Solutions to Stat 710 Problem Set 2

#19.3. \( Z(t) \) is a standard Brownian motion, which implies that for \( 0 \leq t \leq 1 \), 
\( Z(t) - tZ(1) \equiv Y(t) \) is a process Gaussian finite dimensional distributions with mean-0 and covariances for \( 0 \leq s \leq t \leq 1 \) given by 
\[
\text{Cov}(Y(s), Y(t)) = \text{Cov}(Z(s), Z(t)) - s \text{Cov}(Z(1), Z(t)) - t \text{Cov}(Z(s), Z(1)) + st \text{Var}(Z(1))
\]
which is \( s - st - st + st = (s(1 - t)) \), the covariance of Brownian bridge. Since finite dimensional distributions uniquely determine the law of the process on \( l^\infty[0,1] \) or \( C[0,1] \), we are done.

#19.4. Here \( F_m, G_n \) are empirical distribution functions, and via the classical Donsker Theorem, as \( m, n \to \infty \), 
\[
\sqrt{m} (F_m - F) \overset{D}{\to} W^o \circ F, \quad \sqrt{m} (G_n - G) \overset{D}{\to} W^o \circ G \text{ in } l^\infty(R)
\]
From now on, assume that as \( m, n \to \infty \), also \( \frac{m}{m+n} \to \lambda \in (0,1) \).

(i) Then by the Continuous Mapping Theorem, or simply independence of the two empirical processes (for X observations and Y observations respectively), under \( H_0 : F = G \), 
\[
\sqrt{m + n} (F_m(\cdot) - G_n(\cdot)) \overset{D}{\to} \frac{1}{\sqrt{\lambda}} W_1^o \circ F - \frac{1}{\sqrt{1 - \lambda}} W_2^o \circ G \overset{D}{=} \frac{1}{\sqrt{\lambda(1 - \lambda)}} W^o \circ F
\]
where \( W_1^o, W_2^o, W^o \) are Brownian bridge processes, the first two of which are independent. Thus under the null hypothesis the Continuous Mapping Theorem implies 
\[
\sqrt{m+n} K_{m,n} \equiv \sup \left| \sqrt{m+n} (F_m(t) - G_n(t)) \right| \overset{D}{\to} \frac{1}{\sqrt{\lambda(1 - \lambda)}} \sup_s |W^o(s)|
\]
as long as \( F = G \) is continuous.

(ii). By the argument given in (i), for general fixed \( F \neq G \), 
\[
\sqrt{m+n} (F_m(\cdot) - G_n(\cdot) - F + G) \overset{D}{\to} \frac{1}{\sqrt{\lambda}} W_1^o \circ F - \frac{1}{\sqrt{1 - \lambda}} W_2^o \circ G
\]
in \( l^\infty(R) \) as \( m, n \to \infty \). Take \( c \sqrt{m+n} \) equal to the \( 1 - \alpha \) quantile of the (continuously distributed) random variable \( \sup_t |W^o(t)|/\sqrt{\lambda(1 - \lambda)} \), for arbitrarily small fixed \( \epsilon > 0 \), take \( \eta \sqrt{m+n} \) to be the \( 1 - \epsilon \) quantile of the same r.v. It follows that under probabilities with any fixed \( F \neq G \), 
\[
P(K_{m,n} > c) \geq P(\sqrt{m+n} \|F_m - G_n - F + G\|_\infty \leq \eta, \sqrt{m+n} \|F - G\|_\infty > \eta - c)
\]
which converges to \( 1 - \epsilon \) as \( m, n \to \infty \). Since \( \epsilon \) was arbitrary, this shows the test based upon \( K_{m,n} \) is consistent against all fixed alternatives.
(iii) Assume \( F_0 = G_0, F = F_{g/\sqrt{n}}, G = G_{h/\sqrt{n}} \). It is then easy to check by differentiability of the d.f. families with respect to the scalar parameter \( \theta \),

\[
\sqrt{m + n} (F_m(\cdot) - G_n(\cdot)) \overset{D}{=} \frac{1}{\sqrt{\lambda (1 - \lambda)}} W^\alpha + \frac{g}{\sqrt{\lambda}} F'_0 - \frac{h}{\sqrt{1 - \lambda}} G'_0
\]

from which power can readily be calculated (although not in closed form).

\#19.5. Now \( \mathcal{F} = \{ f : [0, 1] \to [0, 1] : \forall x, y, |f(x) - f(y)| \leq |x - y| \} \). Fix \( \epsilon > 0 \) and points \( t_i = \min(i \epsilon / 2, 1) \) for \( i = 0, 1, \ldots, [2/\epsilon] + 1 \). Bracket every \( f \in \mathcal{F} \) (with gap \( \epsilon \) in uniform norm) by functions

\[
h_{L, \tau} = \sum_{i=0}^{[2/\epsilon]+1} I_{[i \epsilon/2, (i+1) \epsilon/2]}(t_i), \quad h_{U, \tau} = \min(h_{L, \tau} + \epsilon, 1)
\]

where the vector \( \tau \) defining these bracketing functions for \( f \) has components \( \tau_i \) defined as \( \max\{t_j : t_j \leq f(t_i)\} \). Moreover, since such \( \tau_i = t_j \) must have \( \tau_{i+1} \) equal to one of \( t_{j-1}, t_j, t_{j+1} \), we can count that the number of such bracketing intervals is \( \leq (3/\epsilon) \cdot 3^{3/\epsilon} \).

