Homework Assignment 3

SDS 384-11 Theoretical Statistics

1. (7=1+3+3 pts) In this question we consider the Jackknife estimate of variance of a symmetrical measurable function of $n-1$ variables S. Let $X_1, \ldots, X_n - 1$ be i.i.d. Consider $S = S(X_1, \ldots, X_{n-1})$. Now let

$$S_i = S(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

So $S = S_n$. If S has finite variance, then the Jackknife estimate of its variance is given by:

$$\mathrm{var}_{JACK}(S) = \sum_i \left(S_i - \frac{\sum_j S_j}{n} \right)^2$$

In Efron and Stein's Annals of Statistics paper in 1981 the following remarkable result was proven.

$$\mathrm{var}(S) \leq E\left(\mathrm{var}_{JACK}(S)\right) \tag{1}$$

This is what we will prove here today. First define $V_i = E[S|X_1, \ldots, X_i] - E[S|X_1, \ldots, X_{i-1}]$.

(a) Prove that $\mathrm{var}(S) = \sum_{i=1}^{n-1} EV_i^2$

(b) Prove that $E\mathrm{var}_{JACK}(S) = (n-1)E[(S_1 - S_2)^2]/2$

(c) Now prove Eq 1.

(a) Pf: WLOG, we assume $i < j$, so $E[V_i V_j] = EE[V_i V_j | X_{1:i}] = E[V_i E[V_j | X_{1:i}]]$ By tower rule, we have

$$E[V_j | X_{1:i}] = E[E[S|X_{1:j}] - E[S|X_{1:j-1}] | X_{1:i}] = 0.$$

So $E[V_i V_j] = 0$. So we have $\mathrm{var}(S) = E[(S - ES)^2] = E(\sum_{i=1}^{n-1} V_i)^2 = \sum_{i=1}^{n-1} EV_i^2$. $\square$.

(b) Pf: For any j, by symmetry, we have $ES_j^2 = ES_1^2$. For any $i \neq j$, we have $E[S_iS_j] = E[S_1S_2]$.

$$\begin{aligned} E\mathrm{var}_{JACK}(S) &= \sum E[S_i^2 + \frac{(\sum S_j)^2}{n^2} - 2S_i\frac{\sum S_j}{n}] \\ &= nE[S_1^2] - \frac{1}{n}E[(\sum S_i)^2] \\ &= nE[S_1^2] - \frac{1}{n}(nE[S_1^2] + 2E[\sum_{i<j} S_iS_j]) \\ &= (n-1)(E[S_1^2] - E[S_1S_2]). \end{aligned}$$

It is easy to check that $(n-1)E[(S_1 - S_2)^2]/2 = (n-1)(E[S_1^2] - E[S_1S_2])$. So we finish the proof. $\square$.

(c) Pf: Let $1 \leq i \leq n-1$. We have

$$E[S_1|X_{3:i+1}] = E[S_2|X_{3:i+1}].$$

Now set $A = E[S_1|X_{1:i+1}] - E[S_1|X_{3:i+1}], B = E[S_2|X_{1:i+1}] - E[S_2|X_{3:i+1}]$. By Jensen's inequality we have

$$\begin{aligned} E[(S_1 - S_2)^2] &= E[E[(S_1 - S_2)^2|X_{1:i+1}]] \\ &\geq E[(E[S_1|X_{1:i+1}] - E[S_2|X_{1:i+1}])^2] \\ &= E(A-B)^2. \end{aligned}$$

A only depends on $X_{2:i+1}$, B only depends on $X_1, X_{3:i+1}$. So A and B are conditionally independent w.r.t. $X_{3:i+1}$. So we have

$$E[AB] = E[E[AB|X_{3:i+1}]] = E[E[A|X_{3:i+1}]E[B|X_{3:i+1}]] = 0.$$

So we have $E(A-B)^2 = EA^2 + EB^2$. By symmetry, $E[A^2] = E[B^2] = E[V_i^2]$. So $E((S_1 - S_2)^2) \geq 2E[V_i^2]$ for any i. By (a) and (b), we get $\mathrm{var}(S) \leq E\left(\mathrm{var}_{JACK}(S)\right)$. $\square$.

2. (2+1+(1+1+1+1+1+2)=10 pts)In this question we will look at the Gaussian Lipschitz theorem. Consider $X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$.

   (a) Prove that the order statistics are 1-Lipschitz.

   (b) Now show that, for large enough n,

   $$c\sqrt{\log n} \leq E[\max_i X_i] \leq \sqrt{2\log n}$$

   where c is some universal constant.

   i. For the upper bound, let $Y = \max_i X_i$. First show that $\exp(tE[Y]) \leq \sum_i E\exp(tX_i)$. Now pick a t to get the right form.

   ii. For the lower bound, do the following steps.

