523 Midterm

Daniel Rui - 5/8/20

Problem 1

Let X be an r.v. and $\phi(t) = \mathbb{E}\big[e^{itX}\big] = \mathbb{E}[\cos(tX) + i\sin(tX)]$ be its ch.f., and so $\operatorname{Re}\phi(t) = \frac{\phi(t)+\phi(-t)}{2}$. Thus to create some r.v. that has ch.f. of $\operatorname{Re}\phi(t)$, we need to somehow weight X equally between "+1" and "−1" — consider a Rademacher r.v. R independent of X, and the ch.f. of the r.v. RX:

$$\mathbb{E}\big[e^{itRX}\big] = \mathbb{E}\big[e^{it\cdot 1\cdot X}\cdot 1_{[R=1]}\big] + \mathbb{E}\Big[e^{it\cdot(-1)\cdot X}\cdot 1_{[R=-1]}\Big] = \frac{\mathbb{E}\big[e^{itX}\big]}{2} + \frac{\mathbb{E}\big[e^{-itX}\big]}{2} = \frac{\phi(t)+\phi(-t)}{2}$$

That is to say, $\operatorname{Re}\phi(t)$ is a ch.f (of the r.v. RX).

Problem 2

Let X be an r.v. with mean 0 and variance $\sigma^2 < \infty$, and consider the r.v. X^* that satisfies the identity $\sigma^2\mathbb{E}[f'(X^*)] = \mathbb{E}[Xf(X)]$ for all absolutely continuous f for which $\mathbb{E}[f'(X^*)]$ and $\mathbb{E}[Xf(X)]$ make sense. The distribution of X^*, called the X-zero bias distribution, is uniquely determined by the identity.

(a) We want to show that $(aX)^* =_d aX^*$, for any $a \neq 0$. By definition, $(aX)^*$ is an r.v. that satisfies $a^2\sigma^2\mathbb{E}[f'((aX)^*)] = \mathbb{E}[aXf(aX)]$ for all appropriate f (described above). On the other hand, if we consider $g(x) = f(ax)$ (for appropriate f), then by the definition of X^*, $\sigma^2\mathbb{E}[g'(X^*)] = \mathbb{E}[Xg(X)] \iff \sigma^2\mathbb{E}[af'(aX^*)] = \mathbb{E}[Xf(aX)] \iff a^2\sigma^2\mathbb{E}[f'(aX^*)] = \mathbb{E}[aXf(aX)]$ and so putting both equalities together we get that $\mathbb{E}[f'((aX)^*)] = \mathbb{E}[f'(aX^*)]$. But we said that the distribution for r.v.'s that satisfied the identity was uniquely determined, so it must be that $aX^* =_d (aX^*)$.

(b) If we consider (the absolutely continuous function) $f_t(x) = (x-t)\cdot 1_{(-\infty,t]}(x)$ with derivative $f_t'(x) = 1_{(-\infty,t]}$, then we get that

$$F_{X^*}(t) = P(X^* \le t) = \mathbb{E}\big[1_{(-\infty,t]}(X^*)\big] = \mathbb{E}[f'(X^*)] = \frac{1}{\sigma^2}\mathbb{E}[Xf(X)] = \frac{\mathbb{E}\big[X(X-t)1_{[X\le t]}\big]}{\sigma^2}$$

In particular, this means that if $|X| \le C$ a.s. (i.e. $1_{[X\le -C-\epsilon]} =_{\text{a.s.}} 0$ and $1_{[X\le C]} =_{\text{a.s.}} 1$),

$$F_{X^*}(-C-\epsilon) = 0 \quad \text{and} \quad F_{X^*}(C) = \frac{\mathbb{E}\big[X^2 - tX\big]}{\sigma^2} = \frac{\sigma^2 - 0}{\sigma^2} = 1$$

(for all $\epsilon > 0$), and so $|X^*| \le C$ almost surely as well.

We can also consider X^s with the X-size bias distribution, where X is non-negative and X^s satisfies $\mathbb{E}[Xf(X)] = \mu\mathbb{E}[f(X^s)]$ (again for all f for which the expectations exist, and $\mu = \mathbb{E}[X]$).

(a) As in (a) above, $(aX)^s =_d aX^s$ for any $a \neq 0$, because $\mathbb{E}[aXf(aX)] = a\mu\mathbb{E}[f((aX)^s)] \iff \mathbb{E}[Xf(aX)] = \mu\mathbb{E}[f((aX)^s)]$ (by definition) and $\mathbb{E}[Xg(X)] = \mu\mathbb{E}[g(X^s)] \iff \mathbb{E}[Xf(aX)] =$

$\mu\mathbb{E}[f(aX^s)]$ where again $g(x) = f(ax)$ (for appropriate f) and so putting the equalities together we get $\mathbb{E}[f(aX^s)] = \mathbb{E}[f((aX)^s)]$. But we said that the distribution for r.v.'s that satisfied the identity was uniquely determined, so it must be that $aX^s =_d (aX^s)$.

(b) As in (b) above, we find the (induced measure of the) c.d.f. of X^s:

$$F_{X^s}(A) = \mathbb{E}[1_A(X^s)] = \frac{1}{\mu}\mathbb{E}[X1_A(X)] = \frac{1}{\mu}\int_{\mathbb{R}} x1_A(x)\,dF_X(x) = \int_A \frac{x}{\mu}\,dF_X(x)$$

Thus if X takes values in $[0, C]$ (a.s.), then taking $A = (C, \infty)$, the above equation makes it clear that $F_{X^s}(A) = 0$, and thus $0 \le X^s \le C$ a.s. as well.

Problem 3

(a) In the framework of Problem 2A, we already established that $F_{X^*}(t) = \frac{1}{\sigma^2}\mathbb{E}\big[X(X-t)1_{[X\le t]}\big]$. If we assume that X has a p.d.f. $f(x)$, then we can write $F_{X^*}(t) = \frac{1}{\sigma^2}\int_{(-\infty,t]} x(x-t)f(x)\,dx$. Differentiating, we get that $F'_{X^*}(t) = \frac{1}{\sigma^2}t(t-t)f(t) + \int_{-\infty}^{t} -xf(x)\,dx$ and $F''_{X^*}(t) = -tf(t)$, which is ≥ 0 for $t \le 0$ and ≤ 0 for $t \ge 0$ (because densities are ≥ 0). Thus, $F_{X^*}(t)$ is convex on $(-\infty, 0]$ and concave on $[0, \infty)$, which means that F_{X^*} is unimodal with 0-mode according to Khinchine's definition.

(b) More generally, in the case where X may not have a p.d.f., we have for $t_1 < t_2$ that

$$\frac{F_{X^*}(t_1) + F_{X^*}(t_2)}{2} = \frac{1}{\sigma^2}\mathbb{E}\left[X\frac{X-t_1}{2}1_{[X\le t_1]} + X\frac{X-t_2}{2}1_{[X\le t_2]}\right]$$

which can be written as $\frac{1}{\sigma^2}\mathbb{E}[Xg(X)]$ if we take $g(x) = \frac{x-t_1}{2}\cdot 1_{(-\infty,t_1]}(x) + \frac{x-t_2}{2}\cdot 1_{(-\infty,t_2]}(x)$. Similarly, $F_{X^*}(\frac{t_1+t_2}{2}) = \frac{1}{\sigma^2}\mathbb{E}[Xh(X)]$ for $h(x) = (x - \frac{t_1+t_2}{2})1_{(-\infty,\frac{t_1+t_2}{2}]}(x)$. But

$$\begin{aligned}[h-g](x) &= \frac{x-t_1}{2}\cdot 1_{(-\infty,\frac{t_1+t_2}{2}]}(x) + \frac{x-t_2}{2}\cdot 1_{(-\infty,\frac{t_1+t_2}{2}]}(x)\\ &\quad - \frac{x-t_1}{2}\cdot 1_{(-\infty,t_1]}(x) - \frac{x-t_2}{2}\cdot 1_{(-\infty,t_2]}(x)\\ &= \frac{x-t_1}{2}\cdot 1_{(t_1,\frac{t_1+t_2}{2}]}(x) - \frac{x-t_2}{2}\cdot 1_{(\frac{t_1+t_2}{2},t_2]}(x)\end{aligned}$$

which is ≥ 0 (because $x \in (t_1, \frac{t_1+t_2}{2}] \implies x - t_1 > 0$ and $x \in (\frac{t_1+t_2}{2}, t_2] \implies x - t_2 \le 0$). Furthermore, $[h-g](x)$ is 0 outside of $[t_1, t_2]$. Thus, for $t_1 < t_2 \le 0$,

$$\frac{F_{X^*}(t_1) + F_{X^*}(t_2)}{2} - F_{X^*}\left(\frac{t_1+t_2}{2}\right) = \frac{1}{\sigma^2}\mathbb{E}[X[g-h](X)] \ge 0$$

because $[g-h](x)$ is ≤ 0, 0 when X is outside $[t_1, t_2]$, so we are only concerned with values of X in $[t_1, t_2]$, a negative interval. Similarly, for $0 \le t_1 < t_2$,

$$\frac{F_{X^*}(t_1) + F_{X^*}(t_2)}{2} - F_{X^*}\left(\frac{t_1+t_2}{2}\right) = \frac{1}{\sigma^2}\mathbb{E}[X[g-h](X)] \le 0$$

because X is positive (i.e. in the range $[t_1, t_2]$) while $[g-h]$ is negative. (One could go back and change the average to a weighted average with λ and $(1-\lambda)$, and the logic would follow much the same way). Thus, $F_{X^*}(t)$ is convex on $(-\infty, 0]$ and concave on $[0, \infty)$, which means that F_{X^*} is unimodal with 0-mode according to Khinchine's definition.

Problem 4

Let $K(t) = \mathbb{E}\big[X(1_{[0\le t\le X]} - 1_{[X\le t<0]})\big]$ where X is a 0-mean r.v.. We have that:

$$\begin{aligned}\int_{\mathbb{R}} K(t)\,dt &= \int_{\mathbb{R}}\int_{\Omega} X\cdot 1_{[0\le t\le X]}\,dP\,dt + \int_{\mathbb{R}}\int_{\Omega} -X\cdot 1_{[X\le t<0]}\,dP\,dt \\ &= \int_{\Omega}\int_{\mathbb{R}} X\cdot 1_{[0\le t\le X]}\,dt\,dP - \int_{\Omega}\int_{\mathbb{R}} X\cdot 1_{[X\le t<0]}\,dt\,dP \\ &= \int_{\Omega} X\cdot 1_{[X\ge 0]}\int_0^X 1\,dt\,dP - \int_{\Omega} X\cdot 1_{[X<0]}\int_X^0 1\,dt\,dP \\ &= \mathbb{E}\big[X^2 1_{[x\ge 0]}\big] - \mathbb{E}\big[-X^2 1_{[x<0]}\big] = \mathbb{E}\big[X^2\big]\end{aligned}$$

where the integral interchange is justified by Fubini-Tonelli because $X\cdot 1_{[0\le t\le X]}$ and $-X\cdot 1_{[X\le t<0]}$ are both non-negative functions. Similarly (again using non-negativity for integral interchange),

$$\begin{aligned}\int_{\mathbb{R}} |t|K(t)\,dt &= \int_{\mathbb{R}}\int_{\Omega} |t|X\cdot 1_{[0\le t\le X]}\,dP\,dt + \int_{\mathbb{R}}\int_{\Omega} -|t|X\cdot 1_{[X\le t<0]}\,dP\,dt \\ &= \int_{\Omega} X\int_{\mathbb{R}} |t|1_{[0\le t\le X]}\,dt\,dP - \int_{\Omega} X\int_{\mathbb{R}} |t|1_{[X\le t<0]}\,dt\,dP \\ &= \int_{\Omega} X\cdot 1_{[X\ge 0]}\int_0^X |t|\,dt\,dP - \int_{\Omega} X\cdot 1_{[X<0]}\int_X^0 |t|\,dt\,dP \\ &= \int_{\Omega} X\cdot 1_{[X\ge 0]}\cdot(\tfrac{1}{2}t^2)\Big|_0^X\,dP - \int_{\Omega} X\cdot 1_{[X<0]}(-\tfrac{1}{2}t^2)\Big|_X^0\,dt\,dP \\ &= \tfrac{1}{2}\mathbb{E}\big[X^3 1_{[X\ge 0]}\big] - \mathbb{E}\big[\tfrac{1}{2}X^3 1_{[X<0]}\big] = \tfrac{1}{2}(\mathbb{E}\big[(X^3)^+\big] + \mathbb{E}\big[(X^3)^-\big]) = \tfrac{1}{2}\mathbb{E}\big[|X|^3\big]\end{aligned}$$

