MATH 507a Final Exam (Takehome Section)

Qilin Ye

December 7, 2022

Problem 1

Let $\{U_i, i \geqslant 1\}$ be i.i.d. uniform in $[0, c]$ and define

$$V_n = \sum_{k=1}^{n} \prod_{i=1}^{k} U_i, \qquad W_n = \sum_{k=1}^{n} \prod_{i=k}^{n} U_i.$$

(1) Show that V_n, W_n have the same distribution for each n.

(2) Show that $\{V_n\}$ has a limit; show that the limit is finite a.s. if $c < e$.

(3) Let Y be a random variable with a density. Show that for every $\epsilon > 0$ there exists $\delta > 0$ such that $\mathbb{P}(|X - Y| < \delta) < \epsilon$ for all X independent of Y.

(4) Show that when $c < 2$, $\mathbb{E}(W_n)$ is bounded in n, but with probability 1, $\{W_n\}$ does not have a limit.

Proof. (1) $V_n = U_1 + U_1U_2 + \ldots + U_1 \ldots U_n$. Since the U_i's are i.i.d., $U_1 \stackrel{d}{=} U_n$, and similarly $\prod_{i=1}^{k} U_i \stackrel{d}{=} \prod_{i=n-k}^{n} U_i$. The claim then follows as V_n, W_n are finite sums of these forms.

(2) Note that $V_{n+1} - V_n = \prod_{i=1}^{n+1} U_i \geqslant 0$, so $\{V_n\}$ is monotone. Thus it has a limit, possibly infinite.

If $c < e$, $k = \mathbb{E} \log U_i = \int_0^c \log t/c \, \mathrm{d}t = c(\log c - 1) < 0$, and so given $\epsilon > 0$, for sufficiently large n, using log transform in conjunction with SLLN yields (as in HW4)

$$\prod_{i=1}^{n} U_i \leqslant e^{(k+\epsilon)n} \leqslant e^{(k/2)n}.$$

Since $\sum_{n \geqslant 1} e^{(k/2)n}$ converges for $k < 0$, the tail of $\{V_n\}$ converges, and we are done.

(3) Let f be the density of Y. Let $\epsilon > 0$ be given. By the hint there exists $\delta > 0$ such that $m(A) < \delta$ (Lebesgue measure on any Borel A) implies $\int_A f \, \mathrm{d}m < \epsilon$. But then

$$\begin{aligned} \mathbb{P}(|X - Y| < \delta/2) &= \int_{\Omega_Y} \int_{\Omega_X} 1_{|X-Y|<\delta/2} f \, \mathrm{d}\nu \, \mathrm{d}\mu \\ &= \int_{\Omega_Y} \int_{(\omega-\delta/2, \omega+\delta/2)} f \, \mathrm{d}\nu \, \mathrm{d}\mu < \int_{\Omega_Y} \epsilon \, \mathrm{d}\mu = \epsilon. \end{aligned}$$

(4) If $c < 2$ then $\mathbb{E}U_1 = c/2 < 1$. then

$$\mathbb{E}(W_n) = \sum_{k=1}^{n} \mathbb{E} \prod_{i=k}^{n} U_i = \sum_{k=1}^{n} (c/2)^{n-k} = \sum_{k=1}^{n} (c/2)^k.$$

As $n \to \infty$ we obtain $\sum_{k=1}^{\infty}(c/2)^k = (c/2)/(1-c/2)$, so indeed $\mathbb{E}W_n$ is bounded in n.

By definition $W_{n+1} = U_{n+1}(W_n + 1)$ so $W_{n+1} - W_n = W_n(U_{n+1} - 1) + U_{n+1}$.

