522 Final Exam

Daniel Rui - 3/17/20

Problem 1

Let $X_1, X_2, \ldots$ be independent r.v.'s satisfying

$$X_n = \begin{cases} n^2 - 1 & \text{with probability } \frac{1}{n^2} \\ -1 & \text{with probability } 1 - \frac{1}{n^2} \end{cases}$$

and let $S_n = X_1 + \ldots + X_n$.

(a) For any $n \in \mathbb{N}$, $\mathbb{E}[X_n] = (n^2-1)\frac{1}{n^2} + (-1)(1 - \frac{1}{n^2}) = \frac{n^2-1}{n^2} - \frac{n^2-1}{n^2} = 0$.

(b) We proceed by Borel-Cantelli (similar to 521-hw10-p1): $\sum_{n=1}^{\infty} P(X_n \neq -1) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$, so $P(X_n \neq -1 \text{ i.o.}) = 0 \implies P(X_n \neq 1 \text{ finitely often}) = 1$, which means that for almost every $\omega \in \Omega$, there is some N_ω s.t. $n > N_\omega \implies X_n(\omega) = -1$. Thus, at a fixed ω, for all n large,

$$\frac{S_n}{n} = \frac{X_1 + \ldots + X_{N_\omega}}{n} + \frac{X_{N_\omega+1} + \ldots + X_n}{n} \leq \frac{N_\omega(N_\omega^2 - 1)}{n} + \frac{-1(n - N_\omega)}{n} = \frac{N_\omega^3}{n} - 1$$

which tends to -1 as $n \to \infty$. Thus, for almost every ω, $\frac{S_n(\omega)}{n} \to -1$, i.e. $\frac{S_n}{n} \to_{\text{a.s.}} -1$.

(c) Defining $\mathscr{A}_n = \sigma[X_1, \ldots, X_n]$, we have that $\mathbb{E}(S_{n+1}|\mathscr{A}_n) = \mathbb{E}(S_n + X_{n+1}|\mathscr{A}_n) =_{\text{a.s.}} S_n + \mathbb{E}[X_{n+1}] = S_n$, so $\{S_n, \mathscr{A}_n\}_{n\in\mathbb{N}}$ is a martingale. By Jensen's inequality, $\{S_n^2, \mathscr{A}_n\}_{n\in\mathbb{N}}$ is a submartingale.

(d) We want to find a ($\mathscr{A}_{n-1}$-measurable) predictable variation process $\langle S_n \rangle$ s.t. $\{S_n^2 - \langle S_n \rangle, \mathscr{A}_n\}_{n\in\mathbb{N}}$ is a martingale:

$$\begin{aligned} \mathbb{E}\big(S_{n+1}^2 - \langle S_{n+1}\rangle | \mathscr{A}_n\big) &= \mathbb{E}\big(S_n^2 + 2S_nX_{n+1} + X_{n+1}^2 - \langle S_{n+1}\rangle | \mathscr{A}_n\big) \\ &=_{\text{a.s.}} S_n^2 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}\big[X_{n+1}^2\big] - \langle S_{n+1}\rangle \\ &= S_n^2 + \mathbb{E}\big[X_{n+1}^2\big] - \langle S_{n+1}\rangle \end{aligned}$$

We want this to equal $S_n^2 - \langle S_n\rangle$, and so $\langle S_{n+1}\rangle = \mathbb{E}\big[X_{n+1}^2\big] + \langle S_n\rangle$. This implies that $\langle S_n\rangle = \mathbb{E}\big[X_n^2\big] + \ldots + \mathbb{E}\big[X_1^2\big]$ (note that this formula is still consistent for $n = 1$: $\mathbb{E}\big(S_1^2 - \langle S_1\rangle|\mathscr{A}_0\big) = \mathbb{E}\big(X_1^2 - \mathbb{E}\big[X_1^2\big]|\{\varnothing, \Omega\}\big) = 0 - 0 = S_0 - \langle S_0\rangle$).

(e) First, observe that $\mathbb{E}\big[X_n^2\big] = (n^2-1)^2\frac{1}{n^2} + 1(1 - \frac{1}{n^2}) = (n^2-1)(\frac{(n^2-1)+1}{n^2}) = n^2 - 1$ and so $\langle S_n\rangle = \sum_{k=1}^n \mathbb{E}\big[X_k^2\big] = \sum_{k=1}^n (k^2 - 1) = \frac{n(n+1)(2n+1)}{6} - n$. Taking $b_n = n^3$, we get that $\frac{\langle S_n\rangle}{b_n}$ converges everywhere (and hence in probability) to $\frac{1}{3}$.