Problem 5

Suppose that $g, h : \mathbb{R}\to\mathbb{R}$ are $\nearrow$, and that X is an r.v such that $\mathbb{E}\big[g^2(X)\big], \mathbb{E}\big[h^2(X)\big] < \infty$. We would like to prove that $\text{Cov}[g(X), h(X)] \ge 0$. As the hint so generously suggests, we consider Y, an independent copy of X so that:

$$\begin{aligned}\mathbb{E}\big[\big(g(Y)-g(X)\big)\big(h(Y)-h(X)\big)\big] &= \mathbb{E}[g(Y)h(Y)] - \mathbb{E}[g(Y)h(X)] - \mathbb{E}[g(X)h(Y)] + \mathbb{E}[g(X)h(X)] \\ &= 2\mathbb{E}[g(X)h(X)] - 2\mathbb{E}[g(X)]\mathbb{E}[h(X)] = 2\text{Cov}[g(X), h(X)]\end{aligned}$$

where we used that $\mathbb{E}[g(X)h(X)] = \mathbb{E}[g(Y)h(Y)]$, $\mathbb{E}[g(X)] = \mathbb{E}[g(Y)]$, and $\mathbb{E}[h(X)] = \mathbb{E}[h(Y)]$ (because $X =_d Y$), and that $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$ by independence of X, Y. Because g, h are both non-decreasing, at any $\omega\in\Omega$, both differences ($g(Y)-g(X)$ and $h(Y)-h(X)$) are either both ≤ 0 if $Y(\omega)\le X(\omega)$ or both ≥ 0 if $Y(\omega)\ge X(\omega)$. Thus, the product of the differences is always ≥ 0, so the expectation is always ≥ 0, meaning of course that $\text{Cov}[g(X), h(X)] \ge 0$

Problem 6

The law of the iterated logarithm states that for $\mathbb{S}(t)$ a standard Brownian motion on $[0,\infty)$,

$$\limsup_{t\to\infty} \frac{\mathbb{S}(t)}{\sqrt{2t\log\log t}} =_{\text{a.s.}} 1$$

(a) We want to show that $\tilde{\mathbb{S}}(t) = t\mathbb{S}(\frac{1}{t})$ is a Brownian motion on $[0,\infty)$. We first verify the expectation and covariance formulas: for any $s,t\in(0,\infty)$,

- $\mathbb{E}\Big[\tilde{\mathbb{S}}(t)\Big] = \mathbb{E}\big[t\mathbb{S}(\frac{1}{t})\big] = t\mathbb{E}\big[\mathbb{S}(\frac{1}{t})\big] = 0$
- $\mathbb{E}\Big[\tilde{\mathbb{S}}(s)\tilde{\mathbb{S}}(t)\Big] = \mathbb{E}\big[st\mathbb{S}(\frac{1}{s})\mathbb{S}(\frac{1}{t})\big] = st\min\{\frac{1}{s},\frac{1}{t}\} = \min\{st\frac{1}{s}, st\frac{1}{t}\} = \min\{t,s\}$

Now from the LIL applied to $-\mathbb{S}(t)$ which is also a Brownian motion (pretty obviously), the lim inf is -1, and so multiplying everything by $\frac{\sqrt{2t\log\log t}}{t}$ (which goes to 0 as $t\to\infty$) yields that both the lim sup and lim inf of $\frac{\mathbb{S}(t)}{t}$ as $t\to\infty$ is a.s. 0. This is the same limit as $\lim_{t\to 0}\tilde{\mathbb{S}}(t)$, and so we see that setting $\tilde{\mathbb{S}}(0)=0$ makes it right continuous a.s. at 0. $\tilde{\mathbb{S}}(t)$ is continuous (a.s.) on $(0,\infty)$ because $\mathbb{S}(t)$ is. Lastly, $\tilde{\mathbb{S}}(t)$ is clearly Gaussian because $\mathbb{S}(t)$ is, and so $\tilde{\mathbb{S}}(t)$ is indeed a Brownian motion on $[0,\infty)$.

(b) Well, taking $\tilde{\mathbb{S}}$ to be the Brownian motion in consideration in the LIL, we have that

$$1 =_{\text{a.s.}} \limsup_{t\to\infty}\frac{t\mathbb{S}(\frac{1}{t})}{\sqrt{2t\log\log t}} = \limsup_{t\to\infty}\frac{\mathbb{S}(\frac{1}{t})}{\sqrt{2\frac{1}{t}\log\log t}} = \limsup_{t'\to 0}\frac{\mathbb{S}(t')}{\sqrt{2t'\log\log(\frac{1}{t'})}}$$

Problem 8

As we discovered in 523hw4p3b, or as given to us in the problem statement, we can represent the sum of the numbers drawn in sampling n balls without replacement from a collection of N balls labeled $c_1,\ldots,c_N$ as

$$Y_N = \sum_{i=1}^{N} b_i c_{\pi_i}$$

where $b_i = 1_{[1,n]}(i)$. From the notation of 523hw4p2 or 523hw4p3a, we have that $\bar{b}_N = \frac{n}{N}$, $B_N^2 = \sum_{i=1}^N (b_i-\bar{b}_N)^2 = n(1-\frac{n}{N})^2 + (N-n)\frac{n^2}{N^2} = n(\frac{N-n}{N})^2 + (N-n)\frac{n^2}{N^2} = n\frac{N-n}{N}(\frac{N-n}{N}+\frac{n}{N}) = n(1-\frac{n}{N})$, and

$$\mathbb{E}[Y_N] = N\bar{b}_N\bar{c}_N = n\bar{c}_N \quad\text{and}\quad \mathrm{Var}[Y_N] = \frac{B_N^2C_N^2}{N-1} = \frac{n}{N}C_N^2\left(1-\frac{n-1}{N-1}\right)$$

A theorem of Hájek, 1961 says that supposing

$$\max_{1\le i\le N}\frac{|b_i-\bar{b}_N|}{B_N}\to 0 \quad\text{and}\quad \max_{1\le i\le N}\frac{|c_i-\bar{c}_N|}{C_N}\to 0,$$

then $\frac{Y_N - \mathbb{E}[Y_N]}{\mathrm{Var}[Y_n]} \to_d \mathrm{Normal}(0,1)$ if and only if

$$\sum_{(i,j): \sqrt{N}|b_i - \bar{b}_N| \cdot |c_j - \bar{c}_N| > \epsilon B_N C_N} \frac{|b_i - \bar{b}_N|^2 |c_j - \bar{c}_N|^2}{B_N^2 C_N^2} \to 0$$

as $N \to \infty$, for every $\epsilon > 0$ (where i, j in the sums are in $\{1, \ldots, N\}$). Substituting what we know about $\{b_i\}$ we see that

$$\max_{1 \le i \le N} \frac{|b_i - \bar{b}_N|}{B_N} = \frac{\max\{1 - \frac{n}{N}, \frac{n}{N}\}}{\sqrt{n(1 - \frac{n}{N})}} = \max\left\{ \frac{\sqrt{1 - \frac{n}{N}}}{\sqrt{n}}, \frac{\sqrt{n}}{\sqrt{N(1 - \frac{n}{N})}} \right\} = \max\left\{ \sqrt{\frac{1}{n} - \frac{1}{N}}, \frac{\sqrt{n}}{\sqrt{N - n}} \right\}$$

which goes to zero as $N \to \infty$ if $n := n_N \to \infty$, but growing much slower than N (i.e $n_N < \epsilon N$ eventually for every $\epsilon > 0$, or $n_N = o(N)$). Thus, the theorem says that if $n_N = o(N) \to \infty$, and

$$\max_{1 \le i \le N} \frac{|c_i - \bar{c}_N|}{C_N} \to 0$$

then $\frac{Y_N - \mathbb{E}[Y_N]}{\mathrm{Var}[Y_n]} \to_d \mathrm{Normal}(0,1)$ if and only if

$$\begin{aligned}
&\sum_{i=1}^{n} \sum_{j: \sqrt{\frac{N}{n}(1 - \frac{n}{N})}|c_j - \bar{c}_N| > \epsilon C_N} \frac{(1 - \frac{n}{N})|c_j - \bar{c}_N|^2}{n C_N^2} + \sum_{i=n+1}^{N} \sum_{j: \sqrt{\frac{n}{N-n}}|c_j - \bar{c}_N| > \epsilon C_N} \frac{\frac{n}{N}|c_j - \bar{c}_N|^2}{(N-n) C_N^2} \\
&= \sum_{j: \sqrt{\frac{N}{n}(1 - \frac{n}{N})}|c_j - \bar{c}_N| > \epsilon C_N} \frac{(N-n)|c_j - \bar{c}_N|^2}{N C_N^2} + \sum_{j: \sqrt{\frac{n}{N-n}}|c_j - \bar{c}_N| > \epsilon C_N} \frac{n|c_j - \bar{c}_N|^2}{N C_N^2} \to 0
\end{aligned}$$

as $N \to \infty$, for every $\epsilon > 0$ (where the j's in the sums are in $\{1, \ldots, N\}$).

Additional observations: It seems that the second term will have no trouble going to 0, given that $\frac{n}{N} \to 0$ and because $\sqrt{\frac{N-n}{n}} \to \infty$ so there will not be many (or any?) j s.t. $\frac{|c_j - \bar{c}_N|}{C_N} > \epsilon \sqrt{\frac{N-n}{n}}$ for large enough N. Thus, if one wants to use this theorem, one should really only be concerned with the first term.

523 Homework 4

Daniel Rui - 4/29/20

Problem 1

We want to show that the following are equivalent:

(i) $\max_{1\le k\le n} P(|X_{nk}|>\epsilon)\to 0$ for all $\epsilon>0$ (uniform asymptotic negligible, or u.a.n. for short)

(ii) $\max_{1\le k\le n} |\phi_{X_{nk}}(t)-1|\to 0$ uniformly on every finite interval of t

(iii) $\max_{1\le k\le n} \mathbb{E}\big[X_{nk}^2\wedge 1\big]\to 0$ (where recall $a\wedge b=\min\{a,b\}$)

Proof: (i) $\implies$ (iii): For any (small) $\epsilon>0$, $\mathbb{E}\big[X_{nk}^2\wedge 1\big]$ can be decomposed as

$$\begin{aligned}\mathbb{E}\big[X_{nk}^2\wedge 1\big]&=\mathbb{E}\big[(X_{n,k}^2\wedge 1)1_{[|X_{nk}|\le\epsilon]}+(X_{nk}^2\wedge 1)1_{[|X_{nk}|>\epsilon]}\big]\\&\le\mathbb{E}\big[\epsilon^2\cdot 1_{[|X_{nk}\le\epsilon]}\big]+\mathbb{E}\big[1\cdot 1_{[|X_{nk}|>\epsilon]}\big]\le\epsilon^2+P(|X_{nk}|>\epsilon)\end{aligned}$$

We can place $\max_{1\le k\le n}$ on both sides, and because $\epsilon>0$ can be chosen arbitrarily small, along with the fact that $\max_{1\le k\le n}P(|X_{nk}|>\epsilon)\to 0$ (from (i)), the LHS must go to 0.
(iii) $\implies$ (i): For any $\epsilon>0$ (and less than 1), observe that on the set $[|X_{nk}|>\epsilon]$, $\frac{X_{nk}^2\wedge 1}{\epsilon^2}\ge 1$ and so

$$P(|X_{nk}|>\epsilon)=\mathbb{E}\big[1_{[|X_{nk}|>\epsilon]}\big]\le\mathbb{E}\Big[\tfrac{X_{nk}^2\wedge 1}{\epsilon^2}1_{[|X_{nk}|>\epsilon]}\Big]\le\frac{1}{\epsilon^2}\mathbb{E}\big[X_{nk}^2\wedge 1\big]$$

We can place $\max_{1\le k\le n}$ on both sides, and fixing any $\epsilon\in(0,1)$, (iii) tells us the RHS goes to 0, so the LHS must as well. Finally for any $\epsilon\ge 1$, just note that $P(|X_{nk}|>\epsilon)\le P(|X_{nk}|>\frac{1}{2})$ which we already know goes to 0 (after we place maximums on both sides).
(i) $\implies$ (ii): Observe that for any $\epsilon>0$,

$$\begin{aligned}|\phi_{X_{nk}}(t)-1|=\big|\mathbb{E}\big[e^{itX_{nk}}-1\big]\big|&=\left|\int_{\mathbb{R}}(e^{itx}-1)\,dF_{X_{nk}}(x)\right|\\&\le\int_{[|x|\le\epsilon]}|e^{itx}-1|\,dF_{X_{nk}}(x)+\int_{[|x|>\epsilon]}|e^{itx}-1|\,dF_{X_{nk}}(x)\\&\le\int_{[|x|\le\epsilon]}|\cos(tx)-1+i\sin(tx)|\,dF_{X_{nk}}(x)+\int_{[|x|>\epsilon]}2\,dF_{X_{nk}}(x)\\&=\int_{[|x|\le\epsilon]}\sqrt{1-2\cos(tx)+1}\,dF_{X_{nk}}(x)+2P(|X_{nk}|>\epsilon)\\&\le\int_{[|x|\le\epsilon]}\sqrt{2\cdot\frac{(tx)^2}{2}}\,dF_{X_{nk}}(x)+2P(|X_{nk}|>\epsilon)\\&\le|t|\int_{[|x|\le\epsilon]}|x|\,dF_{X_{nk}}(x)+2P(|X_{nk}|>\epsilon)\\&\le|t|\epsilon P(|X_{nk}|\le\epsilon)+2P(|X_{nk}|>\epsilon)\le\epsilon|t|+2P(|X_{nk}|>\epsilon)\end{aligned}$$

where we used that $1-\cos x\le\frac{x^2}{2}$ (as can be seen via Taylor series). Again, we can place $\max_{1\le k\le n}$ on both sides. Over any finite interval of t, $|t|$ is bounded by say B, so we can choose $\epsilon>0$ small

enough (and n large enough for (i) to kick in) that the uniform bound $B\epsilon + 2\max_{1\le k\le n} P(|X_{nk}| > \epsilon)$ (not dependent on t) on the RHS goes to 0.
(ii) $\implies$ (i): From the proof of the continuity thm (Lec. 5), $P(|X| \ge \lambda) \le 7\lambda \int_0^{1/\lambda}(1 - \operatorname{Re}\phi(t))\,dt$, so

$$\begin{aligned}
P(|X_{nk}| > \epsilon) &\le 7\epsilon \int_0^{1/\epsilon} (1 - \operatorname{Re}\phi_{X_{nk}}(t))\,dt = 7\epsilon \int_0^{1/\epsilon} |1 - \operatorname{Re}\phi(t)|\,dt \\
&\le 7\epsilon \int_0^{1/\epsilon} \sqrt{(1 - \operatorname{Re}\phi_{X_{nk}}(t))^2 + (\operatorname{Im}\phi_{X_{nk}}(t))^2}\,dt \\
&= 7\epsilon \int_0^{1/\epsilon} |1 - \phi_{X_{nk}}(t)|\,dt \le 7\epsilon \frac{1}{\epsilon} \sup_{0\le t\le\epsilon} |1 - \phi_{X_{nk}}| = 7 \sup_{0\le t\le\epsilon} |1 - \phi_{X_{nk}}|
\end{aligned}$$

We can place $\max_{1\le k\le n}$ on both sides, and because $[0,\epsilon]$ is a finite interval of t, (ii) tells us that the RHS goes to 0.

