2012 Qualifying Exam Section 1

February 21, 2019

Question 1

Let N be a Poisson random variable with parameter μ, and let $X_1, X_2, \ldots,$ be a sequence of i.i.d. Poisson random variables with parameter λ, where $0 < \mu, \lambda < \infty$ Define $U = 1(N > 0) \sum_{i=1}^{N} X_i$. Do the following:

(1.a)

Show that $E(U) = \mu\lambda$ and $\text{Var}(U) = \mu\lambda(1 + \mu\lambda)$

Proof:
Note that $1(N > 0)N = N$. We have

$$\begin{aligned}
\mathbb{E}(U) &= \mathbb{E}\mathbb{E}U|N \\
&= \mathbb{E}\left\{\mathbb{E}1(N > 0)\sum_{i=1}^{N} X_i|N\right\} \\
&= \mathbb{E}\left\{1(N > 0)\mathbb{E}\sum_{i=1}^{N} X_i|N\right\} \\
&= \mathbb{E}\left\{1(N > 0)N\lambda\right\} \\
&= \lambda\mathbb{E}\left\{1(N > 0)N\right\} \\
&= \mu\lambda
\end{aligned}$$

And

$$
\begin{aligned}
\mathbb{E}\mathrm{Var}(U|N) &= \mathbb{E}\mathrm{Var}\left(1(N>0)\sum_{i=1}^{N} X_i\right) \\
&= \mathbb{E}\left\{1(N>0)\mathrm{Var}\left(\sum_{i=1}^{N} X_i|N\right)\right\} \\
&= \lambda\mathbb{E}(1(N>0)N) \\
&= \mu\lambda
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(\mathbb{E}U|N) &= \mathrm{Var}\left(\mathbb{E}\left\{1(N>0)\sum_{i=1}^{N} X_i|N\right\}\right) \\
&= \mathrm{Var}\left(1(N>0)\mathbb{E}\left\{\sum_{i=1}^{N} X_i|N\right\}\right) \\
&= \mathrm{Var}(1(N>0)N\mu) \\
&= \lambda^2\mathrm{Var}(1(N>0)N) \\
&= \mu\lambda^2
\end{aligned}
$$

Hence,

$$
\mathrm{Var}(U) = \mathbb{E}\mathrm{Var}(U|N) + \mathrm{Var}(\mathbb{E}U|N) = \mu\lambda + \mu\lambda^2 = \mu\lambda(1+\lambda)
$$

(1.b)

In this part, we add a subscript k to the Poisson parameters μ and λ defined above to denote dependence on an integer $k \geq 1$. Specifically let $\mu = \mu_k = k$ and $\lambda = \lambda_k = h/k$, where $0 < h < \infty$ is a fixed scalar. We want to study what happens to U as $k \to \infty$. Let $D_i = 1(X_i = 1)$, for all $i \geq 1$, and define

$$T = 1(N > 0) \sum_{i=1}^{N} D_i$$

Do the following:

(i.)

Derive the limits of $\mathbb{E}(U)$ and $\mathrm{Var}(U)$ as $k \to \infty$.

Solution:

$$\begin{aligned}
\mathbb{E}U &= \mu_k \lambda_k = k\frac{h}{k} = h \to h \text{ as } k \to \infty \\
\mathrm{Var}(U) &= \mu_k \lambda_k (1 + \mu_k \lambda_k) \\
&= k\frac{h}{k}\left(1 + \frac{h}{k}\right) \\
&= h\left(1 + \frac{h}{k}\right) \to h(1+0) = h \text{ as } k \to \infty
\end{aligned}$$

(ii.)

Show that $\text{pr}(X_i \neq D_i) = \lambda_k^2\{1 + o(\lambda_k)\}$ as $k \to \infty$.

Solution:

$$\begin{aligned}
P(X_i \neq D_i) &= P(X_i \neq 1_{(X_i=1)}) \\
&= 1 - P(X_i = D_i) \\
&= 1 - P(\{X_i = 0\} \cup \{X_i = 1\}) \\
&= 1 - e^{-\lambda_k} - e^{-\lambda_k}\lambda_k \\
&= 1 - e^{-\lambda_k}(1 + \lambda_k)
\end{aligned}$$

From the Maclaurin Series expansion of $e^{-\lambda_k}$, we have

$$e^{-\lambda_k} = 1 - \lambda_k + \lambda_k^2/2 + O(\lambda_k^3)$$

Hence,

$$\begin{aligned}
P(X_i \neq D_i) &= 1 - (1 - \lambda_k + \lambda_k^2/2 + O(\lambda_k^3))(1 + \lambda_k) \\
&= 1 - (1 + \lambda_k - \lambda_k - \lambda_k^2 + \lambda_k^2/2 + \lambda_k^3/2 + (1 + \lambda_k)O(\lambda_k^3)) \\
&= \frac{\lambda_k^2}{2} + O(\lambda_k^3) \\
&= \frac{\lambda_k^2}{2}(1 + O(\lambda_k))
\end{aligned}$$

(iii.)

Show that $1(U \neq T) \leq 1(N > 0) \sum_{i=1}^{N} 1(X_i \neq D_i)$ and thus $U - T \to 0$, in probability, as $k \to \infty$.

Solution:
Since the left hand side is always exactly 0 or 1 and the right hand side is a non-negative integer, it suffices to show that $1(N > 0) \sum_{i=1}^{N} 1(X_i \neq D_i) \geq 1$ whenever $1(U \neq T) = 1$. Seeking a contradiction, suppose $1(U \neq T) = 1$ but $1(N > 0) \sum_{i=1}^{N} 1(X_i \neq D_i) < 1$. Since this second term must be a non-negative integer, we have $1(N > 0) \sum_{i=1}^{N} 1(X_i \neq D_i) = 0$ This term is zero iff $N = 0$ or $N > 0$ and $X_i = D_i$ for every i

Case 1: If $N = 0$, then $U = 1(N > 0) \sum_{i=1}^{N} X_i = 0$ and $T = 1(N > 0) \sum_{i=1}^{N} D_i = 0$ so that $U = T$, a contradiction.

