Stat 710 12/19/02

Solutions to Selected Problems, HW 5

Ch. 11, #4. The key point in this problem is to use **only** the projection definition, with suitable limiting operations, to obtain the usual definition. It is a backward approach:

(i) Since $E(X \wedge M \,|\, Y)$ is a projection, it satisfies for all square-integrable functions $h(Y)$,

$$E\Big((X \wedge M - E(X \wedge M \,|\, Y)\, h(Y) \Big) \;=\; 0$$

First we check the monotonicity in problem #3: if $U \ge V$ a.s., then $E(U|Y) \ge E(V|Y)$ a.s, because for the square-integrable nonnegative function $h(Y) = I_{[E(U|Y)<E(V|Y)]}$, $E\Big((E(U|Y) - E(Y|Y))h(Y)\Big) = E((U - V)h(Y)) \ge 0$. From this monotonicity, it follows that $E(X \wedge M \,|\, Y)$ is an increasing function of M, and therefore has an a.s. limit Z, finite or infinite, which must in fact be finite if $EX < \infty$, by the motonone convergence theorem, since $E(E(X \wedge M \,|\, Y)) \le E(X) < \infty$.

(ii) If X is square-integrable, then $E(X \wedge M \,|\, Y) \nearrow Z$, and by Fatou's Lemma, $EZ^2 \le \liminf_M \, E(E(X \wedge M \,|\, Y)^2) \le E(X^2) < \infty$, and

$$E((X - Z)\, h(Y)) \;=\; \lim_M \, E((X \wedge M - Z)\, h(Y)) \;=\; 0$$

which implies that Z coincides with the usual definition of $E(X|Y)$ as a projection.

(iii) We know $E(X \wedge M \,|\, Y) \nearrow Z \equiv E(X|Y)$. But if $EX < \infty$, then (again by Fatou) $EZ < \infty$ and by dominated convergence $E|E(X \wedge M \,|Y) - Z| \to 0$, so $0 \;=\; E((X \wedge M - E(X \wedge M \,|\, Y))g(Y) \to E((X - Z)g(Y))$.

(iv) A.s. uniqueness follows when $EX < \infty$, since if Z^* were another positive integrable function of Y satisfying the projection identity, then by choosing the general integrable function $g(Y) = \mathrm{sgn}(Z - Z^*)$, we obtain $E((Z - Z^*)\, \mathrm{sgn}(Z - Z^*)) \;=\; 0$

Finally, $E(X|Y) = E(X^+ \,|\, Y) - E(X^- \,|\, Y)$ for general integrable X.

Ch. 12, #8. In this problem, the difficulty is that the 'kernel' $h_{ij}(x,y) = I_{[i<j,\, x<y]}$ is not and cannot be made symmetric. Nevertheless, the projection idea — not the U-statistic asymptotic normality theorem — applies, as follows.

First, check that (under the assumption of *iid* variables X_i, the term $I_{[X_i<X_j]}$ has mean $1/2$, and its projection onto the space of linear variables (with mean 0) is $(P(x<X)-\frac{1}{2})_{x=X_i} - (P(X<x)-\frac{1}{2})_{x=X_j} = F(X_j) - F(X_i)$. Thus the projection of $T = \sum_{i<j} I_{[X_i<X_j]}$ onto the same linear space gives

$$\sum_{i<j}(F(X_j)-F(X_i)) = \sum_{j=1}^n (j-1)F(X_j) - \sum_{i=1}^n (n-i)F(X_i) = \sum_{k=1}^n (2k-1-n)\, F(X_k)$$

which is a *weighted* sum of *iid* variables. The uniform variable $F(X_k)$ has variance $1/12$, and it is not hard to check that the variances of T and its projection are asymptotically the same, so that the nonidentical-summand CLT applies to show that $(\sqrt{n}/\binom{n}{2})\,(T-\frac{1}{2}\binom{n}{2})$ is asymptotically normally distributed with mean 0 and variance $1/9$, calculated as follows.

$$a.var = \lim_n n\binom{n}{2}^{-2}\sum_{k=1}^n \frac{1}{12}(2k-1-n)^2 = \lim_n \frac{1}{3n}\sum_{k=1}^n (1-\frac{2k-1}{n})^2$$

$$= \frac{1}{3}\int_0^1 (1-2x)^2\,dx = \frac{1}{9}$$

So the test rejects when $T \ge \binom{n}{2}\,(1/2 + z_\alpha/(3\sqrt{n})$.

