Contents

1 Examples

1.1 Decision Theory

1.1.1 Bayes Rules

Example 1.1 (Bayes rule, admissible, minimax rule under modified squared error loss (P4 Theory Exam 2021))**.** *Suppose Z is a random variable with PMF*

$$p_\theta(z) = (1-\theta)\theta^z \qquad z \in \{0,1,\dots\}$$

For $\theta[0,1)$. We sih to study the performance of estimators of θ which will be judged by the risk function

$$R(T,\theta) = \frac{E_{P_\theta}(\{\theta - T(Z)\}^2)}{1-\theta}$$

1. ***Calculate the Bayes rule****. Suppose we have a prior Π with nondegenerate support on $[0,1)$. To find the Bayes rule, we minimize the Bayes risk function wrt the action a.*

$$T_\Pi = \underset{a}{argmin}\ \mathbb{E}_\theta\left[\frac{\{\theta-a\}^2}{1-\theta}|Z=z\right]$$
$$\implies a = \frac{\mathbb{E}_\theta\left[\frac{\theta}{1-\theta}|Z=z\right]}{\mathbb{E}_\theta\left[\frac{1}{1-\theta}|Z=z\right]}$$

   We can write these posterior expectations conditional on $Z=z$ by integrating the value against the PMF.

$$T_\Pi = \frac{\int \frac{\theta}{1-\theta}(1-\theta)\theta^z d\Pi}{\int \frac{1}{1-\theta}(1-\theta)\theta^z d\Pi} = \frac{\mathbb{E}_\Pi(\theta^{z+1})}{\mathbb{E}_\Pi(\theta^z)}$$

   Which is a ratio of posterior expectations.

2. ***Prove T_Π is admissible*** *if Π is a fixed prior with nondegenerate support. We know that all unique Bayes rules are admissible. As shown in part (a), the Bayes rule must satisfy:*

$$T_\Pi = \frac{\mathbb{E}_\Pi(\theta^{z+1})}{\mathbb{E}_\Pi(\theta^z)} \implies [\mathbb{E}_\Pi(\theta^z)]\,T_\Pi - \left[\mathbb{E}_\Pi(\theta^{z+1})\right] = 0$$

   Thus, when Π is fixed, T_Π is a solution in X to the problem $aX - b = 0$ for fixed $a, b \in R$. This is a linear system of equations with only one solution. Thus, T_Π is the unique Bayes rule and therefore is admissible.

3. ***Show constant risk****: consider the estimator $T(z) = 0.5\mathbb{I}(z=0) + \mathbb{I}(z \geq 1)$. Show the risk function is constant over all $\theta \in \Theta$. Notice that $P(Z=0) = (1-\theta)\theta^0 = (1-\theta)$ therefore $P(Z \geq 1) = \theta$.*

$$\begin{aligned} R(T,\theta) &= \frac{E_{P_\theta}(\{\theta - T(Z)\}^2)}{1-\theta} \\ &= \frac{E_{P_\theta}(\{\theta - (0.5\mathbb{I}(z=0)+\mathbb{I}(z\geq 1))\}^2)}{1-\theta} \\ &= \frac{\theta^2 - 2\theta\mathbb{E}_{P_\theta}(0.5\mathbb{I}(z=0)+\mathbb{I}(z\geq 1)) + \mathbb{E}_{P_\theta}((0.5\mathbb{I}(z=0)+\mathbb{I}(z\geq 1))^2)}{1-\theta}]] \\ &= \frac{\theta^2 - \theta(1-\theta) - 2\theta^2 + 0.25(1-\theta) + \theta}{1-\theta} \\ &= 0.25 \end{aligned}$$

   Therefore, this particular form of $T(z)$ ensures that the risk function R is constant over θ.

4. ***Exhibit a minimax estimator****: the idea is to find a minimax estimator by finding a prior such that the Bayes rule developed in part (a) equals the estimator developed in part (c) which has constant risk. Bayes rule + constant risk implies minimax! Setting our bayes estimator equal to our estimator with constant risk, we see*

$$\frac{\mathbb{E}_{\Pi}(\theta^{z+1})}{\mathbb{E}_{\Pi}(\theta^{z})} = \frac{1}{2}\mathbb{I}(z=0) + \mathbb{I}(z \geq 1)$$
$$(z=0 \ \textit{case}) \quad \mathbb{E}_{\Pi}(\theta^{1}) = \mathbb{E}(\theta^{0}) \cdot \frac{1}{2}\mathbb{I}(z=0) = \frac{1}{2}$$
$$(z=1 \ \textit{case}) \quad \mathbb{E}_{\Pi}(\theta^{2}) = \mathbb{E}(\theta^{1}) \cdot \mathbb{I}(z \geq 1) = \frac{1}{2}$$
$$\vdots$$

This implies that all the moments of $\mathbb{E}_{\Pi}[\theta] = \frac{1}{2}$*. The only distribution with constant raw moments is a Bernoulli distribution with* $p = 1/2$*. Thus,* T_{Π} *is minimax!*

Example 1.2 (Bayes, Admisible, Minimax Rules in Poisson-Gamma Model (581 Midterm P3))**.** *Suppose $X \sim Pois(\lambda)$. Consider the weighted squared error loss for λ:*

$$L(T(X), \lambda) := \frac{(T(X) - \lambda)^2}{\lambda}$$

1. ***Compute Bayes Estimator*** *when $\Pi \equiv Gamma(\lambda|a, b)$ with density $b^a \lambda^{a-1} \exp(-b\lambda)/\Gamma(a)$. First calculate the form of the posterior:*

$$\begin{aligned}\lambda|X &\propto X|\lambda \times \Pi \\ &\propto \frac{\lambda^x \exp(-\lambda)}{x!} \times b^a \lambda^{a-1} \exp(-b\lambda)/\Gamma(a) \\ &\propto \lambda^{x+a-1} \exp(-(b+1)\lambda) \equiv Gamma(x + a, b + 1)\end{aligned}$$

   Next, we find the form of the Bayes Estimator by minimizing the Bayes risk function with respect to the action

$$\begin{aligned}\frac{\partial f}{\partial a} &= \frac{\partial}{\partial a} \mathbb{E}\left[\frac{(a-\lambda)^2}{\lambda}\Big|X = x\right] = 0 \\ a &= \frac{1}{\mathbb{E}\left[\frac{1}{\lambda}|X = x\right]}\end{aligned}$$

   Now, since $\lambda \sim Gamma(x+1, b+1)$, $1/\lambda \sim Inv\ Gamma(x+a, b+1)$ which has mean $(b+1)/(x+a-1)$. Therefore, the bayes rule takes value

$$T_\Pi(x) = \frac{1}{(b+1)/(x+a-1)} = \frac{x+a-1}{b+1}$$

2. ***Prove** $T(X) = X$ **is Minimax** under loss. Note that under the specified loss*

$$\mathcal{R}(X, \lambda) = \mathbb{E}\left(\left(\frac{(X-\lambda)}{\lambda^{1/2}}\right)^2\right) = \mathbb{E}(\chi_1^2) = 1$$

   Thus, the risk function is constant over $\lambda \in (0, \infty)$. Our new goal is to construct a sequence of priors Π_k such that

$$\lim_{k\to\infty} r(D_{\Pi_k}, \Pi_k) = \sup_\lambda \mathcal{R}(X, \theta)$$

   We derived the Bayes estimator for $\Pi \sim \Gamma(a, b)$ prior to be $\frac{x+a-1}{b+1}$. To prove $T(x) = x$ is minimax, we can choose the prior sequence $\Pi_k \sim \Gamma(a = 1 + 1/k, b = 1/k)$ such that asymptotically, the Bayes rule $D_{\Pi_k} \to X$ which will attain the constant risk value demonstrated above. Thus, X is minimax.

1.1.2 Minimax Rules

Example 1.3 (Sample mean is minimax)**.** *Consider* $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ *with* σ^2 *known. We claim* $\bar{X}_n$ *is minimax. Under squared error loss, letting* $T : X_1, \ldots, X_n \to \bar{X}_n$

$$R(\bar{X}_n, \theta) = \mathbb{E}[(\bar{X}_n - \theta)^2] = \frac{\sigma^2}{n}$$

Consider the prior sequence $\Pi_k := N(0, k)$*. Under this model, the posterior takes the form*

$$\theta | X \sim N\left(\frac{\bar{x}_n n/\sigma^2}{1/k + n/\sigma^2}, \frac{1}{1/k + n/\sigma^2}\right)$$

Under squared error loss, the Bayes rule is the posterior mean

$$r(T_{\Pi_k}, \Pi_k) - \mathbb{E}[(\bar{x}_n - \theta)^2] = \mathbb{E}\left[\left(\frac{\bar{x}_n n/\sigma^2}{1/k + n/\sigma^2} - \theta\right)^2\right] = \mathbb{E}[(\bar{x}_n - \theta)^2] \to 0$$

Thus, $\sup_{\theta \in \Theta} R(D, \theta) = \frac{\sigma^2}{n} = \lim_{k \to \infty} r(T_{\Pi_k}, \Pi_k)$*. This implies* $\bar{X}_n$ *is minimax in* $P_1 := \{N(\theta, \sigma^2) : \theta \in \mathbb{R}, \sigma^2 \text{ known}\}$ *by Strategy 3 under Finding Minimax Rules.*

We can go further and show that $\bar{X}_n$ *is minimax with respect to distributions with bounded variance. Consider* $P_2 := \{P \in Q^n; support(Q) \subset \mathbb{R}, Var_Q(X) \leq \sigma^2)\}$*. Note that for any distribution in* P_2*, by CLT*

$$R(\bar{X}_n, \theta) = \frac{Var_Q(X)}{n} \leq \frac{\sigma^2}{n}$$

Thus, $\sup_{P \in \mathcal{P}_1} \mathcal{R}(D_1, P) = \sup_{P \in \mathcal{P}_2} \mathcal{R}(D_1, P)$ *implying by Strategy 4 in Finding Minimax Rules that* $\bar{X}_n$ *is minimax over* P_2*.*

Example 1.4 (Lower Bounding the Minimax Risk of a density at a point – Le Cam's Method). *Let $\mathcal{P}(\beta, L)$ be the collection of densities ($q \geq 0, \int q(x)dx = 1$) that belong in a Holder class $\Sigma(\beta, L)$ meaning the density is $(\beta - 1)$-times differentiable with derivative $q^{(\beta-1)}$ that satisfies for all x_1, x_2*

$$\left|q^{(\beta-1)}(x_1) - q^{(\beta-1)}(x_2)\right| \leq L|x_1 - x_2|$$

If our goal is to estimate the density at a point, $p(x_0)$, we can pursue Le Cam's Method.

1. *Propose two candidate distributions with large discrepancy and small KL diveregence. Let ϕ denote the density of a standard normal RV;*

$$p_1 : x \to \sigma^{-1}\phi\left(\frac{x - x_0}{\sigma}\right)$$
$$p_2 : x \to p_1 + Lh_n^{\beta}\left[K\left(\frac{x - x_0}{h_n}\right) - K\left(\frac{x - 1 - x_0}{h_n}\right)\right]$$

   Where for sufficiently small $a > 0$, $K : x \to a\exp\left(-\frac{1}{1-4x^2}\right)\mathbb{I}(|x| \leq 1/2)$.

2. *Verify $p_1, p_2 \in \mathcal{P}$.*

   (a) *p_2: Let $H_\beta(x)$ is the β-the Hermite polynomial.*

$$\frac{d^\beta}{dx^\beta}p_1(x) = (-1)^\beta H_\beta(x)\phi(x)$$

   Since $\lim_{|x|\to\infty} \frac{1}{\sqrt{2\pi}}H_\beta(x)e^{-x^2/2} = 0$ and the derivative is continuous, $\left|\frac{d^\beta}{dx^\beta}p_1(x)\right|$ is bounded uniformly by a constant. We can make this constant $\leq L$ by choosing σ large enough.

   (b) *p_2: clearly integrates to 1 because of bump term cancellations. In order for p_2 to be positive, we need to choose a^* such that*

$$0 < p_1(x) - Lh_n^{\beta}K\left(\frac{x - h_n - x_0}{h_n}\right)$$
$$\implies 0 < p_1(x) - Lh_n^{\beta}a\exp\left(-\frac{1}{1 - 4\left(\frac{x-h_n-x_0}{h_n}\right)^2}\right)\mathbb{I}\left(\left|\frac{x - 1 - x_0}{h_n}\right| \leq 1/2\right)$$
$$\implies 0 < p_1(x) - Lh_n^{\beta}a\exp\left(-\frac{1}{1 - 4\left(\frac{x-h_n-x_0}{h_n}\right)^2}\right)\mathbb{I}\left(x_0 + 1 - \frac{h_n}{2} \leq x \leq x_0 + 1 + \frac{h_n}{2}\right)$$
$$\implies a^* < \inf_{x\in\mathbb{I}(\dots)} \frac{p_1(x)}{Lh_n^{\beta}\exp\left(-\frac{1}{1-4\left(\frac{x-h_n-x_0}{h_n}\right)^2}\right)}$$

   To ensure p_2 is in the Holder class, it is sufficient to show that q is β-times differentiable with bounded derivatives. We note that the Bump functions K and its β derivatives are continuous functions defined on a compact interval, therefore they obtain their maxima and minima. This means that the β-th derivative is upper bounded by a constant, and we can force this constant to be less than L by choosing $\sigma, a > 0$ small enough.

