SDS 384-11 HW 4

Shivam Gupta

April 10, 2019

1 Problem 1

Consider the set of constant distributions $\mathcal{Q}$, that is, for $c \in \mathbb{R}$, there is distribution $Q_c \in \mathcal{Q}$ with corresponding random variable X_c, such that

$$Pr[X_c = c] = 1$$

Now, clearly, every distribution in $\mathcal{Q}$ has finite first moment, and so, $\mathcal{Q} \subseteq \mathcal{P}$. Now, suppose there exists an $f(x)$ such that $E[f(X)] = \mu^2$ for every $P \in \mathcal{P}$ with corresponding r.v. X. This means that for every $Q_c \in \mathcal{Q}$, with corresponding r.v. X_c, $E[f(X_c)] = f(c) = c^2$. Thus, we have that for every $c \in \mathbb{R}$, $f(c) = c^2$. So, our function is $f(x) = x^2$. Now, consider the distribution $N(0, 1) \in \mathcal{P}$ with corresponding random variable X. We have that $E[f(X)] = E[X^2] = 1$ since the variance of X is 1 and mean is 0. However, $\mu^2 = 0$ in this case. Thus, we have a contradiction, and there is no such function f.

2 Problem 2

(a) By definition of estimable parameter we have that $g_1 = E[h_1(X_1, \dots, X_{m_1})]$ and $g_2 = E[h_2(X_1, \dots, X_{m_2})]$ for real valued measurable functions h_1, h_2 and $X_1, \dots, X_{\max(m_1,m_2)} \overset{iid}{\sim} P$.

Now, we can write

$$g_1+g_2 = E[h_1(X_1, \dots, X_{m_1})]+E[h_2(X_1, \dots, X_{m_2})] = E[h_1(X_1, \dots, X_{m_1})+h_2(X_1, \dots, X_{m_2})]$$

(Follows by linearity of expectation)

Thus, if we let $h_3(X_1, \dots, X_{\max(m_1,m_2)}) = h_1(X_1, \dots, X_{m_1}) + h_2(X_1, \dots, X_{m_2})$, we have $g_1 + g_2 = E[h_3(X_1, \dots, X_{\max(m_1,m_2)})]$, and clearly, since h_3 takes $\max(m_1, m_2)$ arguments, the degree of $g_1 + g_2$ is at most $\max(m_1, m_2)$.

(b) Let $X_1, \dots, X_{m_1}, Y_1, \dots, Y_{m_2} \overset{iid}{\sim} P$. Now, let $g_1 = E[h_1(X_1, \dots, X_{m_1})]$ and $g_2 = E[h_2(Y_1, \dots, Y_{m_2})]$. Now, note that

$$g_1 g_2 = E[h_1(X_1, \dots, X_{m_1})]E[h_2(Y_1, \dots, Y_{m_2})]$$

But since the X_is and Y_is are independent, $h_1(X_1, \dots, X_{m_1})$ and $h_2(Y_1, \dots, Y_{m_2})$ are independent random variables. So, the above is equal to

$$E[h_1(X_1, \dots, X_{m_1})h_2(Y_1, \dots, Y_{m_2})]$$

So, if we let $h_3(X_1, \dots, X_{m_1}, Y_1, \dots, Y_{m_2}) = h_1(X_1, \dots, X_{m_1})h_2(Y_1, \dots, Y_{m_2})$, then, $g_1 g_2 = E[h_3(X_1, \dots, X_{m_1}, Y_1, \dots, Y_{m_2})]$ So, clearly, $g_1 g_2$ has degree at most $m_1 + m_2$, since h_3 takes $m_1 + m_2$ arguments.