Problem 2

We have two sequences of real numbers, $\{b_1, \dots, b_N\}$ and $\{c_1, \dots, c_N\}$. Let $\mathbf{R} = (R_1, \dots, R_N)$ be distributed uniformly over the set of permutations of $\{1, \dots, N\}$ (i.e. $P(\mathbf{R} = \sigma_N) = \frac{1}{N!}$ for any permutation σ_N), and denote $S_N = \sum_{j=1}^N b_j c_{R_j}$, $\bar{b}_N = \frac{1}{N}\sum_{j=1}^N b_j = \mathbb{E}\big[b_{R_j}\big]$ (for any $j \in \{1, \dots, N\}$, because of the $N!$ permutations, exactly $(N-1)!$ have $R_j = 1$, exactly $(N-1)!$ have $R_j = 2$, and so on), and $B_N^2 = \sum_{j=1}^N (b_j - \bar{b}_N)^2 = N\operatorname{Var}\big[b_{R_j}\big]$ (again variance can be done over any R_j). $\bar{c}_N$ and C_N^2 are defined similarly.

(a) We compute the variance of S_N as follows:

$$\begin{aligned}
\operatorname{Var}[S_N] &= \sum_{j=1}^N \operatorname{Var}\big[b_j c_{R_j}\big] + 2\sum_{1\le i<j\le n} \operatorname{Cov}\big[b_i c_{R_i}, b_j c_{R_j}\big] \\
&= \frac{1}{N} C_N^2 \sum_{j=1}^N b_j^2 + 2\sum_{1\le i<j\le n} b_i b_j \Big(\mathbb{E}\big[c_{R_i} c_{R_j}\big] - \mathbb{E}[c_{R_i}]\mathbb{E}\big[c_{R_j}\big]\Big) \\
&= \frac{1}{N} C_N^2 \cdot N\mathbb{E}\Big[b_{R_j}^2\Big] + 2\sum_{1\le i<j\le n} b_i b_j \left(\frac{N\bar{c}_N^2 - \mathbb{E}\Big[c_{R_j}^2\Big]}{N-1} - \bar{c}_N^2\right) \\
&= \frac{1}{N} C_N^2 \cdot N\mathbb{E}\Big[b_{R_j}^2\Big] + \left(\frac{\bar{c}_N^2 - \mathbb{E}\Big[c_{R_j}^2\Big]}{N-1}\right)\left(\left(\sum_{j=1}^N b_j\right)^2 - \sum_{j=1}^N b_j^2\right) \\
&= C_N^2 \cdot \mathbb{E}\Big[b_{R_j}^2\Big] + \left(\frac{-\operatorname{Var}\big[c_{R_j}\big]}{N-1}\right)\Big(N^2\bar{b}_N^2 - N\mathbb{E}\Big[b_{R_j}^2\Big]\Big) \\
&= C_N^2 \cdot \mathbb{E}\Big[b_{R_j}^2\Big] + \frac{\frac{1}{N}C_N^2 \cdot N\mathbb{E}\Big[b_{R_j}^2\Big]}{N-1} - \frac{\frac{1}{N}C_N^2 \cdot N^2\bar{b}_N^2}{N-1} \\
&= C_N^2 \left(\mathbb{E}\Big[b_{R_j}^2\Big] + \frac{\mathbb{E}\Big[b_{R_j}^2\Big]}{N-1} - \frac{N\bar{b}_N^2}{N-1}\right) = C_N^2 \left(\frac{N\mathbb{E}\Big[b_{R_j}^2\Big]}{N-1} - \frac{N\bar{b}_N^2}{N-1}\right) \\
&= C_N^2 \left(\frac{N}{N-1}\operatorname{Var}\Big[b_{R_j}^2\Big]\right) = \frac{B_N^2 C_N^2}{N-1}
\end{aligned}$$

where we used the fact that (for $i \neq j$, using 522hw3p3), $P(R_i = k|R_j) = \sum_{k'=1}^{N} \frac{P(R_i=k \cap R_j=k')}{P(R_j=k')} 1_{[R_j=k']}$ $= \frac{1}{N-1} \cdot 1_{[R_j \neq k]}$ (because there are $(N-2)!$ permutations where $R_i = k$ and $R_j = k'$ for $k \neq k'$, and $(N-1)!$ permutations where just $R_j = k'$; for $k = k'$, obviously it's 0), and so

$$\begin{aligned}
\mathbb{E}\big[c_{R_i} c_{R_j}\big] &= \mathbb{E}\big[\mathbb{E}\big(c_{R_i} c_{R_j} | R_j\big)\big] = \mathbb{E}\big[c_{R_j} \mathbb{E}(c_{R_i}|R_j)\big] \\
&= \mathbb{E}\left[c_{R_j} \sum_{k=1}^{N} c_k P(R_i = k|R_j)\right] = \mathbb{E}\left[c_{R_j} \frac{c_1 + \ldots + c_N - c_{R_j}}{N-1}\right] \\
&= \frac{\mathbb{E}\big[c_{R_j} \cdot N\bar{c}_N\big] - \mathbb{E}\left[c_{R_j}^2\right]}{N-1} = \frac{N\bar{c}_N^2 - \mathbb{E}\left[c_{R_j}^2\right]}{N-1}
\end{aligned}$$

(b) From chapter 4 of Chen, Goldstein and Shao (2011), if we have a matrix $\{a_{ij}\}_{i,j=1}^{N}$ and $Y = \sum_{j=1}^{N} a_{j,R_j}$ with mean μ and variance σ^2, and $\gamma := \sum_{i,j=1}^{N} |a_{ij} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}|^3$ where $a_{i\bullet}$ is the mean of the N numbers on the ith row, $a_{\bullet j}$ the mean along the jth column, and $a_{\bullet\bullet}$ the mean over the whole matrix, then for F the d.f. of $\frac{Y-\mu}{\sigma}$ and Φ the d.f. of the standard normal,

$$||F - \Phi||_1 \leq \frac{\gamma}{(N-1)\sigma^3}\left(16 + \frac{56}{N-1} + \frac{8}{(N-1)^2}\right)$$

Applying this to our problem above, we can make the matrix where $a_{ij} = b_i c_j$ and so $Y = S_N = \sum_{j=1}^{N} b_j c_{R_j}$. We've already established that $\mathbb{E}[S_N] = \mu = N\bar{b}_N\bar{c}_N$ and $\mathrm{Var}[S_n] = \sigma^2 = \frac{B_N^2 C_N^2}{N-1}$. Regarding γ, $a_{i\bullet} = b_i\bar{c}_N$, $a_{\bullet j} = c_j\bar{b}_N$, and $a_{\bullet\bullet} = \bar{b}_N\bar{c}_N$ and so

$$\begin{aligned}
\gamma &= \sum_{i,j=1}^{N} |b_i c_j - b_i\bar{c}_N - \bar{b}_N c_j + \bar{b}_N\bar{c}_N|^3 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} |b_i - \bar{b}_N|^3 |c_j - \bar{c}_N|^3 = \left(\sum_{i=1}^{N} |b_i - \bar{b}_N|^3\right)\left(\sum_{j=1}^{N} |c_j - \bar{c}_N|^3\right)
\end{aligned}$$

From the inequality $(x+y)^{3/2} \geq x^{3/2} + y^{3/2}$ (for $x, y \geq 0$, which we can prove by fixing y and then showing that $(x+y_0)^{3/2}$ has derivative $\geq$ to that of $x^{3/2}$, and both sides equal $y_0^{3/2}$ at $x = 0$), we see that

$$\sum_{i=1}^{N} |b_i - \bar{b}_N|^3 = \sum_{i=1}^{N} (|b_i - \bar{b}_N|^2)^{3/2} \leq \left(\sum_{i=1}^{N} |b_i - \bar{b}_N|^2\right)^{3/2} = B_N^3$$

and so doing the same thing for the c's, we can bound γ by $\gamma \leq B_N^3 C_N^3$. Thus

$$\frac{\gamma}{(N-1)\sigma^3} \leq \frac{B_N^3 C_N^3}{(N-1)\frac{B_N^3 C_N^3}{(N-1)^{3/2}}} = \sqrt{N-1}$$

which unfortunately doesn't go to 0. As an alternative perspective, Theorem 6.1 gives another bound for the problem:

$$\sup_{x\in\mathbb{R}} |F(x) - \Phi(x)| \le \frac{16.3 \max_{1\le i,j\le n} |a_{ij} - a_{i\bullet} + a_{\bullet j} - a_{\bullet\bullet}|}{\sigma}$$
$$= \frac{16.3 \max_{1\le i\le n} |b_i - \bar{b}_N| \cdot \max_{1\le i\le n} |c_i - \bar{c}_N|}{B_N C_N}\sqrt{N-1}$$

It looks like in general, it is not true that Y converges to the normal d.f., but we can consider special cases of sequences. Let 3fcm($\{b_i\}$) (3rd folded central moment) denote $\frac{1}{n}\sum_{i=1}^{N} |b_i - \bar{b}_N|^3$ (similarly 2fcm($\{b_i\}$) $= \frac{1}{N}B_N^2$). Thus

$$\frac{\gamma}{(N-1)\frac{B_N^3 C_N^3}{(N-1)^{3/2}}} = \frac{N^2\, \mathrm{3fcm}(\{b_i\})\, \mathrm{3fcm}(\{c_i\})}{(N\, \mathrm{2fcm}(\{b_i\}))^{3/2}(N\, \mathrm{2fcm}(\{c_i\}))^{3/2}}\sqrt{N-1}$$
$$= \frac{\sqrt{N-1}}{N}\frac{\mathrm{3fcm}(\{b_i\})\, \mathrm{3fcm}(\{c_i\})}{[\mathrm{2fcm}(\{b_i\})\, \mathrm{2fcm}(\{c_i\})]^{3/2}}$$

Because 3fcm weights large deviations more than 2fcm, this gives us the intuition that for sequences b_i and c_i that don't vary too much, 3fcm doesn't outgrow 2fcm that much, so we do have convergence to 0. In particular, if both sequences lie within a length 1 interval, then 3fcm $\le$ 2fcm and we get convergence to normal at rate about $\frac{1}{\sqrt{N}}$.

(c) As stated above, the bounds given by Chen, Goldstein and Shao give that the convergence to normal is at rate

$$\frac{1}{\sqrt{N}}\frac{\mathrm{3fcm}(\{b_i\})\, \mathrm{3fcm}(\{c_i\})}{[\mathrm{2fcm}(\{b_i\})\, \mathrm{2fcm}(\{c_i\})]^{3/2}}$$

and so I'll put it as "not faster than $\frac{1}{\sqrt{N}}$".

Problem 3

Let $X_1, \ldots, X_n$ be the numbers resulting from sampling without replacement from a collection on N balls labeled with numbers $a_1, \ldots, a_N$, and define $\bar{a}_N = \frac{1}{N}\sum_{i=1}^{N} a_i$ and $\sigma_a^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a}_N)^2$.

(a) Notice that for any $i \in \{1, \ldots, n\}, j \in \{1, \ldots, N\}$, $P(X_i = a_j) = \frac{1}{N}$ (the number of times X_i is a_j must be the same for all a_j, due to the fact that we can just switch/rename the a_j and keep the problem the same). Thus, $\bar{a}_N = \mathbb{E}[X_i]$ and $\sigma_a^2 = \mathrm{Var}[X_i]$. We now want to find $\mathrm{Cov}[X_j, X_k] = \mathbb{E}[X_j X_k] - \mathbb{E}[X_j]\mathbb{E}[X_k]$ (where $j \ne k$). We perform the same trick as above in Problem 2:

$$\mathbb{E}[X_j X_k] = \mathbb{E}[\mathbb{E}(X_j X_k | X_j)] = \mathbb{E}[X_j \mathbb{E}(X_k | X_j)] = \mathbb{E}\left[X_j \frac{N\bar{a}_N - X_j}{N-1}\right] = \frac{N\bar{a}_N^2 - \mathbb{E}\left[X_j^2\right]}{N-1}$$

so

$$\mathrm{Cov}[X_j, X_k] = \frac{N\bar{a}_N^2 - \mathbb{E}\left[X_j^2\right]}{N-1} - \mathbb{E}[X_j]^2 = \frac{\mathbb{E}[X_j]^2 - \mathbb{E}\left[X_j^2\right]}{N-1} = \frac{-\sigma_a^2}{N-1}$$

Therefore, defining $T_n = X_1 + \dots X_n$, we have that

$$\text{Var}\left[\frac{T_n}{n}\right] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\text{Var}[X_i] + 2\sum_{1\le i<j\le n}\text{Cov}[X_i, X_j]\right)$$
$$= \frac{1}{n^2}\left(n\sigma_a^2 + (n^2-n)\frac{-\sigma_a^2}{N-1}\right) = \frac{\sigma_a^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

(b) This can be framed in terms of Problem 2 by setting the sequences $\{b_i\}_{i=1}^N$ and $\{c_i\}_{i=1}^N$ to be $b_i = 1_{i\in\{1,\dots,n\}}$ and $c_i = a_i$.