[*Given the hint I suspect the proof has something to do with applying (c) to $W_{n+1} - W_n$, but in doing that I need to show that this difference is independent in n, which I have been unable to. I really don't think this is true, but I'll assume it anyways.*] Let $X_n = W_{n+1} - W_n$. Clearly it has a density, so by the previous part, for all $\epsilon > 0$ and all n there exists $\delta = \delta(n, \epsilon)$ with

$$\mathbb{P}(|W_n - W_m| < \delta) < \epsilon \qquad \text{for all } m.$$

Since

$$\{\omega : \{W_n\} \text{ is Cauchy}\} \subset \{\omega : \text{ for all } \delta, |W_n(\omega) - W_m(\omega)| < \delta \text{ for all } n, m > \text{some } N(\delta)\}$$

and later events are contained in the earlier event $\{|W_{N(\delta)} - W_{N(\delta)+1}| < \delta\}$, we therefore have

$$\mathbb{P}(\{W_n\} \text{ is Cauchy}) \leqslant \mathbb{P}(|W_{N(\delta)} - W_{N(\delta)+1}| < \delta) < \epsilon.$$

□

Problem 2

Let $X_1, X_2, \ldots$ be i.i.d. symmetric random varaibles with

$$\mathbb{P}(X_1 > x) = \mathbb{P}(X_1 < x) = \frac{1}{2x^2} \text{ for } x \geqslant 1 \qquad \text{and} \qquad \mathbb{P}(X_1 \in (-1, 1)) = 0.$$

We want to show that X_1 has infinite variance but still satisfies a CLT in the form

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{n \log n}} \implies \mathcal{N}(0,1). \qquad (*)$$

(1) Show that for a general random variable Y and $c > 0$

$$\mathbb{E}(Y^2 1_{\{|Y| \leqslant c\}}) = \int_0^c 2x[\mathbb{P}(|Y| \geqslant x) - \mathbb{P}(|Y| > c)\,\mathrm{d}x.$$

In particular, for X_1,

$$\mathbb{E}(X_1^2 1_{\{|X_1| \leqslant c\}}) = 2 \log c \qquad \text{for} \qquad c \geqslant 1.$$

(2) Let $\{c_n\}$ satisfy $c_n \to \infty$, $c_n/\sqrt{n \log n} \to 0$, and define $Y_{n,i} = X_i 1_{\{|X_i| \leqslant c_n\}}$ for $i \leqslant n$, and let $T_n = \sum_{i=1}^n Y_{n,i}$. Show that $\{c_n\}$ can be chosen so that

$$\frac{T_n - \mathbb{E}T_n}{\sqrt{n \log n}} \implies \mathcal{N}(0,1).$$

(3) Prove (*) above.

Proof. (1) It is well-known that $\mathbb{E}Y^2 = \int_0^\infty 2t\mathbb{P}(Y \geqslant t)\,\mathrm{d}t$. Therefore,

$$\begin{aligned}\mathbb{E}(Y^2 1_{\{|Y|\leqslant c\}}) &= \int_0^\infty 2x\mathbb{P}(|Y|1_{\{|Y|\leqslant c\}} \geqslant x)\,\mathrm{d}x \\ &= \int_0^c 2x\mathbb{P}(|Y| \in [x,c])\,\mathrm{d}x \\ &= \int_0^c 2x[\mathbb{P}(|Y| \geqslant x) - \mathbb{P}(|Y| > c)]\,\mathrm{d}x.\end{aligned}$$

In particular, for X_1 and $c \geqslant 1$,

$$\begin{aligned}\mathbb{E}(X_1^2 1_{\{|X_1|\leqslant c\}}) &= \int_0^1 2x[\mathbb{P}(|X| \geqslant x) - \mathbb{P}(|X| > c)]\,\mathrm{d}x + \int_1^c 2x[\mathbb{P}(|X| \geqslant x) - \mathbb{P}(|X| > c)]\,\mathrm{d}x \\ &= \int_0^1 2x(1 - 1/c^2)\,\mathrm{d}x + \int_1^c 2x(1/x^2 - 1/c^2)\,\mathrm{d}x \\ &= 1 - \frac{1}{c^2} - 1 + \frac{1}{c^2} + 2\log c.\end{aligned}$$

(2) We first note that $Y_{n,i}$ is symmetric, so T_n also is and $\mathbb{E}T_n = 0$. Since

$$\sum_{i=1}^n \mathbb{E}(Y_{n,i}/\sqrt{n\log n})^2 = \frac{1}{\log n}\mathbb{E}(X_1^2 1_{\{|X_1|\leqslant c_n\}}) = \frac{2\log c_n}{\log n},$$

to satisfy the requirements of Lindeberg-Feller, we require $\log(c_n^2)/\log(n) \to 1$. One example is by setting $c_n = \sqrt{n}$.