3 Problem 3

Let

$$h(X_1, X_2, X_3) = (1 - \mathbb{1}\{X_2 \le -X_1\})(1 - \mathbb{1}\{X_3 \le -X_1\})$$
$$-2(1 - \mathbb{1}\{X_2 \le -X_1\})\mathbb{1}\{X_3 \le X_1\} + \mathbb{1}\{X_2 \le X_1\}\mathbb{1}\{X_3 \le X_1\}$$

. We claim that $E_F[h(X_1, X_2, X_3)] = \theta(F)$. Note that by linearity of expectation,

$$E_F[h(X_1, X_2, X_3)] = E_F[(1 - \mathbb{1}\{X_2 \le -X_1\})(1 - \mathbb{1}\{X_3 \le -X_1\})]$$

$$-2E_F[(1 - \mathbb{1}\{X_2 \le -X_1\})\mathbb{1}\{X_3 \le X_1\}] + E_F[\mathbb{1}\{X_2 \le X_1\}\mathbb{1}\{X_3 \le X_1\}]$$

Now,

$$E_F[\mathbb{1}\{X_2 \le X_1\}\mathbb{1}\{X_3 \le X_1\}] = E_{X_1}[E_{X_2,X_3}[\mathbb{1}\{X_2 \le X_1\}\mathbb{1}\{X_3 \le X_1\}|X_1]]$$

But since X_2, X_3 are independent, the indicators $\mathbb{1}\{X_2 \le X_1\}$ and $\mathbb{1}\{X_3 \le X_1\}$ are independent given X_1. Thus, the above is equal to

$$E_{X_1}[E_{X_2}[\mathbb{1}\{X_2 \le X_1\}|X_1]E_{X_3}[\mathbb{1}\{X_3 \le X_1\}|X_1]]$$

But

$$E_{X_2}[\mathbb{1}\{X_2 \le X_1\}|X_1] = Pr[X_2 \le X_1|X_1] = F(X_1)$$

, and similarly for X_3. Thus, the above equals

$$E_{X_1}[F(X_1)F(X_1)] = E_{X_1}[F(X_1)^2]$$

By definition of expectation, the above equals

$$\int_{-\infty}^{\infty} F(x)^2 dF(x)$$

Now, by similar reasoning,

$$E_F[(1-\mathbb{1}\{X_2 \le -X_1\})(1-\mathbb{1}\{X_3 \le -X_1\})] = E_{X_1}[E_{X_2,X_3}[(1-\mathbb{1}\{X_2 \le -X_1\})(1-\mathbb{1}\{X_3 \le -X_1\})|X_1]]$$

By independence, we have that the above is

$$= E_{X_1}[E_{X_2}[(1 - \mathbb{1}\{X_2 \le -X_1\})|X_1]E_{X_3}[(1 - \mathbb{1}\{X_3 \le -X_1\})|X_1]]$$

But $E_{X_2}[(1-\mathbb{1}\{X_2 \leq -X_1\})|X_1] = 1 - Pr[X_2 \leq -X_1|X_1] = 1 - F(-X_1)$. Similarly for X_3.

Thus, the above equals

$$E_{X_1}[(1-F(-X_1))^2] = \int_{-\infty}^{\infty} (1-F(-x))^2 dF(x)$$

By almost identical reasoning,

$$E_F[(1-\mathbb{1}\{X_2 \leq -X_1\})\mathbb{1}\{X_3 \leq X_1\}] = \int_{-\infty}^{\infty} F(x)(1-F(-x))dF(x)$$

Thus, finally,

$$E_F[h(X_1,X_2,X_3)] = \int (1-F(-x))^2 dF(x) - 2\int F(x)(1-F(x))dF(x) + \int F(x)^2 dF(x)$$

To construct the corresponding U statistic, we must first symmetrize our kernel. Let our symmetric kernel

$$f(X_1,X_2,X_3) = \frac{\sum_{\pi\in\Pi_3} h(X_{\pi_1},X_{\pi_2},X_{\pi_3})}{6}$$

be our symmetrized kernel, where Π_3 is the set of all permutations of $\{1,2,3\}$.

Now, our U statistic

$$U_n = \frac{\sum_{\{i_1,i_2,i_3\}\in\mathcal{I}_3} f(X_{i_1},X_{i_2},X_{i_3})}{\binom{n}{3}}$$

where $\mathcal{I}_3$ is the set of subsets of size 3 from $[n]$.

