STAT 210B HWK #2 SOLUTIONS (DUE FEBRUARY 17)

GARVESH RASKUTTI

(1) Consider the V-statistic $V_n = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1} h(x_i, x_j)$ for a symmetric kernel h such that[1] $Eh^2 < \infty$. Show that V_n is asymptotically normal.

By van der Vaart Thm 12.3 (p. 162), if a U-statistic U_n has kernel h with $Eh^2(X_1, X_2) < \infty$ and expectation θ, then $\sqrt{n}(U_n - \theta)$ is asymptotically normal. Our approach is to show that V_n and U_n (with the same kernel) are asymptotically equivalent. Then the result will follow by Slutsky's Theorem.

We first manipulate the expressions for U_n and V_n:

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} h(X_i, X_j) = \frac{1}{n(n-1)} \sum_{i\neq j} h(X_i, X_j) \\ V_n &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1} h(X_i, X_j) = \frac{1}{n^2} \sum_{i\neq j} h(X_i, X_j) + \frac{1}{n^2} \sum_i h(X_i, X_I) \end{aligned}$$

So

$$\begin{aligned} V_n - U_n &= \left(\frac{1}{n^2} - \frac{1}{n(n-1)}\right) \sum_{i\neq j} h(X_i, X_j) + \frac{1}{n^2} \sum_i h(X_i, X_I) \\ &= -\frac{1}{n}\left(\frac{1}{n(n-1)} \sum_{i\neq j} h(X_i, X_j)\right) + \frac{1}{n}\left(\frac{1}{n} \sum_i h(X_i, X_i)\right) \\ &= -\frac{1}{n} U_n + O_P\left(\frac{1}{n}\right) \end{aligned}$$

For the last term we used the LLN to get $\frac{1}{n}\sum h(X_i, X_i) = O_P(1)$. By the asymptotic normality of U-statistics, we know[2] $\sqrt{n}(U_n - \theta) = O_p(1)$, which implies $U_n = \theta + O_p(1/\sqrt{n})$ and $n^{-1}U_n = O_p(1/n)$. This gives $V_n - U_n = O_p(1/n) = o_p(1)$. The result now follows from Slutsky's theorem.

***

Can we just write the V-statistic V_n as a U-statistic? That is, can we find a symmetric function $g(x_i, x_j)$ such that

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1} h(x_i, x_j) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} g(x_i, x_j)$$

No. For any fixed n we can find a g that satisfies the equation, but we need it to be true for all n. The form of $g(x_1, x_2)$ is already determined in the case $n = 2$:

$$\frac{1}{4}\left[h(x_1, x_1) + 2h(x_1, x_2) + h(x_2, x_2)\right] = g(x_1, x_2)$$

Date: February 17, 2011 .

[1] More specifically, in the proof we use $Eh^2(X_1, X_2) < \infty$ and $E|h(X_1, X_1)| < \infty$.

[2] If a sequence converges in distribution, then it is uniformly tight (or "bounded in probability") by the easy part of Prohorov's Theorem (van der Vaart Thm 2.4, p. 8).

Does this work for $n = 3$? Plugging in, we get

$$\begin{aligned}
\frac{1}{\binom{3}{2}} \sum_{1 \le i < j \le 3} g(x_i, x_j) &= \frac{1}{3} \cdot \frac{1}{4} \sum_{1 \le i < j \le 3} [h(x_i, x_i) + 2h(x_i, x_j) + h(x_j, x_j)] \\
&= \frac{1}{12} \left[\sum_{i=1}^{3} \sum_{j=1}^{3} h(x_i, x_j) + \sum_{k=1}^{3} h(x_k, x_k) \right] \\
&= \frac{9}{12} \left(\frac{1}{3^2} \sum_{i=1}^{n} \sum_{j=1} h(x_i, x_j) \right) + \frac{1}{12} \sum_{k=1}^{3} h(x_k, x_k) \\
&\neq \frac{1}{3^2} \sum_{i=1}^{n} \sum_{j=1} h(x_i, x_j)
\end{aligned}$$

(2) Consider the kernel $h(x_1, x_2) = 1(x_1 + x_2 > 0)$. Evaluate $\theta = Eh(X_1, X_2)$ for the mixture $F = (1 - \varepsilon)N(0, 1) + \varepsilon N(\alpha, \beta)$.

First, note that a mixture of two Gaussians is bimodal (and definitely not itself Gaussian).