3. *Study KL divergence, using the Taylor expansion* $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots$

$$\begin{aligned}
-KL(P_1, P_2) &= \int \log\left(\frac{p_2}{p_1}\right) p_1 d\nu \\
&= \int \log\left(\frac{p_1 + bump}{p_1}\right) p_1 d\nu \\
&= \int \left(\sum_{i=1}^{\infty} (-1)^{i+1} \frac{\left(\frac{bump}{p_1}\right)^i}{i}\right) p_1 d\nu \\
\textit{1st order term} &= \int bump\, d\nu = 0 \\
\textit{2nd order term} &= \frac{1}{2}\int \frac{bump^2}{p_1(x)} d\nu \\
&= \frac{1}{2}\int L^2 h^{2\beta} p_1(x)^{-1} \left[K\left(\frac{x-x_0}{h_n}\right) - K\left(\frac{x-1-x_0}{h_n}\right)\right]^2 d\nu \\
&\overset{h_n small}{=} c_1 h_n^{2\beta} \int p^{-1}(x) \left[K\left(\frac{x-x_0}{h_n}\right)^2 - K\left(\frac{x-1-x_0}{h_n}\right)^2\right] \qquad \textit{(h_n small bumps don't overlap)} \\
&= c_1 h_n^{2\beta+1} \int p^{-1}(h_n U + x_0) \left[K(U)^2 - K\left(U - \frac{1}{h_n}\right)^2\right] dU \\
&= c_2 h_n^{2\beta+1} \\
\textit{3rd order term} &= o(h^{3\beta})
\end{aligned}$$

Now the KL-divergence under n-iid draws yields

$$-KL(P_1^n, P_2^n) \geq cnh_n^{2\beta+1}$$

To get a stable lower bound on the KLD, we require $h_n = \mathcal{O}(n^{-\frac{1}{2\beta+1}})$.

4. *Study Discrepancy:*

$$\begin{aligned}
d(P_1, P_2) &= \frac{1}{2}(p_1(x_0) - p_2(x_0))^2 \\
&= \frac{1}{2}\left(p_1(x_0) - p_1(x_0) - Lh_n^{\beta}\left[K(0) - K\left(-\frac{1}{h_n}\right)\right]\right)^2 \\
&= Ch_n^{2\beta}
\end{aligned}$$

We have all the pieces we need now. Applying Le Cam's method, we obtain

$$\inf_{T \in \mathcal{T}} \sup_{P \in \mathcal{P}} R(T, P) \geq \frac{1}{4} d(P_1, P_2) \exp(-KL(P_1, P_2)) \geq c \cdot h_n^{2\beta}$$

In order for the KL divergence to have a stable lower bound, we required $h_n = \mathcal{O}(n^{-\frac{1}{2\beta+1}})$.

$$\inf_{T \in \mathcal{T}} \sup_{P \in \mathcal{P}} R(T, P) \leq c^* n^{-\frac{2\beta}{2\beta+1}}$$

Thus a lower bound on the minimax rate is $\mathcal{O}(n^{-\frac{2\beta}{2\beta+1}})$ *which is a slightly slower than parametric rate.*

Example 1.5 (Lower bounding minimax risk of smooth regression function – Fano's Method)**.** *Suppose we observe* $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} Q \in \mathcal{Q}$*, where* $X \sim U[0,1]$ *and* $Y|X = x \sim N(f_Q(x), 1)$ *where* $f_Q(x) \in \mathcal{F}(\beta, L)$ *a Holder class. Suppose our objective is to estimate* $f_Q(x)$*, with performance quantified by the mean integrated squared error:*

$$L(a, Q^n) = \int_0^1 [a(x) - f_Q(x)]^2 dx$$

We take the following few steps

1. *Define candidate function class. Let* $\mathcal{F}_1$ *denote a convex combination of orthonormal basis functions where the elements of the basis are scaled bump functions:*

$$\mathcal{F}_1 := \left\{ x \to \sum_{j=1}^{m} w_j \phi_j(x) : w \in \{0,1\}^m, \phi_j(x) = Lh^\beta K\left(\frac{x - \frac{j}{m+1}}{h}\right), m \in \left[8, \frac{1}{h-1}\right] \right\}$$

   Where for sufficiently small a*,*

$$K : x \to a \exp\left(-\frac{1}{1-4x^2}\right) \mathbb{I}(|x| < 1/2)$$

   So $\mathcal{F}_1$ *is a collection of functions that are sums of* m *bump functions centered at* $\frac{j}{m+1}$ *for* $j = 1, \ldots, m$*, that are multiplied by 0 or 1, and that do not overlap since* $m \leq \frac{1}{h} = 1 \implies h \leq \frac{1}{m+1}$.
   Recall that $\Omega := \{0,1\}^m$ *indexes the collection of functions in* $\mathcal{F}_1$*. Thus,* $|\mathcal{F}_1| = |\Omega| = 2^m$.

2. *Study the discrepancy:*

$$\begin{aligned}
d(P_w, P_\nu) &= \frac{1}{2} \int [f_w(x) - f_\nu(x)]^2 dx \\
&= \frac{1}{2} \sum_{j=1}^{m} [w_j - \nu_j]^2 \int \phi_j(x)^2 dx \quad \textit{(Bases orthogonal so cross terms cancel)} \\
&= \frac{1}{2} \sum_{j=1}^{m} [w_j - \nu_j]^2 L^2 h^{2\beta+1} \underbrace{\int K(u)^2 du}_{c_2} \quad \textit{(U-sub)} \\
&= \frac{1}{2} c_2 L^2 h^{2\beta+1} \underbrace{\sum_{j=1}^{m} [w_j - \nu_j]^2}_{\textit{Hamming dist}} \\
&= c_3 h^{2\beta+1} H(w, \nu) \quad \left(c_3 := \frac{c_2 L^2}{2}\right)
\end{aligned}$$

   The minimal Hamming distance for two functions in that differ is exactly 1, yielding;

$$\min_{j \neq k} d(P_j, P_k) = c_3 h^{2\beta+1}$$

3. *Study the KL divergence. Turns out KL divergence takes the form:*

$$\begin{aligned}
KL(P_w, P_\nu) &= \frac{n}{2} \int_0^1 [f_w(x) - f_\nu(x)]^2 dx \\
&= c_3 n h^{2\beta+1} H(w, \nu) \quad \textit{(By same logic)} \\
&\leq c_3 n h^{2\beta+1} m \quad \textit{(since } H(w, \nu) \leq m\textit{)}
\end{aligned}$$

4. *Plug into Fano's Bound: recall that* $\Omega := \{0,1\}^m$ *indexes the collection of functions in* $\mathcal{F}_1$.

$$\inf_{T\in\mathcal{T}}\sup_{P\in\mathcal{P}} R(T,P) \geq \frac{\min_{j\neq k} d(P_j,P_k)}{2}\left[1 - \frac{\log 2 + \max_{j\neq k} KL(P_j,\bar{P})}{\log(|\Omega|)}\right]$$
$$\geq \frac{c_3 h^{2\beta+1}}{2}\left(1 - \frac{\log 2 + c_3 n h^{2\beta+1} m}{\log|\Omega|}\right)$$
$$= \frac{c_3 h^{2\beta+1}}{2}\left(1 - \frac{\log 2 + c_3 n h^{2\beta+1} m}{m\log 2}\right)$$

For this bound to be informative, $h = \mathcal{O}(n^{-1/(2\beta+1)})$. *But this produces a lower bound on the minimax risk of* $\mathcal{O}(n^{-1})$, *meaning the problem is as least as difficult as a parametric problem. This suggests that the bound may not be tight.*

5. *Tighten the bound using the Varshamov-Gilbert Lemma. For* $m \geq 8$, *there exista an* $\Omega \subset \Omega$ *s.t.* $|\Omega| \geq 2^{m/8}$ *and* $\min_{w\neq v} H(w,v) \geq \frac{m}{8}$. *If we choose this subset*

$$\inf_{T\in\mathcal{T}}\sup_{P\in\mathcal{P}} R(T,P) \geq \frac{c_3 h^{2\beta+1} m}{16}\left(1 - \frac{\log 2 + c_3 n h^{2\beta+1} m}{\frac{m}{8}\log 2}\right)$$
$$= \frac{c_3 h^{2\beta+1} m}{16}\left(1 - \frac{8}{m} - \frac{8 c_3 n h^{2\beta+1}}{\log 2}\right)$$

Goal is to choose m *as large as possible to provide the tightest bound. If we choose* $m = \lfloor \frac{1}{h} - 1 \rfloor$. *We know that* $\frac{1}{2h} < m < \frac{1}{h}$. *Plugging in the lower bound, we have*

$$\inf_{T\in\mathcal{T}}\sup_{P\in\mathcal{P}} R(T,P) \geq \frac{c_3 h^{2\beta+1} m}{16}\left(1 - \frac{8}{m} - \frac{8 c_3 n h^{2\beta+1}}{\log 2}\right)$$
$$\geq \frac{c_3 h^{2\beta}}{32}\left(1 - 16h - \frac{8 c_3 n h^{2\beta+1}}{\log 2}\right)$$

To ensure that the negative term above is bounded, we require $n = h^{2\beta+1} \implies h = \mathcal{O}(n^{-1/(2\beta+1)})$.

Since the bandwidth is $h = \mathcal{O}(n^{-1/(2\beta+1)})$, *the lower bound on the minimax risk is* $\mathcal{O}(n^{-2\beta/(2\beta+1)})$.

1.1.3 Admissible Rules

Example 1.6 (Posterior Mean is Admissible in Normal Model). *Let* $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ *and* $\theta \sim N(\mu, \tau^2)$*. We will show that the following estimator is admissible*

$$T_\Pi : (X_1, \ldots, X_n) \to \left(1 - \frac{1/\tau^2}{1/\tau^2 + n/\sigma^2}\right) \bar{X}_n + \left(\frac{1/\tau^2}{1/\tau^2 + n/\sigma^2}\right) \mu$$

By Strategy 1 in finding admissible estimators, we are using squared error loss and the Bayes risk is finite because all the random quantities are finite. Also, the normal distribution is absolutely continuous wrt the Lesbegue measure and vice versa. Therefore, T_Π *is unique Bayes and therefore admissible for* $\left(\frac{1/\tau^2}{1/\tau^2+n/\sigma^2}\right) \in (0, 1)$.
When $\left(\frac{1/\tau^2}{1/\tau^2+n/\sigma^2}\right) = 0$, $T : x \to \mu$ *is admissible because it is a constant estimator that achieves risk* 0 *when* $\theta = \mu$*. When* $\left(\frac{1/\tau^2}{1/\tau^2+n/\sigma^2}\right) = 1$*, turns out the sample mean is admissible, but this requires further proof.*

Example 1.7 (Sample mean is Admissible in Normal Model). *Let $X_1, \dots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with σ^2 known. We claim that $\bar{X}_n$ is admissible in the model.*
We will show by demonstrating either

(a) $R(T,\theta) \geq R(\bar{X}_n, \theta)\ \forall \theta \in R$

(b) There exists some θ for which $R(T,\theta) > R(\bar{X}_n, \theta)$

Consider WLOG $\sigma^2 = 1$. Suppose (a) does not hold. We will show that (b) holds. If (a) does not hold, there exists a θ_1 s.t. $R(T, \theta_1) < R(\bar{X}_n, \theta_1)$. By continuity of R, there exists $\epsilon, \delta > 0$ s.t. for all $\theta \in (\theta_1 - \delta, \theta_1 + \delta)$,

$$R(T,\theta) < R(\bar{X}_n, \theta) - \epsilon = \frac{1}{n} - \epsilon$$

Specifying the prior $\Pi = N(0, \tau^2)$ and the Bayes rule T_Π as the posterior mean, we obtain

$$\begin{aligned} r(T_\Pi, \Pi) - R(\bar{X}_n, \theta) &= \int R\left(\frac{n}{1/\tau^2 + n}\bar{X}_n, \theta\right) d\Pi(\theta) - \frac{1}{n} \\ &= \int \left(\frac{n}{1/\tau^2 + n}\bar{X}_n - \theta\right)^2 \cdot \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}\theta^2\right) d\theta - \frac{1}{n} \\ &= \frac{\tau^2}{1 + n\tau^2} - \frac{1}{n} = -\frac{1}{n(1+n\tau^2)} \end{aligned}$$