Case 2: If $N > 0$ and $X_i = D_i$ for every i, then we must have $\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} D_i$ and hence $U = T$, a contradiction.

Now, we have for any $\epsilon > 0$

$$\begin{aligned}
P(|U - T| > \epsilon) &\leq P(U \neq T) = \mathbb{E}1(U \neq T) \\
&\leq \mathbb{E}\left\{1(N > 0) \sum_{i=1}^{N} 1(X_i \neq D_i)\right\} \\
&= \mathbb{E}\left(1(N > 0)\mathbb{E}\left[\sum_{i=1}^{N} 1(X_i \neq D_i|N\right]\right) \\
&= \lambda_k^2(1 + o(\lambda_k))\mathbb{E}(1(N > 0)N) \\
&= \left(\frac{h}{k}\right)^2 (1 + o(\lambda_k))k \\
&= \frac{h^2}{k}(1 + o(\lambda_k)) \\
&\to 0 \text{ as } k \to \infty
\end{aligned}$$

Hence, the result follows.

(iv.)

Show that $T - \sum_{i=1}^{k} D_i \to 0$, in probability, as $k \to \infty$.

Solution
Note that

$$\left| T - \sum_{i=1}^{k} D_i \right| = \begin{cases} \sum_{i=k+1}^{N} D_i & \text{if } N > k \\ 0 & \text{if } N = k \\ \sum_{i=N+1}^{k} D_i & \text{if } N < k \end{cases}$$

Hence, since $D_i = 1(X_i = 1)$ and $P(X_i = 1) = \lambda_k e^{-\lambda_k}$, and since the D_i are i.i.d., we have

$$\left| T - \sum_{i=1}^{k} D_i \right| \sim \text{Bin}\left(|N - k|, \lambda_k e^{-\lambda_k} \right)$$

Let $\epsilon > 0$. Then

$$\begin{aligned} P\left(\left| T - \sum_{i=1}^{k} D_i \right| \geq \epsilon \right) &\leq \frac{\mathbb{E}\left| T - \sum_{i=1}^{k} D_I \right|}{\epsilon} && \text{(Markov's Inequality)} \\ &= \frac{\mathbb{E}\left\{ \mathbb{E}\left| T - \sum_{i=1}^{k} D_I \right| \,\Big|\, N \right\}}{\epsilon} \\ &= \frac{\mathbb{E}|N - k| \lambda_k e^{-\lambda_k}}{\epsilon} \\ &= \frac{\lambda_k e^{-\lambda_k} \mathbb{E}|N - \mu_k|}{\epsilon} \\ &\leq \frac{\lambda_k e^{-\lambda_k} (\mathbb{E}(N - \mu_k)^2)^{1/2}}{\epsilon} && \text{(Cauchy-Schwartz)} \\ &= \frac{\lambda_k e^{-\lambda_k} \text{Var}(N)^{1/2}}{\epsilon} \\ &= \frac{\frac{h}{k} e^{-\frac{h}{k}} k^{1/2}}{\epsilon} \\ &= \frac{h e^{-\frac{h}{k}}}{k^{1/2} \epsilon} \\ &3 \to 0 \text{ as } k \to \infty \end{aligned}$$

(v.)

Show that U converges in distribution to a Poisson random variable with parameter h, as $k \to \infty$.

Solution: Let $Z_k = \sum_{i=1}^k D_i$.

$$\begin{aligned} P(|U - Z_k| \geq \epsilon) &= P(|U - T + T - Z_k| \geq \epsilon) \\ &\leq P(\{|U - T| \geq \epsilon/2\} \cup \{|T - Z_k| \geq \epsilon/2\}) \\ &\leq P(|U - T| \geq \epsilon/2) + P(|T - Z_k| \geq \epsilon/2) \\ &\to 0 \text{ as } k \to \infty \end{aligned}$$

And so we have $U - Z_k \to_p 0$ as $k \to \infty$.

Note that $D_i \sim \text{Bernoulli}(\lambda_k e^{-\lambda_k})$, so that $Z_k \sim \text{Binomial}(k, p_k)$, where $p_k = \lambda_k e^{-\lambda_k}$. Moreover, we have that as $k \to \infty$,

$$\begin{aligned} p_k = \lambda_k e^{-\lambda_k} &= \frac{h}{k} \exp\left\{-\frac{h}{k}\right\} \to 0 \text{ as } k \to \infty \\ kp_k = k\lambda_k e^{-\lambda_k} &= k\frac{h}{k} \exp\left\{-\frac{h}{k}\right\} = h \exp\left\{-\frac{h}{k}\right\} \to h \text{ as } k \to \infty \end{aligned}$$

Thus, we have that $Z_k \to_d Z \sim \text{Poisson}(h)$ as $k \to \infty$. Now, note that

$$U = U - Z_k + Z_k = (U - Z_k) + Z_k = Z_k + o_p(1) \to_d Z$$

by Slutsky's theorem since $U - Z_k \to_p 0$ and $Z_k \to_d Z$.

(1.c)

We now modify the setting in (b) so that $\mu = \mu_k = h/k$ and $\lambda = \lambda_k = k$. Do the following:

(i.)

Derive the limits of $\mathbb{E}(U)$ and $\mathrm{Var}(U)$ as $k \to \infty$.

