x : (\hat{\vartheta} - \vartheta_0)^T [C^T I_n(\hat{\theta})^{-1} C]^{-1} (\hat{\vartheta} - \vartheta_0) \leq \chi^2_{r,\alpha}\}$
  $\hat{\theta} = (\hat{\vartheta}, \hat{\varphi})$ is MLE or RLE, $I_n(\theta)$ is Fisher info,
  $C^T = (I_r, 0)$, 0 is $r \times (k-r)$ matrix of 0's.
  Thm 4.17 $\Rightarrow$ $C(X) = \{\vartheta : \|\hat{V}_n^{-1/2}(\hat{\vartheta} - \vartheta)\|^2 \leq \chi^2_{r,\alpha}\}$
  $\hat{V}_n = C^T [I_n(\hat{\theta})]^{-1} C$.

8

Rao's Score test:

$A(\vartheta_0) = \{x : [S_n(\vartheta_0, \hat{\varphi}_{\vartheta_0})]^T [I_n(\vartheta_0, \hat{\varphi}_{\vartheta_0})]^{-1} [S_n(\ )] \le \chi^2_{\alpha, p}\}$

$S_n(\theta) = \partial \log \ell(\theta) / \partial\theta$.

$C(x)$ is asymptotically correct, but not easy to invert.

- Note in Wald's test, $C^T(\cdot)C$ takes upper sub matrix of $(\cdot)$.

$$\min_{a,b} \left\{ \frac{1}{a} - \frac{1}{b} \ : \ \int_a^b f(x)dx = 1-\alpha \right\}$$

$$0 = \frac{\partial\left(\frac{1}{a} - \frac{1}{b}\right)}{\partial a} = -\frac{1}{a^2} + \frac{1}{b^2}\frac{db}{da}$$

$$\frac{\partial}{\partial a}\left[\int_a^b f(x)dx\right] = \frac{\partial}{\partial a}\left[F(b) - F(a)\right] = 0$$

$$\Rightarrow f(b)\frac{db}{da} - f(a) = 0.$$

$$\Rightarrow \frac{db}{da} = \frac{f(a)}{f(b)}$$

Then min at $0 = -\frac{1}{a^2} + \frac{1}{b^2}\frac{f(a)}{f(b)}$

$\Leftrightarrow a^2 f(a) = b^2 f(b)$

Common Information Matrices

- $\frac{1}{2}(1-\theta^2)e^{\theta x - |x|}$ $\quad I_1(\theta) = \frac{1+\theta^2}{(1-\theta^2)^2}$
- BVN

$$I_n(\theta) = \mathrm{diag}\left(\frac{n_1}{\sigma_1^2}, \frac{n_1}{2\sigma_1^4}, \frac{n_2}{\sigma_2^2}, \frac{n_2}{2\sigma_2^4}\right)$$

If $\sigma_1^2 = \sigma_2^2$

$$I_n(\theta) = \mathrm{diag}\left(\frac{n_1}{\sigma^2}, \frac{n_2}{\sigma^2}, \frac{n_1+n_2}{2\sigma^4}\right).$$

- Linear Model

$$I_n(\beta, \sigma^2) = \begin{pmatrix} \frac{Z^TZ}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

- $DE(\theta)$

$$I_n(\theta) = \frac{3n}{2\theta^2}.$$

- Bernoulli(θ)

$$I_1(\theta) = \frac{1}{\theta(1-\theta)}$$

- Normal(μ, σ^2)

Score$_1$ = $\left(\frac{1}{\sigma^2}(x-\mu), -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(x-\mu)^2\right)$

$$I_1(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

- Geometric (p)

$$I_1(\theta) = \left[\frac{1}{p^2(1-p)}\right]$$

- $N(\mu, \sigma^2)$ w/ $\sigma^2 = \gamma\mu^2$

H_0: MLE is $\hat{\mu} = \frac{-\bar{X} \pm \sqrt{\bar{X}^2 + 4\hat{S}^2}}{2}$

$$S(\mu, \gamma) = \begin{pmatrix} -\frac{n}{\mu} + \frac{n(\bar{x}-\mu)}{\gamma\mu^2} + \frac{\sum(x_i-\mu)^2}{\gamma\mu^3} \\ -\frac{n}{2\gamma} + \frac{\sum(x_i-\mu)^2}{2\mu^2\gamma^2} \end{pmatrix}$$

$$I_n(\mu, \gamma) = n\begin{pmatrix} \frac{1}{\mu^2\gamma} + \frac{2}{\mu^2} & \frac{1}{\mu\gamma} \\ \frac{1}{\mu\gamma} & \frac{1}{2\gamma^2} \end{pmatrix}.$$

Usually $\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, I_1(\theta)^{-1})$

$I_n(\theta) = n I_1(\theta)$.