Problem 2

Let $Y_1, \ldots, Y_n$ be i.i.d. with $\mathbb{E}[Y_i] = 0$ and $\text{Var}[Y_i] = \sigma^2$, and set $X_{ni} = a_{ni}Y_i$ for $i \in \{1, \ldots, n\}$ and constants a_{ni}. As always, $S_n = \sum_{i=1}^n X_{ni}$.

(a) The expectation and variance are respectively $\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_{ni}] = \sum_{i=1}^n a_{ni}\mathbb{E}[Y_i] = 0$ and $\text{Var}[S_n] = \sum_{i=1}^n \text{Var}[X_{ni}] = \sum_{i=1}^n a_{ni}^2 \mathbb{E}[Y_i] = \sigma^2 \sum_{i=1}^n a_{ni}^2 =: \sigma_n^2$ (linearity of variance by independence of the X_{ni}).

(b) Let $A_n^2 = \frac{\max_{1\le i\le n}|a_{ni}|^2}{\sum_{i=1}^n a_{ni}^2} = \frac{\sigma^2}{\sigma_n^2}\max_{1\le i\le n}|a_{ni}|^2$. We want to show that if $A_n^2 \to 0$, then $\frac{S_n}{\sigma_n} \to_d Z \sim \text{Normal(0,1)}$. This immediately rings a bell for the Lindenberg-Feller CLT, which tells us that for X'_{ni} s.t. $\mathbb{E}[X'_{ni}] = 0$ and $\sum_{i=1}^n \text{Var}[X'_{ni}] = 1$, $L_n(\epsilon) := \sum_{i=1}^n \mathbb{E}\big[(X'_{ni})^2 \mathbb{1}_{[|X'_{ni}|>\epsilon]}\big] \to 0$ for any $\epsilon > 0 \implies \sum_{i=1}^n X'_{ni} \to_d Z$.

In our case, take $X'_{ni} = \frac{X_{ni}}{\sigma_n}$ and $S'_n = \sum_{i=1}^n X'_{ni} = \frac{S_n}{\sigma_n} \implies \mathbb{E}[X'_{ni}] = 0, \text{Var}[S'_n] = \frac{\sigma_n^2}{\sigma_n^2} = 1$. Then,

$$\begin{aligned}
L_n(\epsilon) &= \sum_{i=1}^n \frac{1}{\sigma_n^2}\mathbb{E}\big[X_{ni}^2 \mathbb{1}_{[|X_{ni}|>\epsilon\sigma_n]}\big] \\
&= \frac{1}{\sigma_n^2}\sum_{i=1}^n a_{ni}^2 \mathbb{E}\left[Y_i^2 \mathbb{1}_{\left[|Y_i|>\frac{\epsilon\sigma_n}{|a_{ni}|}\right]}\right] \\
&\le \frac{1}{\sigma_n^2}\sum_{i=1}^n a_{ni}^2 \mathbb{E}\left[Y_i^2 \mathbb{1}_{\left[|Y_i|>\frac{\epsilon\sigma_n}{\max_{1\le i\le n}|a_{ni}|}\right]}\right] \\
&= \frac{1}{\sigma_n^2}\mathbb{E}\left[Y_1^2 \mathbb{1}_{\left[|Y_1|>\frac{\epsilon\sigma_n}{\max_{1\le i\le n}|a_{ni}|}\right]}\right]\sum_{i=1}^n a_{ni}^2 \\
&= \frac{1}{\sigma^2}\mathbb{E}\Big[Y_1^2 \mathbb{1}_{[|Y_1|>\frac{\epsilon\sigma}{A_n}]}\Big]
\end{aligned}$$