Solution:

$$\begin{aligned}
\mathbb{E}U &= \mu_k \lambda_k = \frac{h}{k} k = h \to h \text{ as } k \to \infty \\
\mathrm{Var}(U) &= \mu_k \lambda_k (1 + \mu_k \lambda_k) \\
&= \frac{h}{k} k (1 + k) \\
&= h(1 + k) \to \infty \text{ as } k \to \infty
\end{aligned}$$

(ii.)

Show that $U \to 0$ in distribution as $k \to \infty$.

Solution:
Let F_U denote the CDF of U. We have

$$\begin{aligned}
P(U \le u) &= P\left(1(N > 0) \sum_{i=1}^{N} X_i \le u\right) \\
&= P\left(\sum_{i=1}^{N} X_i \le u | N > 0\right) P(N > 0) + P(0 \le u | N = 0) P(N = 0) \\
&= O(1)(1 - e^{-\mu_k}) + 1(u \ge 0) e^{-\mu_k} \\
&= O(1)(1 - e^{-h/k}) + 1(u \ge 0) e^{-h/k} \\
&= O(1) o(1) + 1(u \ge 0) e^{-h/k} \\
&= o(1) + 1(u \ge 0) e^{-h/k} \to 1(u \ge 0) \text{ as } k \to \infty
\end{aligned}$$

Hence, we have

$$F_U(u) \to 1(u \ge 0) = \begin{cases} 0 & u < 0 \\ 1 & u \ge 0 \end{cases}$$

where we recognize the right hand side to be the CDF of the degenerate random variable 0. Hence, $U \to_d 0$ as $k \to \infty$.

PALOMA'S SOLUTION

$$
\begin{aligned}
P(U=0) &= \mathbb{E}P(U=0|N) \\
&= \mathbb{E}[P(X_1=0)^N] \\
&= \mathbb{E}[e^{-N\lambda_k}] \\
&= \sum_{x=0}^{\infty} e^{-x*\lambda_k}\mu_k^x e^{-\mu_k}/x! \\
&= \sum_{x=0}^{\infty} e^{-\mu_k}(\mu_k(e^{-\lambda_k}))^x/x! \\
&= e^{-\mu_k}e^{\mu_k e^{-\lambda_k}} \\
&\to 1 \text{ as } k \to \infty
\end{aligned}
$$

Question 2

Let $Y_1, \ldots Y_n$ be i.i.d random variables from a distribution with mean μ and finite variance. Due to non-response, we may not be able to observe all the $Yi's$ for these n subjects. Let $R_1, \ldots, R_n$ denote indicator of response, i.e., $R_i = 1$ means that Y_i observed and $R_i = 0$ otherwise. Suppose that we also collect additional information $X_1, ..., X_n$, which are i.i.d random variables, from these n subjects. Assume that R_i and Y_i are independent given X_i and that the random vectors (Y_i, R_i, X_i) are i.i.d. for $i = 1, \ldots, n$. Define $\pi(x) = P(R_i = 1|X_i = x)$ and assume $\pi(x)$ is known and bounded by a positive constant from below for any x in the support of X_i.

(2.a)

A simple estimator for μ is the average of the observed $Y_i's$

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n R_i Y_i}{\sum_{i=1}^n R_i}$$

Derive the asymptotic limit of $\hat{\mu}_1$, denoted by μ_* and give the asymptotic distribution of $\sqrt{n}(\hat{\mu}_1 - \mu_*)$. Leave expressions in the result.

Solution:
Note that $\frac{1}{n}\sum_{i=1}^n R_i Y_i \to_p \mathbb{E}R_1 Y_1 := \mu_{RY}$ as $n \to \infty$ by the weak law of large numbers. Similarly, $\frac{1}{n}\sum_{i=1}^n R_i \to_p \mathbb{E}R_1 := \mu_R$ as $n \to \infty$ By the continuous mapping theorem, it follows that

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n R_i Y_i}{\sum_{i=1}^n R_i} = \frac{\frac{1}{n}\sum_{i=1}^n R_i Y_i}{\frac{1}{n}\sum_{i=1}^n R_i} \to_P \frac{\mu_{RY}}{\mu_R} := \mu_*$$

as $n \to \infty$.

Where

$$\begin{aligned}
\mu_{RY} &= \mathbb{E}R_1 Y_1 \\
&= \mathbb{E}\{\mathbb{E}R_1 Y_1 | X_1\} \\
&= \mathbb{E}\{(\mathbb{E}R_1 | X_1)\mathbb{E}(Y_1 | X_1)\} \because \text{ conditional independence} \\
&= \mathbb{E}\{P(R_1 = 1 | X_1)\mathbb{E}(Y_1 | X_1)\} \\
&= \mathbb{E}\{\pi(X_1)\mathbb{E}(Y_1 | X_1)\}
\end{aligned}$$

and

$$\begin{aligned}
\mu_R &= \mathbb{E}R_1 \\
&= \mathbb{E}\mathbb{E}R_1 | X_1 \\
&= \mathbb{E}P(R_1 = 1 | X_1) \\
&= \mathbb{E}\pi(X_1)
\end{aligned}$$

DELTA METHOD SOLUTION (possibly only proper way to do it).

Let $Z_n = (\frac{1}{n}\sum_{i=1}^n R_iY_i, \frac{1}{n}\sum_{i=1}^n Y_i)^T$. By the weak law of large numbers, we have

$$Z_n \to_p \mu := (\mu_{RY}, \mu_R)$$

as $n \to \infty$, where $\mu_{RY} = \mathbb{E}R_1Y_1$ and $\mu_R = \mathbb{E}R_1$

By the continuous mapping theorem, we have

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n R_iY_i}{\sum_{i=1}^n R_i} = \frac{\frac{1}{n}\sum_{i=1}^n R_iY_i}{\frac{1}{n}\sum_{i=1}^n R_i} \to_P \frac{\mu_{RY}}{\mu_R} = \mu_*$$

since $\pi(x) > 0$.