By optimality of the Bayes rule

$$r(T_\Pi, \Pi) - R(\bar{X}_n, \theta) \leq r(T_1, \Pi) - R(\bar{X}_n, \theta)$$

$$\implies \frac{\tau^2}{1+n\tau^2} - \frac{1}{n} = -\frac{1}{n(1+n\tau^2)} \leq \int \left[\mathcal{R}(T_1, \theta) - \frac{1}{n}\right]^+ \Pi(d\theta) - \int \left[\mathcal{R}(T_1, \theta) - \frac{1}{n}\right]^- \Pi(d\theta)$$

Recall that for $\theta \in (\theta_1 - \delta, \theta_1 + \delta)$ and $R(T_1, \theta) < \frac{1}{n} - \epsilon$ implying $[R(T_1, \theta) - 1/n]^- > \epsilon$. Then simple bounding yields

$$\begin{aligned} \int \left[\mathcal{R}(T_1, \theta) - \frac{1}{n}\right]^- \Pi(d\theta) &\leq \int_{\theta_1 - \delta}^{\theta_1 + \delta} \left[\mathcal{R}(T_1, \theta) - \frac{1}{n}\right]^- \Pi(d\theta) \\ &\leq \epsilon \int_{\theta_1 - \delta}^{\theta_1 + \delta} d\Pi(\theta) \\ &= \epsilon \Pi(\theta_1 - \delta \leq \theta \leq \theta_1 + \delta) \end{aligned}$$

Implying

$$\int \left[\mathcal{R}(T_1, \theta) - \frac{1}{n}\right]^+ d\Pi(\theta) \geq -\frac{1}{n(1+n\tau^2)} + \epsilon\Pi(\theta_1 - \delta \leq \theta \leq \theta_1 + \delta)$$

Noting that

$$\sqrt{2\pi}\tau\left(-\frac{1}{n(1+n\tau^2)} + \epsilon\Pi(\theta_1 - \delta \leq \theta \leq \theta_1 + \delta)\right) \overset{\tau \to \infty}{\longrightarrow} 2\epsilon\delta$$

Thus, choosing τ_0 s.t. $\sqrt{2\pi}\tau_0\left(-\frac{1}{n(1+n\tau_0^2)} + \epsilon\Pi(\theta_1 - \delta \leq \theta \leq \theta_1 + \delta)\right) > \epsilon\delta$ we obtain

$$\int \left[\mathcal{R}(T_1, \theta) - \frac{1}{n}\right]^+ d\Pi(\theta) \leq \left(-\frac{1}{n(1+n\tau_0^2)} + \epsilon\Pi(\theta_1 - \delta \leq \theta \leq \theta_1 + \delta)\right) > \frac{\epsilon\delta}{\sqrt{2\pi}\tau_0} > 0$$

Thus, ther exists θ for which $R(T, \theta) > R(\bar{X}_n, \theta)$ implying condition (b) holds. Thus, the sample mean is admissible.

Example 1.8 (Sample mean is inadmissible in $d \geq 3$)**.** *Suppose $X_1, \dots, X_n \sim N(\theta, I_d)$ for $d \geq 3$. Let T^{JS} be the James-Stein estimator:*

$$T^{JS} : x \to \begin{cases} \left(1 - \frac{(d-2)}{n||\bar{x}_n||^2}\right)\bar{x}_n & \text{if } \bar{x}_n \neq (0,\dots,0) \\ 0 & \text{if } \bar{x}_n = (0,\dots,0) \end{cases}$$

Under MSE loss, letting T denote the sample mean

$$\begin{aligned} R(T^{JS},\theta) &= \mathbb{E}[||T^{JS}(||X||)X - \theta||^2] \qquad (T \text{ is spherically symmetric est}) \\ &= \mathbb{E}\left[||[T^{JS}(||X||) - 1]X + [X-\theta]||^2\right] \\ &= \mathbb{E}\left[||[T^{JS}(||X||) - 1]X||^2\right] + \mathbb{E}[||X-\theta||^2] - 2\mathbb{E}\left[\langle[1 - T^{JS}(||X||)]X,\ X - \theta\rangle\right] \\ &= \mathbb{E}\left[\frac{(d-2)^2}{||X||^2}\right] + R(T,\theta) - 2(d-2)\mathbb{E}\left[\left\langle \frac{X}{||X||^2}, X - \theta \right\rangle\right] \end{aligned}$$

To show the third term in the above display is $-2\mathbb{E}\left[||[T^{JS}(||X||) - 1]X||^2\right]$, we appeal to Stein's Lemma.

Stein's Lemma: *Letting $Y \sim N(\mu, \sigma^2 I_d)$ and $g_1, \dots, g_d$ be functions from $\mathbb{R}^d \to \mathbb{R}$ s.t. for all $j = 1, \dots, d$, $\mathbb{E}\left|\frac{\partial}{\partial y_j} g_j(y)|_{y=Y}\right| < \infty$. Letting $g : y \to (g_1(y), \dots, g_d(y))$, we have*

$$\mathbb{E}[\langle g(Y), Y - \mu\rangle] = \sigma^2 \mathbb{E}\left[\sum_{j=1}^d \frac{\partial}{\partial y_j} g_j(y)\right]$$

Define $g_j : z \to \frac{z_j}{||z||^2}$. Then we see that

$$\begin{aligned} \mathbb{E}\left[\left\langle \frac{X}{||X||^2}, X - \theta \right\rangle\right] &= \mathbb{E}\left[\langle g(X), X - \theta\rangle\right] \\ &= \mathbb{E}\left[\sum_{j=1}^d \frac{\partial}{\partial y_j} g_j(y)\right] \qquad (\text{Stein's Lemma}) \\ &= \mathbb{E}\left[\sum_{j=1}^d \frac{1}{||X||^2} - \frac{2X_j}{||X||^4}\right] \qquad (\text{Quotient Rule}) \\ &= \mathbb{E}\left[\frac{d}{||X||^2} - \frac{2||X||^2}{||X||^4}\right] \\ &= \mathbb{E}\left[\frac{d-2}{||X||^2}\right] \end{aligned}$$

Therefore,

$$\begin{aligned} R(T^{JS},\theta) &= \mathbb{E}\left[\frac{(d-2)^2}{||X||^2}\right] + R(T,\theta) - 2(d-2)\mathbb{E}\left[\left\langle \frac{X}{||X||^2}, X - \theta \right\rangle\right] \\ &= \mathbb{E}\left[\frac{(d-2)^2}{||X||^2}\right] + R(T,\theta) - 2\mathbb{E}\left[\frac{(d-2)^2}{||X||^2}\right] \\ &= R(T,\theta) - \mathbb{E}\left[\frac{(d-2)^2}{||X||^2}\right] \end{aligned}$$

Therefore, $R(T^{JS},\theta) < R(T,\theta)$ for all θ.

1.2 Hypothesis Testing

Example 1.9 (Power under local alternatives for location family)**.** *Suppose $X_1^n \sim P_\theta$ for location family where (i) P_θ has density $f(x-\theta)$, (ii) f is symmetric about 0, (iii) f is positive and continuously differentiable with finite second moment.*
Suppose we wish to test $H_0 : \theta = 0$ and $H_1 : \theta > 0$ with the following sign and t-statistics:

1. *Sign: $S_n = \frac{1}{n}\sum I(X_i > 0)$*
2. *t-statistic: $T_n = \frac{1}{n}\sum \frac{X_i}{\hat{\sigma}_n}$ where $\hat{\sigma}_n$ is the empirical standard deviation.*

The estimators are both asymptotically linear:

$$\sqrt{n}\left(S_n - \frac{1}{2}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(I(X_i > 0) - \frac{1}{2}\right) \rightsquigarrow N(0, 1/4)$$
$$\sqrt{n}T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i}{\sigma} + o_p(1)$$

Let's show that both estimators are regular. For the sign statistic, $\mu(\theta) := P_\theta(X > 0)$

$$\begin{aligned}
\dot{\mu}(0) &= \frac{\partial}{\partial\theta}P_\theta(X > 0)\Big|_{\theta=0} = \frac{\partial}{\partial\theta}\int_0^\infty f(x-\theta)dx\Big|_{\theta=0} \\
&= \int_0^\infty \frac{\partial}{\partial\theta}f(x-\theta)\Big|_{\theta=0}dx \\
&= -\int_0^\infty \dot{f}(x)dx = \int\left[-\frac{\dot{f}(x)}{f(x)}\right]I(x > 0)dP_0(x) \\
&= \int \dot{\ell}_0(x)I(x > 0)dP_0(x) \\
&= P_0(\dot{\ell}_0 s_0)
\end{aligned}$$

Recall regularity is equivalent to $P_0(g_\theta\dot{\ell}_\theta) = \dot{\mu}(\theta)$. Note that $\theta = 0$ under H_0 implies S_n is regular. Now for the t-statistic, let $\mu(\theta) := \theta/\sigma$. We have

$$\begin{aligned}
\int \dot{\ell}_0 t_0(x)dP_0(x) &= \int \dot{\ell}_0(x)\frac{x}{\sigma}dP_0(x) \\
&= -\sigma^{-1}\int \frac{\dot{f}(x)}{f(x)}xdP_0(x) \\
&= -\sigma^{-1}\int \dot{f}(x)\cdot xdx \\
&= \sigma^{-1}\int\left(f(x) - \frac{d}{dx}[xf(x)]\right)dx \qquad \textit{(Product rule and add subtract)} \\
&= \sigma^{-1} - \lim_{a\to\infty}\int_{-a}^{a}\left(\frac{d}{dx}[xf(x)]\right)dx \qquad \textit{(Pdf integrates to 1)} \\
&= \sigma^{-1} - \lim_{a\to\infty}a[f(a) - f(-a)] \\
&= \sigma^{-1} = \dot{\mu}(0)
\end{aligned}$$

Thus, T_n is also regular.

Knowing S_n and T_n are regular ALEs, we know that their corresponding tests

$$\mathbb{I}(\sqrt{n}(2S_n - 1) > z_{1-\alpha})$$
$$\mathbb{I}(\sqrt{n}T_n > z_{1-\alpha})$$

have power functions under local alternatives take the form for all h:

$$\pi_n\left(\frac{h}{\sqrt{n}}\right) \overset{n\to\infty}{\rightsquigarrow} 1-\Phi\left(z_{1-\alpha} - h^T\frac{\dot{\mu}(0)}{\sigma(0)}\right)$$
$$\implies P_{\theta+h/\sqrt{n}}(\sqrt{n}(2S_n-1) > z_{1-\alpha}) = 1-\Phi\left(z_{1-\alpha}-2hf(0)\right)$$
$$\implies P_{\theta+h/\sqrt{n}}(\sqrt{n}(T_n) > z_{1-\alpha}) = 1-\Phi\left(z_{1-\alpha}-h\sigma^{-1}\right)$$

Thus, we can compare the relative power of the sign test to the t-test under local alternatives by the ratio of their two slopes:

1. *If* $2f(0)\sigma > 1$, *the sign test has greater local power.*
2. *If* $2f(0)\sigma < 1$, *the* t *test has greater local power.*

This indicates when $f(0)$ *is very large relative to the variance* σ, *the sign test is more powerful under local alternatives asymptotically. For instance, if we consider* f *to be a density that for small* $\epsilon > 0$, *places mass* $(1-\epsilon)$ *on Unif*$(-1,1)$ *and* ϵ *mass at* $N(0,\epsilon^4)$, *then the sign test will have much greater power.*

1.3 Empirical Process Theory

1.3.1 Concentration Inequalities

Example 1.10 (Bivariate U statistic (McDiarmind's Inequality)). *A good use case of McDiarmind's inequality is in the study of the concentration of U-statistics, where $g : \mathbb{R}^2 \to \mathbb{R}$ and*

$$U := \binom{n}{2}^{-1} \sum_{j \leq k} g(X_j, X_k)$$

If g is bounded, say $||g||_\infty \leq b$, then McDiarmind's inequality yields for a given coordinate k:

$$\begin{aligned} |f(x) - f(x^{\setminus k})| &\leq \binom{n}{2}^{-1} \sum_{j \neq k} |g(x_j, x_k) - g(x_j, x'_k)| \\ &\leq \frac{(n-1)(2b)(2)}{(n)(n-1)} = \frac{4b}{n} \end{aligned}$$

So the bounded differences property holds with parameter $c_i = \frac{4b}{n}$ in each coordinate. By McDiarmind's Inequality

$$P(|U - \mathbb{E}(U)| \geq t) \leq 2 \exp\left(-\frac{nt^2}{8b^2}\right)$$

Example 1.11 (Gaussian Order Statistics (Lipschitz Transformation of Gaussian))**.** *Let $X_{(k)}$ denote the k-th order statistic of a Gaussian random vector. Let $Y_{(k)}$ denote the k-th order statistic from an iid ghost sample from the same Gaussian distribution. Turns out*

$$|X_{(k)} - Y_{(k)}| \leq ||X - Y||_2$$

so each order statistic is 1*-Lipschitz. Based on the concentration result for lipschitz transformations of Gaussian random vectors*

$$P[|X_{(k)} - \mathbb{E}[X_{(k)}]| \geq \delta] \leq 2\exp\left(-\frac{\delta^2}{2}\right)$$