   A. Show that $E[Y] \geq \delta P(Y \geq \delta) + E[\min(Y, 0)]$

B. Now show that $E[\min(Y,0)] \geq E[\min(X_1,0)]$
C. Finally, relate $P(Y \geq \delta)$ to $P(X_1 \geq \delta)$ by using independence.
D. Now show that $P(X_1 \geq \delta) \geq \exp(-\delta^2/\sigma^2)/c$, for some universal constant c.
E. Choose the parameter δ carefully to have $P(X_1 \geq \delta) \geq 1/n$, for large enough n.

(a) Pf: WLOG, we assume $X_{(k)} \geq Y_{(k)}$. So $|X_{(k)} - Y_{(k)}| = X_{(k)} - Y_{(k)}$. There are at least $n-(k-1)$ components of X that are greater than or equal to $X_{(k)}$ and at least k components of Y that are smaller than or equal to $Y_{(k)}$. So there exists an l, such that $X_{(k)} \leq X_l, Y_{(k)} \geq Y_l$. So $X_{(k)} - Y_{(k)} \leq X_l - Y_l = |X_l - Y_l| \leq \|X-Y\|_2$. So it is 1-Lipschitz. $\square$.

(b) Pf: (i) By Jensen's inequality, we have

$$\exp(tE[Y]) \leq E[\exp(tY)] \leq \sum E\exp(tX_i).$$

If $t > 0$, we have

$$E[Y] \leq \frac{\log(\sum E[\exp(tX_i)])}{t} = \frac{\log(n\exp(\frac{1}{2}t^2))}{t}.$$

Let $t = \sqrt{2\log n}$. So $E[Y] \leq \sqrt{2\log n}$.
(ii) (A) For any $\delta > 0$,

$$\begin{aligned} E[Y] &\geq \delta P(Y \geq \delta) + E[Y\mathbf{1}\{Y < \delta\}] \\ &= \delta P(Y \geq \delta) + E[Y\mathbf{1}\{Y < 0\}] \\ &= \delta P(Y \geq \delta) + E[\min(Y,0)]. \end{aligned}$$

(B) $Y \geq X_1$, so $\min(Y,0) \geq \min(X_1,0)$. So $E\min(Y,0) \geq E\min(X_1,0)$.

(C) By i.i.d. of X_i, we have

$$\begin{aligned} P(Y \geq \delta) &= 1 - P(Y < \delta) \\ &= 1 - P\left(\max_i X_i < \delta\right) \\ &= 1 - \prod_{i=1}^{n} P\left(X_i \leq \delta\right) \\ &= 1 - \left(1 - P\left(X_1 \geq \delta\right)\right)^n. \end{aligned}$$

(D) Because $\frac{(y+\delta)^2}{2} \le y^2+\delta^2$, we have

$$\begin{aligned}
P\left(X_1 \ge \delta\right) &= \int_{X\ge\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{\frac{-(y+\delta)^2}{2}} dy \\
&\ge \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\left(y^2+\delta^2\right)} dy \\
&= \frac{1}{\sqrt{2\pi}} e^{-\delta^2} \int_0^\infty e^{-y^2} dy \\
&= \frac{1}{\sqrt{2\pi}} e^{-\delta^2} \frac{1}{2} \int_{-\infty}^\infty e^{-y^2} dy \\
&= \frac{1}{2\sqrt{2}} e^{-\delta^2}.
\end{aligned}$$

So we have $c = 2\sqrt{2}$.

(E) Now we choose δ, such that

$$P\left(X_1 \ge \delta\right) \ge \frac{1}{2\sqrt{2}} \cdot e^{-\delta^2} \ge \frac{1}{n}.$$

So we have $\delta \le \sqrt{\log\left(\frac{n}{2\sqrt{2}}\right)}$. So we choose $\delta = \sqrt{\log\left(\frac{n}{2\sqrt{2}}\right)}$.

(F) We know $\left(1-\frac{1}{n}\right)^n \to 1/e$, so we have

$$P(Y \ge \delta) = 1 - \left(1 - P\left(X_1 \ge \delta\right)\right)^n \ge 1 - \left(1 - \frac{1}{n}\right)^n.$$

We take limits on the both sides, so we have

$$P(Y \ge \delta) \ge 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = 1 - e^{-1},$$

and we have

$$E[\min(Y,0)] \ge E\left[\min\left(X_1, 0\right)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 x \cdot e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{x=0}^{x=-\infty} = -\frac{1}{\sqrt{2\pi}}.$$

So we have

$$EY \ge \sqrt{\log\left(\frac{n}{2\sqrt{2}}\right)} \left(1 - \frac{1}{e}\right) - \frac{1}{\sqrt{2\pi}} = \sqrt{\log n - \log 2\sqrt{2}} \left(1 - \frac{1}{e}\right) - \frac{1}{\sqrt{2\pi}} = \Theta(\sqrt{\log n}).$$

So there is a universal constant c, such that $c\sqrt{\log n} \le E[\max_i X_i]$ when n large. $\square$.