By the Central Limit Theorem,

$$\sqrt{n}(Z_n - \mu) \to_d N(0, \Sigma)$$

where $\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{pmatrix}$ and where $\sigma_{11}^2 = \mathrm{Var}(R_1Y_1)$, $\sigma_{22}^2 = \mathrm{Var}(R_1)$, and $\sigma_{12} = \mathrm{Cov}(R_1Y_1, R_1)$

Let $g : R^2 \to R$ be given by $g(z_1, z_2) = z_1/z_2$. We have

$$\left.\frac{\partial g}{\partial z_1}\right|_\mu = \left.\frac{1}{z_2}\right|_\mu = \frac{1}{\mu_R}$$
$$\left.\frac{\partial g}{\partial z_2}\right|_\mu = -\left.\frac{z_1}{z_2^2}\right|_\mu = -\frac{\mu_{RY}}{\mu_R^2}$$

By the Delta Method, we have

$$\sqrt{n}(\mu_1 - \mu_*) \to_d N(0, \tau^2)$$

where

$$\tau^2 = \frac{1}{\mu_R^2}\sigma_{11}^2 + \left(\frac{\mu_{RY}}{\mu_R^2}\right)^2 \sigma_{22}^2 - 2\frac{\mu_{RY}}{\mu_R^3}\sigma_{12}$$

and where

$$\begin{aligned}
\sigma_{11}^2 &= \mathrm{Var}(R_1Y_1) \\
&= \mathbb{E}\mathrm{Var}(R_1Y_1|X_1) + \mathrm{Var}(\mathbb{E}R_1Y_1|X_1) \\
&= \mathbb{E}\{\mathbb{E}R_1^2Y_1^2|X_1 - (\mathbb{E}R_1Y_1|X_1)^2\} + \mathrm{Var}\{(\mathbb{E}R_1|X_1)\mathbb{E}(Y_1|X_1)\} \\
&= \mathbb{E}\{(\mathbb{E}R_1|X_1)\mathbb{E}(Y_1^2|X_1) - (\mathbb{E}R_1|X_1)^2(\mathbb{E}Y_1|X_1)^2\} + \mathrm{Var}\{\pi(X_1)\mathbb{E}Y_1|X_1\} \\
&= \mathbb{E}\{\pi(X_1)[\mathbb{E}(Y_1^2|X_1) - \pi(X_1)(\mathbb{E}Y_1|X_1)^2]\} + \mathrm{Var}\{\pi(X_1)\mathbb{E}Y_1|X_1\}
\end{aligned}$$

$$
\begin{aligned}
\sigma_{22}^2 &= \mathrm{Var}(R_1)\\
&= \mathrm{Var}(\mathbb{E}R_1|X_1) + \mathbb{E}\mathrm{Var}(R_1|X_1)\\
&= \mathrm{Var}(\pi(X_1)) + \mathbb{E}\{\pi(X_1)(1-\pi(X_1))\}\\
&= \mathbb{E}\pi^2(X_1) - (\mathbb{E}\pi(X_1))^2 + \mathbb{E}\pi(X_1) - \mathbb{E}\pi^2(X_1)\\
&= \mathbb{E}\pi(X_1)[1-\pi(X_1)]
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{12} &= \mathrm{Cov}\{R_1, Y_1, R_1\}\\
&= \mathbb{E}\mathrm{Cov}\{R_1Y-1, R_1|X_1\} + \mathrm{Cov}\{\mathbb{E}R_1Y_1|X_1, \mathbb{E}R_1|X_1\}\\
&= \mathbb{E}\{\mathbb{E}R_1^2Y_1|X_1 - (\mathbb{E}R_1Y_1|X_1)(\mathbb{E}R_1|X_1)\} + \mathrm{Cov}\{\pi(X_1)\mathbb{E}(Y_1|X_1, \pi(X_1)\}\\
&= \mathbb{E}\{\pi(X_1)\mathbb{E}Y_1|X_1 - (\mathbb{E}R_1|X_1)^2(\mathbb{E}Y_1|X_1)\} + \mathbb{E}\left\{\pi^2(X_1)\mathbb{E}Y_1|X_1 - \mathbb{E}(\pi(X_1)\mathbb{E}Y_1|X_1)\mathbb{E}\pi(X_1)\right\}\\
&= \mathbb{E}[\pi(X_1)\mathbb{E}Y_1|X_1] - \mathbb{E}\{\pi^2(X_1)\mathbb{E}(Y_1|X_1\} + \mathbb{E}\{\pi^2(X_1)\mathbb{E}Y_1|X_1\} - \{\mathbb{E}(\pi(X_1)\mathbb{E}Y_1|X_1)\}\mathbb{E}\pi(X_1)\\
&= \mathbb{E}\{\pi(X_1)\mathbb{E}Y_1|X_1\} - \mathbb{E}\{\pi(X_1)\mathbb{E}Y_1|X_1)\}\mathbb{E}\{\pi(X_1)\}\\
&= \mathbb{E}\{\pi(X_1)\mathbb{E}Y_1|X_1\}\{1-\pi(X_1)\}\\
&= \mathbb{E}\{\pi(X_1)\mathbb{E}Y_1|X_1[1-\mathbb{E}\pi(X_1)]\}
\end{aligned}
$$