1.3.2 Establish Uniform LLN and Upper Bounding Empirical Process Terms

Example 1.12 (Establish Uniform LLN in Lipschitz Function Class – Dudley)**.** *If $\mathcal{F}$ denotes a class of $[0,1] \to \mathbb{R}$-valued Lipschitz functions s.t., $|f(x) - f(y)| \leq L|x-y|$.*
Let's first derive the metric entropy (log covering number) of the function class $\mathcal{F}$. Create $M = \lfloor \frac{1}{\epsilon} \rfloor$ grid points $x_i = (i-1)\epsilon$ for $i = 1, \dots, M$ on $[0,1]$. Defining ϕ as

$$\phi(u) := \begin{cases} 0 & \text{if } u < 0 \\ u & \text{if } 0 \leq u \leq 1 \\ 1 & \text{else} \end{cases}$$

For any binary sequence $\beta = \{-1, +1\}^M$, define a function f_β such that

$$f_\beta(y) = \sum_{i=1}^{M} \beta_i L \epsilon \phi\left(\frac{y - x_i}{\epsilon}\right)$$

consider the interval from $y \in (0, x_1)$. ϕ increases linearly in $y - x_i$ the interval with slope $\pm L$. Thus, $f_\beta(y)$ is piecewise linear with slope $\pm L$ over each pair of gridpoints. For any two functions $f_\beta, f_{\beta'}$, there is at least one interval where the two functions start at the same point and have opposite slopes, implying that $||f_\beta - f_{\beta'}||_\infty \geq 2L\epsilon$. Thus, $\{f_\beta, \beta \in \{-1,+1\}^M\}$ forms a $2L\epsilon$-packing in the supnorm. By relationships between covering and packing numbers

$$2^M = |f_\beta| \leq M(2L\epsilon, \mathcal{F}, ||\cdot||_\infty) \leq N(L\epsilon, \mathcal{F}, ||\cdot||_\infty)$$

Defining $\delta = \epsilon L$, and recalling that $M = \lfloor \frac{1}{\epsilon} \rfloor$, we have

$$C \cdot \frac{L}{\delta} \leq \log N(\delta, \mathcal{F}, ||\cdot||_\infty)$$

Also by plotting $\{f_\beta, \beta \in \{-1,+1\}^M\}$, one can see that the farthest an element of $\mathcal{F}$ can be from a given f_β pointwise is $L\epsilon$. Thus, $\{f_\beta, \beta \in \{-1,+1\}^M\}$ is a δ-cover for $\mathcal{F}$. The covering number (size of smallest cover), then:

$$N(\delta, \mathcal{F}, ||\cdot||_\infty) \leq |f_\beta| = C^* \cdot \frac{L}{\delta}$$

Therefore, going back to $\epsilon > 0$ notation

$$\sup_Q \log(N(\epsilon, \mathcal{F}, L^2(Q))) = \log(N(\epsilon, \mathcal{F}, ||\cdot||_\infty) = \mathcal{O}\left(\frac{L}{\epsilon}\right)$$

Recognizing that $D = 2L < \infty$, Dudley's entropy integral gives:

$$\mathbb{E}||R_n||_{\mathcal{F}} \leq \frac{8}{\sqrt{n}} \sup_Q \left[\int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}, L^2(P_n))} d\epsilon\right] \equiv \frac{8}{\sqrt{n}} \left[\int_0^D \mathcal{O}\left(\frac{L}{\epsilon}\right) d\epsilon\right] = \mathcal{O}(n^{-1/2})$$

Therefore, the entropy integral is satisfied, and the empirical process term is controlled. Also, $\mathcal{F}$ is Donsker since is satisfies the entropy integral.

Example 1.13 (Establish Uniform LLN in Class of Functions Lipschitz in Indexing Parameters – Dudley)**.** *Let $\mathcal{F} := \{g_\beta : \beta \in \mathbb{R}^p; ||\beta||_2 \leq 1\}$ be a collection of functions indexed by parameter β where $|g_{\beta_1}(x) - g_{\beta_2}(x)| \leq L||\beta_1 - \beta_2||$.*
Step 1: Note that the indexing parameter set $B = \{\beta \in \mathbb{R}^p : ||\beta||_2 = 1\}$ is a sphere of radius 1. We previously proved that the ϵ-covering number of a ball of radius r has the upper bound

$$N(\epsilon, B(0,r), ||\cdot||_{L^p(P)}) \leq \left(\frac{2r}{\epsilon} + 1\right)^p$$

Step 2: we also know that functions Lipschitz in their indexing parameters also satisfy the following covering number bound on their function space $\mathcal{F}$

$$N(\epsilon, \mathcal{F}, ||\cdot||_{\mathcal{F}}) \leq N(\epsilon/L, B, ||\cdot||_B)$$

Step 3: bringing these two together

$$N(\epsilon, \mathcal{F}, ||\cdot||_{\mathcal{F}}) \leq N(\epsilon/L, B(0,1), ||\cdot||_2) \leq \left(\frac{2\cdot 1}{\epsilon/L} + 1\right)^p$$
$$\implies \log N(\epsilon, \mathcal{F}, ||\cdot||_{\mathcal{F}}) \leq p\log\left(\frac{2L}{\epsilon} + 1\right) \approx p\log\left(\frac{L}{\epsilon}\right)$$

And the Dudley integral is:

$$\begin{aligned}\frac{8}{\sqrt{n}}\left[\int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}, L^2(P_n))}d\epsilon\right] &\leq \frac{8}{\sqrt{n}}\int_0^{2L}\sqrt{p\log\left(\frac{L}{\epsilon}\right)}d\epsilon \\ &\lessapprox \frac{8}{\sqrt{n}}L\sqrt{p}\int_0^1 \log(1/\delta)d\delta \\ &\lessapprox \frac{8}{\sqrt{n}}L\sqrt{p} \\ \implies \mathbb{E}||P_n - P||_{\mathcal{F}} \lessapprox \mathbb{E}||R_n||_{\mathcal{F}} = \mathcal{O}\left(\frac{L\sqrt{p}}{\sqrt{n}}\right) &\end{aligned}$$

Thus, a function that is Lipschitz in its 1-dimensional indexing parameter controls the empirical process term at a $\mathcal{O}(n^{-1/2})$ rate! However, as the dimension of the indexing parameter increases, we get slower convergence.

Example 1.14 (Establish Uniform LLN in Sobolev Class)**.** *Let $\mathcal{F}$ be a collection of functions $f : [0,1] \to \mathbb{R}$ such that*

1. *Uniformly bounded:* $||f||_\infty \leq 1$
2. *Absolutely continuity of $(k-1)$-th derivative*
3. $\int f^{(k)}(x)^2 dx \leq 1$ *for some* $k \in \mathbb{N}$

There exists a constant C such that the log bracketing number wrt the supnorm metric takes form for all $\epsilon \in [0,1]$:

$$\log N_{[]}(\epsilon, \mathcal{F}, ||\cdot||_\infty) \leq C \left(\frac{1}{\epsilon}\right)^{1/k}$$

Suppose $k \geq 1$, then by the bracketing integral bound is finite;

$$\begin{aligned}\mathbb{E}||P_n - P||_{\mathcal{F}} &\leq \frac{C}{\sqrt{n}} \int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, ||\cdot||_\infty)} d\epsilon \\ &\leq \frac{C^*}{\sqrt{n}} \int_0^1 \sqrt{\epsilon^{-1/k}} d\epsilon \\ &= \mathcal{O}(n^{-1/2})\end{aligned}$$

Thus, we can control the empirical process term at $\mathcal{O}(n^{-1/2})$ rates. Also since the function class satisfies the uniform entropy integral, it is Donsker.

1.4 M/Z Estimation

Example 1.15 (Limiting distributions and Regularity of Mean and Median (581 HW 8 P3))**.** *Let μ_n and m_n be the sample mean and median respectively. Find the limit distributions of each under P_{θ_0} and $P_{\theta_0+h/\sqrt{n}}$ given $P_{\theta'} \equiv N(\theta', 1)$.*

We start with the sample mean, μ_n. By WLLN, the sample mean is consistent and by the central limit theorem,

$$\sqrt{n}(\mu_n - \theta_0) \overset{P_{\theta_0}}{\rightsquigarrow} N(0,1)$$

Recalling that the normal distribution is QMD and both distributions are mutually contiguous, local asymptotic normality holds, the log likelihood ratio affords a taylor expansion, and is asymptotically normal. The joint distribution between the sample mean and log likelihood ratio is normal with covariance h. Le Cam's third lemma says that the distribution of the MLE under sampling from the local alternative is

$$\sqrt{n}(\mu_n - \theta_0) \overset{P_{\theta_0+h/\sqrt{n}}}{\rightsquigarrow} N(h,1) \implies \sqrt{n}\left(\mu_n - \left(\theta_0 + \frac{h}{\sqrt{n}}\right)\right) \overset{P_{\theta_0+h/\sqrt{n}}}{\rightsquigarrow} N(0,1)$$

Thus, the MLE is invariance to local perturbations in the parameter, implying that it is a regular estimator.

We now turn our attention to the sample median. The sample median can be defined as a z-estimator that solves the estimating equation $P_n z_\theta = 0$ where $z_\theta(x) = \mathbb{I}(x \leq \theta) - \frac{1}{2}$.

We start by proving consistency. This is a 1-dimensional Z-estimator, where the estimating function is decreasing in the parameter θ and has exactly one root. The sample estimating equation converges point wise to the population estimating equation by WLLN. We also know that for the population median equal to θ_0 and small $\epsilon > 0$,

$$P_0(\mathbb{I}(x \leq \theta_0 + \epsilon) - 0.5) < 0 < P_0(\mathbb{I}(x \leq \theta_0 - \epsilon) - 0.5)$$

Thus, $m_n \overset{p}{\to} \theta_0$.

Now we characterize the asymptotic distribution of the sample median under P_{θ_0}. Checking the conditions for asymptotic normality, we know that

1. *The estimating function z_θ is squared differentiable because it is bounded.*
2. *Pz_θ is differentiable at θ_0:*

$$\frac{\partial}{\partial\theta} P z_\theta\Big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta} P\left(\mathbb{I}(x \leq \theta) - \frac{1}{2}\right)\Big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta} F_x(\theta) - \frac{1}{2}\Big|_{\theta=\theta_0} = f_x(\theta_0)$$

   Recalling the form of the normal density, $f_x(\theta_0) = \frac{1}{\sqrt{2\pi}}$.

3. *$\{z_\theta(x) : \theta \in \mathbb{R}\}$ forms a Donsker class because it is a shifted indicator function, and indicator functions are VC class.*

Under these conditions

$$\sqrt{n}(m_n - \theta_0) = -V_{\theta_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{\theta_0}(X_i) + o_P(1) \overset{P_{\theta_0}}{\rightsquigarrow} N\left(0, V_{\theta_0}^{-1} P_0[z_{\theta_0} z_{\theta_0}^T](V_{\theta_0}^{-1})^T\right)$$

$$\equiv N\left(0, \frac{\mathbb{E}((\mathbb{I}(X \leq \theta_0) - 0.5)^2}{\left(\frac{1}{\sqrt{2\pi}}\right)^2}\right) \equiv N(0, \pi/2)$$

Lastly, we investigate the distribution of m_n under the local alternative. This relies on the applying Le Cam's third lemma for asymptotic linear estimators. For an asymptotic linear estimator with influence function ϕ_θ, the asymptotic distribution under the local alternative is

$$\sqrt{n}(m_n - \theta_0) \overset{P_{\theta_0+h/\sqrt{n}}}{\rightsquigarrow} N(P_0(\phi_{\theta_0} \cdot \dot{\ell})h, P_0\phi_{\theta_0}^2)$$

Thus, we must evaluate the $P_0(\phi_{\theta_0} \cdot \dot{\ell})$ to learn the limiting distribution.

1. *Recall the influence function of m_n is given by $\phi_0(x) = \frac{\mathbb{I}(x \le \theta_0) - \frac{1}{2}}{f(\theta_0)} = \sqrt{2\pi}\left(\mathbb{I}(x \le \theta_0) - \frac{1}{2}\right) = \sqrt{2\pi}\left(\frac{1}{2} - \mathbb{I}(x \ge \theta_0)\right)$.*
2. *Recall the score is given by $\dot{\ell}(x) = \frac{\partial}{\partial\theta} - \frac{1}{2}\sum_{i=1}^n (X_i - \theta_0)^2 = \sum_{i=1}^n (X_i - \theta_0)$.*

Writing the inner product of these quantities we obtain by using the mean of a positive-restricted normal.

$$\begin{aligned} P_0(\phi_{\theta_0} \cdot \dot{\ell}) &= \int \sqrt{2\pi}\left(\frac{1}{2} - \mathbb{I}(x \le \theta_0)\right)(x - \theta_0)dP_0(x) \\ &= \int \sqrt{2\pi}\left(\mathbb{I}(x \ge \theta_0) - \frac{1}{2}\right)(x - \theta_0)dP_0(x) \\ &= \sqrt{2\pi}\int_0^\infty (x - \theta_0)dP_0(x) = \sqrt{2\pi}\frac{1}{\sqrt{2\pi}} = 1 \end{aligned}$$

Therefore, the sample median m_n is also a regular estimator

$$\sqrt{n}(m_n - \theta_0) \overset{P_{\theta_0+h/\sqrt{n}}}{\rightsquigarrow} N(h, \pi/2) \implies \sqrt{n}\left(m_n - \left(\theta_0 + \frac{h}{\sqrt{n}}\right)\right) \overset{P_{\theta_0+h/\sqrt{n}}}{\rightsquigarrow} N(0, \pi/2)$$

1.5 Calculating Influence Functions

Each of these examples are taken from Chapter 20 in van der Vaart.