Problem 4

We suppose that $Y_1, \dots$ are i.i.d. with d.f. G and ch.f. $\phi(t)$. Let N_λ be $\sim$ Poisson(λ) and independent of the $\{Y_i\}$'s, and let $S_\lambda = \sum_{j=1}^{N_\lambda} Y_j$. Then,

$$\phi_{S_\lambda}(t) = \mathbb{E}\big[e^{itS_\lambda}\big] = \mathbb{E}\big[\mathbb{E}\big(e^{itS_\lambda}|N_\lambda\big)\big] = \mathbb{E}\big[\mathbb{E}\big(e^{itY_1}\cdots e^{itY_{N_\lambda}}|N_\lambda\big)\big]$$
$$= \mathbb{E}\big[\phi^{N_\lambda}(t)\big] = \sum_{k=0}^{\infty}\phi^k(t)\cdot e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda\phi(t))^k}{k!} = e^{\lambda(\phi(t)-1)}$$

Problem 5

In the K-function approach for sums of independent r.v.'s, the analogous identity to the Stein identity for normal r.v.'s $\mathbb{E}[f'(W)] = \mathbb{E}[Wf(W)]$ is

$$\mathbb{E}[Wf(W)] = \mathbb{E}\Big[f'\Big(W^{(I)} + \xi_I^*\Big)\Big]$$

where the notation is as follows:

- $W = \sum_{i=1}^n \xi_i$, where the ξ_i are all independent and satisfy $\mathbb{E}[\xi_i] = 0$ and $\sum_{i=1}^n \mathbb{E}\big[\xi_i^2\big] = 1$
- $W^{(i)} = W - \xi_i$
- $K_i(t) = \mathbb{E}\big[\xi_i\cdot(1_{[0,\xi_i]}(t) - 1_{[\xi_i,0)}(t))\big]$
- ξ_i^* has density $K_i(t)/\mathbb{E}\big[\xi_i^2\big]$, and independent of all other ξ_j and ξ_j^* for $j \neq i$
- I is a index r.v., independent of all ξ_i and ξ_i^*, satisfying $P(I = i) = \mathbb{E}\big[\xi_i^2\big]$

Alternatively/equivalently, the identity may be written as

$$\mathbb{E}[Wf(W)] = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\mathbb{E}\Big[f'\Big(W^{(i)} + t\Big)\Big]K_i(t)\,dt$$

Remember, the point of the Stein identity and equation $f'(w) - wf(w) = h(w) - \mathbb{E}[h(Z)]$ was to find f_h for every h we care about (recall the definition of weak convergence $\mathbb{E}[h(X)] \to \mathbb{E}[h(Z)] \iff X \to_d Z$ for some classes of h) such that $\mathbb{E}[h(W) - h(Z)] = \mathbb{E}[f_h'(W) - Wf_h(W)]$, thereby allowing us to understand the LHS by understanding the more easily understood RHS. This K-function approach provides another perspective on the same quantity; i.e.

$$\mathbb{E}[h(W) - h(Z)] = \mathbb{E}[f_h'(W) - Wf(W)] = \mathbb{E}\left[f_h'(W) - f_h'\left(W^{(I)} + \xi_I^*\right)\right]$$

523 Homework 3

Daniel Rui - 4/22/20

Problem 1

Consider a random variable X with p.d.f. $f_X(x) = \frac{1}{x\log^2 x} \cdot 1_{[e,\infty)}$ (where this is a p.d.f. because it's ≥ 0 and has integral 1). However, for any $r > 0$ (using the substitution $e^u = x \iff u = \log x \implies du = \frac{1}{x}\,dx$,

$$\mathbb{E}[X^r] = \int_e^\infty x^r \frac{1}{x\log^2 x}\,dx = \int_1^\infty \frac{e^{ru}}{u^2}\,du = \infty$$

(because $\frac{e^{ru}}{u^2} \to \infty$ as $u \to \infty$. However, if we take $g(x) = \sqrt{\log x}$ (which does go to ∞ as $x \to \infty$), we get that

$$\mathbb{E}[g(X)] = \int_e^\infty \frac{1}{x\log^{1.5} x}\,dx = \int_1^\infty \frac{1}{u^{1.5}}\,du = 2 < \infty$$

Problem 2

If we have $X \sim \text{Poisson}(\lambda)$, then

$$\begin{aligned}\lambda\mathbb{E}[f(X+1)] &= \lambda\sum_{n=0}^\infty f(n+1)\frac{e^{-\lambda}\lambda^n}{n!} = \sum_{n=0}^\infty (n+1)f(n+1)\frac{e^{-\lambda}\lambda^{n+1}}{(n+1)!} \\ &= \sum_{n=1}^\infty nf(n)\frac{e^{-\lambda}\lambda^n}{n!} = \sum_{n=0}^\infty nf(n)\frac{e^{-\lambda}\lambda^n}{n!} = \mathbb{E}[f(X)]\end{aligned}$$

Problem 3

Let $\xi_1, \ldots, \xi_n$ be i.i.d. Unif(0,1) r.v.s, and let $\mathbf{X}_i = (1_{[\xi_i \leq t_1]} - t_1, \ldots, 1_{[\xi_i \leq t_k]} - t_k)$ for $0 < t_1 < \ldots < t_k < 1$. Also, define $\mathbb{U}_n(t) = \sqrt{n}(\mathbb{G}_n(t) - t)$ where $\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{\xi_i \leq t}$. Clearly all the $\mathbf{X}_i$ are i.i.d., and so by the multivariate CLT, we have that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{X}_i = \left(\frac{1}{\sqrt{n}}\left[\Big(\sum_{i=1}^n 1_{[\xi_i \leq t_1]}\Big) - nt_1\right], \ldots, \frac{1}{\sqrt{n}}\left[\Big(\sum_{i=1}^n 1_{[\xi_i \leq t_k]}\Big) - nt_k\right]\right) = (\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))$$

converges in distribution to a Normal$(0, \Sigma)$ distribution where Σ is the covariance matrix, with the ith row and jth column entry being $\mathbb{E}\big[(1_{[\xi_1 \leq t_i]} - t_i)(1_{[\xi_1 \leq t_j]} - t_j)\big] = \mathbb{E}\big[1_{[\xi_1 \leq t_i]}1_{[\xi_1 \leq t_j]}\big] - t_it_j = \min\{t_i, t_j\} - t_it_j$. If we denote $\mathbb{U}$ to be a standard Brownian bridge on $[0,1]$, then $(\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))$ is normally distributed (because Brownian bridges are Gaussian processes) and has covariance matrix with entries exactly equal to the covariance matrix from above (due to the fact that for Brownian bridges, $\mathbb{E}[\mathbb{U}(s)\mathbb{U}(t)] = \min\{s,t\} - st$). Thus, we have that

$$(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k)) \to_d (\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))$$

Problem 4

Let $X_1, \ldots, X_n$ be i.i.d. with mean 0 and variance 1, and let $\mathbb{S}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i$. We want to prove that (again for $0 < t_1 < \ldots < t_k < 1$):

$$(\mathbb{S}_n(t_1), \ldots, \mathbb{S}_n(t_k)) \to_d (\mathbb{S}(t_1), \ldots, \mathbb{S}(t_k))$$

By the Cramér-Wold device, this is equivalent to $a_1\mathbb{S}_n(t_1)+\ldots+a_k\mathbb{S}_n(t_k) \to_d a_1\mathbb{S}(t_1)+\ldots+a_k\mathbb{S}_n(t_k)$, $\forall \mathbf{a} \in \mathbb{R}^k$. But this is equivalent to $a'_1\mathbb{S}_n(t_1) + a'_2(\mathbb{S}_n(t_2) - \mathbb{S}_n(t_1)) + \ldots + a'_k(\mathbb{S}_n(t_k) - \mathbb{S}_n(t_{k-1})) \to_d a'_1\mathbb{S}(t_1) + a'_2(\mathbb{S}(t_2) - \mathbb{S}(t_1)) + \ldots + a'_k(\mathbb{S}(t_k) - \mathbb{S}(t_{k-1}))$ for every $(a'_1, \ldots, a'_k) \in \mathbb{R}^k$ (because for every $\mathbf{a}$, there is corresponding $\mathbf{a'}$ s.t. everything is equal, and vice versa).

Lévy's continuity theorem gives that this is if and only if the characteristic functions converge, but because $\mathbb{S}_n(t_1), (\mathbb{S}_n(t_2) - \mathbb{S}_n(t_1)), \ldots, (\mathbb{S}_n(t_k) - \mathbb{S}_n(t_{k-1}))$ are all independent (and same with $\mathbb{S}$), the characteristic functions of the whole thing will be the products of the characteristic functions of the pieces, and so it will suffice to prove that the ch.f. of $a'_i(\mathbb{S}_n(t_i) - \mathbb{S}_n(t_{i-1}))$ converges to that of $a'_i(\mathbb{S}(t_i) - \mathbb{S}(t_{i-1}))$, which is iff $(\mathbb{S}_n(t_i) - \mathbb{S}_n(t_{i-1}))$ is distributed Normal$(0, t_i - t_{i-1})$. This is simply an application of the CLT and Slutsky's theorem:

$$\begin{aligned}
\mathbb{S}_n(t_i) - \mathbb{S}_n(t_{i-1}) &= \frac{1}{\sqrt{n}} \sum_{i=\lfloor nt_{i-1} \rfloor + 1}^{\lfloor nt_i \rfloor} X_i =_d \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor} X_i \\
&= \sqrt{\frac{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor}{n}} \frac{1}{\sqrt{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor}} \sum_{i=1}^{\lfloor nt_i \rfloor - \lfloor nt_{i-1} \rfloor} X_i \\
&\to_d \sqrt{t_i - t_{i-1}} \text{Normal}(0,1) \sim \text{Normal}(0, t_i - t_{i-1})
\end{aligned}$$

Now this holds for $i \in \{1, \ldots, k\}$ (setting $t_0 = 0$), and as I said above, we can go from this to convergence of ch.f.s by the continuity theorem, multiply everything together from independence, and get back what we wanted to prove.

Problem 5

We have $\{U_n\}$ and $\{V_n\}$ s.t. U_n and V_n are independent for every n, and that $U_n \to_d U$ and $U_n + V_n \to_d U$ for some U independent of the V_n. We want to prove that this implies that $V_n \to_p 0$. Using Skorokhod's convergence in distribution to almost sure convergence theorem, there is another sequence $\{U'_n\}$ s.t. $U'_n =_d U_n$ and $U' =_d U$, and $U'_n \to_{\text{a.s.}} U'$ (where again U' is independent of the V_n). Then by Slutsky's theorem (and the fact that $U'_n - U' \to_{\text{a.s.}} 0 \implies U'_n - U' \to_p 0$), have that

$$U' + V_n = (U'_n + V_n) + (U' - U'_n) =_d (U_n + V_n) + (U' - U'_n) \to_d U =_d U'$$

We can add an absolutely continuous function A to U' in the expression (which keeps the $\to_d$ if we make A independent from everything; can see this because independence means ch.f.s can be multiplied, and if we multiply same thing on both sides the ch.f.s still converge to each other), we get that

$U' + A$ is absolutely continuous (see here; a quick gist is that $f_{X+Y}(t) = \int_{\mathbb{R}} f_X(t-y)\,dF_Y(y)$ which can be checked via integrating). So we can go ahead and assume that U' is absolutely continuous and thus has a density.

Suppose that $V_n \not\to_p 0$, which means that there is $\epsilon_1, p > 0$ s.t. $P(|V_n| \geq \epsilon_1) = P(V_n \geq \epsilon_1) + P(V_n \leq -\epsilon_1) \geq 2p$ for infinitely many n (denote this infinite sequence of n's by S); w.l.o.g., assume that $P(V_n \geq \epsilon_1) \geq p$ for all $n \in S$ (the case for $P(V_n \leq -\epsilon_1) \geq p$ is just everything here flipped).

Now that U' has a density, $F_{U'}(t)$ is continuous, and so $s(t) = P(t < U' \leq t+1) = F_{U'}(t+1) - F_{U'}(t)$ is continuous as well with $s(\infty-) = 0$ and $s(-\infty+) = 0$, implying that it attains its maximum s at some (or many) finite t. Let t_0 be the left-most point at which $s(t)$ attains its maximum. By the continuity of $s(t)$, there is some small enough $\epsilon_2 > 0$ s.t. $\sup_{t \leq t_0 - \epsilon_2} s(t) = r < s$. Take $\epsilon = \min\{\epsilon_1, \epsilon_2\}$ and denote $p_n := P(V_n \geq \epsilon)$ (which means that for all $n \in S$, $p_n \geq p > 0$).