Let $\delta, \delta_1, \delta_2$ be i.i.d. $P(\delta = 1) = 1 - \varepsilon$, $P(\delta = 0) = \varepsilon$. Then let Z_{11}, Z_{21} be i.i.d. $N(0, 1)$ and let Z_{12}, Z_{22} be i.i.d. $N(\alpha, \beta)$. Finally, define

$$\begin{aligned}
X_1 &= \delta_1 Z_{11} + (1 - \delta_1) Z_{12} \\
X_2 &= \delta_2 Z_{21} + (1 - \delta_2) Z_{22}
\end{aligned}$$

Then X_1 and X_2 are i.i.d. F.

We want to find $\theta = Eh(x_1, x_2) = P(X_1 + X_2 > 0)$. We can condition on the mixing variables δ_1 and δ_2:

$$\begin{aligned}
P(X_1 + X_2 > 0) &= E\left[P\left(X_1 + X_2 > 0 | \delta_1, \delta_2\right)\right] \\
&= P\left[(\delta_1, \delta_2) = (0, 0)\right] P(Z_{12} + Z_{22} > 0) \\
&\quad + P\left[(\delta_1, \delta_2) = (0, 1)\right] P(Z_{12} + Z_{21} > 0) \\
&\quad + P\left[(\delta_1, \delta_2) = (1, 0)\right] P(Z_{11} + Z_{22} > 0) \\
&\quad + P\left[(\delta_1, \delta_2) = (1, 1)\right] P(Z_{11} + Z_{21} > 0)
\end{aligned}$$

The sum of two independent Gaussians, say $N(a_1, b_1)$ and $N(a_2, b_2)$ is again Gaussian with distribution $N(a_1 + a_2, b_1 + b_2)$. So

$$\begin{aligned}
P(X_1 + X_2 > 0) &= \varepsilon^2 \left[1 - \Phi\left(\frac{-2\alpha}{\sqrt{2\beta}} \right) \right] + 2\varepsilon(1 - \varepsilon) \left[1 - \Phi\left(\frac{-\alpha}{\sqrt{1 + \beta}} \right) \right] \\
&\quad + (1 - \varepsilon)^2 \left[1 - \Phi(0) \right] \\
&= \varepsilon^2 \Phi\left(\frac{\alpha\sqrt{2}}{\sqrt{\beta}} \right) + 2\varepsilon(1 - \varepsilon)\Phi\left(\frac{\alpha}{\sqrt{1 + \beta}} \right) + \frac{1}{2}(1 - \varepsilon)^2
\end{aligned}$$

(3) Let $X_i \in \mathbb{R}^p$ be i.i.d. random variables $i = 1, 2, ..., n$ such that $X_i \sim \mathcal{N}(0, I_{p \times p})$. Define the class $\mathcal{F}$ as:

$$\mathcal{F} = \{f : \mathbb{R}^p \to \mathbb{R} \mid f(x_1, x_2, .., x_p) = \sum_{j=1}^{p} \beta_j x_j \ ; \ \sum_{j=1}^{p} |\beta_j| \leq R\}.$$

Show that:

$$\sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]| \xrightarrow{P} 0.$$

Note that

$$\begin{aligned}
\sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]| &= \sup_{\|\beta\|_1 \leq R} |\frac{1}{n} \sum_{i=1}^{n} \beta^T X_i - \mathbb{E}[\beta^T X]|] \\
&= \sup_{\|\beta\|_1 \leq R} |\frac{1}{n} \sum_{i=1}^{n} \beta^T X_i| \\
&= \frac{1}{\sqrt{n}} \sup_{\|\beta\|_1 \leq R} |\beta^T (\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}})| \\
&= \frac{R}{\sqrt{n}} \| \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}} \|_\infty.
\end{aligned}$$

where we recall that $X_i \sim \mathcal{N}(0, I_{p \times p})$. The final equality follows from Hölder's inequality. It is straightforward to show that $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}} \sim \mathcal{N}(0, I_{p \times p})$ and hence

$$\sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]| = \frac{R}{\sqrt{n}} \max_{1 \leq j \leq p} Z_j,$$

where $Z_j \sim \mathcal{N}(0, 1)$. It remains to show that $\frac{R}{\sqrt{n}} \max_{1 \leq j \leq p} Z_j \xrightarrow{P} 0$. This can be proven in a number of ways. The following approach allows us to find a rate of convergence.