Example 1.16 (Mean functional). *Suppose the sample mean $\psi(P_n)$ is the plug-in estimator of the mean functional $\psi(P) = \int x dP(x)$. By the von-Mises expansion, the influence function is*

$$\begin{aligned}\psi'_P(\delta_x - P) &= \frac{d}{d\epsilon}\int x\, d[(1-\epsilon)P + \epsilon\delta_x](x)\Big|_{\epsilon=0}\\ &= x - \int x dP(x)\end{aligned}$$

Example 1.17 (Wilcoxon Mann-Whitney Statistic). *Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ are random sample from a bivariate distribution with empirical distributions F_n and G_n for each margin. The Mann-Whitney Statistic is a plug-in estimator of the functional $\psi_0 = \psi(P, G) = \int F dG$:*

$$\psi(P_n, G_n) = \int F_n dG_n = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathbb{I}(X_i \le Y_j)$$

The influence function of the Mann-Whitney statistic can also be calculated from the von-Mises expansion

$$\begin{aligned}\psi'_P(\delta_x - F, \delta_y - G) &= \frac{d}{d\epsilon}\int ((1-\epsilon)F + \epsilon\delta_x) d\left[(1-\epsilon)G + \epsilon\delta_y\right]\Big|_{\epsilon=0}\\ &= \frac{d}{d\epsilon}\int (1-\epsilon)^2 F dG + \int (1-\epsilon)\epsilon F d\delta_y + \int \epsilon(1-\epsilon)\delta_x dG + \int \epsilon^2 \delta_x d\delta_y\Big|_{\epsilon=0}\\ &= F(y) + 1 - G_-(x) - 2\int F dG\end{aligned}$$

Example 1.18 (Z estimators). *The Z-estimator $\psi(P_0)$ is the solution to the population-based estimating equation $P_0 z_{\psi(P_0)} = 0$. Differentiating with respect to ϵ across the identity*

$$0 = (P + \epsilon\delta_x) z_{\psi(P+\epsilon\delta_x)} = P z_{\psi(P+\epsilon\delta_x)} + \epsilon z_{\psi(P+\epsilon\delta_x)}(x)$$

Assumin the derivatives exist and z_ψ is continuous, we have that

$$0 = \left(\frac{\partial}{\partial\theta} P z_\theta\right)_{\theta=\psi(P)} \left[\frac{d}{dt}\psi(P + t\delta_x)\right]_{t=0} + z_{\psi(P)}(x)$$

Where the expression in parentheses is the influence function and is given by

$$-\left(\frac{\partial}{\partial\theta} P z_\theta\right)^{-1}_{\theta=\psi(P)} z_{\psi(P)}(x)$$

Example 1.19 (Quantiles). *The p-th quantile of distribution function F is $\psi(F) = F^{-1}(p)$. We set $F_\epsilon = (1-\epsilon)F + \epsilon\delta_x$ and differentiate wrt ϵ the identity*

$$p = F_\epsilon F_\epsilon^{-1}(p) = (1-t)F(F_\epsilon^{-1}(p)) + \epsilon\delta_x(F_t^{-1}(p))$$

We find that

$$0 = -F(F^{-1}(p)) + f(F^{-1}(p))\left[\frac{d}{d\epsilon}F_\epsilon^{-1}(p)\right]_{t=0} + \delta_x(F^{-1}(p))$$

Where the influence function is given by

$$\left[\frac{d}{d\epsilon}F_\epsilon^{-1}(p)\right]_{t=0} = \psi'_F(\delta_x - F) = -\frac{\mathbb{I}(x \le F^{-1}(p)) - p}{f(F^{-1}(p))}$$

This implies that the sequence of empirical quantiles is asymptotically normal

$$\sqrt{n}(F_n^{-1}(t) - F^{-1}(t)) \rightsquigarrow N\left(0, P_0\left[-\frac{\mathbb{I}(x \le F^{-1}(p)) - p}{f(F^{-1}(p))}\right]^2\right) \equiv N\left(0, \frac{p(1-p)}{f(F^{-1}(p))^2}\right)$$

Example 1.20 (Cramer-von Mises statistic: higher order expansion)**.** *The Cramer-von Mises statistic* $\psi(F_n)$ *estimates the following parameter* $\psi(F) = \int (F - F_0)^2 dF_0$ *for some fixed* F_0*. The von-Mises expansion yields*

$$\psi(F + \epsilon H) = \int (F + \epsilon H - F_0)^2 dF_0 = \int (F - F_0)^2 dF_0 + 2\epsilon \int (F - F_0) H dF_0 + \epsilon^2 \int H^2 dF_0$$

The first derivative of the above form takes the form from F *along path* H *is given by:*

$$\frac{\partial}{\partial \epsilon} \psi(F + \epsilon H) = 2 \int (F - F_0) H dF_0$$

Plugging in $\epsilon = 1/\sqrt{n}$ *and* $H = \mathbb{G}_n = \sqrt{n}(F - F_0)$*, we have*

$$\psi'(F) \equiv \psi'(F_0 + \epsilon H) \equiv 2 \int (F_0 - F_0) H dF_0 = 0$$

Therefore, first order expansion is degenerate. To determine the asymptotic distribution, we must go to the second order derivative

$$\psi''_{F_0}(H) = \frac{\partial^2}{\partial \epsilon^2} \psi(F + \epsilon H)\Big|_{\epsilon=0} = 2 \int H^2 dF_0$$

Which for $\epsilon = 1/\sqrt{n}$ *and* $H = \mathbb{G}_n = \sqrt{n}(F - F_0)$ *produces*

$$\psi''_{F_0}(\mathbb{G}_n) = 2 \int \mathbb{G}_n^2 dF_0$$

The von Mises expansion suggests the following approximation

$$\begin{aligned} \psi(F_n) - \psi(F) &= \cancelto{0}{\frac{1}{\sqrt{n}} \psi'_{F_0}(\mathbb{G}_n)} + \frac{1}{2!} \frac{1}{n^{2/2}} \psi''_{F_0}(\mathbb{G}_n) + \ldots \\ &\approx \frac{1}{n} \int \mathbb{G}_n^2 dF_0 \end{aligned}$$

1.6 Semiparametric/Nonparametric Inference

1.6.1 Function-valued parameters

Example 1.21 (Uniform confidence bands for CDF)**.** *Suppose our goal is to construct confidence bands for the CDF $F_0(t)$ uniformly over all $t \in \mathbb{R}$. We estimate $F_0(t)$ with the class of functions $\mathcal{F} := \{x \to \mathbb{I}(x \leq t) : t \in \mathbb{R}\}$. By Donsker's Theorem and the continuous mapping theorem,*

$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \textit{ in } \ell^\infty(\mathcal{F})$$
$$||\mathbb{G}_n||_\mathcal{F} \rightsquigarrow ||\mathbb{G}||_\mathcal{F}$$

Our goal of constructing valid confidence bands is equivalent to finding $\{L_n(t), U_n(t)\}$ such that

$$\lim_{n\to\infty} P(L_n(t) \leq F_0(t) \leq U_n(t)) \geq 1-\alpha \quad \forall\, t \in \mathbb{R}$$

We propose the following bounds where c is the $(1-\alpha)$-quantile of $||\mathbb{G}||_\mathcal{F}$

$$L_n(t) := F_n(t) - \frac{c}{\sqrt{n}} \quad U_n(t) := F_n(t) - \frac{c}{\sqrt{n}}$$

These bounds are asymptotically valid because

$$\begin{aligned}
&\lim_{n\to\infty} P_0(L_n(t) \leq F_0(t) \leq U_n(t)) \qquad \forall t \in \mathbb{R}\\
&= \lim_{n\to\infty} P_0\left(F_n(t) - \frac{c}{\sqrt{n}} \leq F_0(t) \leq F_n(t) - \frac{c}{\sqrt{n}}\right) \qquad \forall t \in \mathbb{R}\\
&= \lim_{n\to\infty} P_0\left(-c \leq \sqrt{n}(F_0(t) - F_n(t)) \leq c\right) \qquad \forall t \in \mathbb{R}\\
&= \lim_{n\to\infty} P_0\left(\sqrt{n}|F_n(t) - F_0(t)| \leq c\right) \qquad \forall t \in \mathbb{R}\\
&= \lim_{n\to\infty} P_0\left(\sup_t \sqrt{n}|F_n(t) - F_0(t)| \leq c\right)\\
&= \lim_{n\to\infty} P_0\left(\sup_{f\in\mathcal{F}} \sqrt{n}|(P_n - P_0)h| \leq c\right)\\
&= \lim_{n\to\infty} P_0\left(||\mathbb{G}_n||_\mathcal{F} \leq c\right)\\
&= P_0\left(||\mathbb{G}||_\mathcal{F} \leq c\right)\\
&= (1-\alpha)
\end{aligned}$$

1.6.2 Establishing Asymptotic Linearity

Example 1.22 (Coefficient of Variation is ALE)**.** *Let* $C_n := \frac{\sigma_n}{\mu_n}$ *be the plug-in estimator for* $C_0 := \frac{\sigma_0}{\mu_0}$*. Let* $h(u,v) = u^{1/2}v^{-1}$ *and let* $C_0 := h(\sigma_0^2, \mu_0), C_n := h(\sigma_n^2, \mu_n)$*. We know that:*

$$\begin{pmatrix}\sigma_n^2\\ \mu_n\end{pmatrix} - \begin{pmatrix}\sigma_0^2\\ \mu_0\end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix}(X_i-\mu_0)^2-\sigma_0^2\\ (X_i-\mu_0)\end{pmatrix} + o_P(n^{-1/2})$$

By the Delta Method for ALEs/Influence Functions, we have

$$\begin{aligned}
C_n - C_0 = h(\sigma_n^2,\mu_n) - h(\sigma_0^2,\mu_0) &= \frac{1}{n}\sum_{i=1}^n \left\langle \nabla h(\sigma_0^2,\mu_0)^T, \begin{pmatrix}(X_i-\mu_0)^2-\sigma_0^2\\ (X_i-\mu_0)\end{pmatrix}\right\rangle + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^n \left\langle \left(\frac{1}{2\sigma_0\mu_0}, -\frac{\sigma_0}{\mu_0^2}\right)^T \begin{pmatrix}(X_i-\mu_0)^2-\sigma_0^2\\ (X_i-\mu_0)\end{pmatrix}\right\rangle + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^n \frac{(X_i-\mu_0)^2-\sigma_0^2}{2\mu_0\sigma_0} - \frac{\sigma_0(X_i-\mu_0)}{\mu_0^2} + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^n \frac{\mu_0(X_i-\mu_0)^2 - \mu_0\sigma_0^2 - 2\sigma_0^2(X_i-\mu_0)}{2\mu_0^2\sigma_0} + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^n C_0\left[\frac{\mu_0^2(X_i-\mu_0)^2 - \mu_0^2\sigma_0^2 - 2\mu_0\sigma_0^2(X_i-\mu_0)}{2\mu_0^2\sigma_0^2}\right] + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^n C_0\left[\frac{1}{2}\left(\frac{X_i-\mu_0}{\sigma_0}\right)^2 - \frac{X_i}{\mu_0} + \frac{1}{2}\right] + o_P(n^{-1/2})
\end{aligned}$$

So C_n *is asymptotically linear with influence function* $\phi_{P_0}(x) := C_0\left[\frac{1}{2}\left(\frac{x-\mu_0}{\sigma_0}\right)^2 - \frac{x}{\mu_0} + \frac{1}{2}\right]$

Example 1.23 (Average Absolute Deviation from Mean is ALE). *Suppose we wish to infer about* $\psi_0 := \int P_0|x-\mu_0|$ *for* $\mu_0 = \int x dP_0(x)$. *Consider the plug-in estimator:*

$$\psi_n := \frac{1}{n}\sum_{i=1}^n |X_i - \bar{X}_n|$$

Noting that $\psi_n = P_n f_n$ *and* $\psi_0 = P_0 f_0$ *for* $f_n(x) = |x-\bar{X}_n|$ *and* $f_0(x) = |x-\mu_0|$, *we write the following expansion*

$$\psi_n - \psi_0 = (P_n - P_0)f_0 + P_0(f_n - f_0) + (P_n - P_0)(f_n - f_0)$$

Where term 1 is linear. The other two terms require further inspection. Let's study term 2. Letting $h(u) : u \to \int |x-u| dP_0(x)$, *we have that*

$$P_0(f_n - f_0) = h(\bar{X}_n) - h(\mu_0) = h'(\mu_0)(\bar{X}_n - \mu_0) + o_P(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^n h'(\mu_0)(X_i - \mu_0) + o_P(n^{-1/2})$$