We assumed that $A_n^2 \to 0 \implies P(|Y_1| > \frac{\epsilon\sigma}{A_n}) = P(Y_1^2 > \frac{\epsilon^2\sigma^2}{A_n^2}) \le \frac{\mathbb{E}[Y_1]}{\epsilon^2\sigma^2/A_n^2} = \frac{\sigma^2 A_n^2}{\epsilon^2\sigma^2}$ (by Markov) $\to 0$, and we know that $Y_1^2\mathbb{1}_{[|Y_1|>\frac{\epsilon\sigma}{A_n}]} \le Y_1^2$, which is integrable ($\mathbb{E}\big[Y_1^2\big] = \sigma^2 < \infty$) and so by the DCT,

$$\lim_{n\to\infty}\mathbb{E}\Big[Y_1^2\mathbb{1}_{[|Y_1|>\frac{\epsilon\sigma}{A_n}]}\Big] = \mathbb{E}\Big[\lim_{n\to\infty}\mathbb{E}\Big[Y_1^2\mathbb{1}_{[|Y_1|>\frac{\epsilon\sigma}{A_n}]}\Big]\Big] = 0$$

Thus, for any $\epsilon > 0$, we see that $L_n(\epsilon) \to 0$, and so by Lindenberg-Feller, $S'_n = \frac{S_n}{\sigma_n} \to_d Z$.

(c) As an explicit example, take $a_{ni} = \frac{1}{\sqrt{n}}(\frac{i}{n})^\alpha$ for $\alpha \in \mathbb{R}$, so $\sigma_n^2 = \sigma^2\sum_{i=1}^n a_{ni}^2 = \frac{1}{n}\sum_{i=1}^n(\frac{i}{n})^{2\alpha}$. This is a Riemann sum (MSE), and so

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\frac{i}{n}\right)^{2\alpha} = \int_0^1 t^{2\alpha}\,dt = \left(\frac{t^{2\alpha+1}}{2\alpha+1}\right)\Big|_0^1 = \begin{cases}\frac{1}{2\alpha+1} & \text{if } \alpha > -\frac{1}{2}\\ \infty & \text{otherwise}\end{cases}$$

Also, $\max_{1\le i\le n}|a_{ni}|^2 = \frac{1}{n}(\frac{n}{n})^{2\alpha} = \frac{1}{n}$, so for all α, $A_n^2 \to 0$ (and so (b) applies). We want to find all the values α s.t. $S_n \to v_\alpha Z \sim \text{Normal}(0, v_\alpha^2)$ for $v_\alpha^2 < \infty$. I claim that $\sigma_n \to v_\alpha$ for some $v_\alpha \ne 0, \pm\infty$ (i.e. $\frac{\sigma_n}{v_\alpha} \to 1$) $\iff S_n \to_d v_\alpha Z$:

Proof: perhaps a little non-rigorously, ($\implies$): $\frac{\sigma_n}{v_\alpha} \to 1$ and $\frac{S_n}{\sigma_n} \to_d Z \implies \frac{S_n}{\sigma_n}\frac{\sigma_n}{v_\alpha} \to Z \implies S_n \to v_\alpha Z$, and ($\impliedby$): $\frac{S_n}{v_\alpha} \to_d Z$ and $\frac{S_n}{\sigma_n} \to_d Z \implies \frac{S_n}{v_\alpha}\frac{\sigma_n}{\sigma_n} = \frac{S_n}{v_\alpha}\frac{v_\alpha}{\sigma_n} \to_d Z \implies \frac{\sigma_n}{v_\alpha} \to 1$. See here (MSE) and here (Billingsley) for a more rigorous treatment.

Thus finally, $\alpha > -\frac{1}{2} \iff S_n \to_d v_\alpha Z$, where $v_\alpha^2 = \frac{1}{2\alpha+1}$.

Problem 3

Consider a bounded continuous function f on $[0, \infty)$, and define $L(\lambda) := \int_0^\infty e^{-\lambda x} f(x)\,dx$ for $\lambda \in (0, \infty)$. Let $X_1, X_2, \ldots$ be i.i.d $\text{Exp}(\lambda)$ so $\mathbb{E}[X_i] = \frac{1}{\lambda}, \text{Var}[X_i] = \frac{1}{\lambda^2}$, and let $S_n = \sum_{i=1}^n X_i$.

(a) The distribution of S_n is $\text{Gamma}(n, \lambda)$ (equiv. $\text{Erlang}(n, \lambda)$) (Wiki), w/ p.d.f. $\frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} 1_{(0,\infty)}$.

