Statistics 330b/600b, Math 330b spring 2017
Homework # 5
Due: Thursday 23 February

*[1] Suppose $U_1, U_2, \dots$ is a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each distributed Unif$(0,1)$. For each θ in $[0,1]$ define a random variable $X_i(\theta) = \{U_i \le \theta\}$ and $S_n(\theta) = \sum_{i \le n} X_i(\theta)$. You may use the fact that $S_n(\theta)$ has a Binomial distribution, with

$$\mathbb{P}\{S_n(\theta) = k\} = \binom{n}{k} \theta^k (1-\theta)^{n-k} \qquad \text{for } k = 0, 1, \dots, n.$$

(i) If you are not already aware of the fact, prove that $\mathbb{P}S_n(\theta) = n\theta$ and $\mathbb{P}\left(S_n(\theta) - n\theta\right)^2 = n\theta(1-\theta)$. Hint: Don't forget that $S_n(\theta)$ is a sum of independent random variables. Kill the orphans!

(ii) Let g be a continuous function defined on $[0,1]$. Remember that g must also be uniformly continuous: for each fixed $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that

$$|g(s) - g(t)| \le \epsilon \qquad \text{whenever } |s-t| \le \delta_\epsilon, \text{ for } s,\, t \text{ in } [0,1].$$

Remember also that $|g|$ must be uniformly bounded, say, $\sup_t |g(t)| = M < \infty$. Show that $|g(x/n) - g(\theta)| \le \epsilon + 2M|x - n\theta|^2/(n\delta_\epsilon)^2$ for $0 \le x \le n$.

(iii) Show that $p_n(\theta) := \mathbb{P}g(S_n(\theta)/n)$ is a polynomial in θ.

(iv) Deduce that $\sup_{0 \le \theta \le 1} |p_n(\theta) - g(\theta)| < 2\epsilon$ for n large enough. That is, deduce that $g(\cdot)$ can be uniformly approximated by polynomials over the range $[0,1]$, a result known as the ***Weierstrass approximation theorem***.

*[2] Suppose $X_1, X_2, \dots$ are independent random variables with $\mathbb{P}X_i = 0$ for each i and $M := \sup_i \mathbb{P}X_i^6 < \infty$. Show that $\mathbb{P}|X_1 + \dots + X_n|^6 = O(n^3)$ as $n \to \infty$.

*[3] Define functions $S : [0,1) \to [0,1)$ and $T : [0,1) \to \{0,1\}$ by

$$T(y) = \lfloor 2y \rfloor \quad \text{AND} \quad S(y) = 2y - T(y).$$

(i) Explain why both S and T are Borel measurable functions.

(ii) Suppose Y_0 is a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution Unif$[0,1)$. Show that the random variables $X_1 = T(Y_0)$ and $Y_1 = S(Y_0)$ are independent with $Y_1 \sim$ Unif$[0,1)$ and $X_1 \sim$ Ber$(1/2)$. Hint: Consider $\mathbb{P}\{Y_1 \le y, X_1 = x\}$ for $x \in \{0,1\}$.

(iii) Recursively define $Y_i = S(Y_{i-1})$ and $X_i = T(Y_{i-1})$. Show that X_1, X_2, X_3, Y_3 are independent. Hint: Note that X_2, X_3 and Y_3 are measurable functions of Y_1, which is independent of X_1.

(iv) In fact the random variables $\{X_i : i \in \mathbb{N}\}$ are independent, each distributed Ber$(1/2)$. Proof? Define a new random variable $Z := \sum_{i \in \mathbb{N}} 2X_i/3^i$. Let P denote the distribution of Z. For each $k \in \mathbb{N}$, explain why there exist 2^k disjoint intervals of length 3^{-k}, each with P probability 2^{-k}.

(v) Deduce that there exists a closed subset A of $[0,1)$ with zero Lebesgue measure and $PA = 1$.

(vi) Explain why $P\{y\} = 0$ for each $y \in [0,1)$.

(vii) Define $F(y) := P[0,y]$ for $0 \le y \le 1$. Show that F is a continuous function with derivative $F'(y)$ defined and equal to zero on the open set A^c.

Remark. The function F is continuous with derivative zero almost everywhere [Lebesgue] on $[0,1)$. Nevertheless $F(1) - F(0) \ne \int F'(y)\,dy$. How does that fit with what you learned in Calculus?