Stat 710, HW6 12/19/02

ARE's & Sample $\hat{\rho}$ vs. Kendall τ

In a few of the problems in HW6, it is important to ascertain not only the variancce of the test-statistics under H_0, but also the asymptotic expectation under contiguous alternatives. Through consideration of the 'slopes' $\mu'(\vartheta)/\sigma(\vartheta)$, whose squares are called 'efficicacy' in other books, we compare different test-statistics not all of which are asymptotically unbiased for the same parameter ϑ, with respect to Asymptotic Relative Efficiency. Recall that for a normalized test-statistic which under contiguous alternatives $\vartheta = \vartheta_0 + c/\sqrt{n}$ has asymptotic expectation ah and variance σ_0^2, the asymptotic power for a one-sided size-α test is $1 - \Phi(z_\alpha - ah/\sigma_0)$, so that different test-statistics are compared via ARE which is the ratio of their quantities a^2/σ_0^2.

Ch. 13, #3. Based on the idea given in the problem, we check that the scaled and centered Spearman's Correlation is

$$\sqrt{n}\,\rho_n \;=\; \frac{12\sqrt{n}}{n(n^2-1)} \sum_{i=1}^n R_i^X\, R_i^Y \;-\; 3\sqrt{n}\,\frac{n+1}{n-1}$$

is asymptotically equivalent to the U-statistic (without symmetrized kernel)

$$\sqrt{n}\,\frac{3}{n^{5/2}} \sum_{i=1}^n \sum_{k \neq l} \mathrm{sgn}(X_i - X_k)\,\mathrm{sgn}\,(Y_i - Y_l) \;-\; 3\sqrt{n}\,\frac{n+1}{n-1}$$

and, via projections, is is turn asymptotically equivalent to

$$\frac{3}{\sqrt{n}} \sum_{i=1}^n (2F_X(X_i) - 1)\,(2F_Y(Y_i) - 1)$$

Finally, by the central limit theorem (for iid summands), this statistic under contiguous alternatives $f_Y(y) = f_X(y - h/\sqrt{n})$ is asymptotically normally distributed with variance 1 .

Ch. 14, #3. In this problem, it is necessary to consider the signed-rank statistic

$$\sqrt{n}\left(\frac{1}{\binom{n}{2}} \sum_{i<j} I_{[X_i + X_j > 0]} \;-\; \frac{1}{2}\right)$$

Under the contiguous alternatives $f_X(x) = f(x - h/\sqrt{n})$, for symmetric density f, the statistic centered for the null hypothesis (h=0) is

$$\sqrt{n}\hat{U} \;=\; -\frac{2}{\sqrt{n}} \sum_{i=1}^{n} (F(-X_i) - \frac{1}{2})$$

which via projection is asymptotically equivalent (under the contiguous alternative) to

$$-\frac{2}{\sqrt{n}} \sum_{i=1}^{n} (F(-X_i) - \int f(x - \frac{h}{\sqrt{n}})\, F(-x)\, dx) + \sqrt{n} \int\; f(x - \frac{h}{\sqrt{n}})\, (F(-x) - \frac{1}{2})$$

$$\sim \mathcal{N}\Big(h \int f(x)\, f(-x)\, dx,\; \frac{1}{3}\Big)$$

The result is that the 'slope' for this statistic is $\sqrt{3} \int f^2$.

Ch. 14, #5. It is given in the book, Chapter 14, page 30, that the scaled sample correlation coefficient $\sqrt{n}\, r_n$, is asymptotically unbiased for $\rho \sqrt{n}$ and is normally distributed with variance $(1-\rho^2)^2$. In the present setting, of alternatives contiguous to the independent case, with $\rho = c/\sqrt{n}$, we have $\sqrt{n}\, r_n \sim \mathcal{N}(c,\, 1)$.

Next, for $\tau = \frac{4}{n(n-1)} \sum_{i<j} I_{[(X_i - X_j)(Y_i - Y_j) > 0]}\, 1$, the book gives asymptotic normality for $\sqrt{n}\,\tau$ with variance $4/9$ and (under bivariate-normal distribution with correlation $c/\sqrt{n}$)

$$E(\sqrt{n}\,\tau) \;=\; \sqrt{n}\, (4 P_{c/\sqrt{n}}(X > 0, Y > 0) \;-\; 1)$$

$$=\; \sqrt{n} \int_0^{\infty} \phi(x)\, (\frac{1}{2} - \Phi(-\frac{cx}{\sqrt{n}}))\, dx \;\sim\; \frac{2c}{\pi}$$

where in the last line we have approximated $\Phi(-cx/\sqrt{n})$ by $cx\phi(0)/\sqrt{n}$.

As a result, the 'slope' for the τ statistic is $(2/\pi)/(3/2) = 3/\pi$, and the ARE of tau to r_n becomes $(3/\pi)^2$.