Statistics 330b/600b, Math 330b spring 2018
Homework # 5
Due: Thursday 22 February

Please attempt at least the starred problems. Please explain your reasoning. Keep looking at latex.pdf. Please try to avoid strings of $\therefore$ (???) $\because$ (???) $\therefore$ (???) $\because$ (???) . . . . They really, really can be hard to decipher.

*[1] Suppose $\{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_5\}$ are independent sub-σ-fields for a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Define $\mathcal{G}_1 = \sigma(\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$ and $\mathcal{G}_2 = \sigma(\mathcal{F}_4 \cup \mathcal{F}_5)$.

(i) Show that $\mathcal{E}_1 = \{F_1F_2F_3 : F_i \in \mathcal{F}_i \text{ for } i = 1, 2, 3\}$ is a generating class for $\mathcal{G}_1$.

(ii) Prove that $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent.

*[2] 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$(0, 1)$. That is, the distribution of each U_i is Lebesgue measure on $\mathcal{B}[0, 1]$. Define $S_n(\omega, \theta) = \sum_{i\le n} \mathbb{1}\{U_i(\omega) \le \theta\}$, for each θ in $[0, 1]$. (Formally speaking, $\{S_n(\theta) : 0 \le \theta \le 1\}$ is a stochastic process.) You may use the fact that $S_n(\theta)$ has a Binomial distribution, with

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

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

(iv) Deduce that $\sup_{0\le\theta\le1} |p_n(\theta) - g(\theta)| < 2\epsilon$ 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***.

*[3] (Hoeffding's inequality)

(i) Suppose X is a random variable with $\mathbb{P}X = 0$ and $\mathbb{P}\{a \le X \le b\} = 1$, where a and b are constants. Prove that

$$\operatorname{var}(X) = \min\nolimits_{t\in\mathbb{R}} \mathbb{P}|X - t|^2 \le (b - a)^2/4.$$

(ii) With X as in (i), use HW2.3 to show that the function $L(\theta) := \log \mathbb{P}e^{\theta X}$ has $L''(\theta) \le (b - a)^2/4$.

(iii) Suppose $X_1, \dots, X_n$ are independent random variables for which $\mathbb{P}X_i = 0$ and $\mathbb{P}\{a_i \le X_i \le b_i\} = 1$ for constants a_i and b_i. Prove that

$$\mathbb{P}\exp(\theta(X_1 + \dots + X_n)) \le \exp\left(\theta^2 \sum\nolimits_i (b_i - a_i)^2/8\right).$$

(iv) Prove that

$$\mathbb{P}\{\sum\nolimits_i X_i \ge t\} \le \inf_{\theta>0} \mathbb{P}\exp\left(-\theta t + \theta \sum\nolimits_i X_i\right) \le \exp\left(-2t^2/\sum\nolimits_i (b_i - a_i)^2\right).$$

for each $t \ge 0$.