(b) We want to show that

$$\mathbb{E}[f(S_n)] = (-1)^{n-1} \frac{\lambda^n L^{(n-1)}(\lambda)}{(n-1)!}$$

Observe that

$$\mathbb{E}[f(S_n)] = \int_0^\infty f(x) \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}\,dx = \frac{\lambda^n}{(n-1)!} \int_0^\infty f(x) x^{n-1} e^{-\lambda x}\,dx$$

and so showing that $\int_0^\infty f(x) x^{n-1} e^{-\lambda x}\,dx = (-1)^{n-1} L^{(n-1)}(\lambda)$ will suffice. We proceed by induction: base case $n = 1$: $\int_0^\infty f(x) e^{-\lambda x}\,dx = L(\lambda)$ ✓

Now assume n works and so $\int_0^\infty f(x) x^{n-1} e^{-\lambda x}\,dx = (-1)^{n-1} L^{(n-1)}(\lambda)$. Differentiating both sides w.r.t. λ and pulling it under the integral (which we can do because $f(x) x^{n-1} e^{-\lambda x}$ is continuous on $\{(x,t) : [0,\infty) \times (0,\infty)\}$, and the λ-partial is also continuous on that region), yielding

$$\int_0^\infty f(x)(-1) x^n e^{-\lambda x}\,dx = (-1)^{n-1} L^{(n)}(\lambda)$$

and rearranging the (-1) gives the desired result.

(c) With parameter $\frac{n}{y}$ we have

$$\mathbb{E}_{n/y}[f(S_n)] = \int_0^\infty f(x) \frac{(\frac{n}{y})^n x^{n-1} e^{-(n/y)x}}{(n-1)!}\,dx$$

while with parameter 1,

$$\mathbb{E}_1\left[f\left(\frac{yS_n}{n}\right)\right] = \int_0^\infty f\left(\frac{y}{n}x\right) \frac{x^{n-1} e^{-x}}{(n-1)!}\,dx = \int_0^\infty f(u) \frac{(\frac{n}{y}u)^{n-1} e^{-(n/y)u}}{(n-1)!} \frac{n}{y}\,du$$

(where we made the substitution $u = \frac{y}{n}x \implies du = \frac{y}{n}\,dx$). This matches with $\mathbb{E}_{n/y}[f(S_n)]$ from above, so $\mathbb{E}_{n/y}[f(S_n)] = \mathbb{E}_1[f(\frac{yS_n}{n})]$

(d) Finally, observe the following:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}_{n/y}[f(S_n)] &= \lim_{n\to\infty} \mathbb{E}_1\left[f\left(\frac{yS_n}{n}\right)\right] = \lim_{n\to\infty} \int_\Omega f\left(y\frac{S_n}{n}\right) dP = \int_\Omega \lim_{n\to\infty} f\left(y\frac{S_n}{n}\right) dP \\ &= \int_\Omega f\left(y \lim_{n\to\infty} \frac{S_n}{n}\right) dP = \int_\Omega f(y \cdot \mathbb{E}_1[X_1])\,dP = \int_\Omega f(y)\,dP = f(y)\end{aligned}$$

where we pulled the limit inside the integral by the DCT (f is bounded $\implies$ DCT applies), inside f by continuity, and used $\mathbb{E}[X_1] = 1 < \infty \implies \frac{S_n}{n} \to_{\text{a.s.}} 1$ (SLLN). Thus, we have the following formula for f in terms of the derivatives of its Laplace transform:

$$f(y) = \lim_{n\to\infty} (-1)^{n-1} \frac{(\frac{n}{y})^n L^{(n-1)}(\frac{n}{y})}{(n-1)!}$$

Problem 4

Let $Z_1, Z_2, \dots$ be i.i.d. Normal(0,1) and $S_n = Z_1 + ... + Z_n$ ($\implies S_n \sim \text{Normal}(0, n)$; see Wiki) , and define $Y_n = e^{aS_n - bn}$.

(a) For $r \geq 1 : Y_n \to_r 0 \iff r < \frac{2b}{a^2}$: using the MGF $\mathbb{E}\left[e^{tS_n}\right] = e^{\mu t + \sigma^2 t^2/2} = e^{nt^2/2}$ (Wiki), we have

$$\int_\Omega |Y_n - 0|^r \, dP = \int_\Omega e^{raS_n - rbn} \, dP = \frac{1}{e^{rbn}} \mathbb{E}\left[e^{raS_n}\right] = \frac{1}{e^{rbn}} e^{nr^2a^2/2} = e^{nr(\frac{ra^2}{2} - b)}$$

which goes to 0 iff $r < \frac{2b}{a^2}$ (i.e. iff the exponent is negative).