4 Problem 4

(a) Note that since the distribution is symmetric about 0, this means that $f(x) = f(-x)$ for all x where f is the pdf of X, which means that

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{\infty} xf(x)dx$$

$$= \int_{0}^{\infty} -xf(-x)dx + \int_{0}^{\infty} xf(x)dx = -\int_{0}^{\infty} xf(x)dx + \int_{0}^{\infty} xf(x)dx = 0$$

Now,

$$E[h(X_1,X_2)|X_1] = E[X_1X_2+(X_1^2-\sigma^2)(X_2^2-\sigma^2)|X_1] = E[X_1X_2|X_1]+E[(X_1^2-\sigma)^2(X_2^2-\sigma^2)|X_1]$$

Since X_1, X_2 are independent, we have that the above is equal to

$$E[X_1|X_1]E[X_2|X_1] + E[X_1^2-\sigma^2|X_1]E[X_2^2-\sigma^2|X_1]$$

Now since the mean of X_i is 0, $E[X_2|X_1] = E[X_2] = 0$. Furthermore, since the variance of X_2 is σ^2, we have that $E[X_2^2 - \sigma^2|X_1] = 0 = E[X_2^2] - \sigma^2 = 0$. Thus, the above equals 0.

Now

$$\xi_1 = cov(E[h(X_1, X_2)|X_1], E[h(X_1, X_3)|X_1]) = cov(0, 0) = 0$$

Now note that by independence of X_1 and X_2,

$$E[X_1X_2 + (X_1^2 - \sigma^2)(X_2^2 - \sigma^2)] = E[X_1]E[X_2] + E[X_1^2 - \sigma^2]E[X_2^2 - \sigma^2] = 0$$

Also, note that

$$E[h(X_1, X_2)^2] = E[(X_1X_2 + (X_1^2 - \sigma^2)(X_2^2 - \sigma^2))^2]$$

$$= E[(X_1X_2)^2] + 2E[X_1X_2(X_1^2 - \sigma^2)(X_2^2 - \sigma^2)] + E[(X_1^2 - \sigma^2)^2(X_2^2 - \sigma^2)^2]$$

By independence and identical distribution of X_1, X_2, the above equals

$$E[X_1^2]^2 + 2E[X_1(X_1^2 - \sigma^2)]^2 + E[X_1^2 - \sigma^2]^2$$

Now $E[X_1^2 - \sigma^2] = E[X_1^2] - \sigma^2 = 0$. Also, note that by symmetry,

$$E[X^3] = \int_{-\infty}^{\infty} x^3 f(x)dx = \int_{-\infty}^{0} x^3 f(x)dx + \int_{0}^{\infty} x^3 f(x)dx$$

$$= \int_{0}^{\infty} -x^3 f(-x)dx + \int_{0}^{\infty} x^3 f(x)dx = -\int_{0}^{\infty} x^3 f(x)dx + \int_{0}^{\infty} x^3 f(x)dx = 0$$

So, we have that the above expression

$$E[h(X_1, X_2)^2] = E[X_1^2]^2 = \sigma^4$$

But

$$\xi_2 = Var(E[h(X_1, X_2)|X_1, X_2]) = Var[h(X_1, X_2)]$$

$$= E[(X_1X_2 + (X_1^2 - \sigma^2)(X_2^2 - \sigma^2))^2] - E[X_1X_2 + (X_1^2 - \sigma^2)(X_2^2 - \sigma^2)]^2 = \sigma^4 > 0$$

(b) Note that

$$U = \frac{\sum_{i<j} h(X_i, X_j)}{\binom{n}{2}} = \frac{\sum_{i\neq j} h(X_i, X_j)}{n(n-1)} = \frac{\sum_{i\neq j} X_iX_j + (X_i^2 - \sigma^2)(X_j^2 - \sigma^2)}{n(n-1)}$$

$$= \frac{[(\sum_i X_i)^2 - \sum_i X_i^2] + [(\sum_i (X_i^2 - \sigma^2))^2 - \sum_i (X_i^2 - \sigma^2)^2]}{n(n-1)}$$