$$\begin{aligned}
\mathbb{P}(\frac{R}{\sqrt{n}} \max_{1 \leq j \leq p} Z_j > \epsilon) &= \mathbb{P}(\max_{1 \leq j \leq p} Z_j > \frac{\sqrt{n}\epsilon}{R}) \\
&\leq \sum_{j=1}^{p} \mathbb{P}(Z_j > \frac{\sqrt{n}\epsilon}{R}),
\end{aligned}$$

by standard union bound. Finally, by applying the Chernoff bound

$$\mathbb{P}(Z_j > \frac{\sqrt{n}\epsilon}{R}) \leq \exp\big(-\frac{n\epsilon^2}{2R^2}\big),$$

for each j. Therefore

$$\begin{aligned}
\mathbb{P}(\frac{R}{\sqrt{n}} \max_{1 \leq j \leq p} Z_j > \epsilon) &\leq p \exp\big(-\frac{n\epsilon^2}{2R^2}\big) \\
&= \exp\big(\log p - \frac{n\epsilon^2}{2R^2}\big).
\end{aligned}$$

Therefore as long as $R\sqrt{\frac{\log p}{n}} \to 0$ (which is trivially true if both R and p are finite),

$$\sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)]| \xrightarrow{P} 0.$$

(4) Let $(\sigma_i)_{i=1}^n$ be i.i.d symmetric random variables such that $\mathbb{P}(\sigma_i = +1) = \mathbb{P}(\sigma_i = -1) = 1/2$. Furthermore let $(X_i)_{i=1}^n$ be i.i.d. with $X_i \sim \mathbb{P}$ and independent of $(\sigma_i)_{i=1}^n$. Letting conv$(\mathcal{F})$ be the convex hull of $\mathcal{F}$, show that:

$$\textbf{(a)} \quad \sup_{f \in \textbf{conv}(\mathcal{F})} |\frac{1}{n}\sum_{i=1}^n \sigma_i f(X_i)| = \sup_{f \in \mathcal{F}} |\frac{1}{n}\sum_{i=1}^n \sigma_i f(X_i)|.$$

(b) If $\mathcal{F}$ satisfies the ULLN, then so does conv$(\mathcal{F})$.

Note that any $g \in \text{conv}(\mathcal{F})$ may be expressed as $g = \sum_{k=1}^K \alpha_k f_k \in \text{conv}(\mathcal{F})$, where $f_k \in \mathcal{F}$, $\alpha_k > 0$ and $\sum_{k=1}^K \alpha_k = 1$ for some finite K.

(a) For any fixed $g \in \text{conv}(\mathcal{F})$:

$$\begin{aligned}
|\sum_{i=1}^n \sigma_i g(X_i)| &= |\sum_{i=1}^n \sum_{k=1}^K \sigma_i \alpha_k f_k(X_i)| \\
&= |\sum_{k=1}^K \alpha_k \sum_{i=1}^n \sigma_i f_k(X_i)| \\
&\leq \sum_{k=1}^K \alpha_k |\sum_{i=1}^n \sigma_i f_k(X_i)| \\
&\leq \max_k |\sum_{i=1}^n \sigma_i f_k(X_i)| \\
&\leq \sup_{f \in \mathcal{F}} |\sum_{i=1}^n \sigma_i f(X_i)|.
\end{aligned}$$

Since the right-hand side is independent of g, we can conclude that

$$\sup_{g \in \text{conv}(\mathcal{F})} |\sum_{i=1}^n \sigma_i g(X_i)| \leq \sup_{f \in \mathcal{F}} |\sum_{i=1}^n \sigma_i f(X_i)|.$$

The reverse inequality is trivial since $\mathcal{F} \subset \text{conv}(\mathcal{F})$

(b) The proof is almost identical in this case. Consider any function $g = \sum_{k=1}^K \alpha_k f_k \in \text{conv}(\mathcal{F})$, where $f_k \in \mathcal{F}$, $\alpha_k > 0$ and $\sum_{k=1}^K \alpha_k = 1$. Then

$$\begin{aligned}
|P_n g - Pg| &= \left|\sum_{k=1}^K \alpha_k (P_n - P) f_k\right| \\
&\leq \sum_{k=1}^K \alpha_i |(P_n - P) f_k| \\
&\leq \max_k |(P_n - P) f_k| \\
&\leq \sup_{f \in \mathcal{F}} |P_n f - Pf|.
\end{aligned}$$

Since the RHS is independent of g, we conclude that

$$\sup_{g \in \text{conv}(\mathcal{F})} |P_n g - Pg| \leq \sup_{f \in \mathcal{F}} |P_n f - Pf|$$

.