(b) For $b = \frac{a^2}{2}$,

$$Y_n = \frac{e^{aS_n}}{e^{a^2n/2}} = \prod_{i=1}^n \frac{e^{aZ_i}}{e^{a^2/2}}$$

where $X_i := \frac{e^{aZ_i}}{e^{a^2/2}} \implies \mathbb{E}[X_i] = \frac{e^{0 \cdot t + 1 \cdot a^2/2}}{e^{a^2/2}} = 1$, and i.i.d. because the Z_i were.

(c) Kakutani's martingale theorem tells us that for independent non-negative mean-1 $X_1, X_2, \dots$, $\{M_n, \mathscr{A}_n\}_{n\geq1}$ is a mean-1 martingale (where $M_n = \prod_{i=1}^n X_i$), and $M_n \to_{\text{a.s.}} M_\infty \in \mathscr{L}_1$ (this part is actually given by the (sub)-martingale convergence theorem). Furthermore, among many other things, it says that if $\prod_{i=1}^\infty \mathbb{E}\left[X_i^{1/2}\right]$ is NOT > 0, then necessarily $M_\infty = 0$ almost surely. Well, our X_i (defined above) satisfy these conditions, but have

$$\mathbb{E}\left[X_n^{1/2}\right] = \mathbb{E}\left[\frac{e^{aZ_n/2}}{e^{a^2/4}}\right] = \frac{e^{0 + \frac{(a^2/4)\cdot 1}{2}}}{e^{a^2/4}} = e^{-a^2/8} \implies \prod_{i=1}^\infty \mathbb{E}\left[X_i^{1/2}\right] = 0$$

and so $Y_n = M_n \to_{\text{a.s.}} 0$.

Problem 5

(a) Suppose that Y is a r.v. with values in $[-c, c]$ with $\mathbb{E}[Y] = 0$. For any $\theta \in \mathbb{R}$, $f_\theta(z) = e^{\theta z}$ for $z \in [-c, c]$ is convex, and so $f_\theta(y) \leq \ell_\theta(y)$ on $[-c, c]$ where $\ell_\theta(y)$ is the line between $(-c, f_\theta(-c))$ and $(c, f_\theta(c))$. Then, we have $\mathbb{E}[f_\theta(Y)] \leq \mathbb{E}[\ell_\theta(Y)] = \ell_\theta(\mathbb{E}[Y])$ (by linearity of expectation) $= \ell_\theta(0)$. 0 is the average of $-c$ and c, and so $\ell_\theta(0)$ will be the average of $f_\theta(-c)$ and $f_\theta(c)$, i.e. $\mathbb{E}\left[e^{\theta Y}\right] \leq \frac{e^{-\theta c} + e^{\theta c}}{2} = \cosh(\theta c)$.

Furthermore, noting that $2^n n! = (2n)(2n-2)\cdots(2) \leq (2n)(2n-1)\cdots(2)(1) = (2n)!$, we have

$$\cosh(x) = \frac{\sum_{n=1}^\infty (-1)^n \frac{x^n}{n!} + \sum_{n=1}^\infty \frac{x^n}{n!}}{2} = \sum_{n=1}^\infty \frac{x^{2n}}{(2n)!} \leq \sum_{n=1}^\infty \frac{x^{2n}}{2^n n!} = e^{x^2/2}$$

and so

$$\mathbb{E}\big[e^{\theta Y}\big] \le \cosh(\theta c) \le e^{\theta^2 c^2/2}$$

(b) Suppose we have a martingale $\{M_n, \mathscr{A}_n\}_{n\ge 0}$ with $M_0 = 0$ satisfying $|M_n - M_{n-1}| \le c_n$ for some constants $c_n > 0$ for all $n \ge 1$. By Doob's maximal submartingale inequality, we know that

$$P\left(\max_{0\le k\le n} M_k \ge x\right) \le \inf_{r>0} \frac{\mathbb{E}\big[e^{rM_n}\big]}{e^{rx}}$$

