Solutions to Homework #2
Math 6070-1, Spring 2006

Let $X_1, X_2, \ldots, X_n$ be random variables, all i.i.d., and with the same distribution function F that has density $f := F'$. Define for all $p \geq 1$,

$$Q_n^{(p)}(F) := \left\{ \int_{-\infty}^{\infty} \left| \hat{F}_n(x) - F(x) \right|^p f(x)\, dx \right\}^{1/p}.$$

This is a kind of "distance" between $\hat{F}_n$ and F, although it is different from $D_n(F)$.

1. *Prove that $Q_n^{(p)}(F) \leq 2$, so $Q_n^{(p)}(F)$ is always finite.*

   Solution: The inequality for $Q_n^{(p)}(F)$ follows readily from the facts that: (i) $|\hat{F}_n(x) - F(x)| \leq \hat{F}_n(x) + F(x) \leq 2$; and (ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

2. *Compute $Q_n^{(p)}(F_u)$ where $F_u(x) = x$ for $0 \leq x \leq 1$, $F_u(x) = 0$ if $x < 0$, and $F_u(x) = 1$ if $x \geq 1$.*

   Solution: Because $f_u(x) := F_u'(x) = 1$ for $0 \leq x \leq 1$,

$$Q_n^{(p)}(F) := \left\{ \int_0^1 \left| \hat{F}_n(x) - x \right|^p dx \right\}^{1/p}.$$

3. *Prove that $Q_n^{(p)}(F)$ is distribution-free. That is, if F and G are strictly increasing distribution functions such that F' and G' both exist, then the distribution of $Q_n^{(p)}(F)$ is the same as that of $Q_n^{(p)}(G)$.*

   Solution: Let F^{-1} denote the inverse function to F. Then, a change of variables $[x = F^{-1}(x)]$ yields the following:

$$\begin{aligned} Q_n^{(p)}(F) &= \left\{ \int_0^1 \left| \hat{F}_n\left(F^{-1}(z)\right) - z \right|^p f\left(F^{-1}(z)\right) dF^{-1}(z) \right\}^{1/p} \\ &= \left\{ \int_0^1 \left| \hat{F}_n\left(F^{-1}(z)\right) - z \right|^p dz \right\}^{1/p} \end{aligned}$$

   In the last line we used the fact that if $z = F(x)$ then $dz = f(x)\, dx = f(F^{-1}(z))\, dx$; whence it follows that $dF^{-1}(z) = dz/f(F^{-1}(z))$. Finally, note that

$$\begin{aligned} \hat{F}_n\left(F^{-1}(z)\right) &= \frac{1}{n} \sum_{j=1}^n \mathbf{I}\left\{X_j \leq F^{-1}(z)\right\} \\ &= \frac{1}{n} \sum_{j=1}^n \mathbf{I}\left\{F(X_j) \leq z\right\}, \end{aligned}$$

and this is the empirical distribution function based on i.i.d. uniform-$(0\,,1)$ variables $F(X_1),\ldots,F(X_n)$. Hence, $Q_n^{(p)}(F)$ has the same distribution as $Q_n^{(p)}(F_u)$. The asserted distribution-free-ness follows readily from this, because $Q_n^{(p)}(G)$ has the same distribution as $Q_n^{(p)}(F_u)$ too, and for the same reasons.

4. *Prove that under F, as $n \to \infty$, $Q_n^{(2)}(F) \xrightarrow{\mathrm{P}} 0$, provided that F is strictly increasing and has a density. Do this by first proving that*

$$\mathrm{E}_F\left[\left|Q_n^{(2)}(F)\right|^2\right] = \frac{1}{6n}.$$

Solution: Because of distribution-free-ness we can and will assume that $F = F_u$. Because $\mathrm{E}_{F_u}[\hat{F}_n(x)] = F_u(x) = x$ for $x \in [0\,,1]$,

$$\mathrm{E}_{F_u}\left[\left|Q_n^{(2)}(F_u)\right|^2\right] = \mathrm{E}_{F_u}\left[\int_0^1 \left|\hat{F}_n(x) - x\right|^2\,dx\right] = \int_0^1 \mathrm{Var}_{F_u}\left(\hat{F}_n(x)\right)\,dx.$$