Now because U' and V_n are independent, $P(t_0 < U' + V_n \leq t_0 + 1 | V_n) = P(t_0 - V_n < U' \leq t_0 - V_n + 1 | V_n) = s(t_0 - V_n)$ for all n, which means that $P(t_0 < U' + V_n \leq t_0 + 1 | V_n \geq \epsilon) \leq r$. But we know that $P(t_0 < U' + V_n \leq t_0 + 1) = P(t_0 < U' + V_n \leq t_0 + 1 | V_n \geq \epsilon) P(V_n \geq \epsilon) + P(t_0 < U' + V_n \leq t_0 + 1 | V_n < \epsilon) P(V_n < \epsilon)$, and so for $n \in S$, $P(t_0 < U' + V_n \leq t_0 + 1) \leq r p_n + s(1 - p_n) \leq rp + s(1-p)$ (because $p_n \geq p > 0$ and $r \leq s$). Thus,

$$\begin{aligned}\liminf_{n\to\infty} P(t_0 < U' + V_n \leq t_0 + 1) &\leq rp + s(1-p) \\ &< s = P(t_0 < U' \leq t_0 + 1) \\ &= \lim_{n\to\infty} P(t_0 < U' + V_n \leq t_0 + 1)\end{aligned}$$

where the last step follows due to absolute continuity of U' (hence every real number is a continuity point of $F_{U'}$). This is of course impossible, and so it must be that $V_n \to_p 0$.

Problem 6

The authors of this beautiful paper: `https://arxiv.org/pdf/1810.01768.pdf` (titled "Three remarkable properties of the Normal distribution", authored by Eric Benhamou, Beatrice Guez, and Nocolas Paris), have explained the proof of the theorem in great detail. As an overview, they prove 4 lemmas, the latter 2 of which are used to prove the theorem. Their third lemma states that:

- If X is a real random variable with d.f. F_X s.t. there exists $\eta > 0$ s.t. $f(\eta) = \int_{\mathbb{R}} e^{\eta^2 x^2}\,dF_X(x)$ is finite, and that the ch.f. of X has no zeroes in $\mathbb{C}$, then X has a normal distribution.

The fourth lemma gives another way of computing $f(\eta)$:

- $f(\eta)$ can be written as

$$f(\eta) = 1 + \int_0^\infty 2x\eta^2 e^{\eta^2 x^2} P(|X| \geq x)\,dx$$

For the first 2 lemmas and the proofs of all 4, please read the paper in its fine detail.

Now in the proof of the theorem, they begin by ruling out the case where $X_1 + X_2$ has 0 variance as trivial (which it is). They then prove that X_1 and X_2 can not both have atoms because then

$$0 < P(X_1 = a)P(X_2 = b) = P(X_1 = a, X_2 = b) \leq P(X_1 + X_2 = a + b) = 0$$

which is impossible. Thus w.l.o.g. they take X_2 to be without atoms and so there is some m s.t. $P(X_2 \leq m) = P(X_2 \geq m) = \frac{1}{2}$. Then they have the inequalities (for $x > 0$):

$$\begin{aligned} P(|X_1| \geq x) &= 2P(X_1 \geq x, X_2 \geq m) + 2P(X_1 \leq -x, X_2 \leq m) \\ &\leq 2P(X_1 + X_2 \geq x + m) + 2P(X_1 + X_2 \leq -x + m) \\ &= 2P(|X_1 + X_2 - m| \geq x) \end{aligned}$$

Using the fourth lemma, we see that for the $f(\eta)$ corresponding to X_1,

$$\begin{aligned} 0 \leq f(\eta) &= 1 + \int_0^\infty 2x\eta^2 e^{\eta^2 x^2} P(|X_1| \geq x)\, dx \\ &\leq 1 + 2\int_0^\infty 2x\eta^2 e^{\eta^2 x^2} P(|X_1 + X_2 - m| \geq x)\, dx \\ &< 2 + 2\int_0^\infty 2x\eta^2 e^{\eta^2 x^2} P(|X_1 + X_2 - m| \geq x)\, dx \end{aligned}$$

But this expression is 2 times the $f(\eta)$ corresponding to the r.v. $(X_1 + X_2 - m)$ (let's call it $g(\eta)$). Because $(X_1 + X_2 - m)$ is normal, $g(\eta)$ is finite and so $f(\eta)$ is finite as well. Finally, $\phi_{X_1}\phi_{X_2} = \phi_{X_1+X_2}$ by independence of X_1 and X_2, and because $\phi_{X_1+X_2}$ is never 0 in $\mathbb{C}$ (another property of the normal distribution), ϕ_{X_1} can't either. And thus by the third lemma, X_1 is normal. Disregarding the trivial case that X_1 is 0 variance, we now know that X_1 has no atoms, so we flip X_1 and X_2 to get that X_2 is normal as well.

The crux of the proof lies in the third lemma; basically the key to cracking the problem was to isolate enough properties of normal distributions that we could have some sort of checklist that would guarantee that we had a normal distributed r.v..

523 Homework 2

Daniel Rui - 4/15/20

Problem 1

Define a sequence of partitions $\mathscr{P}_n = \{(t_{n,k-1}, t_{nk}] : k = 1, \ldots, n\}$ of $[0,1]$, where $0 = t_{n0} < \ldots < t_{nn} = 1$, and define the corresponding rth variation of $\mathbb{S}$ to be $V_n(r) = \sum_{k=1}^n |\mathbb{S}(t_{nk}) - \mathbb{S}(t_{n,k-1})|^r$. The mesh of a partition $\mathscr{P}_n$ is $||\mathscr{P}_n|| = \sup_{1\le k\le n} |t_{nk} - t_{n,k-1}|$.

Let us now look at a special case: $\mathscr{P}_n = \{(\frac{k-1}{2^n}, \frac{k}{2^n}] : k = 1, \ldots, 2^n\}$. We want to prove that as $n \to \infty$ (i.e. as the mesh $= \frac{1}{2^n}$ goes to 0), $V_n(1) \to_{\text{a.s.}} \infty$. With this partition and because of Brownian motion's stationary and independent increment property, $\mathbb{S}(t_{nk}) - \mathbb{S}(t_{n,k-1})$ are all Gaussian, mean 0 and variance $\frac{1}{2^n}$, and so can be thought of as $\frac{1}{\sqrt{2^n}} Z_k$ for Z_k i.i.d. Normal(0,1). Thus $\mathbb{E}[V_n(1)] = \sum_{k=1}^{2^n} \frac{1}{\sqrt{2^n}} \mathbb{E}[|Z_k|]$ and $\text{Var}[V_n(1)] = \sum_{k=1}^{2^n} \frac{1}{2^n} \text{Var}[|Z_k|]$. The mean and variance of a folded normal are $\sqrt{\frac{2}{\pi}}$ and $1 - \frac{2}{\pi}$ respectively, and so putting everything together we have

$$\mathbb{E}[V_n(1)] = \frac{1}{\sqrt{2^n}}(2^n)\sqrt{\frac{2}{\pi}} = \sqrt{\frac{2^{n+1}}{\pi}} \quad \text{and} \quad \text{Var}[V_n(1)] = \frac{1}{2^n}(2^n)\left(1 - \frac{2}{\pi}\right) = 1 - \frac{2}{\pi}$$

As the hint suggests, we use the Paley-Zygmund inequality: for $\lambda \in [0,1]$,

$$P\big(V_n(1) > \lambda \mathbb{E}[V_n(1)]\big) \ge (1-\lambda)^2 \frac{(\mathbb{E}[V_n(1)])^2}{\text{Var}[V_n(1)] + (\mathbb{E}[V_n(1)])^2} = (1-\lambda)^2 \frac{2^{n+1}}{(\pi - 2) + 2^{n+1}} \to (1-\lambda)^2$$

Now we fix an arbitrary (small) $\lambda \in (0,1)$, (small) $\epsilon > 0$, and (large) $c > 0$. The above limit can be written as: there exists some N_1 s.t. $n > N_1 \implies P(V_n(1) > \lambda\mathbb{E}[V_n(1)]) \ge (1-\lambda)^2 - \epsilon$. Clearly, $\mathbb{E}[V_n(1)]$ grows without bound as $n \to \infty$, and so there will exist some N_2 large enough that $\lambda\mathbb{E}[V_n(1)] > c$, which means that for all $n > \max\{N_1, N_2\}$, $P(V_n(1) > c) \ge P(V_n(1) \ge \lambda\mathbb{E}[V_n(1)]) \ge (1-\lambda)^2 - \epsilon$. Now because the partitions are nested so nicely, we can use the triangle inequality to get that $V_{n+1}(1) \ge V_n(1)$, and so $[V_{n+1}(1) > c] \supseteq [V_n(1) > c]$. If we denote $V_\infty(1) := \lim_{n\to\infty} V_n(1)$, we get that $[V_\infty(1) > c] = \bigcup_{n=1}^\infty [V_n(1) > c]$, and due to limit-measure commutativity for monotone sets,

$$P([V_\infty(1) > c]) = P\left(\bigcup_{n=1}^\infty [V_n(1) > c]\right) = \lim_{n\to\infty} P(V_n(1) > c) \ge (1-\lambda)^2 - \epsilon$$

And because λ and ϵ can get arbitrarily close to 0 and this inequality will still hold, we have that for any $c > 0$, $P(V_\infty(1) > c) = 1$, which means that $V_\infty(1) =_{\text{a.s.}} \infty$.

Problem 3

Let $Y_1, \ldots, Y_{n+1}$ i.i.d. Exp(1), and define $S_k = Y_1 + \ldots + Y_k$ for $1 \le k \le n+1$. We want to show that

$$\left(\frac{S_1}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}}\right) =_d (\xi_{n:1}, \ldots, \xi_{n:n})$$

where $\xi_{n:i}$ are the order statistics of n i.i.d. Uniform(0,1) r.v.'s (as above in Problem 2). The joint density of order statistics in general is $f_{(X_{n:1},\ldots,X_{n:n})}(x_1,\ldots,x_n) = n! f_X(x_1)\cdots f_X(x_n) 1_{[x_1<\ldots<x_n]}$, and so applying this to our case we have that $f_{(\xi_{n:1},\ldots,\xi_{n:n})}(x_1,\ldots,x_n) = n! 1_{[0<x_1\ldots x_n<1]}$.

Now we just have to show that $(\frac{S_1}{S_{n+1}},\ldots,\frac{S_n}{S_{n+1}})$ has the above density. First, let us establish that $(Y_1,\ldots,Y_n)$ has density $f_{(Y_1,\ldots,Y_n)}(y_1,\ldots,y_n) = f_{Y_1}(y_1)\cdots f_{Y_n}(y_n) = \prod_{k=1}^n e^{-y_k} 1_{[y_1\geq 0,\ldots,y_n\geq 0]}$ due to independence. The change of variable theorem for p.d.f.'s (whose proof stems from the basic change of variables theorem for multi-dimensional integration) says that if there is a continuous(ly differentiable?) 1-1 function $\mathbf{G}:\mathbb{R}^n\to\mathbb{R}^n$ s.t. $(B_1,\ldots,B_n) = \mathbf{G}(A_1,\ldots,A_n)$ with $\mathbf{H}$ as its inverse, we have that

$$f_{(B_1,\ldots,B_n)}(b_1,\ldots,b_n) = |J_{\mathbf{H}}(b_1,\ldots,b_n)| f_{(A_1,\ldots,A_n)}(\mathbf{H}(b_1,\ldots,b_n))$$

for all $(b_1,\ldots,b_n)\in\mathbf{G}(S)$ where $S\subseteq\mathbb{R}^n$ is a set for which $P((A_1,\ldots,A_n)\in S)=1$ (and of course $J_{\mathbf{H}}$ is the Jacobian matrix of $\mathbf{H}$). Outside of $\mathbf{G}(S)$, $f_{(B_1,\ldots,B_n)}$ will be 0.

We now find our $\mathbf{G}$: denote $\Sigma_k = \frac{S_k}{S_{n+1}}$ (for $k\in\{1,\ldots,n\}$), and $\Sigma_{n+1} = S_{n+1}$; then we have $\mathbf{\Sigma}_{n+1} := (\Sigma_1,\ldots,\Sigma_{n+1}) = \mathbf{G}(\mathbf{Y}_{n+1})$ where $\mathbf{G}(\mathbf{y}_{n+1}) = (\frac{y_1}{y_1+\ldots+y_{n+1}}, \frac{y_1+y_2}{y_1+\ldots+y_{n+1}},\ldots,\frac{y_1+\ldots+y_n}{y_1+\ldots+y_{n+1}}, y_1+\ldots+y_{n+1})$. Inversely, $\mathbf{H}(s_1,\ldots,s_{n+1}) = (s_1 s_{n+1}, (s_2-s_1)s_{n+1},\ldots,(s_n-s_{n-1})s_{n+1}, s_{n+1}-s_n s_{n+1})$. The Jacobian of $\mathbf{H}=(H_1,\ldots,H_{n+1})$ is therefore

$$|J_{\mathbf{H}}(\mathbf{s}_{n+1})| = \begin{vmatrix} [\partial_1 H_1](\mathbf{s}_{n+1}) & \cdots & [\partial_1 H_{n+1}](\mathbf{s}_{n+1}) \\ \vdots & \ddots & \vdots \\ [\partial_{n+1} H_1](\mathbf{s}_{n+1}) & \cdots & [\partial_{n+1} H_{n+1}](\mathbf{s}_{n+1}) \end{vmatrix} = \begin{vmatrix} s_{n+1} & -s_{n+1} & 0 & \cdots\cdots & 0 & 0 \\ 0 & s_{n+1} & -s_{n+1} & \cdots\cdots & 0 & 0 \\ 0 & 0 & s_{n+1} & \ddots\ \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots\ \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots\ \ddots & -s_{n+1} & 0 \\ 0 & 0 & \cdots & \cdots\cdots & s_{n+1} & -s_{n+1} \\ s_1 & (s_2-s_1) & \cdots & \cdots\cdots & (s_n-s_{n-1}) & 1-s_n \end{vmatrix}$$

Because adding columns to each other does not affect the determinant (see (iii) of here), we can add column 1 to column 2, then 2 to 3, and so on (making sure to keep track of changes to the bottom row) until we have a lower left triangular matrix, which of course has determinant equal to the product of the entries along the diagonal, which are n of s_{n+1} and one 1, yielding $|J_{\mathbf{H}}(\mathbf{s}_{n+1})| = s_{n+1}^n$.