By the delta method. Let $F_0(u) := \int \mathbb{I}(x<u)dP_0(x)$ *and* $G_0(u) = \int \mathbb{I}(x<u)x dP_0(x)$. *Then* $h(u)$ *is given by*

$$\begin{aligned}
h(u) &= \int |x-u| dP_0(x) = \int (u-x)\mathbb{I}(x<u)dP_0(x) + \int (x-u)\mathbb{I}(x>u)dP_0(x) \\
&= \int (u-x)\mathbb{I}(x<u)dP_0(x) + \int (x-u)(-\mathbb{I}(x<u)+1)dP_0(x) \\
&= uF_0(u) - G_0(u) + \int (u-x)\mathbb{I}(x<u)dP_0(x) + \int (x-u)dP_0(x) \\
&= 2uF_0(u) - 2G_0(u) - u + \mu_0 \\
&= u[2F_0(u) - 1] + [\mu_0 - 2G_0(u)]
\end{aligned}$$

Therefore

$$h'(u) = 2F_0(u) - 1 \implies h'(\mu_0) = 2F_0(\mu_0) - 1$$

Now we study term 3. Note that $|(f_n - f_0)| = ||x-\bar{X}_n| - |x-\mu_0|| \le |\mu_0 - \bar{X}_n|$. *Therefore, the total variational norm of* $(f_n - f_0) \le 2|\mu_0 - \bar{X}_n|$. *The WLLN says that there will exist a constant* $K < \infty$ *such that* $|\mu_0 - \bar{X}_n| < K$ *w.p. 1. Thus, the function class is bounded in total variation and is therefore Donsker. Therefore,*

$$(P_n - P_0)(f_n - f_0) = o_P(n^{-1/2})$$

The result is that

$$\psi_n - \psi_0 = \frac{1}{n}\sum_{i=1}^n [|X_i - \mu_0| - \psi_0 + [2F_0(\mu_0) - 1](X_i - \mu_0)] + o_P(n^{-1/2})$$

Example 1.24 (IPW Estimator is ALE)**.** *Suppose $X = (Y, \Delta, W)$ with Y the outcome of interest only observed when $\Delta = 1$. W are covariates. Suppose we wish to infer about the mean of Y. If the missingness mechanism only depends on W (MAR), the mean outcome is*

$$\psi_0 = E_0[E_0(Y|\Delta = 1, W)]$$

Let $\tilde{Q}_0(w) := E_0(Y|\Delta = 1, W = w), g_0(w) := P_0(\Delta = 1|W = w), Q_{W,0}(w) := P_0(W \leq w)$. We can now write

$$\psi_0 = E_0[\tilde{Q}_0(W)] = E_0\left[E_0\left[\frac{\Delta Y}{g_0(W)}|Y\right]\right]$$

Case 1: If g_0 is known, this motivates the following plug-in estimator:

$$\psi_{0,n} := \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i Y_i}{g_0(W_i)} = P_n f_0$$

Which is linear with influence function $\phi_{P_0}(x) : x \to \frac{\delta y}{g_0(w)} - \psi_0$.
Case 2: If the missingness probability is unknown, but is known to lie in a parametric model $\{g_\theta : \theta \in \Theta\}$ with $g_0 = g_{\theta_0}$. Suppose we have an ALE θ_n for θ_0 with influence function φ_{θ_0}. Letting $g_n := g_{\theta_n}$ and $f_n(x) := \frac{\delta y}{g_n(w)}$, we can consider the new plug-in estimator

$$\psi_n := \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i Y_i}{g_n(W_i)} = P_n f_n$$

To show this estimator is asymptotically linear, examine the expansion

$$\psi_n - \psi_0 = (P_n - P_0)f_0 + P_0(f_n - f_0) + (P_n - P_0)(f_n - f_0)$$

Study term 2. First note that $g_n(w) - g_0(w) = \frac{\partial}{\partial\theta}g_\theta(w)\Big|_{\theta=\theta_0}(\theta_n - \theta) + o_P(n^{-1/2})$ by Taylor expansion. Now Term 2 takes form

$$\begin{aligned}
P_0(f_n - f_0) &= \int \tilde{Q}_0(w,1)g_0(w)\left[\frac{1}{g_n(w)} - \frac{1}{g_0(w)}\right]Q_{W,0}(dw)\\
&= -\int \tilde{Q}_0(w,1)\frac{1}{g_0(w)}\left[g_n(w) - g_0(w)\right]Q_{W,0}(dw) + o_P(n^{-1/2})\\
&= -\int \tilde{Q}_0(w,1)\frac{1}{g_0(w)}\left[\frac{\partial}{\partial\theta}g_\theta(w)\Big|_{\theta=\theta_0}(\theta_n - \theta)\right]Q_{W,0}(dw) + o_P(n^{-1/2})\\
&= -\int \tilde{Q}_0(w,1)\frac{1}{g_0(w)}\left[\frac{\partial}{\partial\theta}g_\theta(w)\Big|_{\theta=\theta_0}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_{\theta_0}(X_i)\right)\right]Q_{W,0}(dw) + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n}\gamma_0\varphi_{\theta_0}(X_i) + o_P(n^{-1/2})
\end{aligned}$$

For $\gamma_0 = -\int \tilde{Q}_0(w,1)\frac{1}{g_0(w)}\frac{\partial}{\partial\theta}g_\theta(w)\Big|_{\theta=\theta_0}Q_{W,0}(dw)$.
Term 3 requires that $(f_n - f_0)$ falls in a Donsker class with probability approaching 1.
Under this condition ψ_n is asymptotically linear with influence function

$$\phi^*_{P_0}(x) := \phi_{P_0}(x) + \gamma_0\varphi_{\theta_0}(X_i)$$

Thus if θ_n is an asymptotically linear (and parametric efficient) estimator of θ_0, we can obtain smaller variance than the Case 1 estimator!

Example 1.25 (Robust Mean is ALE (P6 Theory Exam 2021))**.** *Let $X_1, \dots, X_n{}^{iid}P_0 \in M$ where M is the nonparametric model with finite second moment and strictly positive density on the nonnegative real numbers. We wish to estimate $\psi_0 = \psi(F_0)$*

$$\psi(F) := \mathbb{E}_F[X\mathbb{I}(X \leq Q_\beta(F))] = \int_0^{Q_\beta(F)} udF(u)$$

Where $Q_\beta(F)$ is the β-quantile of F. Let $\mu_0 := \mu(F_0), \mu_n := \mu(F_n)$ and $q_0 := Q_\beta(F_0)$. Also note that the Gateaux derivative of Q_β at F in direction h is given by $\dot{Q}_\beta(F;h) = \frac{-h(Q_\beta(F))}{f(Q_\beta(F))}$.

1. ***Calculate Gauteaux derivative of** ψ. Using the fundamental theorem of calculus and product rule, the Gauteaux derivative is defined as*

$$\begin{aligned}
\dot{\psi}(F;h) &= \frac{d}{d\epsilon}\psi(F+\epsilon h)\Big|_{\epsilon=0} \\
&= \frac{d}{d\epsilon}\left[\int_0^{Q_\beta(F+\epsilon h)} udF(u) + \epsilon\int_0^{Q_\beta(F+\epsilon h)} udh(u)\right]\Big|_{\epsilon=0} \\
&= \frac{d}{d\epsilon}\left[\int_0^{Q_\beta(F+\epsilon h)} uf(u)du + \epsilon\int_0^{Q_\beta(F+\epsilon h)} udh(u)\right]\Big|_{\epsilon=0} \\
&= Q_\beta(F+\epsilon h)f(Q_\beta(F+\epsilon h))\cdot\dot{Q}_\beta(F;h) + \int_0^{Q_\beta(F+\epsilon h)} udh(u) + \epsilon\left(Q_\beta(F+\epsilon h)\dot{Q}_\beta(F;h)\right)\Big|_{\epsilon=0} \\
&= Q_\beta(F+\epsilon h)\left(f(Q_\beta(F+\epsilon h))\right)\left(\frac{-h(Q_\beta(F))}{f(Q_\beta(F))}\right) + \int_0^{Q_\beta(F+\epsilon h)} udh(u) + \epsilon\left(Q_\beta(F+\epsilon h)\dot{Q}_\beta(F;h)\right)\Big|_{\epsilon=0} \\
&= \int_0^{Q_\beta(F)} udh(u) - Q_\beta(F)h(Q_\beta(F))
\end{aligned}$$

2. ***Asymptotic Linearity and Influence Function**. By the Functional Delta method, we know that $\psi_n = \psi(F_n)$ is asymptotically linear with influence function equal to the Gauteax derivative (under Hadamard differentiability wrt supremum norm):*

$$\psi(F_n) - \psi(F_0) = \frac{1}{n}\sum_{i=1}^n \dot{\psi}(F_0; \mathbb{I}(X_i \leq \cdot) - F_0) + o_P(n^{-1/2})$$

To calculate the influence function, we look to part (a). Recalling $q_0 = Q_\beta(F_0)$:

$$\begin{aligned}
\dot{\psi}(F_0; \mathbb{I}(X_i \leq \cdot) - F_0) &= \int_0^{Q_\beta(F_0)} u\ (\mathbb{I}(X_i \leq \cdot) - F_0)(du) - Q_\beta(F_0)\left[\mathbb{I}(X_i \leq Q_\beta(F_0)) - F_0(Q_\beta(F))\right] \\
&= \int_0^{q_0} u\ (\mathbb{I}(X_i \leq \cdot) - F_0)(du) - q_0\left[\mathbb{I}(X_i \leq q_0) - \beta\right] \\
&= \int_0^{q_0} u\ d(\mathbb{I}(X_i \leq u)) - \psi_0 - q_0\mathbb{I}(X_i \leq q_0) + \beta q_0 \\
&= (x - q_0)\mathbb{I}(X_i \leq q_0) - \psi_0 + \beta q_0
\end{aligned}$$

3. *Show that $\sqrt{n}(\mu_n - \mu_0)$ where μ_n is the sample mean. We solve for the variance of X using the law of total variance with $A_1 = \mathbb{I}(x \leq q_0), A_2 = \mathbb{I}(x > q_0)$ which partition the outcome space.*

$$\begin{aligned}
Var(X) &= \sum_{i=1}^2 P(A_i)\cdot Var(X|A_i) + \left[\sum_{i=1}^2 \mathbb{E}[X|A_i]^2[1-P(A_i)][P(A_i)]\right] - 2\mathbb{E}[X|A_1]P(A_1)\mathbb{E}[X|A_2]P(A_2) \\
&= \beta\, Var(X|X \leq q_0) + (1-\beta)\, Var(X|X > q_0) + \beta(1-\beta)\left(\mathbb{E}[X|X \leq q_0] - \mathbb{E}[X|X > q_0]\right)^2
\end{aligned}$$

4. *Compare the asymptotic variances of $\sqrt{n}(\psi_n - \psi_0)$ to $\sqrt{n}(\mu_n - \mu_0)$ via the asymptotic relative efficiency: comparing the squares of each of the influence functions:*

$$\frac{\mathbb{E}\left([(x-q_0)\mathbb{I}(x \leq q_0)]^2 - 2(x-q_0)\mathbb{I}(x \leq q_0)(\psi_0 - \beta q_0) + (\psi_0 - \beta q_0)^2\right)}{\mathbb{E}\left([x-\mu_0]^2\right)}$$

$$= \frac{\mathbb{E}\left(\{x\mathbb{I}(x \leq q_0) - \psi_0 - (q_0\mathbb{I}(x \leq q_0) - \beta q_0)\}^2\right)}{\mathbb{E}\left([x-\mu_0]^2\right)}$$

$$= \frac{\mathbb{E}\left(\{x\mathbb{I}(x \leq q_0) \ - \psi_0\}^2 - 2\{x\mathbb{I}(x \leq q_0) \ - \psi_0\}\{q_0\mathbb{I}(x \leq q_0) - \beta q_0)\} + \{q_0\mathbb{I}(x \leq q_0) - \beta q_0)\}^2\right)}{\mathbb{E}\left([x-\mu_0]^2\right)}$$

$$= \frac{Var(X|X \leq q_0) - 2\psi_0 q_0 + 2\psi_0\beta q_0 + 2\psi_0\beta q_0 - 2\psi_0\beta q_0 + q_0^2\beta - 2\beta^2 q_0^2 + \beta^2 q_0^2}{Var(X)}$$

$$= \frac{Var(X|X \leq q_0) - 2\psi_0 q_0(1-\beta) + q_0^2\beta(1-\beta)}{Var(X)} = \frac{Var(X|X \leq q_0) - q_0(1-\beta)(2\psi_0 - q_0\beta)}{Var(X)}$$

Thus, as long as $(2\psi - q_0\beta) > 0$ then we are assured a reduction in variance compared to the sample mean. This makes sense because the influence function of the trimmed mean is bounded and therefore is robust to outliers.

