University of Illinois Spring 2001	**Econ 476** **Final Exam Answers**	Department of Economics Professor Roger Koenker

(a.) For $k \notin \{i,j\}$ $E((-)(-)) = E(-)E(-)$ by $\perp\!\!\!\perp$ of $\{X_i\}$ but both factors have mean zero. For $k = i \neq j$, condition on $X_k = X_i$ so we have

$$\begin{aligned}
&E_{X_j}((H_{k\cdot} - \theta)(H_{kj} - H_{k\cdot} - H_{j\cdot} + H_{\cdot\cdot})|X_k) \\
&= (H_{k\cdot} - \theta)(H_{k\cdot} - H_{k\cdot} - H_{j\cdot} + \theta) \\
&= (H_{k\cdot} - \theta)(-H_{j\cdot} + \theta)
\end{aligned}$$

Now take expectations wrt to X_k or X_j. To make some heuristic connection to regression, we can view the decomposition of U_n into the $\bar{Y}_n$ and V_n parts as analogous to the usual decomposition of a random variable (regression response variable y) into a conditional mean component ($\hat{y} = x'\hat{\beta}$) and a residual $\hat{u}$ component. In such decompositions the conditional mean component is the orthogonal projection of y into the lower dimensional subspace spanned by the x's. And we have: $\hat{u} \perp \hat{y}$, since $\hat{u}'\hat{y} = u'(I - P_X)P_X y = 0$ by the idempotency of P_X.

(b.) $V_n = (n(n-1))^{-1}\sum\sum_{i\neq j} V_{ij}$ but $EV_{ij} = 0$ so

$$\begin{aligned}
\text{Var }(V_n) &= (n(n-1))^{-2}\sum\sum_{i\neq j}\sum\sum_{k\neq \ell} E(V_{ij}V_{k\ell}) \\
&= (n(n-1))^{-2} 2EV_{12}^2 n(n-1)
\end{aligned}$$

only the terms with ij matching $k\ell$ contribute. There are $n(n-1)$ ij pairs, for each we have a matching $k\ell$ *pair* one as $k\ell = ij$, one as $k\ell = ji$ so we get the expression above. Note that this term is $O(n^{-2})$ and therefore negligible relative to the conditional mean contribution of the Y_i's.

(c.) $\sqrt{n}(U_n - \theta) = \sqrt{n}2(\bar{Y}_n - \theta) + \sqrt{n}V_n$

By (b.) $\sqrt{n}V_n$ has mean zero and variance converging to zero, in fact, $Var(\sqrt{n}V_n) \simeq O_p(1/n)$, so $\sqrt{n}V_n \to 0$, since it converges in quadratic mean. On the other hand $\sqrt{n}(\bar{Y}_n - \theta)$ is a classical iid CLT situation with limiting $\mathcal{N}(0, \sigma^2)$ behavior. Thus by the Slutsky lemma,

$$\sqrt{n}(U_n - \theta) \rightsquigarrow \mathcal{N}(0, 4\sigma^2)$$

(d.)

$$\begin{aligned}
Y_i &= (X_i - \mu)^2 + E(X_j - \mu)^2 - 2(X_i - \mu)E((X_j - \mu)|X_i) \\
&= (X_i - \mu)^2 + \sigma^2
\end{aligned}$$

(e.)

$$Y_j = E(|X_i - X_j| \; |X_j)$$

$$
\begin{aligned}
&= \int_{-\infty}^{X_j}(X_j - x)dF(x) + \int_{X_j}^{\infty}(x - X_j)dF(x) \\
&= X_jF(X_j) - X_j(1 - F(X_j) + \int_{X_j}^{\infty} xdF(x) - \int_{-\infty}^{X_j} xdF(x) \\
&= X_j(2F(X_j) - 1) + \int_{F(X_j)}^{1} F^{-1}(t)dt - \int_{0}^{F(X_j)} F^{-1}(t)dt \\
&= X_j(2F(X_j) - 1) + \mu - 2\psi(F(X_j))
\end{aligned}
$$

where $\mu = \int_0^1 F^{-1}(t)dt = \int_{-\infty}^{\infty} xdF(x)$, and $\psi(t) = \int_0^t F^{-1}(s)ds$. Thus,

(f.)