Finally, take our set S to be $S=\{(y_1,\ldots,y_n): y_i>0, y_i=y_j \implies i=j\}$ (probability that different Y_i take the same value is 0, so $P((Y_1,\ldots,Y_{n+1})\in S)=1$) so that $\mathbf{G}(S) = \{(s_1,\ldots,s_{n+1}): 0<s_1<\ldots<s_n<1, s_{n+1}>0\}$. Putting this into the formula above we get

$$\begin{aligned} f_{(\Sigma_1,\ldots,\Sigma_{n+1})}(s_1,\ldots,s_{n+1}) &= s_{n+1}^n f_{(Y_1,\ldots,Y_{n+1})}(\mathbf{H}(s_1,\ldots,s_{n+1}))\cdot 1_{\mathbf{G}(S)} \\ &= s_{n+1}^n e^{-(s_1 s_{n+1}+(s_2-s_1)s_{n+1}+\ldots+(s_n-s_{n-1})s_{n+1}+s_{n+1}-s_n s_{n+1})}\cdot 1_{\mathbf{G}(S)} \\ &= s_{n+1}^n e^{-s_{n+1}}\cdot 1_{\mathbf{G}(S)} \end{aligned}$$

Finally, to release our dependency on s_{n+1}, we find the marginal p.d.f. via integration:

$$\begin{aligned} f_{(\Sigma_1,\ldots,\Sigma_n)}(s_1,\ldots,s_n) &= \int_{-\infty}^{\infty} s_{n+1}^n e^{-s_{n+1}} \cdot 1_{\mathbf{G}(S)}\, ds_{n+1} \\ &= \int_0^{\infty} s_{n+1}^n e^{-s_{n+1}} \cdot 1_{[0<s_1<\ldots<s_n<1]}\, ds_{n+1} \\ &= \Gamma(n+1)\cdot 1_{[0<s_1<\ldots<s_n<1]} = n!\cdot 1_{[0<s_1<\ldots<s_n<1]} \end{aligned}$$

which is indeed the distribution we got from the order statistics of the uniform r.v.'s.

Problem 2

Let $\xi_1,\ldots,\xi_n$ be i.i.d. Uniform(0,1), and let $0=\xi_{n:0}\le\xi_{n:1}\le\ldots\le\xi_{n:n}\le\xi_{n:n+1}=1$ be the order statistics, and let $\delta_{n:i}=\xi_{n:i}-\xi_{n:i-1}$ for $i\in\{1,\ldots,n+1\}$ be the spacings. We want to prove that $\sqrt{n}\max\limits_{1\le i\le n+1}\delta_{n:i}\to_p 0$.

From Problem 3, we saw that $(\frac{S_1}{S_{n+1}},\ldots,\frac{S_n}{S_{n+1}}) =_d (\xi_{n:1},\ldots,\xi_{n:n})$, and so intuitively, by taking differences, $(\frac{Y_1}{S_{n+1}},\ldots,\frac{Y_n}{S_{n+1}},\frac{Y_{n+1}}{S_{n+1}}) =_d (\delta_{n:1},\ldots,\delta_{n:n},\delta_{n:n+1})$ (a rigorous proof could be carried out using the p.d.f. transformation method from above). Thus, $\max_{1\le i\le n+1}\delta_{n:i} =_d \max_{1\le i\le n+1} Y_i/S_{n+1}$. Because $\frac{Y_i}{S_{n+1}}$ is independent to S_{n+1}, $\max_{1\le i\le n+1} Y_i/S_{n+1}$ is also independent to S_{n+1}, and so

$$\mathbb{E}\left[\max_{1\le i\le n+1} Y_i\right] = \mathbb{E}\left[\frac{\max_{1\le i\le n+1} Y_i}{S_{n+1}} S_{n+1}\right] = \mathbb{E}\left[\max_{1\le i\le n+1}\delta_{n:i}\right]\mathbb{E}[S_{n+1}] = (n+1)\mathbb{E}\left[\max_{1\le i\le n+1}\delta_{n:i}\right]$$

The order statistics of samples from an exponential distribution are well known; we will just use the fact that $\max_{1\le i\le n+1} Y_i =_d \sum_{i=1}^{n+1}\frac{Y_1}{i}$ (see here; proved using p.d.f. transformations like we did in Problem 3). This of course means that the expectation is

$$(n+1)\mathbb{E}\left[\max_{1\le i\le n+1}\delta_{n:i}\right] = \mathbb{E}\left[\max_{1\le i\le n+1} Y_i\right] = \sum_{i=1}^{n+1}\frac{1}{i} \iff \mathbb{E}\left[\max_{1\le i\le n+1}\delta_{n:i}\right] = \frac{H_{n+1}}{n+1}$$

where H_n is the nth harmonic number. Markov's inequality gives us that for any (small) $c>0$,

$$P\left(\sqrt{n}\max_{1\le i\le n+1}\delta_{n:i} > c\right) = P\left(\max_{1\le i\le n+1}\delta_{n:i} > \frac{c}{\sqrt{n}}\right) \le \frac{\sqrt{n}H_{n+1}}{c(n+1)} \to 0$$

which implies that $\sqrt{n}\max_{1\le i\le n+1}\delta_{n:i} > c \to_p 0$. As for other constants c_n s.t. $c_n\max_{1\le i\le n+1}\delta_{n:i}$ is bounded in probability up to an exceptional event of arbitrarily small probability (i.e. for any ϵ, there is C_ϵ and N_ϵ s.t. for $n\ge N_\epsilon$, $P(c_n\max_{1\le i\le n+1}\le C_\epsilon) > 1-\epsilon$), we could e.g. take $c_n=\frac{n}{\log n}$, which would yield

$$P\left(\frac{n+1}{\log(n+1)}\max_{1\le i\le n+1}\delta_{n:i} > c\right) = P\left(\max_{1\le i\le n+1}\delta_{n:i} > \frac{c\log(n+1)}{n+1}\right) \le \frac{(n+1)H_{n+1}}{c(n+1)\log(n+1)} \to \frac{1}{c}$$

and so for every $\epsilon>0$, take $C_\epsilon=\frac{1}{\epsilon}$ as the bound for $c_n\max_{1\le i\le n+1}\delta_{n:i}$ (up to exceptional event).

Problem 4

Strassen's function LIL gives that for $\mathbb{Z}_n(t) = \mathbb{S}(nt)/\sqrt{2n\log\log n}$ and continuous (w.r.t. the uniform metric) $g : C[0,1] \to \mathbb{R}$, that

$$\limsup_{n\to\infty} g(\mathbb{Z}_n) = \sup_{f\in\mathscr{K}} g(f) \text{ a.s.}$$

where $\mathscr{K}$ is defined as

$$\mathscr{K} = \left\{ f \in C[0,1] : f(0) = 0, f(t) = \int_0^t f'(s)\,ds, \int_0^1 [f'(s)]^2\,ds \le 1 \right\}$$

In our case, we have $g(f) = \int_0^1 f_0(t)f(t)\,dt$ for some fixed $f_0(t)$ s.t. $f_0(0) = 0, f_0(t) = \int_0^t f_0'(s)\,ds$, and $\int_0^1 [f_0'(s)]^2\,ds < \infty$. This is continuous w.r.t. the uniform metric because $\sup_{0\le t\le 1}|f_1(t) - f_2(t)| < \frac{3}{2}\epsilon||f_0'||_2$ implies that

$$\left| \int_0^1 f_0(t)(f_1(t) - f_2(t)) \right| dt \le \int_0^1 |f_0(t)| \frac{3}{2}\epsilon||f_0'||_2\,dt \le \frac{3}{2}\epsilon||f_0'||_2 \int_0^1 \sqrt{t}\,dt = \epsilon$$

because

$$|f_0(t)| = \left| \int_0^1 1_{[0,t]} f_0'(s)\,ds \right| \le \sqrt{\int_0^1 1_{[0,t]}^2\,ds \int_0^1 [f_0'(s)]^2\,ds} = \sqrt{t}||f_0'||_2$$

Similarly, we have for $f \in \mathscr{K}$ that $|f(t)| \le \sqrt{t}$. We now establish several different bounds:

$$\int_0^1 f_0(t)f(t)\,dt \le \left| \int_0^1 f_0(t)f(t)\,dt \right| \le \int_0^1 |f_0(t)|\sqrt{t}\,dt$$

Second:

$$|f_0(t)f(t)| \le \sqrt{t}||f_0'||_2\sqrt{t} = t||f_0'||_2$$

implying that

$$\int_0^1 f_0(t)f(t)\,dt \le \int_0^1 |f_0(t)f(t)|\,dt \le ||f_0'||_2 \int_0^1 t\,dt = \frac{||f_0'||_2}{2}$$

Third:

$$\int_0^1 f_0(t)f(t)\,dt \le \sqrt{\int_0^1 f_0^2(t)\,dt \int_0^1 f^2(t)\,dt} \le ||f_0||_2 \sqrt{\int_0^1 t\,dt} = \frac{||f_0||_2}{\sqrt{2}}$$

The third bound is in fact stronger than the second:

$$\int_0^1 f_0^2(t)\,dt \le \int_0^1 t||f_0'||_2^2\,dt = \frac{||f_0'||_2^2}{2} \implies ||f_0||_2 \le \frac{||f_0'||_2}{\sqrt{2}} \iff \frac{||f_0||_2}{\sqrt{2}} \le \frac{||f_0'||_2}{2}$$

The first bound is rather mysterious. I could not find any functions $f \in \mathscr{K}$ that could actually attain these bounds. However, consulting with other people in the class led to this bound: denote

$\tilde{f}_0(s) = \int_s^1 f_0(t)\,dt$; then

$$\int_0^1 f_0(t)f(t)\,dt = \int_0^1 \int_0^t f_0(t)f'(s)\,ds\,dt = \int_0^1 \int_s^1 f_0(t)f'(s)\,dt\,ds$$
$$= \int_0^1 f'(s)\tilde{f}_0(s)\,ds \leq \sqrt{\int_0^1 [f'(s)]^2\,ds \int_0^1 \tilde{f}_0^2(s)\,ds} \leq ||\tilde{f}_0^2||_2$$

with equality holding for f defined by $f(t) = \int_0^t f'(s)\,ds$ with $f'(s) = \frac{\tilde{f}_0(s)}{||\tilde{f}_0||_2}$:

$$\int_0^1 f_0(t)f(t)\,dt = \int_0^1 f_0(t)\int_0^t f'(s)\,ds\,dt = \int_0^1 \int_s^1 f_0(t)f'(s)\,dt\,ds$$
$$= \int_0^1 \frac{\tilde{f}_0(s)}{||\tilde{f}_0||_2}\int_s^1 f_0(t)\,ds\,dt = \frac{1}{||\tilde{f}_0^2||_2}\int_0^1 \tilde{f}_0^2(s)\,ds = ||\tilde{f}_0^2||_2$$

$f \in \mathscr{K}$ because the derivative obviously has square integral of 1, and the integral definition satisfies the other two properties. Therefore, we have found our supremum:

$$\limsup_{n\to\infty} g(\mathbb{Z}_n) =_{\text{a.s.}} \sup_{f\in\mathscr{K}} g(f) = ||\tilde{f}_0^2||_2$$

Problem 5

Let $\phi(t)$ be a characteristic function (so for some r.v. X, $\phi(t) = \mathbb{E}\left[e^{itX}\right]$). Fix $c > 0$, $t \in \mathbb{R}$, and define $g(u) = \phi(tu)$ and $U \sim \text{Unif}(0, c)$. The law of the unconscious statistician gives that

$$\mathbb{E}[g(U)] = \int_{-\infty}^{\infty} g(u)f_U(u)\,du = \int_0^c \frac{\phi(tu)}{c}\,du$$

Thus for any $t \in \mathbb{R}$, we have that $\mathbb{E}\left[e^{itXU}\right] = \int_0^c \frac{\phi(tu)}{c}\,du$ and so $\varphi(t) = \int_0^c \frac{\phi(tu)}{c}\,du$ is the characteristic function of the random variable XU.