Example 1.26 (Absolute Mean Difference and Gini Index are ALE (Theory Exam 2020 P7)). *Suppose $X_1, \ldots, X_n \overset{iid}{\sim} P \in M$ where M is nonparametric model on $(0, \infty)$. Let target of inference be $\delta_0 = \Delta(P_0)$ defined as*

$$\Delta(P) = \mathbb{E}_P|X_1 - X_2|$$

With X_1, X_2 independent draw from P.

(a) Show $\delta_n = \Delta(P_n)$ is asymptotically linear, determine its influence function, and derive its large sample distribution.

A convenient way of writing the estimand is

$$\Delta(P) := \int \int |X_1 - X_2| dP(X_1) dP(X_2)$$

A natural place to start is the following V statistic

$$V_n := \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i - X_j|$$

Which by linearization argument has an asymptotic distribution that is dominated by

$$2(P_n - P_0) \int |X - u| dP_0(x)$$

$$\implies V_n - \delta_0 = \frac{1}{n} \sum_{i=1}^{n} 2 \left(\int |X - u| dP_0(x) - \delta_0 \right) + o_P(n^{-1/2})$$

However, the more appropriate estimator, and the estimator that is actually equivalent to $\delta_n = \Delta(P_n)$ would be the U-statistic

$$U_n := \binom{n}{2} \sum_{i=1}^{n} \sum_{i<j} |X_i - X_j|$$

Which by a linearization argument has the same influence function as the V-statistic

$$\phi_{P_0}(x) = 2 \left(\int |x - u| dP_0(u) - \delta_0 \right)$$

By asymptotic linearity large sample asymptotic distribution of δ_n is normal with variance given by the variance of the influence function

$$\sqrt{n}(\delta_n - \delta_0) \rightsquigarrow N \left(0, 4\mathbb{E} \left[\left(\int |x - u| dP_0(u) - \delta_0 \right)^2 \right] \right)$$

$$\equiv N \left(0, 4 \left[\mathbb{E} \left[\left(\int |x - u| dP_0(u) \right)^2 \right] - \delta_0^2 \right] \right)$$

(b) Define the following parameter, the Gini Index

$$\Psi(P) := \frac{\mathbb{E}_P|X_1 - X_2|}{2\mathbb{E}_P(X)}$$

We can show $\Psi(P_n)$ is asymptotically linear for Ψ_0 by using the delta method for asymptotic linear estimators. We lay out the following ingredients that will be useful in our calculation *(a)* $\phi_1(x) = 2\left(\mathbb{E}_0|X-x| - \Delta(P_0)\right)$ *(part (a)*

(b) $\phi_2(x) = (x-\mu_0)$

(c) $\frac{\partial \Psi(P_0)}{\partial \mathbb{E}_P|X_1 - X_2|} = \frac{1}{2\mu_0}$

(d) $\frac{\partial \Psi(P_0)}{\partial \mu_0} = -\frac{\mathbb{E}_P|X_1-X_2|}{2\mu_0^2}$

From the delta method for ALEs we have

$$\begin{aligned}
\tilde{\phi}(x) &= \langle \nabla f(\Psi(P_0)), \phi_{P_0}\rangle \\
&= \left\langle \left(\frac{1}{2\mu_0}, -\frac{\mathbb{E}_P|X_1-X_2|}{2\mu_0^2}\right), \left(2\left(\mathbb{E}_0|X-x| - \Delta(P_0)\right), (x-\mu_0)\right)\right\rangle \\
&= \frac{2\left(\mathbb{E}_0|X-x| - \Delta(P_0)\right)}{2\mu_0} - \frac{(x-\mu_0)\mathbb{E}_P|X_1-X_2|}{2\mu_0^2} \\
&= \frac{2\mu_0\left(\mathbb{E}_0|X-x| - \Delta(P_0)\right)}{2\mu_0^2} - \frac{(x-\mu_0)\mathbb{E}_P|X_1-X_2|}{2\mu_0^2} = \frac{2\mu_0\left(\mathbb{E}_0|X-x| - \Delta(P_0)\right)}{2\mu_0^2} - \frac{(x-\mu_0)\Delta(P_0)}{2\mu_0^2} \\
&= \frac{2\mu_0\left(\mathbb{E}_0|X-x|\right) - (\mu_0 + x)\Delta(P_0)}{2\mu_0^2} \\
&= \frac{\mathbb{E}_0|X-x| - (\mu_0+x)\psi_0}{\mu_0}
\end{aligned}$$

(c) *Next, consider the submodel* $M_0 = \{P \in M : E_P(X) = \mu_0\}$ *for some known* μ_0*. Derive the tangent space* $T_{M_0}(P_0)$*.*
The following model just imposes a moment restriction $P_0(g_0) = 0$ *where* $g_0(x) = x - \mu_0$*. We can determine the form of the tangent space by exploring this constraint along a linear submodel of the form* $p_\theta(x) = [1+\theta h]p_0(x)$*.*

$$\begin{aligned}
&\int g_0[1+\theta h]f_0 dx = 0 \\
\implies &\cancelto{0}{\mathbb{E}_0[g_0]} + \theta\mathbb{E}_0[g_0 h] = 0 \\
\implies &\mathbb{E}_0[(x-\mu_0)h] = 0 \implies \mathbb{E}_0[xh] = 0 \qquad \textit{(Recalling scores mean-0)}
\end{aligned}$$

(d) *Consider the model* $P_0 \sim Exp(1)$ *with density* $\exp(-x)\mathbb{I}(0 < x < \infty)$*. Using the fact that* $\mathbb{E}_0|X-x| = 2e^{-x} + x - 1$*, determine if the influence function of* ψ_n *is an element of the tangent space of the restricted model* T_{M_0}*.*

Step 1 is to see if $\phi_0(x)$ *lives in* $L_0^2(P_0)$*. First we write some useful facts*

(a) $\mathbb{E}[X] = 1$ *by properties of Exponential.*

(b) $\mathbb{E}(\Psi(P_0)) = \mathbb{E}\left[\frac{\mathbb{E}_P|X_1-X_2|}{2\mathbb{E}_P(X)}\right] = \frac{1}{2}$ *by properties of Exponential.*

(c) $\mathbb{E}_0|X-x| = 2e^{-x} + x - 1$ *(given in problem)*

Therefore,

$$\begin{aligned}
\mathbb{E}[\phi_0(x)] &= \int_0^\infty 2e^{-x}e^{-x}dx + \mathbb{E}_0[X] - 1 - (\mathbb{E}_0[X]+1)\mathbb{E}(\Psi_0) \\
&= \int Exp(2)dx + 1 - 1 + (2)\mathbb{E}[\Psi_0] \\
&= \int 1 - 2\frac{1}{2} = 0
\end{aligned}$$

Step 2 is verify that $\mathbb{E}[x\phi_0(x)] = 0$*, which proves that the influence function lives in the tangent space.*

$$\begin{aligned}\mathbb{E}[x\phi_0(x)] &= \int_0^\infty x2e^{-x}e^{-x}dx + \mathbb{E}_0[X^2] - \mathbb{E}_0[X] - (\mathbb{E}_0[X^2] + \mathbb{E}_0[X])\mathbb{E}(\Psi_0) \\ &= \underbrace{\mathbb{E}_0[Y]}_{(2)} dx + 1 - 1 + (2)\mathbb{E}[\Psi_0] \\ &= \int 1 - 2\frac{1}{2} = 0\end{aligned}$$

Thus, ϕ_0 *is in the tangent space under the model with the moment restriction:* T_{M_0}.

Example 1.27 (Difference in Conditional Mean estimators are ALE (P5 Theory Exam 2020))**.** *Suppose there are n iid draws (A_i, Y_i) from P_0 where $A_i = \{0,1\}$ and $P_0(A=1) = 0.5$. We wish to contrast*

$$\psi_0 = \mathbb{E}_{P_0}(Y|A=1) - \mathbb{E}_{P_0}(Y|A=0)$$

Consider the following two estimators

$$\psi_{1n} = \frac{\sum_{i=1}^n A_i Y_i}{\sum_{i=1}^n A_i} - \frac{\sum_{i=1}^n (1-A_i) Y_i}{\sum_{i=1}^n (1-A_i)}$$
$$\psi_{2n} = 2\left[\frac{1}{n}\sum_{i=1}^n A_i Y_i - \frac{1}{n}\sum_{i=1}^n (1-A_i) Y_i\right]$$

1. *Show both estimators are unbiased and consistent. We demonstrate both conditions first for ψ_{1n}. We focus on the first term and the second holds WLOG:*

$$\mathbb{E}\left[\frac{\sum_{i=1}^n A_i Y_i}{\sum_{i=1}^n A_i}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{i=1}^n A_i Y_i}{\sum_{i=1}^n A_i}\Big|A=1\right]\right] = \mathbb{E}\left[\frac{\sum_{i=1}^n \mathbb{E}[Y|A=1]}{n}\right] = \mathbb{E}[Y|A=1]$$

   To prove consistency, we note that WLLN yields $\frac{1}{n}\sum_{i=1}^n A_i Y_i \xrightarrow{p} \frac{1}{2}\mathbb{E}[Y|A=1]$, $\frac{1}{n}\sum_{i=1}^n A_i \xrightarrow{p} \frac{1}{2}$, implying by the continuous mapping theorem that $\frac{\sum_{i=1}^n A_i Y_i}{\sum_{i=1}^n A_i} \xrightarrow{p} \mathbb{E}[Y|A=1]$.
   To show ψ_{2n} is unbiased and consistent

$$\mathbb{E}\left[2\left[\frac{1}{n}\sum_{i=1}^n A_i Y_i\right]\right] = 2\mathbb{E}[AY] = 2\mathbb{E}[A\mathbb{E}[Y|A=1]] = \mathbb{E}[Y|A=1]$$

   Consistency follows from the WLLN and CMT.

2. *Derive the large sample distributions of $\sqrt{n}(\psi_{1n} - \psi_0)$ and $\sqrt{n}(\psi_{2n} - \psi_0)$. Which has the smaller asymptotic variance?*
   To solve this question we apply the delta method for ALEs. IN the case of ψ_{1n}, we consider the following function $f(a,b,c,d) = \frac{a}{b} - \frac{c}{d}$ and the following estimators with the following influence functions

$$\frac{1}{n}\sum_{i=1}^n A_i Y_i \qquad \phi_1 = ay - 0.5\mathbb{E}[Y|A=1]$$
$$\frac{1}{n}\sum_{i=1}^n A_i \qquad \phi_2 = a - 0.5$$
$$\frac{1}{n}\sum_{i=1}^n (1-A_i) Y_i \qquad \phi_3 = (1-a)y - 0.5\mathbb{E}[Y|A=0]$$
$$\frac{1}{n}\sum_{i=1}^n (1-A_i) \qquad \phi_2 = (1-a) - 0.5$$

   Applying the delta method for ALEs yields the influence function for ψ_{1n}

$$\begin{aligned}\tilde{\phi}_1 &= \langle \nabla f(\psi_0), \phi_{P_0}\rangle \\ &= \frac{ay - 0.5\mathbb{E}[Y|A=1]}{0.5} - \frac{(a-0.5)0.5\mathbb{E}[Y|A=1]}{0.25} \\ &- \frac{(1-a)y - 0.5\mathbb{E}[Y|A=0]}{0.5} - \frac{((1-a)-0.5)0.5\mathbb{E}[Y|A=0]}{0.25} \\ &= 2\left[a(y - \mathbb{E}[Y|A=1])\right] - 2\left[(1-a)(y - \mathbb{E}(Y|A=0))\right]\end{aligned}$$

To solve for the influence function of ψ_{2n} we'll note the estimator is linear so it has influence function

$$\tilde{\phi}_2 = 2(ay - 0.5\mathbb{E}[Y|A=1]) - 2\left([(1-a)y - 0.5\mathbb{E}[Y|A=0]]\right)$$

Asymptotic linearity ensures both distributions will be asymptotically normal with variance equal to the variance of the influence function. Let's compare the asymptotic variances

$$\begin{aligned}
Var(\tilde{\phi}_1) &= 4\,Var(a(y - \mathbb{E}[Y|A=1]) - (1-a)(y - \mathbb{E}(Y|A=0))) \\
&= 4\mathbb{E}[A^2(Y - \mathbb{E}[Y|A=1])^2 + (1-A)^2(Y - \mathbb{E}(Y|A=0))^2] \\
&= 2\left[Var(Y|A=1) + Var(Y|A=0)\right] \\
Var(\tilde{\phi}_2) &= 4\,Var((ay - 0.5\mathbb{E}[Y|A=1]) - ((1-a)y - 0.5\mathbb{E}[Y|A=0])) \\
&= 4\mathbb{E}((ay)^2 + (1-a)^2y^2) \qquad \textit{(Ignore constants)} \\
&= 2(\mathbb{E}[Y^2|A=1] + \mathbb{E}[Y^2|A=0])
\end{aligned}$$

Note that the variance is upper bounded by the second moments. Therefore $Var(\tilde{\psi}_1) < Var(\tilde{\psi}_2)$.