TRYING TO USE SLUTSKY BUT DIDN'T COME OUT SOLUTION: We have

$$
\begin{aligned}
\sqrt{n}(\hat{\mu}_1 - \mu_*) &= \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^n R_iY_i}{\frac{1}{n}\sum_{i=1}^n R_i} - \frac{\mu_{RY}}{\mu_R}\right)\\
&= \frac{1}{n^{-1}\sum_{i=1}^n R_i}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n R_iY_i - \mu_{RY}\frac{n^{-1}\sum_{i=1}^n R_i}{\mu_R}\right)\\
&= \frac{1}{n^{-1}\sum_{i=1}^n R_i}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n R_iY_i - \mu_{RY} + \mu_{RY} - \mu_{RY}\frac{n^{-1}\sum_{i=1}^n R_i}{\mu_R}\right)\\
&= \frac{1}{n^{-1}\sum_{i=1}^n R_i}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n R_iY_i - \mu_{RY}\right) + \frac{-\mu_{RY}}{n^{-1}\sum_{i=1}^n R_i}\sqrt{n}\left(\frac{n^{-1}\sum_{i=1}^n R_i}{\mu_R} - 1\right)
\end{aligned}
$$

2.b

A Horwitz-Thompson estimator for μ is given as

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)}$$

Show that $\hat{\mu}_2$ is a consistent estimator for μ and derive the asymptotic distribution of $\sqrt{n}(\hat{\mu}_2 - \mu)$. Leave expressions in the result.

Solution:
Since (R_i, Y_i, X_i) are i.i.d. triples, $\frac{R_iY_i}{\pi(X_i)}$ are i.i.d. random variables. Hence, by the weak law of large numbers,

$$\begin{aligned}
\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)} &\to_p \mathbb{E}\frac{R_1Y_1}{\pi(X_1)} \\
&= \mathbb{E}\left\{\mathbb{E}\frac{R_1Y_1}{\pi(X_1)}|X_1\right\} \\
&= \mathbb{E}\left\{\frac{1}{\pi(X_1)}\mathbb{E}R_1Y_1|X_1\right\} \\
&= \mathbb{E}\left\{\frac{1}{\pi(X_1)}(\mathbb{E}R_1|X_1)(\mathbb{E}Y_1|X_1)\right\} \\
&= \mathbb{E}\left\{\frac{1}{\pi(X_1)}\pi(X_1)(\mathbb{E}Y_1|X_1)\right\} \\
&= \mathbb{E}(\mathbb{E}Y_1|X_1) \\
&= \mathbb{E}Y_1 \\
&= \mu
\end{aligned}$$

Now, note that

$$\begin{aligned}
\sigma_2^2 := \mathrm{Var}\left(\frac{R_1Y_1}{\pi(X_1)}\right) &= \mathrm{Var}\left(\mathbb{E}\frac{R_1Y_1}{\pi(X_1)} \mid X_1\right) + \mathbb{E}\mathrm{Var}\left(\frac{R_1Y_1}{\pi(X_1)} \mid X_1\right) \\
&= \mathrm{Var}\left(\frac{1}{\pi(X_1)}(\mathbb{E}R_1|X_1)(\mathbb{E}Y_1|X_1)\right) + \mathbb{E}\frac{1}{\pi^2(X_1)}\mathrm{Var}(R_1Y_1 \mid X_1) \\
&= \mathrm{Var}(Y_1) + \mathbb{E}\frac{1}{\pi^2(X_1)}\mathrm{Var}(R_1Y_1 \mid X_1)
\end{aligned}$$

We must show that the $\sigma_2^2 < \infty$. By Hoelder's Inequality,

$$\mathbb{E}\left\{\frac{1}{\pi^2(X_i)}\mathrm{Var}(R_1Y_1|X_1)\right\} \leq \left(\mathbb{E}\left(\frac{1}{\pi^2(X_i)}\right)^2\right)^{1/2} (\mathbb{E}\mathrm{Var}(R_1Y_1|X_1)^2)^{1/2}$$

Since $0 < c < \pi(X_1) \leq 1$, $1 \leq \frac{1}{\pi^2(X_i)} < \infty$ and so $0 < \mathbb{E}\frac{1}{\pi^2(X_i)} < \infty$. Moreover, $(R_1Y_1)^2 \leq Y_1^2$ so $\mathbb{E}(R_1Y_1)^2|X_1 \leq \mathbb{E}Y_1^2|X_1$. Hence, $\mathbb{E}\mathbb{E}(R_1Y_1)^2|X_1 \leq \mathbb{E}\mathbb{E}(Y_1^2|X_1) = \mathbb{E}Y_1^2 < \infty$. Thus, $\sigma_2^2 < \infty$.

By the Central Limit Theorem,

$$\sqrt{n}(\hat{\mu}_2 - \mu) \to_d N(0, \sigma_2^2)$$

2.c

For any measurable function $g(X_i)$ with finite second moment, we define

$$\hat{\mu}_g = n^{-1}\left\{\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)} + \sum_{i=1}^{n}\left(1-\frac{R_i}{\pi(X_i)}\right)g(X_i)\right\}$$

Show that $\hat{\mu}_g$ is a consistent estimator for μ and derive the asymptotic distribution of $\sqrt{n}(\hat{\mu}_g-\mu)$

Leave expressions in the result.

Solution:
By the weak law of large numbers and the continuous mapping theorem,

$$\begin{aligned}
\hat{\mu}_g &= n^{-1}\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)} + n^{-1}\sum_{i=1}^{n}g(X_i) - n^{-1}\sum_{i=1}^{n}\frac{R_i}{\pi(X_i)}g(X_i)\\
&\to_p \mu + \mathbb{E}g(X_1) - \mathbb{E}\frac{R_1}{\pi(X_1)}g(X_1)\\
&= \mu + \mathbb{E}g(X_1) - \mathbb{E}\left\{\frac{g(X_1)}{\pi(X_1)}(\mathbb{E}R_1|X_1)\right\}\\
&= \mu + \mathbb{E}g(X_1) - \mathbb{E}\frac{g(X_1)}{\pi(X_1)}\pi(X_1)\\
&= \mu + \mathbb{E}g(X_1) - \mathbb{E}g(X_1)\\
&= \mu
\end{aligned}$$

Hence, $\hat{\mu}_g$ is consistent for μ.