Problem 6

The p.d.f. of the Logistic(0,1) distribution is $\frac{e^x}{(1+e^x)^2}$ and so the characteristic function is $\phi(t) = \int_{-\infty}^{\infty} e^{itx}\frac{e^x}{(1+e^x)^2}\,dx$. Looking at ϕ only along the line $t = iy$ for $y \in \mathbb{R}$:

$$\phi(iy) = \int_{-\infty}^{\infty} e^{-yx}\frac{e^x}{(1+e^x)^2}\,dx = \int_{-\infty}^{\infty} (e^x)^{-y}\frac{e^x}{(1+e^x)^2}\,dx$$

Using the substitution $u = \frac{1}{1+e^x} \implies du = -\frac{e^x}{(1+e^x)^2}\,dx$ and also $\implies e^x = \frac{1}{u} - 1 = \frac{1-u}{u}$, we have

$$\phi(iy) = \int_1^0 \left(\frac{1-u}{u}\right)^{-y}(-1)\,du = \int_0^1 u^y(1-u)^{-y}\,du = B(1+y, 1-y)$$

where B is the beta function defined as $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$. Finally, we use the identity relating the beta and gamma functions: $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, the fact that $\Gamma(1+z) = z\Gamma(z)$, and the gamma reflection formula: $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$, to get that

$$\phi(iy) = \frac{\Gamma(1+y)\Gamma(1-y)}{\Gamma(2)} = \frac{y\Gamma(y)\Gamma(1-y)}{1} = \frac{\pi y}{\sin(\pi y)}$$

Defining $\psi(z) = \frac{\pi z}{\sinh(\pi z)}$, we get that $\psi(iy) = \frac{i\pi y}{\sinh(i\pi y)} = \frac{i\pi y}{i\sin(\pi y)} = \frac{\pi y}{\sin(\pi y)}$, and so $\phi(iy) = \psi(iy)$. Because both functions are analytic and agree on the imaginary axis, they must be identical, and so $\frac{\pi z}{\sinh(\pi z)}$ is the characteristic function of the Logistic(0,1) distribution.

523 Homework 1

Daniel Rui - 4/8/20

Problem 1

Given X with $\mathbb{E}[X]=0$ ($\implies \mathbb{E}[X^+]=\mathbb{E}[X^-]$) with d.f. F, We want to prove that

$$H(a,b)=\frac{1}{\mathbb{E}[X^+]}\int_{[0,a]}\int_{(0,b]} u+v\,dF(-u)\,dF(v)=\frac{1}{\mathbb{E}[X^+]}\int_{[0,a]}\int_{[-b,0)} -u+v\,dF(u)\,dF(v)$$

is a bivariate distribution function on $[0,\infty)\times(0,\infty)$. First we prove it has mass 1:

$$\begin{aligned}
H(\infty,\infty)&=\frac{1}{\mathbb{E}[X^+]}\left(\int_{[0,\infty)}\int_{(-\infty,0)} -u\,dF(u)\,dF(v)+\int_{[0,\infty)}\int_{(-\infty,0)} v\,dF(u)\,dF(v)\right)\\
&=\frac{1}{\mathbb{E}[X^+]}\left(\int_{[0,\infty)}\int_{X^{-1}((-\infty,0))} -X\,dP\,dF(v)+\int_{[0,\infty)} vF((-\infty,0))\,dF(v)\right)\\
&=\frac{1}{\mathbb{E}[X^+]}\left(\int_{[0,\infty)}\mathbb{E}\big[X^-\big]\,dF(v)+F((-\infty,0))\int_{X^{-1}([0,\infty))} X\,dP\right)\\
&=\frac{1}{\mathbb{E}[X^+]}\Big(\mathbb{E}\big[X^-\big]F([0,\infty))+\big(1-F([0,\infty))\big)\mathbb{E}\big[X^+\big]\Big)\\
&=\frac{1}{\mathbb{E}[X^+]}\Big(\mathbb{E}\big[X^+\big]+F([0,\infty))(\mathbb{E}\big[X^-\big]-\mathbb{E}\big[X^+\big])\Big)=1
\end{aligned}$$

Furthermore, H is increasing w.r.t. both variables (with the other fixed) because $u+v$ is non-negative and the measure $F(-u)$ and $F(v)$ are of course also non-negative. Lastly, it is "right-continuous" in each of the its variables (with the other fixed): I will show the one where b is held constant here; the other case follows similarly.

$$\begin{aligned}
\lim_{a\searrow a_0} H(a,b_0)&=\lim_{a\searrow a_0}\frac{1}{\mathbb{E}[X^+]}\int_{[0,a]}\int_{(0,b_0]} u+v\,dF(-u)\,dF(v)\\
&=\lim_{a\searrow a_0}\frac{\Big(\mathbb{E}\big[-X\cdot 1_{X^{-1}([-b_0,0))}\big]F([0,a])+F([-b_0,0))\mathbb{E}\big[X\cdot 1_{X^{-1}([0,a])}\big]\Big)}{\mathbb{E}[X^+]}
\end{aligned}$$

which converges to $H(a_0,b_0)$ because $F(a)$ converges to $F(a_0)$ by right continuity of F, and because $\mathbb{E}\big[X\cdot 1_{X^{-1}((a_0,a])}\big]$ goes to 0 by the DCT (dominated by $|X|$ which must have finite integral because $\mathbb{E}[X]$ exists, and because $1_{X^{-1}((a_0,a])}$ goes to 0 everywhere as $a\searrow a_0$).

Problem 2

(a) From lecture one, the sums of i.i.d. r.v.'s have the strong Markov property; in particular, if we have n balls, R of them red, letting X_i be 1 if the ith sample is red and 0 otherwise (i.i.d. because we are sampling with replacement), $S_n=\sum_{i=1}^n X_i$ (which counts the number of red balls drawn) is a strong Markov process (and hence also a Markov process).

(b) see (a).

(c) What if we do not replace? Is it still a Markov process then? Let us remind ourselves what a Markov process is: $\{S_t\}$ is a Markov process if $\forall B \in \mathscr{B}$ and all $s, t \in \mathbb{N}$, $s < t$, $P(S_t \in B|\mathscr{A}_s) = P(S_t \in B|S_s)$ (where in this case $\mathscr{A}_s = \sigma[S_1, \ldots, S_s]$) — alternative definition here. We utilize the formula from last quarter: if $\Omega = \bigsqcup_{i=I} D_i$ for finite/countable I, and $\mathscr{D} = \sigma[\{D_1, \ldots\}]$, we have that

$$P(A|\mathscr{D}) = \sum_{i \in I} \frac{P(A \cap D_i)}{P(D_i)} 1_{D_i}$$

In our case, we have that $\sigma[S_s] = \sigma[\{[S_s = 0], [S_s = 1], \ldots\}]$ and $\mathscr{A}_s = \sigma[X_1, \ldots, X_s] = \sigma[\{[\mathbf{X}_s = (0, \ldots, 0)], [\mathbf{X}_s = (1, \ldots, 0)], \ldots\}]$ (running over every vector of $\{0,1\}^s$ and where $\mathbf{X}_s = (X_1, \ldots, X_s)$). Thus,

$$\begin{aligned}
P(S_t = k|S_{t-1}) &= \sum_{i=1}^{\infty} \frac{P(S_t = k \cap S_{t-1} = i)}{P(S_{t-1} = i)} 1_{[S_{t-1}=i]} \\
&= \frac{P(S_t = k \cap S_{t-1} = k)}{P(S_{t-1} = k)} 1_{[S_{t-1}=k]} + \frac{P(S_t = k \cap S_{t-1} = k-1)}{P(S_{t-1} = k-1)} 1_{[S_{t-1}=k-1]} \\
&= P(S_t = k|S_{t-1} = k) 1_{[S_{t-1}=k]} + P(S_t = k|S_{t-1} = k-1) 1_{[S_{t-1}=k-1]} \\
&= \left(1 - \frac{R-k}{n-(t-1)}\right) 1_{[S_{t-1}=k]} + \left(\frac{R-(k-1)}{n-(t-1)}\right) 1_{[S_{t-1}=k-1]}
\end{aligned}$$

On the other hand, denoting $\mathbf{x}_{t-1} = (x_1, \ldots, x_{t-1})$ and $|\mathbf{x}_{t-1}| = x_1 + \ldots + x_{t-1}$, we have

$$\begin{aligned}
P(S_t = k|\mathscr{A}_{t-1}) &= \sum_{\substack{\mathbf{x}_{t-1} \in \{0,1\}^{t-1} \\ |\mathbf{x}_{t-1}|=k}} \frac{P([X_t = 0] \cap [\mathbf{X}_{t-1} = \mathbf{x}_{t-1}])}{P([\mathbf{X}_{t-1} = \mathbf{x}_{t-1}])} \cdot 1_{[\mathbf{X}_{t-1}=\mathbf{x}_{t-1}]} \\
&\quad + \sum_{\substack{\mathbf{x}_{t-1} \in \{0,1\}^{t-1} \\ |\mathbf{x}_{t-1}|=k-1}} \frac{P([X_t = 1] \cap [\mathbf{X}_{t-1} = \mathbf{x}_{t-1}])}{P([\mathbf{X}_{t-1} = \mathbf{x}_{t-1}])} \cdot 1_{[\mathbf{X}_{t-1}=\mathbf{x}_{t-1}]} \\
&= \sum_{\substack{\mathbf{x}_{t-1} \in \{0,1\}^{t-1} \\ |\mathbf{x}_{t-1}|=k}} \left(1 - \frac{R-k}{n-(t-1)}\right) \cdot 1_{[\mathbf{X}_{t-1}=\mathbf{x}_{t-1}]} \\
&\quad + \sum_{\substack{\mathbf{x}_{t-1} \in \{0,1\}^{t-1} \\ |\mathbf{x}_{t-1}|=k-1}} \left(\frac{R-(k-1)}{n-(t-1)}\right) \cdot 1_{[\mathbf{X}_{t-1}=\mathbf{x}_{t-1}]} \\
&= \left(1 - \frac{R-k}{n-(t-1)}\right) 1_{[S_{t-1}=k]} + \left(\frac{R-(k-1)}{n-(t-1)}\right) 1_{[S_{t-1}=k-1]}
\end{aligned}$$

because

$$[S_{t-1} = k] = \bigsqcup_{\substack{\mathbf{x}_{t-1} \in \{0,1\}^{t-1} \\ |\mathbf{x}_{t-1}|=k}} [\mathbf{X}_{t-1} = \mathbf{x}_{t-1}]$$

(which can be verified using a rather trivial $\subseteq$-$\supseteq$ argument). Thus even without replacement this process is a Markov process.

Problem 3

Consider $C_\infty := C^0([0,\infty))$, i.e. the set of a all continuous functions on $[0,\infty]$. We want to define the following metric:

$$\rho_\infty(x,y) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{\rho_k(x,y)}{1+\rho_k(x,y)}$$

where $\rho_k(x,y) := \sup\limits_{0\le t\le k} |x(t)-y(t)|$.

(a) We verify that (C_∞, ρ_∞) is indeed a metric space:

- $\rho_\infty(x,y) \geq 0$ obviously because everything we are working with is ≥ 0.
- $\rho_\infty(x,y) = \rho_\infty(y,x)$ obviously because $\rho_k(x,y) = \rho_k(y,x)$ for all k.
- $\rho_\infty(x,y) = 0 \iff x = y$: ($\Longleftarrow$) is obvious because then $\rho_k(x,y) = 0$ for all k; ($\Longrightarrow$) is also pretty easy because the only way for the sum to be 0 is if every term is 0, which means for all k, x must equal y everywhere on $[0,k]$, which of course means that $x = y$ on $[0,\infty)$.
- Finally, the triangle inequality:

$$\begin{aligned}
\frac{\rho_k(x,y)}{1+\rho_k(x,y)} &= \left(1 - \frac{1}{1+\rho_k(x,y)}\right) = \left(1 - \frac{1}{1+\sup_{0\le t\le k}|x(t)-y(t)|}\right) \\
&= \sup_{0\le t\le k}\left(1 - \frac{1}{1+|x-y|}\right) = \sup_{0\le t\le k}\left(1-\frac{1}{1+|x-z+z-y|}\right) \\
&\le \sup_{0\le t\le k}\left(1 - \frac{1}{1+|x-z|+|z-y|}\right) = \sup_{0\le t\le k}\left(\frac{|x-z|+|z-y|}{1+|x-z|+|z-y|}\right) \\
&= \sup_{0\le t\le k}\left(\frac{|x-z|}{1+|x-z|+|z-y|} + \frac{|z-y|}{1+|x-z|+|z-y|}\right) \\
&\le \sup_{0\le t\le k}\left(\frac{|x-z|}{1+|x-z|} + \frac{|z-y|}{1+|z-y|}\right) \\
&\le \sup_{0\le t\le k}\left(\frac{|x-z|}{1+|x-z|}\right) + \sup_{0\le t\le k}\left(\frac{|z-y|}{1+|z-y|}\right) = \rho_k(x,z) + \rho_k(z,y)
\end{aligned}$$

Yay!

(b) We now want to show that $\rho_\infty(x,y) \to 0 \iff \rho_k(x,y) \to 0$ for all $k \in \mathbb{N}$. Because $|C_\infty| = |\mathbb{R}|$ (see here), we just have to concern ourselves with sequences of functions x_t where $t \in [0,\infty)$.

($\Longrightarrow$) Because all the terms are positive, $\rho_\infty(x_t,y)$ is of course $\geq \frac{1}{2^k}\frac{\rho_k(x_t,y)}{1+\rho_k(x_t,y)}$ for any k. Thus, as $t\to\infty$, $\rho_\infty(x_t,y) \to 0 \implies \frac{1}{2^k}\frac{\rho_k(x_t,y)}{1+\rho_k(x_t,y)} \to 0$. The only way for this to approach 0 is if $\rho_k(x_t,y)\to 0$, and so we're done with this direction.