Before invoking Lindeberg-Feller we need to check the second condition; for all $\epsilon > 0$ we want to show

$$\sum_{i=1}^n \mathbb{E}((Y_{n,i}/\sqrt{n\log n})^2 : |Y_{n,i}/\sqrt{n\log n}| > \epsilon) \to 0.$$

But this is obvious, once we plug in the definition of $Y_{n,i}$:

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}((Y_{n,i}/\sqrt{n\log n})^2 : |Y_{n,i}/\sqrt{n\log n}| > \epsilon) &= \frac{1}{\log n}\mathbb{E}(Y_{n,1}^2 : |Y_{n,1}| > \epsilon\sqrt{n\log n}) \\ &= \frac{1}{\log n}\mathbb{E}(X_1^2 : |X_1| \leqslant c_n \text{ and } |X_1| > \epsilon\sqrt{n\log n}).\end{aligned}$$

By assumption $c_n/\sqrt{n\log n} \to 0$, so for sufficiently large n, the expression above vanishes. Therefore Lindeberg-Feller gives the desired limiting distribution.

(3) Since $\mathbb{E}S_n = 0$ as well it suffices to ensure $\mathbb{P}(S_n \neq T_n) \to 0$, so that the limits of $S_n/\sqrt{n\log n}$ and $T_n/\sqrt{n\log n}$ agree. That is, we want

$$\mathbb{P}(S_n \neq T_n) \leqslant \sum_{i=1}^n \mathbb{P}(X_i \neq Y_{n,i}) = n\mathbb{P}(X_1 \geqslant c_n) = \frac{n}{2c_n^2} \to 0.$$

One such example is $c_n = n^{1/2}\log\log n$: clearly $c_n \to \infty$, $c_n/\sqrt{n\log n} = \log\log n/(\log n)^{1/2} \to 0$, $n/(2c_n^2) = 1/(\log\log n)^2 \to 0$, and all claims in (2) still hold. □

Problem 3

(1) Suppose $(\Omega, \mathcal{F}_0, \mathbb{P})$ is a probability space, $\mathcal{F} \subset \mathcal{F}_0$, and X_1, X_2, Y_1, Y_2 are positive r.v.'s satisfying

$$\mathbb{E}(\log X_2 \mid \mathcal{F}) \geqslant \log X_1 \qquad \text{and} \qquad \mathbb{E}(\log Y_2 \mid \mathcal{F}) \geqslant \log Y_1.$$

Show that

$$\mathbb{E}(X_2 Y_2 \mid \mathcal{F}) \geqslant X_1 Y_1.$$

(2) Let $(\Omega, \mathcal{F}_0, \mathbb{P})$ be a probability space and let $\mathcal{F}, \mathcal{G}$ be independent sub-σ-fields. Let X be a r.v. with $\mathbb{E}|X| < \infty$. What is

$$\mathbb{E}\left(\mathbb{E}(X \mid \mathcal{F}) \mid \mathcal{G}\right)?$$

Proof. (1) By linearity

$$\mathbb{E}(\log X_2 Y_2 \mid \mathcal{F}) = \mathbb{E}(\log X_1 + \log X_2 \mid \mathcal{F}) = \mathbb{E}(\log X_2 \mid \mathcal{F}) + \mathbb{E}(\log Y_2 \mid \mathcal{F}) \geqslant \log X_1 + \log Y_1 = \log(X_1 Y_1).$$

Then Jensen's inequality implies

$$\mathbb{E}(X_2 Y_2 \mid \mathcal{F}) = \exp(\mathbb{E} \log X_2 Y_2 \mid \mathcal{F}) = X_1 Y_1.$$

(2) For $G \in \mathcal{G}$,

$$\int_G \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}) \mid \mathcal{G})\, \mathrm{d}\mathbb{P} = \int_G \mathbb{E}(X \mid \mathcal{F})\, \mathrm{d}\mathbb{P} = \int_\Omega \mathbb{E}(X \mid \mathcal{F}) 1_G\, \mathrm{d}\mathbb{P}$$
$$[\text{by independence}] = \mathbb{P}(G) \int_\Omega \mathbb{E}(X \mid \mathcal{F})\, \mathrm{d}\mathbb{P} = \mathbb{P}(G)\mathbb{E}X,$$

so $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}) \mid \mathcal{G}) = \mathbb{E}X$.

□