3. (3 pts) Let $\mathcal{P}$ be the set of all distributions on the real line with finite first moment. Show that there does not exist a function $f(x)$ such that $Ef(X) = \mu^2$ for all $P \in \mathcal{P}$ where μ is the mean of P, and X is a random variable with distribution P.

We must have $h(x)dP(x) = \mu^2$ for all distributions on the real line with mean μ. If P is degenerate at a point y, this implies that $h(y) = y^2$ for all y. But if P has mean zero ($\mu = 0$) and is not degenerate, then $h(x)dP(x) = x^2 dP(x) > 0 = \mu^2$. which is a contradiction.

4. (4=2+2 pts) Let g_1 and g_2 be estimable parameters within $\mathcal{P}$ with respective degrees m_1 and m_2.

   (a) Show $g_1 + g_2$ is an estimable parameter with degree $\leq \max(m_1, m_2)$.

   (b) Show $g_1 g_2$ is an estimable parameter with degree at most $m_1 + m_2$.

   (a) Pf: There are h_1, h_2 such that $Eh_1(X_1, \cdots, X_{m_1}) = g_1, Eh_2(X_1, \cdots, X_{m_2}) = g_2$. So $E[h_1(X_1, \cdots, X_{m_1}) + h_2(X_1, \cdots, X_{m_2})] = g_1 + g_2$. So $g_1 + g_2$ is estimable. By definition of degree, we have the degree $\leq \max(m_1, m_2)$. $\square$.

   (b) Pf: Let $X_1, \cdots, X_{m_1}, X_{m_1+1}, \cdots, X_{m_1+m_2}$ be i.i.d. random variables. So there are h_1, h_2 such that $Eh_1(X_1, \cdots, X_{m_1}) = g_1, Eh_2(X_{m_1+1}, \cdots, X_{m_1+m_2}) = g_2$. And $h_1(X_1, \cdots, X_{m_1})$ and $h_2(X_{m_1+1}, \cdots, X_{m_1+m_2})$ are independent. So

$$E[h_1(X_1, \cdots, X_{m_1})h_2(X_{m_1+1}, \cdots, X_{m_1+m_2})]$$

$$= E[h_1(X_1, \cdots, X_{m_1})]E[h_2(X_{m_1+1}, \cdots, X_{m_1+m_2})] = g_1 g_2.$$

   So $g_1 g_2$ is estimable, and by definition of degree, we have the degree at most $m_1 + m_2$. $\square$.

5. (6=2+2+1.5+.5) A continuous distribution with CDF $F(x)$, on the real line is symmetric about the origin if, and only if, $1 - F(x) = F(-x)$ for all real x. This suggests using the parameter,

$$\theta(F) = \int (1 - F(x) - F(-x))^2 dF(x) \tag{2}$$

$$= \int ((1 - F(-x))^2 dF(x) - 2\int (1 - F(-x))F(x)dF(x) + \int F(x)^2 dF(x) \tag{3}$$

   as a nonparametric measure of how asymmetric the distribution is. Find a kernel h, of degree 3, such that $E_F h(X_1, X_2, X_3) = \theta(F)$ for all continuous F. Find the corresponding U statistic.

   Write for independent X_1, X_2, and X_3,

$$\theta(F) = \int P(X_1 > -x, X_2 > -x)dF(x) - 2\int P(X_1 > -x, X_2 < x)dF(x) + 1/3$$
$$= P(X_1 + X_3 > 0, X_2 + X_3 > 0) - 2P(X_1 + X_3 > 0, -X_2 + X_3 > 0) + 1/3$$

   This leads to the unbiased estimate of θ, $f(x_1, x_2, x_3) = I(x_1 + x_3 > 0, x_2 + x_3 > 0) - 2I(x_1 + x_3 > 0, -x_2 + x_3 > 0) + 1/3$. This is not symmetric in its arguments, so the symmetrized version has six terms, $h(x_1, x_2, x_3) = [f(x_1, x_2, x_3) + f(x_1, x_3, x_2) + f(x_2, x_1, x_3) + f(x_2, x_3, x_1) + f(x_3, x_1, x_2) + f(x_3, x_2, x_1)]/6$ The corresponding U-statistic is $U_n = \frac{1}{\binom{n}{3}} \sum_{i_1 < i_2 < i_3} h(X_{i_1}, X_{i_2}, X_{i_3})$.

Many of you also expanded the last term out as $P(X_1 \leq X_3, X_2 \leq X_3)$. But note that since we have i.i.d random variables, this quantity is 1/3. I have given full score for this.