Now we could use the triangle inequality to get that $|M_n| \le \sum_{i=1}^n c_i$, but this bound is not sharp enough. Instead, let's take the following approach (notice that we are not given the independence of the M_n):

$$\begin{aligned}\mathbb{E}\big[e^{rM_n}\big] &= \mathbb{E}\Big[e^{r(M_n - M_{n-1})} e^{rM_{n-1}}\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\Big(e^{r(M_n - M_{n-1})} e^{rM_{n-1}}\Big|\mathscr{A}_{n-1}\Big)\Big] \\ &= \mathbb{E}\Big[e^{rM_{n-1}}\mathbb{E}\Big(e^{r(M_n - M_{n-1})}\Big|\mathscr{A}_{n-1}\Big)\Big]\end{aligned}$$

We can't directly use (a), but we can quickly prove a conditional version of that inequality:

$$\begin{aligned}\mathbb{E}\Big(f_\theta(M_n - M_{n-1})\Big|\mathscr{A}_{n-1}\Big) &\le_{\text{a.s.}} \mathbb{E}\Big(\ell_\theta(M_n - M_{n-1})\Big|\mathscr{A}_{n-1}\Big) \\ &=_{\text{a.s.}} \ell_\theta(\mathbb{E}(M_n - M_{n-1}|\mathscr{A}_{n-1})) =_{\text{a.s.}} \ell_\theta(0) = \cosh(\theta c_n)\end{aligned}$$

Thus, continuing our chain of equalities above,

$$\begin{aligned}\mathbb{E}\big[e^{rM_n}\big] &= \mathbb{E}\Big[e^{rM_{n-1}}\mathbb{E}\Big(e^{r(M_n - M_{n-1})}\Big|\mathscr{A}_{n-1}\Big)\Big] \\ &\le \mathbb{E}\big[e^{rM_{n-1}}\cosh(rc_n)\big] \\ &\le \mathbb{E}\Big[e^{rM_{n-1}}\exp\Big(\tfrac{r^2c_n^2}{2}\Big)\Big] = \exp\Big(\tfrac{r^2c_n^2}{2}\Big)\mathbb{E}\big[e^{rM_{n-1}}\big]\end{aligned}$$

Continuing inductively, we get that

$$\mathbb{E}\big[e^{rM_n}\big] \le \exp\Big(\tfrac{r^2c_n^2}{2}\Big)\exp\Big(\tfrac{r^2c_{n-1}^2}{2}\Big)\cdots\exp\Big(\tfrac{r^2c_1^2}{2}\Big)\mathbb{E}\big[e^{rM_0}\big] = \exp\Big(\tfrac{r^2}{2}\textstyle\sum_{i=1}^n c_i^2\Big)$$

Plugging this into Doob's inequality, we have

$$P\left(\max_{0\le k\le n} M_k \ge x\right) \le \inf_{r>0}\frac{\exp\Big(\frac{r^2}{2}\sum_{i=1}^n c_i^2\Big)}{e^{rx}}$$

To minimize the RHS, let's calculate the first and second derivatives w.r.t r:

$$\begin{aligned}\partial_r[\text{RHS}] &= \left(\frac{2r}{2}\sum_{i=1}^n c_i^2 - x\right)\frac{\exp\Big(\frac{r^2}{2}\sum_{i=1}^n c_i^2\Big)}{e^{rx}} \\ \partial_{rr}[\text{RHS}] &= \left(\sum_{i=1}^n c_i^2\right)\frac{\exp\Big(\frac{r^2}{2}\sum_{i=1}^n c_i^2\Big)}{e^{rx}} + \left(r\sum_{i=1}^n c_i^2 - x\right)^2\frac{\exp\Big(\frac{r^2}{2}\sum_{i=1}^n c_i^2\Big)}{e^{rx}}\end{aligned}$$

The first derivative is 0 at $r = \frac{x}{\sum_{i=1}^n c_i^2}$ and the second derivative is always ≥ 0, so we've found r that minimizes the RHS. Plugging it in yields

$$P\left(\max_{0\leq k\leq n} M_k \geq x\right) \leq \exp\left(\frac{x^2}{2\sum_{i=1}^n c_i^2} - \frac{x^2}{\sum_{i=1}^n c_i^2}\right) = \exp\left(-\frac{x^2}{2\sum_{i=1}^n c_i^2}\right)$$

Problem 6

Let $X_1, X_2, \dots$ be i.i.d. Poisson(1), $S_n = X_1 + ... + X_n$, and $Z_n = \dfrac{S_n - n}{\sqrt{n}}$.