\#19.6. (i) Here \( \mathcal{C} = \{(a, b) : -\infty < a \leq b < \infty\} \). Such intervals obviously pick out individual points from among 2 but cannot separate the middle of 3 ordered points on the line. Therefore \( VC(\mathcal{C}) > 2 \) but \( \leq 3 \) and therefore is equal to 3.

(ii) Now \( \mathcal{C} = \{(-\infty, a_1] \times (-\infty, a_2] : a_1, a_2 \in \mathbb{R}\} \subset \mathbb{R}^2 \). Again, obviously \( VC(\mathcal{C}) > 2 \) since \( \mathcal{C} \) picks out all subsets of two points \( (a_1, a_2), (b_1, b_2) \) which satisfy \( a_1 < b_1, a_2 > b_2 \). Now consider sets of three points \( a, b, c \) in the plane, and without loss of generality let \( c_1 \leq \max(a_1, b_1) \) and \( c_2 \leq \max(a_2, b_2) \). Then any set \( C \in \mathcal{C} \) containing \( a, b \) necessarily contains \( c \) also. Therefore \( VC(\mathcal{C}) = 3 \).

(iii) Now fix a monotonic function \( \psi \), with \( \mathcal{C} \) equal to the set of subgraphs for functions \( \psi(\cdot - \theta) \) as \( \theta \) ranges over the whole real line. Obviously \( VC(\mathcal{C}) = 2 \), since for any two points \( (x_i, t_i) \), the point with smaller value of \( \psi(x_i) - t_i \) is necessarily contained in any subgraph which contains the larger value \( \psi(x_i) - t_i \).

\#19.7. Let \( \mathcal{F} \) be VC, i.e. the collection of sets \( \{(x, t) : f(x) > t\} \) with \( f \) ranging over all of \( \mathcal{F} \), is VC.

(i) \( \{x_1, \ldots, x_n\} \) is shattered by \( \{f \geq 0\}_{f \in \mathcal{F}} \) whenever \( \{(x_1, 0), \ldots, (x_n, 0)\} \) is shattered by \( \mathcal{F} \)-subgraphs, denoted \( SG_\mathcal{F} \).

(ii) Now fix a function \( g \), and consider whether \( \mathcal{G} = \{\{(x, t) : f(x) + g(x) > t\} : f \in \mathcal{F}\} \) shatters \( \{(x_1, t_1), \ldots, (x_n, t_n)\} = S_n \). Note

\[
\{(x_{i_k}, t_{i_k}) \mid k = 1, \ldots, r\} = \{(x_i, t_i) \in S_n : f(x_i) > t_i - g(x_i)\}
\]
which says that these indices \( i_k \) are those \( i \) for which \( f(x_i) > t_i - g(x_i) \). Hence \( \mathcal{G} \) shatters \( S_n \) if and only if \( SG_F \) shatters \( \{ (x_i, t_i - g(x_i)) \} \). Thus the VC indices of \( \mathcal{G} \) and \( SG_F \) are the same!

(iii) The argument and result are similar to that in (ii) except that now, we consider second coordinates \( \frac{t_i}{g(x_i)} \), we must consider separately points \( x_i \) with \( g(x_i) < 0 \), \( = 0 \), and \( > 0 \). It is easy to argue that within any set of \( 3n - 2 \) points \((x_i, t_i)\) there must be at least \( n \) satisfying one of the conditions \( g(x_i) < 0 = 0 \), or \( > 0 \). Then if \( n = VC(SG_F) \), at least one of the three sets \( \{ (x_i, t_i) : g(x_i) < 0, f(x_i) < \frac{t_i}{g(x_i)} \} \) or \( \{ (x_i, t_i) : g(x_i) > 0, f(x_i) > \frac{t_i}{g(x_i)} \} \), or \( \{ (x_i, t_i) : g(x_i) = 0 > t_i \} \) fails to be shattered by subgraphs in \( SG_F \).

#19.10. Now \( \hat{m} = \text{med}(X_1, \ldots, X_n) \) is a near root of \( \sum_{i=1}^{n} \text{sgn}(X_i - \theta) \). We are asked for the asymptotic distribution of \( n^{-1} \sum_{i=1}^{n} |X_j - \hat{m}|. \) We assume the distribution of \( X_i \) is continuous, with unique median \( m_0 \). (That is, \( m_0 \) is a point of left and right increase for the d.f. \( F \) of \( X_i \).)

First use the Donsker property of \( \mathcal{F} = \{ \text{sgn}(x - \theta), |x - \theta| : \theta \in \mathbb{R} \} \) to conclude from Lemma 19.24 that as \( n \to \infty \)

\[
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left( |X_j - \hat{m}| - (E|X_1 - \theta|)_{\theta=\hat{m}} - |X_j - m_0| + E|X_1 - m_0| \right) \overset{p}{\to} 0
\]

Also, near \( m_0 \) we know (from p.55) that \( E|X_1 - \theta| - E|X_1| = 2 \int_{0}^{\hat{m}} F(x) dx - \hat{m} \), which implies that

\[
\sqrt{n} \left( E|X_1 - \theta|_{\theta=\hat{m}} - E|X_1 - m_0| \right) \overset{p}{\to} \sqrt{n} (\hat{m} - m_0) (2F(m_0) - 1) = 0
\]

Therefore

\[
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left( |X_j - \hat{m}| - |X_j - m_0| \right) \overset{p}{\to} 0
\]

which implies

\[
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left( |X_j - \hat{m}| - E|X_1 - m_0| \right) \overset{p}{\to} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left( |X_j - m_0| - E|X_1 - m_0| \right)
\]

which converges in distribution by the usual CLT to \( \mathcal{N}(0, \text{Var}(|X_1 - m_0|)) \).