($\Longleftarrow$) Fix an $\epsilon > 0$, and fix an N s.t. $\frac{1}{2^N} < \frac{\epsilon}{2}$. Now because $\rho_k(x_t,y)\to 0$ for every k, we know there is t_k s.t. $\rho_k(x_t,y) < \frac{\epsilon}{2}$ for every $t \geq t_k$. This means that for $t \geq \max\{t_1,\dots,t_N\}$, we have that

$$\rho_\infty(x_t,y) = \sum_{k=1}^{N} \frac{1}{2^k}\frac{\rho_k(x_t,y)}{1+\rho_k(x_t,y)} + \sum_{k=N+1}^{\infty} \frac{1}{2^k}\frac{\rho_k(x_t,y)}{1+\rho_k(x_t,y)}$$
$$\le \sum_{k=1}^{N}\frac{1}{2^k}\rho_k(x_t,y) + \sum_{k=N+1}^{\infty}\frac{1}{2^k}\cdot 1 < \sum_{k=1}^{N}\frac{1}{2^k}\frac{\epsilon}{2} + \frac{1}{2^N} < 1\cdot\frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon$$

which means of course that $\rho_\infty(x_t,y)\to 0$ as $t\to\infty$.

Problem 4

Let $X_0 = 0$ and $X_1,\dots$ be i.i.d. with mean 0 and variance $\sigma^2<\infty$. Of course, $S_k = \sum_{i=1}^k X_i$ for $k\in\mathbb{Z}_{\ge 0}$.

(a) We wish to find the asymptotic distribution of $\frac{1}{c_n}\sum_{i=1}^n S_i$ for "appropriate" c_n. We will proceed using the machinery developed in 12.8: define $\mathbb{S}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor} X_i$. Writing S_i in terms of $\mathbb{S}_n$, we have that $S_i = \sqrt{n}\mathbb{S}_n(\frac{i}{n})$, and so $\frac{1}{c_n}\sum_{i=1}^n S_i = \frac{\sqrt{n}}{c_n}\sum_{i=1}^n \mathbb{S}_n(\frac{i}{n})$. This is kind of looking like a Riemann sum; we can make the connection more obvious by setting $c_n = n^{3/2}$. Intuitively, it feels like the sum $\frac{1}{n}\sum_{i=1}^n \mathbb{S}_n(\frac{i}{n})$ would go to $\int_0^1 \mathbb{S}(t)\,dt$. We prove this rigorously below:

Letting $\mathbb{S}'_n(t) = \frac{\mathbb{S}_n(t)}{\sigma}$ (so it has variance 1), the Skorokhod embedding theorem gives that $||\mathbb{S}'_n - \mathbb{S}|| := \sup_{0\le t\le 1}|\mathbb{S}'_n(t)-\mathbb{S}(t)| \to_p 0$ as $n\to\infty$. Thus $|\frac{1}{n}\sum_{i=1}^n \mathbb{S}'_n(\frac{i}{n}) - \frac{1}{n}\sum_{i=1}^n \mathbb{S}(\frac{i}{n})| \le \frac{1}{n}\sum_{i=1}^n |\mathbb{S}'_n(\frac{i}{n}) - \mathbb{S}(\frac{i}{n})| \le \frac{1}{n}\sum_{i=1}^n ||\mathbb{S}'_n - \mathbb{S}|| = ||\mathbb{S}'_n - \mathbb{S}|| \to_p 0$. Because $\mathbb{S}$ is almost surely everywhere continuous, $\frac{1}{n}\sum_{i=1}^n \mathbb{S}(\frac{i}{n}) \to_{\text{a.s.}} \int_0^1 \mathbb{S}(t)\,dt$, and so putting things together, we can say that $\frac{1}{n}\sum_{i=1}^n \mathbb{S}'_n(\frac{i}{n}) \to_p \int_0^1 \mathbb{S}(t)\,dt \iff \frac{1}{n}\sum_{i=1}^n \mathbb{S}_n(\frac{i}{n}) \to_p \sigma\int_0^1 \mathbb{S}(t)\,dt$.

$\int_0^1 \mathbb{S}(t)$ is $\sim$ Normal$(0,1)$ (see here). It has mean 0, and variance that we calculate below:

$$\mathbb{E}\left[\left(\int_0^1 \mathbb{S}(t)dt\right)^2\right] = \mathbb{E}\left[\int_0^1 \mathbb{S}(t)dt\int_0^1 \mathbb{S}(s)\,ds\right] = \mathbb{E}\left[\int_0^1\int_0^1 \mathbb{S}(t)\mathbb{S}(s)\,dt\,ds\right]$$
$$= \int_0^1\int_0^1 \mathbb{E}[\mathbb{S}(t)\mathbb{S}(s)]\,dt\,ds = \int_0^1\int_0^1 \min\{t,s\}\,dt\,ds$$
$$= 2\int_0^1\int_0^1 t\cdot 1_{t\le s}\,dt\,ds = 2\int_0^1\int_0^s t\,dt\,ds = 2\cdot\frac{1}{6} = \frac{1}{3}$$

Thus all in all we have $\frac{1}{n}\sum_{i=1}^n \mathbb{S}_n(\frac{i}{n}) \to_p \sigma\int_0^1 \mathbb{S}(t) \implies \frac{1}{n}\sum_{i=1}^n \mathbb{S}_n(\frac{i}{n}) \to_d \sigma\int_0^1 \mathbb{S}(t) \sim$ Normal$(0,\sigma^2/3)$.

(b) If we take the absolute value of $|S_i$ above, we will still have $|S_i| = \sqrt{n}|\mathbb{S}_n(\frac{i}{n})|$ and so $\frac{1}{c_n}\sum_{i=1}^n |S_i| = \frac{\sqrt{n}}{c_n}\sum_{i=1}^n |\mathbb{S}_n(\frac{i}{n})|$. Furthermore, we have $\left|\frac{1}{n}\sum_{i=1}^n |S'_n(\frac{i}{n})| - \frac{1}{n}\sum_{i=1}^n |\mathbb{S}(\frac{i}{n})|\right| \le \frac{1}{n}\sum_{i=1}^n \left||\mathbb{S}'_n(\frac{i}{n})| - |\mathbb{S}(\frac{i}{n})|\right| \le \frac{1}{n}\sum_{i=1}^n |\mathbb{S}'_n(\frac{i}{n}) - \mathbb{S}(\frac{i}{n})| \le \frac{1}{n}\sum_{i=1}^n ||\mathbb{S}'_n - \mathbb{S}|| = ||\mathbb{S}'_n - \mathbb{S}|| \to_p 0$. $\frac{1}{n}\sum_{i=1}^n |\mathbb{S}(\frac{i}{n})| \to_{\text{a.s.}} \int_0^1 |\mathbb{S}(t)|\,dt$ with the same reasoning as above, and so $\frac{1}{n}\sum_{i=1}^n |\mathbb{S}'_n(\frac{i}{n})| \to_p \int_0^1 |S(t)|\,dt \iff \frac{1}{n}\sum_{i=1}^n |\mathbb{S}_n(\frac{i}{n})| \to_p \sigma\int_0^1 |\mathbb{S}(t)|\,dt$.

Problem 5

I've narrowed things down to three potential topics, from which you can help me pick the one that I'll do my project on:

- Fleming-Viot particle model: problem 6 from https://sites.math.washington.edu/~burdzy/open.pdf, and https://arxiv.org/pdf/0905.1999.pdf. I could study this problem, try to understand the Arxiv paper, consider variations on it, or even (with astronomically small probability) crack the open problem given by Burdzy at the end of his discussion of problem 6.
- Maybe a variant/further study of "longest increasing subsequences", https://www.stat.washington.edu/jaw/RESEARCH/PAPERS/lis.pdf? I mean, you wrote the paper so I would assume that you have some ideas on further paths to pursue (assuming of course I am of the ability to do the pursuing).
- In class a couple days ago you mentioned a topic that you said your colleague from grad school should write a book on (something do with optimal transport and Brownian motion?). Perhaps there is something from this area that I could write a little bit on?
- These past few days contained a lot of material (Markov and strong Markov, Skorokhod embedding, the LIL) that I would like to go over in depth (the pacing was too fast for me to fully understand what was going on). Perhaps the paper could allow me the time and place to go through everything and write up proofs in detail of the things I mentioned above. My only concern is that this topic is too "simple" in that you already covered the material in class, albeit at a quick pace. Do you have any suggestions about how I could make this topic work?

Problem 6

Let us remind ourselves with the reflection principle and Mill's ratio: for $a > 0$ and $\tau_a = \inf\{t : \mathbb{S}(t) = a\}$, and $\mathbb{X}(t) = \mathbb{S}(t+\tau_a) - \mathbb{S}(\tau_a) = \mathbb{S}(t+\tau_a) - a$,

$$\begin{aligned} P\left(\left[\sup_{0\le s\le t} \mathbb{S}(s) \ge a\right]\right) &= P\left(\left[\sup_{0\le s\le t} \mathbb{S}(s) \ge a\right] \cap [\mathbb{S}(t) \ge a]\right) + P\left(\left[\sup_{0\le s\le t} \mathbb{S}(s) \ge a\right] \cap [\mathbb{S}(t) < a]\right) \\ &= P([\mathbb{S}(t) \ge a]) + P\left(\left[\sup_{0\le s\le t} \mathbb{S}(s) \ge a\right] \cap [\mathbb{X}(t-\tau_a) < 0]\right) \end{aligned}$$

but the second term is (because $\mathbb{X}(t)$ is independent of $\mathscr{F}_{\tau_a}$ and $[\sup_{0\le s\le t} \mathbb{S}(s) \ge a] = [\tau_a \le t] \in \mathscr{F}_{\tau_a}$):

$$\begin{aligned} P\left(\left[\sup_{0\le s\le t} \mathbb{S}(s) \ge a\right] \cap [\mathbb{X}(t-\tau_a) < 0]\right) &= \mathbb{E}\left[P\left(\left[\sup_{0\le s\le t} \mathbb{S}(s) \ge a\right] \cap [\mathbb{X}(t-\tau_a) < 0] \Big| \mathscr{F}_{\tau_a}\right)\right] \\ &= \mathbb{E}\left[1_{[\sup_{0\le s\le t} \mathbb{S}(s) \ge a]} \cdot P\left(\mathbb{X}(t-\tau_a) < 0 \Big| \mathscr{F}_{\tau_a}\right)\right] \\ &= \mathbb{E}\left[1_{[\sup_{0\le s\le t} \mathbb{S}(s) \ge a]} \cdot P([\mathbb{X}(t-\tau_a) < 0])\right] \\ &= \frac{1}{2} P\left(\sup_{0<s<t} \mathbb{S}(s) \ge a\right) \end{aligned}$$

where the $\frac{1}{2}$ comes from the fact that the probability of a Brownian motion being < 0 at any given time is $\frac{1}{2}$. Combining everything together gives that

$$P\left(\left[\sup_{0\le s\le t}\mathbb{S}(s)\ge a\right]\right)=P([\tau_a\le t])=2P([\mathbb{S}(t)\ge a])$$

Mill's ratio is that

$$\frac{\lambda}{\lambda^2+1}\phi(\lambda)<1-\Phi(\lambda)<\frac{1}{\lambda}\phi(\lambda)$$

We want to show that

$$\frac{|\mathbb{S}(\tau_n)-\mathbb{S}(n)|}{\sqrt{2n\log\log n}}\to_{\text{a.s.}}0$$

where the stopping times τ_n are such that $S_n=\sum_{k=1}^n X_k$ is embedded in $\mathbb{S}$: $S_n=\mathbb{S}(\tau_n)$. Fixing $c>0$ and $r>1$ s.t. $\frac{c^2}{r-1}\ge 1+\delta$ (for some fixed $\delta>0$), defining $\mathbb{X}(t)=\mathbb{S}(t+r^n)-\mathbb{S}(r^n)$ (which is a standard Brownian motion on $t\in[0,\infty)$), and recalling that $\mathbb{S}(t)\sim\text{Normal}(0,t)$, we have

$$P\left(\sup_{r^n\le t\le r^{n+1}}\frac{|\mathbb{S}(t)-\mathbb{S}(r^n)|}{\sqrt{2r^n\log\log(r^n)}}>c\right)\le P\left(\sup_{0\le t\le r^{n+1}-r^n}\mathbb{X}(t)>c\sqrt{2r^n\log\log(r^n)}\right)$$

$$=2P\Big(\mathbb{X}(r^n(r-1))>c\sqrt{2r^n\log\log(r^n)}\Big)=2P\Big(\sqrt{r^n(r-1)}Z>c\sqrt{2r^n\log\log(r^n)}\Big)$$

$$=2P\left(Z>c\sqrt{\tfrac{2\log\log(r^n)}{r-1}}\right)\le\tfrac{2}{\sqrt{2\pi}}\sqrt{\tfrac{r-1}{2c^2\log(n\log r)}}\exp\Big(-\tfrac{c^2}{r-1}\log n\Big)$$

$$\le\tfrac{4}{\sqrt{2\pi}}\sqrt{\tfrac{r-1}{2c^2\log n}}n^{-c^2/(r-1)}\text{ for }n\text{ large.}$$