(a) The distribution of S_n is Poisson(n) (Wiki), with p.d.f. $\dfrac{\lambda^i e^{-\lambda}}{i!}$ for $i \in \mathbb{N}_0 := \{0, 1, \dots\}$.

(b) The expectation of Z_n^- is

$$\begin{aligned}\mathbb{E}\big[Z_n^-\big] &= \frac{1}{\sqrt{n}}\mathbb{E}\big[(S_n - n)^-\big] = \frac{1}{\sqrt{n}}\sum_{i=0}^n |i-n|\cdot\frac{n^i e^{-n}}{i!} = \frac{e^{-n}}{\sqrt{n}}\sum_{i=0}^n (n-i)\frac{n^i}{i!} \\ &= \frac{e^{-n}}{\sqrt{n}}\sum_{i=0}^n \frac{n^{i+1} - in^i}{i!} = \frac{e^{-n}}{\sqrt{n}}\left(\frac{n^{0+1} - 0}{0!} + \sum_{i=1}^n\left(\frac{n^{i+1}}{i!} - \frac{n^i}{(i-1)!}\right)\right)\end{aligned}$$

which telescopes to leave just $\mathbb{E}[Z_n^-] = \dfrac{e^{-n}}{\sqrt{n}}\dfrac{n^{n+1}}{n!}$.

(c) Take $X_i' = X_i - 1$, so $\mathbb{E}[X_i'] = 0$ and $\text{Var}[X_i'] = 1$. The classical CLT gives that $\frac{S_n'}{\sqrt{n}} = \frac{S_n - n}{\sqrt{n}} = Z_n \to_d Z \sim$ Normal(0,1).

(d) Per the definition of convergence in distribution, $Z_n \to_d Z \iff \mathbb{E}[f(Z_n)] \to \mathbb{E}[f(Z)]$ for any $f \in C_b(\mathbb{R})$. In particular, consider the sequence of functions $f_m(x) := m \cdot 1_{(-\infty,-m)} + (-x)\cdot 1_{[-m,0]} + 0 \cdot 1_{(0,\infty)}$, all of which are in $C_b(\mathbb{R})$. Let us denote $f_\infty := (-x)\cdot 1_{(-\infty,0]} + 0\cdot 1_{(0,\infty)}$ (which unfortunately is not in $C_b(\mathbb{R})$). Now,

$$\begin{aligned}\big|\mathbb{E}\big[Z_n^-\big] - \mathbb{E}\big[Z^-\big]\big| &= \big|\mathbb{E}[f_\infty(Z_n)] - \mathbb{E}[f_\infty(Z)]\big| \\ &\leq \big|\mathbb{E}[f_\infty(Z_n)] - \mathbb{E}[f_m(Z_n)]\big| + \big|\mathbb{E}[f_m(Z_n)] - \mathbb{E}[f_m(Z)]\big| + \big|\mathbb{E}[f_m(Z)] - \mathbb{E}[f_\infty(Z)]\big|\end{aligned}$$

Let's look at the first term first:

$$\big|\mathbb{E}[f_\infty(Z_n)] - \mathbb{E}[f_m(Z_n)]\big| = \big|\mathbb{E}\big[(-Z_n - m)1_{[Z_n<-m]}\big]\big| = \big|\mathbb{E}\big[|Z_n + m|\cdot 1_{[Z_n<-m]}\big]\big|$$

The r.v. $|Z_n + m|$ is integrable, because $|Z_n + m| \leq |Z_n| + m \leq \max\{1, |Z_n|^2\} + m \leq Z_n^2 + 1 + m$ (MSE), and we know $\mathbb{E}\big[Z_n^2 + 1 + m\big] = 1 + 1 + m < \infty$. Furthermore, $P(Z_n < -m) \leq P(|Z_n| > m) = P(Z_n^2 > m^2) \leq \frac{1}{m^2} \to 0$ as $m \to \infty$ (Markov). Thus, by the DCT,

$$\lim_{m\to\infty}\mathbb{E}\big[|Z_n + m|\cdot 1_{[Z_n<-m]}\big] = \mathbb{E}\Big[\lim_{m\to\infty}|Z_n + m|\cdot 1_{[Z_n<-m]}\Big] = 0$$

and so the first term gets arbitrarily small for large enough m. The third term is dealt with in exactly the same way. The second term gets arbitrarily small for large enough n because

$\mathbb{E}[f_m(Z_n)] \to \mathbb{E}[f_m(Z)]$ by weak convergence. Thus, $\mathbb{E}[Z_n^-] \to \mathbb{E}[Z^-] = \frac{1}{2}\sqrt{\frac{2}{\pi}} = \frac{1}{\sqrt{2\pi}}$ (Wiki).