Let

$$\bar{X}_n = \sum_i X_i/n$$

and let

$$\overline{X_n^2 - \sigma^2} = \sum_i (X_i^2 - \sigma^2)/n$$

Now,

$$U = \frac{[(\sqrt{n}\overline{X_n})^2 - \sum_i X_i^2/n] + [(\sqrt{n}\overline{X_n^2 - \sigma^2})^2 - \sum_i (X_i^2 - \sigma^2)^2/n]}{n-1}$$

Now note that $E[X_i] = 0$ and $Var[X_i] = \sigma^2$. So, by central limit theorem for the first term, and law of large numbers for the second term, the terms in the first bracket above converges in distribution to

$$(Z^2 - 1)\sigma^2$$

where $Z \sim N(0, 1)$.

For the terms in the second bracket, note that

$$E[(X_i^2 - \sigma^2)] = 0$$

since X_i has mean 0 and variance σ^2. Also note that

$$Var[X_i^2 - \sigma^2] = E[(X_i^2 - \sigma^2)^2] - E[X_i^2 - \sigma^2]^2 = E[X_i^4] - \sigma^4$$

Thus, the third term (first term in second bracket) converges to $Z^2 * (E[X_i^4] - \sigma^4)$ by central limit theorem. By law of large numbers, the last term (second term in first bracket) converges to $E[(X_i^2 - \sigma^2)^2] = E[X_i^4] - \sigma^4$. Thus, finally, we have that $(n-1)U$ converges in distribution to

$$\sigma^2(Z^2 - 1) + (E[X_i^4] - \sigma^4)(Z^2 - 1)$$

So,

$$nU \to \frac{n}{n-1}\sigma^2(Z^2 - 1) + (E[X_i^4] - \sigma^4)(Z^2 - 1) \to \sigma^2(Z^2 - 1) + (E[X_i^4] - \sigma^4)(Z^2 - 1)$$

where the convergence is in distribution.

5 Problem 5

First note that for a random variable $T - E[T]$ that can be written as $\sum_{i=1}^N p_i T_i$ where $\sum_{i=1}^N p_i = 1$, $p_i \geq 0$, and T_i are the sum of independent random variables, then we have the following tail bound:

$$Pr[T - E[T] \geq t] \leq e^{-\lambda t} E[e^{\lambda T}] = e^{-\lambda t} E[\exp(\lambda \sum_{i=1}^{N} p_i T_i)]$$

By Jensen's inequality,

$$\exp(\lambda \sum_{i=1}^{N} p_i T_i) \leq \sum_{i=1}^{n} p_i \exp(\lambda T_i)$$

So, we have

$$Pr[T - E[T] \geq t] \leq \sum_{i=1}^{N} p_i E[\exp(-\lambda t + \lambda T_i)]$$

Now, for an order r U statistic, say U, let $g(X_1, \ldots, X_r)$ be its kernel function. Let

$$V(X_1, \ldots, X_n) = \frac{1}{k}\{g(X_1, \ldots, X_r) + g(X_{r+1}, \ldots, X_{2r}) + \cdots + g(X_{kr-r+1}, \ldots, X_{kr})\}$$

where $k = n/r$. Then, the U statistic U can be written as

$$U = \frac{\sum_{\pi \in \Pi} V(X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n})}{n!}$$

Now, $U - E[U]$ can be viewed in the form above, with each term $T_\pi = V(X_{\pi_1}, \ldots, X_{\pi_n}) - E[V(X_{\pi_1}, \ldots, X_{\pi_n})]$ a sum of independent random variables, $N = n!$ and $p_i = 1/n!$. Now, if g is bounded so that

$$a \leq g(x_1, \ldots, x_r) \leq b$$

then, by Hoeffding's inequality, since V is the average of $k = n/r$ independent random variables,

$$E[\exp(\lambda(T_\pi - t))] \leq \exp(-\lambda t + \lambda^2 (b-a)^2/2k) \leq \exp(-\frac{2kt^2}{(b-a)^2})$$

The claim follows.