Let σ_g^2 be the asymptotic variance. Then

$$\begin{aligned}
\sigma_g^2 &= \mathrm{Var}\left\{\frac{R_1Y_1}{\pi(X_1)} + \left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)\right\}\\
&= \mathrm{Var}\left(\frac{R_1Y_1}{\pi(X_1)}\right) + \mathrm{Var}\left\{\left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)\right\} + 2\mathrm{Cov}\left\{\frac{R_1Y_1}{\pi(X_1)}, \left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)\right\}\\
&= A + B + 2C
\end{aligned}$$

Note that $A = \sigma_2^2$. We have

$$\begin{aligned}
B &= \mathbb{E}\mathrm{Var}\left\{\left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)|X_1\right\} + \mathrm{Var}\left\{\mathbb{E}\left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)|X_1\right\}\\
&= \mathbb{E}g^2(X_1)\mathrm{Var}\left\{1-\frac{R_1}{\pi(X_1)}|X_1\right\} + \mathrm{Var}\left\{g(X_1)\mathbb{E}\left(1-\frac{R_1}{\pi(X_1)}\middle| X_1\right)\right\}\\
&= \mathbb{E}\frac{g^2(X_1)}{\pi^2(X_1)}\pi(X_1)(1-\pi(X_1)+0\\
&= \mathbb{E}\frac{g^2(X_1)(1-\pi(X_1))}{\pi(X_1)}
\end{aligned}$$

Finally,

$$C = \mathrm{Cov}\left\{\mathbb{E}\left.\frac{R_1Y_1}{\pi(X_1)}\right| X_1, \mathbb{E}\left.\left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)\right| X_1\right\} + \mathbb{E}\mathrm{Cov}\left\{\frac{R_1Y_1}{\pi(X_1)}, \left.\left(1-\frac{R_1}{\pi(X_1)}\right)g(X_1)\right| X_1\right\}$$

$$\begin{aligned}
&= \mathrm{Cov}\left\{\frac{1}{\pi(X_1)}(\mathbb{E}R_1|X_1)(\mathbb{E}Y_1|X_1), g(X_1)\mathbb{E}\left(1-\left.\frac{R_1}{\pi(X_1)}\right| X_1\right)\right\} \\
&\quad + \mathbb{E}\frac{g(X_1)}{\pi(X_1)}\mathrm{Cov}\left\{R_1Y_1, 1-\frac{R_1}{\pi(X_1)} \mid X_1\right\}
\end{aligned}$$

$$\begin{aligned}
&= 0 - \mathbb{E}\frac{g(X_1)}{\pi(X_1)}\mathrm{Cov}\left\{R_1Y_1, \frac{R_1}{\pi(X_1)} \mid X_1\right\} \\
&= -\mathbb{E}\frac{g(X_1)}{\pi^2(X_1)}\mathrm{Cov}\{R_1Y_1, R_1 \mid X_1\}
\end{aligned}$$

$$\begin{aligned}
&= -\mathbb{E}\frac{g(X_1)}{\pi^2(X_1)}\left\{\mathbb{E}(R_1^2Y_1|X_1) - (\mathbb{E}R_1Y_1|X_1)(\mathbb{E}R_1|X_1)\right\} \\
&= -\mathbb{E}\frac{g(X_1)}{\pi^2(X_1)}\left\{(\mathbb{E}R_1^2|X_1)(\mathbb{E}Y_1|X_1) - (\mathbb{E}R_1|X_1)^2(\mathbb{E}Y_1|X_1)\right\} \\
&= -\mathbb{E}\frac{g(X_1)\mathbb{E}(Y_1|X_1)(1-\pi(X_1)}{\pi(X_1)}
\end{aligned}$$

It is easy to see that σ_g^2 is finite since σ_2^2 is finite and π is bounded from below by a positive constant, so $1/\pi(x)$ is finite for all x. Since g is measurable, all expectations involving $g(X_1)$ are finite, and the product of two measurable functions is measurable.

2.d.

In order to minimize the variance, we only have to minimize the $B-2C$ term in the variance expression above since the A term does not depend on g. We have

$$\begin{aligned}
B - 2C &= \mathbb{E}\frac{g^2(X_1)(1-\pi(X_1))}{\pi(X_1)} - 2\mathbb{E}\frac{g(X_1)\mathbb{E}(Y_1|X_1)(1-\pi(X_1))}{\pi(X_1)} \\
&= \mathbb{E}\frac{1-\pi(X_1)}{\pi(X_1)}g(X_1)\left[g(X_1) - 2\mathbb{E}(Y_1|X_1)\right]
\end{aligned}$$

Take $g(x) = \mathbb{E}(Y_1|X_1 = x)$.

Question 3

3.a.

In this part, let T_0 be an unbiased estimator of an unknown parameter θ and consider the properties of T_0 under squared error loss.

3.a.i.

Show that $T_0 + c$ is not a minimax estimator under squared error loss, where $c \neq 0$ is a known constant.

<u>*Solution:*</u>
Note that

$$\begin{aligned}
R(\theta, T_0 + c) &= \mathbb{E}_\theta(T_0 + c - \theta)^2 \\
&= \mathbb{E}_\theta(T_0 - \theta + c)^2 \\
&= \mathbb{E}_\theta(T_0 - \theta)^2 - 2c\mathbb{E}_\theta(T_0 - \theta) + c^2 \\
&= R(\theta, T_0) + c^2
\end{aligned}$$

since T_0 is unbiased for θ i.e., $\mathbb{E}_\theta(T_0 - \theta) = 0$.