But recall that the preceding variance is equal to $x(1-x)/n$. Therefore,

$$\mathrm{E}_{F_u}\left[\left|Q_n^{(2)}(F_u)\right|^2\right] = \frac{1}{n}\int_0^1 x(1-x)\,dx = \frac{1}{6n}.$$

Warning: The original version of this exercise asked you to prove that $Q_n^{(p)}(F)$ goes to zero in probability for any p [not just $p = 2$]. The reason is this: Because $|\hat{F}_n(x) - F(x)|^p \le \{D_n(F)\}^p$, it follows that $Q_n^{(p)}(F) \le D_n(F)$, and so the Glivenko–Cantelli theorem does the rest.

5. *Suppose F is strictly increasing has a density. Then provide a heuristic justification of the fact that, under F, as $n \to \infty$,*

$$\sqrt{n}\; Q_n^{(p)}(F) \xrightarrow{d} \left\{\int_0^1 |B^\circ(x) - x|^p\,dx\right\}^{1/p},$$

where B° denotes the Brownian bridge on $[0\,,1]$. Later on in a Project we will see how to simulate the distribution of the latter limiting object.

Solution: We can alternatively prove that as $n \to \infty$,

$$n^{p/2}\int_{-\infty}^{\infty}\left|\hat{F}_n(x) - F(x)\right|^p f(x)\,dx \xrightarrow{d} \int_0^1 |B^\circ(x) - x|^p\,dx.$$

Because we are interested in the asymptotic distribution of $Q_n^{(p)}(F)$, we can and will assume that $F = F_u$ [the distribution-free property]. Now, the CLT argument in the lecture notes provides a rigorous justification of the following: For all $k \geq 1$,

$$n^{p/2}\frac{1}{k}\sum_{j=1}^{k}\left|\hat{F}_n\left(\frac{j}{k}\right)-\frac{j}{k}\right|^p = \frac{1}{k}\sum_{j=1}^{k}\left|\sqrt{n}\left(\hat{F}_n\left(\frac{j}{k}\right)-\frac{j}{k}\right)\right|^p \xrightarrow{d} \frac{1}{k}\sum_{j=1}^{k}\left|B^\circ\left(\frac{j}{k}\right)-\frac{j}{k}\right|^p.$$

On the other hand, as $k \to \infty$, then Riemann-sum approximations reveal that:

$$\frac{1}{k}\sum_{j=1}^{k}\left|\hat{F}_n\left(\frac{j}{k}\right)-\frac{j}{k}\right|^p \to \int_0^1 \left|\hat{F}_n(x)-x\right|^p\,dx; \text{ and}$$

$$\frac{1}{k}\sum_{j=1}^{k}\left|B^\circ\left(\frac{j}{k}\right)-\frac{j}{k}\right|^p \to \int_0^1 \left|B^\circ(x)-x\right|^p\,dx.$$

So the assertion of the exercise is feasible. [It is in fact correct, although our "proof" falls short of completely proving it.]

6. *Use 4 to test $H_0: F = F_0$ versus $H_1: F \neq F_0$ for a known distribution function F_0 that is strictly increasing and has a density.*

   Solution: Find c such that

   $$\mathrm{P}_F\left\{Q_n^{(p)}(F) \geq c/\sqrt{n}\right\} = 1-\alpha.$$

   By the distribution-free property, this c does not depend on F. So it can be either found by simulation, or by approximation via

   $$\mathrm{P}\left\{\int_0^1 \left|B^\circ(x)-x\right|^p\,dx \geq c^p\right\} = 1-\alpha.$$

   [You would do well to check the arithmetic!] Then, we opt to reject H_0 if and only if $Q_n^{(p)}(F) \geq c/\sqrt{n}$.

7. *Suppose F has a density f which satisfies $f(x) > 0$ for all x. Then prove that F is strictly increasing.*

   Solution: This follows from the identity, $F(x) = \int_{-\infty}^{x} f(u)\,du$. [To be completely honest we need f to have some minimal regularity properties. For instance, "f = piece-wise continuous" will do. Remember that this sort of regularity is needed even to define the integral of f via Riemann-sum approximations. So assuming this sort of regularity is natural as well as inevitable.]