$$EY_i = 2E[XF(X) - \psi(F(X))]$$

But note that

$$
\begin{aligned}
EXF(X) &= \int_{-\infty}^{\infty} xF(x)dF(x) \\
&= \int_0^1 F^{-1}(t)tdt \\
&= -\int_0^1 \psi(t)dt + \psi(t)t|_0^1 \\
&= -\int \psi(t)dt + \mu
\end{aligned}
$$

and

$$
\begin{aligned}
E\psi F(X) &= \int_{-\infty}^{\infty}\int_0^{F(x)} F^{-1}(s)dsdF(x) \\
&= \int_0^1\int_0^t F^{-1}(s)dsdt \\
&= \int_0^1 \psi(t)dt
\end{aligned}
$$

so

$$EY_i = 2[\mu - 2\int_0^1 \psi(t)dt]$$

Now the Lorenz curve corresponding to the distribution, F, is given by

$$\lambda(t) = \mu^{-1}\psi(t) = \mu^{-1}\int_0^t F^{-1}(s)ds$$

so we have

$$EY_i = 2\mu[1 - 2\lambda(1)]$$

So, $EY_i = 2\mu\gamma$ where γ is the usual Gini coefficient of inequality. So a nice way to write γ, is

$$\gamma = \frac{E(|X_i - X_j|)}{E(X_i + X_j)}$$

For the variance we have,

$$\begin{aligned} V(Y_i) &= EY_i^2 - (EY_i)^2 \\ &= E(E(|X_i - X_j|X_j))^2 - \theta^2 \\ &= \int (\int (|x-y|dF(x)))^2 dF(y) - \theta^2 \end{aligned}$$

(g.) This part caused more difficulty than I had intended. Given the form of the $R_{ij}(\delta)$ terms in the exam, we can compute the conditional mean (projection) bits to get our Y_i's. Note that the Y_i's here depend on δ since the U-statistic is a *function*. The first term yields,

$$\begin{aligned} \text{sgn }((X_i + X_j)|X_j) &= \int sgn(x - X_j)dF(x) \\ &= -\int_{\infty}^{X_j} dF(x) + \int_{X_j}^{\infty} dF(x) \\ &= 1 - 2F(X_j) \end{aligned}$$

Note that this is a random variable, but very conveniently $F(X_j)$ are iid standard uniform. For the second term,

$$\begin{aligned} &E[\int_0^{2\delta/\sqrt{n}} (I(X_i + X_j \leq s) - I(X_i + X_j < 0)ds|X_j] \\ &= \int_0^{2\delta/\sqrt{n}} (F(s - X_j) - F(-X_j))ds \\ &= \frac{1}{\sqrt{n}} \int_0^{2\delta} (F(t/\sqrt{n} - X_j) - F(-X_j))dt \\ &= n^{-1} f(X_j) \int_0^{2\delta} tdt \\ &= n^{-1} 2f(X_j)\delta^2 \end{aligned}$$

So we can write the objective function

$$\begin{aligned} D_n(\delta) &= \binom{n}{2}^{-1} \sum R_{ij}(\delta) \\ &= 2\bar{Y}_n + (n(n-1))^{-1} \sum\sum V_{ij} \end{aligned}$$

where the later term is asymptotically negligible by the same arguments we have made earlier, and

$$\sum Y_i = -\delta W_n + \delta^2 \xi_n$$

where (why?)

$$W_n = n^{-1/2} \sum (1 - 2F(X_j)) \rightsquigarrow \mathcal{N}(0, 1/3)$$

and (why?)

$$\xi_n = n^{-1} \sum f(X_j) \to \int_{-\infty}^{\infty} f^2(x)dx \equiv \xi_0.$$

Now consider the minimizer of the limiting form of the objective function,

$$\tilde{\delta}_n = W_n/(2\xi_n).$$

This is a pseudo-estimator. It obviously can't be computed from the data, it is the minimizer of the limiting form of the objective function, but the crucial thing is that asymptotically it behaves like the Hodges-Lehmann estimator because the two objective functions are (eventually) nearly the same. By Slutsky's lemma, we have,

$$\tilde{\delta}_n \rightsquigarrow \mathcal{N}(0, (12\xi_0^2)^{-1}).$$

Since $D_n(\delta)$ converges uniformly in δ, it follows that $\hat{\delta}_n$ has the same limiting distribution as $\tilde{\delta}_n$.

(h.) We would expect that the median of the pairwise averages would behave more robustly than the mean, and perhaps someless less robustly than the sample median. This is the case. As F becomes heavier tailed, the contribution of $\sigma^2(F)$ in the ARE expression blows up, but the asymptotic variance of the Hodges-Lehmann estimator is affected very little. In the extreme Cauchy case we have inconsistency of the sample mean, and therefore the ARE is unbounded.

(i.) Finding the least favorable distribution involves minimizing the expression for the ARE. If we fix the scale of the distribution F we can set $\sigma = 1$, and then the problem reduces to minimizing $\int_{-\infty}^{\infty} f^2(x)dx$. This is just the problem of finding the optimal kernel in density estimation, and leads to the Epanechnikov quadratic kernel. Plugging that into the ARE expression yields worst case ARE of .864. This is rather remarkable: the sample mean can be arbitrarily worse than Hodges Lehmann, but it can't be any better than about 15 percent.