Now,

$$\begin{aligned}
\sup_\theta R(\theta, T_0 + c) &= \sup_\theta \{R(\theta, T_0) + c^2\} \\
&= \sup_\theta \{R(\theta, T_0)\} + c^2 \\
&> \sup_\theta \{R(\theta, T_0)\}
\end{aligned}$$

provided $R(\theta, T_0) < \infty$ since $c \neq 0$.

3.a.ii.

Show that the estimator cT_0 is not minimax under squared error loss unless $\sup_\theta R_T(\theta) = \infty$ for any estimator T of θ, where $c \in (0, 1)$ is a known constant and $R_T(\theta)$ is the frequentist risk function for T.

<u>*Solution:*</u> Note that

$$\begin{aligned}
R_{cT_0}(\theta) &= R(\theta, cT_0) \\
&= \mathbb{E}_\theta(cT_0 - \theta)^2 \\
&= \mathbb{E}_\theta\{cT_0 - T_0 + T_0 - \theta\}^2 \\
&= \mathbb{E}_\theta\{(c-1)T_0 + T_0 - \theta\} \\
&= \mathbb{E}_\theta(T_0 - \theta)^2 + (c-1)^2\mathbb{E}_\theta(T_0^2) \\
&= R_{T_0}(\theta) + (c-1)^2\mathbb{E}_\theta(T_0^2)
\end{aligned}$$

Hence,

$$\begin{aligned}
\sup_\theta R_{cT_0}(\theta) &= \sup_\theta \left\{R_{T_0}(\theta) + (c-1)^2\mathbb{E}_\theta(T_0^2)\right\} \\
&\geq \sup_\theta R_{T_0}(\theta)
\end{aligned}$$

with strict inequality holding if T_0 is a nonconstant estimator since $c \neq 0$. Thus, if cT_0 is minimax, then we must have $\sup_\theta R_{T_0}(\theta) = \infty$.

(Why does this mean we must have the risk for EVERY estimator be infinity?)

3.b.

In this part, let $X = 1$ or 0 with probabilities p and q respectively, and consider the estimation of p with loss function $L(p, a)$ equal to 1 when $|a - p| \geq 0.25$ and equal to 0 otherwise. The most general randomized estimator is $T_0 = U$ when $X = 0$ and $T_0 = V$ when $X = 1$, where U and V are two random variables with known distributions.

3.b.i.

Evaluate the risk function and the maximum risk of T_0 when U and V are uniform on $(0, 0.5)$ and $(0.5, 1)$, respectively.

<u>*Solution:*</u>
Note that

$$\begin{aligned}
R(p, T_0) &= \mathbb{E}_p L(p, T_0) \\
&= \mathbb{E}_p \mathbf{1}_{|T_0 - p| > 0.25} \\
&= P_p(|T_0 - p| > 0.25) \\
&= P_p(|T_0 - p| > 0.25 | X = 0) P_p(X = 0) + P_p(|T_0 - p| > 0.25 | X = 1) P_p(X = 1) \\
&= (1 - p) P_p(|U - p| > 0.25) + p P_p(|V - p| > 0.25)
\end{aligned}$$

Now,

$$\begin{aligned}
|U - p| \geq 0.25 &\iff U - p \geq 0.25 \quad \text{or} \quad U - p < -0.25 \\
&\iff U \geq p + 0.25 \quad \text{or} \quad U < p - 0.25
\end{aligned}$$

$$P(U \geq p + 0.25) = \begin{cases} 0 & \text{if } p \geq 0.25 \\ 0.5 - 2p & \text{if } p \leq 0.25 \end{cases}$$

$$P(U \leq p - 0.25) = \begin{cases} 0 & \text{if } p \leq 0.25 \\ 2p - 0.5 & \text{if } 0.25 \leq p \leq 0.75 \\ 1 & \text{if } p > 0.75 \end{cases}$$

Thus,

$$\begin{aligned}
P(|U - p| \geq 0.25) &= \begin{cases} 0.5 - 2p & \text{if } p < 0.25 \\ 2p - 0.5 & \text{if } 0.25 \leq p < 0.75 \\ 1 & \text{if } p \geq 0.75 \end{cases} \\
&= \min\{1, 2|p - 0.25|\}
\end{aligned}$$

Similarly, we have $|V - p| > 0.25 \iff V > p + 0.25$ or $V < p - 0.25$

$$P(V > p + 0.25) = \begin{cases} 1, & p < 0.25 \\ 1 - 2(p - 0.25) = 2(0.75 - p), & 0.25 \leq p \leq 0.75 \\ 0, & p > 0.75 \end{cases}$$

$$P(V < p - 0.25) = \begin{cases} 0, & p < 0.75 \\ 2(p - 0.75), & p \geq 0.75 \end{cases}$$

Summing these together, we get

$$\begin{aligned} P(|V - p| > 0.25) &= \begin{cases} 1, & p < 0.25 \\ 2(0.75 - p), & 0.25 \leq p \leq 0.75 \\ 2(p - 0.75), & p \geq 0.75 \end{cases} \\ &= \min\{1, 2|p - 0.75|\} \end{aligned}$$

It follows that the risk function is

$$R(p, T_0) = (1 - p) \min\{1, 2|p - 0.25|\} + p \min\{1, 2|p - 0.75|\}$$

The worst case scenario is $p = 1/2$ as this is the most difficult to estimate since the two random variables are cut at $1/2$, it represents an extreme for either U or V. Thus, the maximum risk is $R(1/2, T_0) = 1/2$.