3. *We can construct a 95% asymptotic confidence interval for ψ_0 using ψ_{1n}.*

$$\psi_{1n} \pm 2z_{1-\alpha/2}\frac{\sqrt{2}\times\sqrt{\frac{\sum_{i=1}^n A_i\left(Y_i - \frac{\sum_{j=1}^n A_jY_j}{\sum_{j=1}^n A_j}\right)^2}{\sum_{i=1}^n A_i} + \frac{\sum_{i=1}^n (1-A_i)\left(Y_i - \frac{\sum_{j=1}^n (1-A_j)Y_j}{\sum_{j=1}^n (1-A_j)}\right)^2}{\sum_{i=1}^n (1-A_i)}}}{n}$$

Where the things in parentheses are the plugin formulas for the conditional variances.

4. *Show that the asymptotic covariance between $\sqrt{n}(\psi_{1n} - \psi_0)$ and $\sqrt{n}(\psi_{2n} - \psi_0)$ is equal to the variance of $\sqrt{n}(\psi_{1n} - \psi_0)$.*

$$\begin{aligned}
\mathbb{E}[\tilde{\phi}_1 \cdot \tilde{\phi}_2] &= 4\mathbb{E}[(AY)^2 - A^2\mathbb{E}[Y|A=1]Y - 0.5AY\mathbb{E}[Y|A-1] + 0.5A\mathbb{E}[Y|A=1]^2] \\
&+ 4\mathbb{E}[((1-A)Y)^2 - (1-A)^2\mathbb{E}[Y|A=0]Y - 0.5(1-A)Y\mathbb{E}[Y|A-1] + 0.5(1-A)\mathbb{E}[Y|A=0]^2] \\
&= 4\left(0.5\mathbb{E}[Y^2|A=1] - 0.5\mathbb{E}[Y|A=1]^2 - 0.25\mathbb{E}[Y|A=1] + 0.25\mathbb{E}[Y|A=1]\right) \\
&+ 4\left(0.5\mathbb{E}[Y^2|A=-] - 0.5\mathbb{E}[Y|A=0]^2 - 0.25\mathbb{E}[Y|A=0] + 0.25\mathbb{E}[Y|A=0]\right) \\
&= 2\,Var(Y|A=1) + 2\,Var(Y|A=0)
\end{aligned}$$

1.7 Efficiency Theory and Efficient Estimators

Example 1.28 (Efficient Estimators Under Moment Restriction (P7 Theory Exam 2021))**.** *Suppose* $X_1, \ldots, X_n \overset{iid}{\sim} P_0 \in M$ *where* M *is the nonparametric model of each distribution* P *satisfying* $Pf_0^2 < \infty$ *with support in* $(-B, +B)$ *for a fixed* f_0*. Suppose we wish to estimate the mean of* f_0*:* $\psi_0 = P_0 f_0$*.*

Consider a multivariate $g_0 : \mathbb{R}^m \to \mathbb{R}$ *that is bounded and consider the model containing the collection of distributions with the moment restriction based on the multivariate function* $M_0 := \{P \in M : P^m g_0 = 0\}$*.*

The tangent space of M_0 *at* P *is given by*

$$T_{M_0}(P) := \{h \in L_0^2(P) : \int h(x)\bar{g}_P(X)dP(x) = 0\}$$

Where $\bar{g}_P = P^{m-1}g_0$*.*

(a) ***Derive form of projection onto tangent space****. Consider an arbitrary element* $s^* \in L_0^2(P)$*. The projection* $\Pi(s|T_M)$ *onto the tangent space satisfies the following property.*

$$\langle s - \Pi(s|T_M), a\bar{g}_p(X)\rangle = 0$$

This property ensures that the set of allowable scores $s - \Pi(s|T_M)$ *satisfies the desired moment restriction. Note that the model space is a linear span, so* $\Pi(s|T_M)$ *is given by*

$$\begin{aligned} a &= \underset{a}{argmin} ||s - a\bar{g}_p||^2_{L^2(P)} \\ \implies \frac{\partial}{\partial a} &\left[P(s^2) - 2aP(s\bar{g}_p) + a^2 P(\bar{g}_p^2)\right] = 0 \\ \implies a^* &= \frac{\int s\bar{g}_P dP}{\int \bar{g}_P^2 dP} \end{aligned}$$

Taken together implying the form of the projection onto the tangent space is

$$\begin{aligned} s^* &= s(x) - a^*\bar{g}_p \\ &= s(x) - \frac{\int s\bar{g}_P dP}{\int \bar{g}_P^2 dP}\bar{g}_p \end{aligned}$$

The last condition we need to check is that s^* *actually lives in the tangent space. To verify this*

$$\begin{aligned} \int s^*(x)\bar{g}_P(x)dP(x) &= \int \left(s(x) - \frac{\int s\bar{g}_P dP}{\int \bar{g}_P^2 dP}\bar{g}_P\right)\bar{g}_P(x)dP(x) \\ &= \int (s\bar{g}_P(x) - s\bar{g}_P)dP = 0 \end{aligned}$$

(b) ***Canonical gradient****: Note that* ψ_0 *has nonparametric influence function* $\phi(x) = f_0(x) - \psi_0$ *where* $\mu_0 := \mathbb{E}[X]$*. Using the fact above, we have that the canonical gradient/EIF is obtained by projecting* $\phi(x)$ *onto* T_M*.*

$$\begin{aligned} \phi^*(x) &= f_0(x) - \psi_0 - \frac{\int (f_0(x) - \psi_0)\bar{g}_P dP}{\int \bar{g}_P^2(x)dP}\bar{g}_p(x) \\ &= f_0(x) - \psi_0 - \frac{\int f_0(x)\bar{g}_P dP - \overbrace{\int \psi_0\bar{g}_P dP}^{0}}{\int \bar{g}_P^2(x)dP}\bar{g}_p(x) \\ &= f_0(x) - \psi_0 - \frac{\int f_0(x)\bar{g}_P dP}{\int \bar{g}_P^2(x)dP}\bar{g}_p(x) \end{aligned}$$

Where the cancellation occurred because $\int \bar{g}_P dP = 0$ *in* M_0*.*

(c) ***Efficient One-Step Estimator****: an asymptotically efficient estimator in M_0 can be obtained by taking the plug-in estimator and adding the empirical mean of the EIF.*

$$\begin{aligned}\psi^*(P_n) &= \psi(P_n) + P_n\phi^*(x)\\ &= P_n f_0(X_i) + \cancelto{0}{P_n(f_0(X_i) - \psi(P_n))} - P_n\left(\frac{P_n(f_0(X)P_n^{m-1}g_0)}{P_n(P_n^{m-1}g_0)^2}P_n^{m-1}g_0\right)\\ &= P_n f_0(X_i) - \frac{P_n(f_0(X)\bar{g}_n)}{P_n(\bar{g}_n)^2}P_n(P_n^{m-1}g_0)\end{aligned}$$

Let's inspect the last term. Recognizing g_0 is P_0-mean-zero, by linearization we have:

$$\begin{aligned}P_n^m g_0 &= (P_n^m - P_0^m)g_0\\ &= m(P_n - P)\bar{g}_P(x) + o_P(n^{-1/2}) = mP_n\bar{g}_P(x) + o_P(n^{-1/2})\\ \implies P_n\bar{g}_P &= \frac{P_n^m g_0}{m}\end{aligned}$$

Substituting in the original expression, we have an

$$\psi^*(P_n) = P_n f_0(X_i) - \frac{1}{m}\frac{P_n(f_0(X)\bar{g}_n)}{P_n(\bar{g}_n)^2}P_n^m g_0$$

Which is the efficient one-step estimator.

(d) ***Example:*** *If $\mathbb{E}[(X-\mu_0)^3] = 0$, show the sample mean is efficient for the population mean in a population with known variance.*
We know the sample mean P_nX has nonparametric influence function of $\phi(x) = x - \mu_0$. To show it is efficient in M_0, we must show that the second term in the EIF derived above is 0. Define $g_0(x_1, x_2) = \frac{1}{2}(x_1 - x_2) - \sigma^2$ (known) such that $P^2 g_0 = 0$.

$$\begin{aligned}&\int(x_1-\mu_0)\left[\int\frac{1}{2}(x_1-x_2)^2 - \sigma^2 dP(x_2)\right]dP(x_1)\\ &= \int(x_1-\mu_0)\left[\int\frac{1}{2}((x_1-\mu_0)-(x_2-\mu_0))^2 dP(x_2)\right]dP(x_1) - \cancelto{0}{\sigma^2\int(x_1-\mu_0 dP(x_1)}\\ &= \frac{1}{2}\int\int(x_1-\mu_0)\left((x_1-\mu_0)^2 - \cancelto{0}{2(x_1-\mu_0)(x_2-\mu_0)} + (x_2-\mu_0)^2\right)dP(x_1)dP(x_2)\\ &= \frac{1}{2}\left(\int(x_1-\mu_0)^3 dP(x_1) + \int\left(\cancelto{0}{\int(x_1-\mu_0)dP_1(x_1)}\right)(x_2-\mu_0)^2 dP(x_2)\right)\\ &= 0\end{aligned}$$

Where the last step holds because $\mathbb{E}[(X-\mu_0)^3] = 0$. Thus, the nonparametric influence function equals the EIF under this model, so the sample mean is efficient for the population mean.

Example 1.29 (Gateaux Derivatives and Functional Delta Method (583 Final Problem 2))**.** *Suppose $X_1, \ldots, X_n \overset{iid}{\sim} P_0$ with support on $0 < [a, b] < \infty$. We are interested in estimating the harmonic mean parameter*

$$\Psi(F) := \frac{1}{\int \frac{1}{u} dF(u)}$$

1. *Calculate the Gateaux derivative of the functional.*

$$\begin{aligned} \frac{d}{d\epsilon}\Psi(F + \epsilon h)\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}\left[\int \frac{1}{u} d(F + \epsilon h)(u)\right]^{-1}\Big|_{\epsilon=0} \\ &= -\int \frac{1}{u} dh(u) \left[\int \frac{1}{u} d(F + \epsilon h)(u)\right]^{-2}\Big|_{\epsilon=0} \\ &= -[\Psi(F_0)]^2 \int \frac{1}{u} dh(u) \end{aligned}$$

2. *Show $r(F_n - F_0) = o_P(n^{-1/2})$ where $r(h)$ is given by*

$$r(h) = \Psi(F_0 + h) - \Psi(F_0) - \dot{\Psi}(F_0; h)$$

   We could do this by showing the parameter is Hadamard differentiability of Ψ. But we pursue a more simple approach. Letting $h = F_n - F_0$ we have

$$r(F_n - F_0) = \Psi(F_n) - \Psi(F_0) - \dot{\Psi}(F_0; F_n - F_0)$$

   Let's invoke a Taylor expansion about $\Psi(F_0)$ on the first two terms

$$\begin{aligned} \Psi(F_n) - \Psi(F_0) &\approx \dot{\Psi}(F_0; F_n - F_0) \\ &= \left[P_0\left(\frac{1}{X}\right)\right]^{-2} (P_n - P_0)\left(\frac{1}{X}\right) + \mathcal{O}\left(\left((P_n - P_0)\left(\frac{1}{X}\right)\right)^2\right) \end{aligned}$$

   Notice that since the function $1/X$ is monotone and bounded, it is in a Donsker class and therefore $||(\sqrt{n}(P_n - P_0)(1/X))^2|| = \mathcal{O}_P(1)$, implying the term above is $\mathcal{O}(n^{-1})$ or $o_P(n^{-1/2})$. Therefore the desired result holds

$$\Psi(F_n) - \Psi(F_0) - \dot{\Psi}(F_0; F_n - F_0) = o_P(n^{-1/2})$$

3. *Now we can apply the functional delta method to prove that $\Psi(F_n)$ is an asymptotically linear function with influence function*

$$\begin{aligned} \dot{\Psi}(F_0; \delta(x) - F_0) &= -[\Psi(F_0)]^2 \int \frac{1}{X} d(\delta(x) - F_0) \\ &= -[\Psi(F_0)]^2 \left(\frac{1}{x} - \mathbb{E}_{P_0}\left[\frac{1}{X}\right]\right) \end{aligned}$$

4. *To develop a 95% asymptotically valid confidence interval for the harmonic mean $\Psi(F_0)$, we create a confidence interval using the plug-in estimator with a consistent estimator for the variance*

$$\Psi(F_n) \pm z_{1-\alpha/2} \frac{\sqrt{\widehat{Var}(\dot{\Psi})}}{n}$$

$$\widehat{Var}(\dot{\Psi}) := \Psi(F_n)^{-4} \cdot \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{X_i} - \left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{X_j}\right]\right)^2$$