STOR 654 Homework 10

1. *Extreme value theory for the Gaussian.* Let a_n and b_n be the extreme value scaling and centering constants for the maximum M_n of n independent standard Gaussian random variables.

(a) Fix $x \in \mathbb{R}$ and let $x_n = x/a_n + b_n$. Show that $n\,\phi(x_n)/x_n \to e^{-x}$ as n tends to infinity. [In your calculations, identify and pay careful attention to the leading order terms.]

(b) Using the result of part (a) and the standard Gaussian tail bound from an earlier homework, show that $n(1 - \Phi(x_n)) \to e^{-x}$.

(c) Use part (b) and the lemma from lecture to show that as n tends to infinity

$$\mathbb{P}(a_n(M_n - b_n) \leq x) \to G(x) = e^{-e^{-x}}$$

(d) Show that $G(x)$ is the CDF of $-\log V$ where $V \sim \text{Exp}(1)$.

2. Let M_n be the maximum of n iid $\mathcal{N}(0,1)$ random variables. Use the Gaussian extreme value theorem to establish the following limiting results.

a. $\mathbb{P}(M_n \geq \sqrt{2 \log n}) \to 0$ as $n \to \infty$

b. $M_n/\sqrt{2 \log n} \to 1$ in probability as $n \to \infty$

3. Let $V \subseteq \mathbb{R}^n$ be a finite set of vectors $v = (v_1, \ldots, v_n)^t$ with $L = \max_{v \in V} ||v||_2$, and let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher (sign) variables.

(a) Use Hoeffding's MGF inequality to bound the moment generating functions of the random variables $\sum_{i=1}^n \varepsilon_i v_i$ in terms of the constant L.

(b) Show that

$$\mathbb{E}\left[\max_{v \in V} \sum_{i=1}^n \varepsilon_i v_i\right] \;\leq\; \sqrt{2L^2 \log |V|}$$

4. Let $U_{(r)}$ be the r'th order statistics of $U_1, \ldots, U_n$ iid $\sim$ U$(0,1)$. Show that $\mathbb{E}\, U_{(r)} = r/(n+1)$. (A formula for the density of $U_{(r)}$ was given in an earlier lecture.)

5. Define the function $g_s(z) = \exp(sz - s^2/2)$, where s is a constant, and let $Z \sim \mathcal{N}(0,1)$. Show that for $t \in \mathbb{R}$

a. $\mathbb{E} g_s(Z)\,\mathbb{I}(Z \leq t) = \Phi(\epsilon t)$

b. $\mathbb{E} g_s^2(Z)\,\mathbb{I}(Z \leq t) = e^{s^2}\Phi(t - 2s)$

c. $\Phi(t-2s) \leq \phi(2s - t)$ when $t - 2s \leq -1$

Fix $\epsilon \in (0, 1/2)$. For $n \geq 1$ let $t_n = \sqrt{2\log n}$, $s_n = (1-\epsilon)t_n$, and define

$$\tilde{L}_n = n^{-1}\sum_{i=1}^{n} g_{s_n}(Z_i)\,\mathbb{I}(Z_i \leq t_n)$$

where $Z_1, \ldots, Z_n$ iid $\sim \mathcal{N}(0,1)$. Using (a)-(c) above as needed, establish the following:

d. $\mathbb{E}\tilde{L}_n = \Phi(\epsilon t) \to 1$ as $n \to \infty$

e. $\mathrm{Var}(\tilde{L}_n) \leq (2\pi)^{-1} e^{-\epsilon^2 t^2} \to 0$ as $n \to \infty$

f. $\tilde{L}_n \to 1$ in probability as $n \to \infty$

6. Let $x < y$ be real numbers. Carefully justify each of the following (in)equalities.

$$\frac{e^y - e^x}{y - x} \;=\; \int_0^1 e^{tx + (1-t)y}\,dt \;\leq\; \int_0^1 (te^x + (1-t)e^y)\,dt \;=\; \frac{1}{2}(e^x + e^y)$$

7. Let X and Y be random variables, possibly defined on different probability spaces, with CDFs F and G, respectively. We say that X is stochastically larger than Y, written $X \stackrel{\mathrm{d}}{\geq} Y$ if $F(x) \leq G(x)$ for each $x \in \mathbb{R}$. Explain the intuition behind the definition. Recall that if X is a random variable with continuous CDF F then $F(X) \stackrel{\mathrm{d}}{=} U(0,1)$. In general, even if F is not continuous, it is still true that $F(X) \stackrel{\mathrm{d}}{\geq} U(0,1)$. Establish this fact. Hint: Let φ be the percentile function of F. Argue that $F(\varphi(u)) \geq u$ for $0 < u < 1$, and recall that $X \stackrel{\mathrm{d}}{=} \varphi(U)$ where $U \sim U(0,1)$.

8. *Independent Copies.* Let X, X' be independent random variables with the same distribution. In this case we say that X' is an independent copy of X.

(a) Show that $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}(X - X')^2$

(b) Argue formally or informally that $\mathbb{E}(X' \,|\, X) = \mathbb{E}X$

(c) Using the result of part (b) and Jensen's inequality for conditional expectations, show that $\mathbb{E}|X - \mathbb{E}X| \leq \mathbb{E}|X - X'|$. This is a key step in establishing a number of important bounds in empirical process theory.

9. (Using symmetry) Let $X_1, \cdots, X_n$ be iid positive random variables such that $\mathbb{E}X_k = a$ and $\mathbb{E}X_k^{-1} = b$ are finite. For $1 \leq k \leq n$, let $S_k = X_1 + \cdots + X_k$ and let $V_k = X_k/S_n$.

(a) Show that $\mathbb{E}S_m^{-1} < \infty$ for $m = 1, \cdots, n$.

(b) Argue informally that $V_1, \cdots, V_n$ have the same distribution.

(c) Conclude from part (b) that $\mathbb{E}(S_m/S_n) = m/n$ for $1 \leq m \leq n$.

(d) Show that $\mathbb{E}(S_n/S_m) = 1 + (n-m)\, a\, \mathbb{E}S_m^{-1}$ for $1 \leq m \leq n$.

(e) Verify the inequality $x + x^{-1} \geq 2$ for $x > 0$.

(f) Use the inequality of part (e) to show that $\mathbb{E}(S_n/S_m) \geq n/m$ for $1 \leq m \leq n$.

10. Let $U \sim \mathrm{U}(0,1)$. Show that $\max(U, 1-U) \overset{\mathrm{d}}{=} 1/2 + U/2$.