

## 1. SUPPLEMENTAL APPENDIX

### 1.1. An Example of conservative Kolmogorov-Smirnov and Kuiper tests.

Our examples are as follows. Let  $X_i$  be a Bernoulli random variable with parameter  $p = 1/2$ . Under  $H_0$ , the empirical cdf of a sample of size  $N$ ,  $F_N$  is  $F_N(x) = [1 - \bar{X}]\mathbf{1}_{[0,1)}(x) + \mathbf{1}_{[1,\infty)}(x)$  while the true cdf is  $F(x) = (1/2)\mathbf{1}_{[0,1)}(x) + \mathbf{1}_{[1,\infty)}(x)$ . The definitions of the Kolmogorov-Smirnov ( $D_N$ ) and Kuiper ( $V_N \equiv D_N^+ + D_N^-$ ) are given by

$$\begin{aligned} D_N &= \sup_x |F_N(x) - F(x)| = |1/2 - \bar{X}| \\ D_N^+ &= \sup_x [F_N(x) - F(x)]_+ = \max\{1/2 - \bar{X}, 0\} \\ D_N^- &= \sup_x [F_N(x) - F(x)]_- = \max\{\bar{X} - 1/2, 0\} \end{aligned}$$

So that  $D_N = V_N = |1/2 - \bar{X}|$ . By the CLT,  $\sqrt{N}D_N$  and  $\sqrt{N}V_N$  both converge in distribution to a  $N(0, 1/4)$ , giving asymptotic test values for a .99 level test of  $\approx 2.58/2 = 1.29$ . This shows that the respective test levels based on the assumption of a continuous  $F$ , namely 1.628 for  $D_N$  and 2.001 for  $V_N$  are much too large. In particular for  $V_N$  and large  $N$ ,  $\Pr(|\sqrt{N}(\bar{X} - 1/2)| \leq 2.001) \approx .99994$ . In other words instead of falsely rejecting the null 1% of the time, by using the 2.001 cutoff rule will falsely reject it only .006% of the time which is far too conservative.

### 1.2. Proofs.

**Lemma.** *Suppose  $\int |f(x)| dx < \infty$ . Then  $c^* \equiv \int_{[0,1]} f(y) dy$  solves*

$$\min_c \sup_{A \text{ measurable}} \left| \int_{[0,1] \cap A} [f(x) - c] dx \right|$$

*and the minimum attained is  $\frac{1}{2} \int_{[0,1]} |f(x) - c^*| dx$ .*

*Proof.* We define two set mappings  $A^+(c), A^-(c)$  respectively by

$$A^+(c) \equiv \{x : f(x) - c > 0\}, \quad A^-(c) \equiv \{x : f(x) - c < 0\}$$

and since  $f$  is measurable, both  $A^+(c)$  and  $A^-(c)$  are measurable. For any fixed  $c$  we also have

$$\sup_{A \text{ measurable}} \left| \int_{[0,1] \cap A} [f(x) - c] dx \right| = \max \left\{ \int_{[0,1] \cap A^+(c)} [f(x) - c] dx, - \int_{[0,1] \cap A^-(c)} [f(x) - c] dx \right\}$$

Define functions  $B^+(c), B^-(c)$  corresponding to the sets  $A^+(c), A^-(c)$  by

$$B^+(c) \equiv \int_{[0,1] \cap A^+(c)} [f(x) - c] dx, \quad B^-(c) \equiv - \int_{[0,1] \cap A^-(c)} [f(x) - c] dx$$

Since  $c' > c$  implies  $A^+(c') \subset A^+(c)$ ,  $[f(x) - c] \mathbf{1}_{A^+(c)}$  is decreasing in  $c$  so that  $B^+(c)$  is decreasing and similarly  $B^-(c)$  is increasing. Since we have

$$(1.1) \quad \sup_{A \text{ measurable}} \left| \int_{[0,1] \cap A} [f(x) - c] dx \right| = \max \{ B^+(c), B^-(c) \}$$

any  $\tilde{c}$  s.t.  $B^+(\tilde{c}) = B^-(\tilde{c})$  minimizes Equation (1.1). Note that identically we have

$$(1.2) \quad B^+(c) - B^-(c) = \int_{[0,1]} [f(x) - c] dx = \int_{[0,1]} f(x) dx - c$$

so that  $c^* \equiv \int_{[0,1]} f(x) dx$  minimizes Equation (1.1) and  $|c^*| < \infty$  since  $\int |f(x)| dx < \infty$ . Furthermore,  $c^*$  is unique (proof omitted). The second claim follows from Equation (1.1) and  $B^+(c^*) = [B^+(c^*) + B^-(c^*)] / 2$ .  $\square$

**Lemma.**  $Y \in I(\epsilon)$  iff  $aY + b \in I(\epsilon)$  for all integers  $a, b$  with  $a \neq 0$ .

*Proof.* One direction is obvious by taking  $a = 1, b = 0$ . Considering the other direction, fix  $Y \in I(\epsilon)$  and by assumption there exist positive constants  $c_i$  s.t. for

every measurable set  $A$ ,

$$(1.3) \quad \left| \int_A f(y)dy - \int_A \sum c_i \mathbf{1}_{[i,i+1)}(y)dy \right| \leq \epsilon$$

and for any strictly monotone transformation  $T$  of  $Y$  with differentiable inverse we have  $\int_A f(y)dy = \int_{TA} f \circ T^{-1}(y) \cdot (T^{-1})'(y)dy$  where  $g(y) \equiv f \circ T^{-1}(y) \cdot (T^{-1})'(y)$  is the pdf of  $T(Y)$ . Referring to Equation (1.3), we also have

$$\int_A \sum c_i \mathbf{1}_{[i,i+1)}(y)dy = \int_{TA} \sum c_i \mathbf{1}_{[T(i),T(i+1))}(y) \cdot (T^{-1})'(y)dy$$

Assuming  $T$  is measurable, since Equation (1.3) holds for any  $A$ , in particular  $T^{-1}(A)$ , we have for any measurable  $A$  that

$$(1.4) \quad \left| \int_A g(y)dy - \int_A \sum c_i \mathbf{1}_{[T(i),T(i+1))}(y) \cdot (T^{-1})'(y)dy \right| \leq \epsilon$$

Considering  $T(x) \equiv ax + b$  for  $a, b \in \mathbb{Z}$  and appealing to Equation (1.4), we have for every  $A$  that

$$\left| \int_A g(y)dy - \int_A \sum ac_i \mathbf{1}_{[ai+b, a(i+1)+b)}(y)dy \right| \leq \epsilon$$

Defining  $d_j \equiv \sum ac_i \mathbf{1}_{[ai+b, a(i+1)+b)}(j)$ , from the last equation we have

$$\left| \int_A g(y)dy - \int_A \sum d_j \mathbf{1}_{[j,j+1)}(y)dy \right| \leq \epsilon$$

so that  $T(Y) \in I(\epsilon)$  as claimed. □