3.c.

In this part, one has a sample of n iid normal random variables with mean θ and variance σ^2, $X_1, \ldots, X_n$.

3.c.i.

Assume $0 < \sigma^2 < K$ is known, where K is a finite positive constant. Is the sample mean $\bar{X}$ minimax with respect to the loss function $L(\theta, a) = (\theta - a)^2/\sigma^2$? Justify your answer rigorously.

: Solution:
Note that

$$R(\theta, \bar{X}) = \mathbb{E}L(\theta, \bar{X}) = \mathbb{E}(\theta - \bar{X})^2/\sigma^2 = \frac{\sigma^2/n}{\sigma^2} = \frac{1}{n}$$

Thus, if we can find a sequence of Bayes rules converging to $\bar{X}$ whose limit Bayes risk is $1/n$, $\bar{X}$ will be minimax. Let $\sigma_n^2 = \sigma^2/n$

Suppose $\theta \sim N(\mu, \tau^2)$. Then

$$\begin{aligned}
p(\theta|x) &\propto p(x|\theta)\lambda(\theta) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \theta)^2\right\} \exp\left\{-\frac{1}{2\tau^2}(\theta - \tau)^2\right\} \\
&\propto \exp\left\{\frac{1}{2\sigma_n^2}(\theta - \bar{X})^2 - \frac{1}{2\tau^2}(\theta - \tau)^2\right\} \\
&\propto \exp\left\{\frac{1}{2}\left[\left(\frac{1}{\sigma_n^2} + \frac{1}{\tau^2}\right)\theta^2 - 2\theta\left(\frac{\bar{X}}{\sigma_n^2} + \frac{\mu}{\tau^2}\right)\right]\right\} \\
&= \exp\left\{-\frac{1}{2\left(\frac{1}{\sigma_n^2} + \frac{1}{\tau^2}\right)^{-1}}\left[\theta^2 - 2\theta\left(\frac{1}{\sigma_n^2} + \frac{1}{\tau^2}\right)^{-1}\left(\frac{\bar{X}}{\sigma_n^2} + \frac{\mu}{\tau^2}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\left(\frac{1}{\sigma_n^2} + \frac{1}{\tau^2}\right)^{-1}}\left[\theta - \left(\frac{1}{\sigma_n^2} + \frac{1}{\tau^2}\right)^{-1}\left(\frac{\bar{X}}{\sigma_n^2} + \frac{\mu}{\tau^2}\right)\right]^2\right\}
\end{aligned}$$

We recognize this as the kernel of a $N(\mu_*, \tau_*^2)$ distribution, where

$$\mu_* = \frac{\bar{X}/\sigma_n^2 + \mu/\tau^2}{1/\sigma_n^2 + 1/\tau^2}$$

$$\tau_*^2 = (1/\sigma_n^2 + 1/\tau^2)^{-1}$$

Note that the loss function is simply weighted squared loss. Thus, we have that the Bayes rule for this prior and for this loss function is

$$\begin{aligned} d_\Lambda(x) &= \arg\min_a \frac{\mathbb{E}_\theta \frac{1}{\sigma^2}(\theta - a)^2 | X = x}{\mathbb{E}_\theta \frac{1}{\sigma^2} | X = x} \\ &= \arg\min_a \mathbb{E}_\theta (\theta - a)^2 | X = x \end{aligned}$$

where the σ^2's cancel because $0 < \sigma^2 \leq M < \infty$. This is simply the Bayes rule for squared error loss, which is known to be the posterior mean. Thus, we have

$$d_\Lambda(x) = \mu_*$$

Now, let Λ_k be a sequence of priors with $\Lambda_k = N(0, \tau_k^2)$ where $\tau_k^2 = k$. Then the Bayes Risk is

$$\begin{aligned} \mathcal{R}(\Lambda_k, d_{\Lambda_k}) &= \mathbb{E}_X \mathbb{E}_{\theta|X} L(\theta, d_{\Lambda_k}) \\ &= \mathbb{E}_X \mathbb{E}_{\theta|X} \frac{1}{\sigma^2}(\theta - \mathbb{E}\theta|X)^2 \\ &= \frac{1}{\sigma^2} \mathbb{E}_X \mathrm{Var}(\theta|X) \\ &= \frac{1}{\sigma^2} \mathbb{E}_X (1/\sigma_n^2 + 1/\tau_k^2)^{-1} \\ &\to \frac{1}{\sigma^2}(1/\sigma_n^2 + 0)^{-1} = \frac{1}{\sigma^2}\frac{\sigma^2}{n} = \frac{1}{n} \end{aligned}$$

Hence, we found a sequence of priors whose limit Bayes risk of Bayes rules converges to the risk of $\bar{X}$. By Theorem 1.13, $\bar{X}$ is a minimax rule.

3.c.ii

Redo part (i) without assuming σ^2 is known.

<u>*Solution:*</u>
Let $P_0 = \{N(\theta, \sigma^2) : 0 < \sigma^2 < K\}$ and $P_1 = \{N(\theta, \sigma^2) : \sigma^2 \in \mathbb{R}^+\}$.

Recall that $\bar{X}$ was minimax for P_0. If $\sigma^2 < \infty$,

$$\begin{aligned}
\sup_{P \in P_1} R(P, \bar{X}) &= \sup_{P \in P_1} \mathbb{E}_\theta(\theta - \bar{X})^2/\sigma^2 \\
&= \mathrm{Var}(\bar{X})/\sigma^2 \\
&= 1/n \\
&= \sup_{P \in P_0} R(P, \bar{X})
\end{aligned}$$

so that $\bar{X}$ is minimax for all $\sigma^2 < \infty$.