(e) Finally, putting parts (b) and (d) together we have

$$\frac{e^{-n}}{\sqrt{n}} \frac{n^{n+1}}{n!} \to \frac{1}{\sqrt{2\pi}} \implies \frac{\sqrt{2\pi n} n^n e^{-n}}{n!} \to 1 \implies n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

522 Homework 8

Daniel Rui - 3/11/20

Problem 1

For real-valued $\{X, X_n\}$, we want to prove that $F_n(x) = P(X_n \le x) \to P(X \le x) = F(x)$ for all x s.t. $P(X = x) = 0$ (i.e. for all x in the continuity set C_F) $\iff \mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in C^\infty(\mathbb{R})$.

($\implies$) Fix any $f \in C^\infty(\mathbb{R})$ and $\epsilon > 0$. Now observe that for a non-negative r.v. X s.t. $\mathbb{E}[X] < \infty$, and g s.t. $||g||_\infty$ and $||g'||_\infty$ are both $< \infty$ (where $||g||_\infty := \inf\{c \in \mathbb{R} : |g| \le c \text{ a.e.}\}$), we have that

$$\begin{aligned}\int_0^\infty g'(x)P(X > x)\,dP &= \int_0^\infty \int_\Omega g'(x)1_{[X>x]}\,dP\,dx = \int_\Omega \int_0^\infty g'(x)1_{[X>x]}\,dx\,dP \\ &= \int_\Omega \int_0^X g'(x)\,dx\,dP = \int_\Omega g(X) - g(0)\,dP = \mathbb{E}[g(X)] - g(0)\end{aligned}$$

where the integral interchange is justified because

$$\int_\Omega \int_0^\infty |g' 1_{[X>x]}|\,dx\,dP \le \int_\Omega \int_0^X ||g'||_\infty\,dx\,dP = ||g'||_\infty \mathbb{E}[X] < \infty$$

In general, we have that

$$\begin{aligned}\int_b^\infty g'(x)P(X > x)\,dx &= \int_0^\infty g'(x) \int_\Omega 1_{[X>x]}\,dx\,dP \\ &= \int_\Omega 1_{[X>b]} \int_b^X g'(x)\,dx\,dP = \mathbb{E}\big[g(X)1_{[X>b]}\big] - g(b)P(X > b)\end{aligned}$$

and

$$\begin{aligned}\int_{-\infty}^a g'(x)P(X \le x)\,dx &= \int_{-\infty}^a g'(x) \int_\Omega 1_{[X\le x]}\,dx\,dP \\ &= \int_\Omega 1_{[X\le a]} \int_X^a g'(x)\,dx\,dP = g(a)P(X \le a) - \mathbb{E}\big[g(X)1_{[X\le a]}\big]\end{aligned}$$

where the integral interchanges are justified in much the same way as in the $X \ge 0$ case. Setting $a = b = 0$ and subtracting the bottom from the top, we get

$$\mathbb{E}[g(X)] = g(0) + \int_0^\infty g'(x)P(X > x)\,dx - \int_{-\infty}^0 g'(x)P(X \le x)\,dx$$

and so

$$\begin{aligned}\mathbb{E}\big[g(X)1_{[a<X\le b]}\big] &= \mathbb{E}[g(X)] - \mathbb{E}\big[g(X)1_{[X\le a]}\big] - \mathbb{E}\big[g(X)1_{[X>b]}\big] \\ &= g(0) - g(a)P(X \le a) - g(b)P(X > b) \\ &\quad + \int_0^b g'(x)P(X > x)\,dx - \int_a^0 g'(x)P(X \le x)\,dx\end{aligned}$$

Now in the context of this problem, $f \in C^\infty(\mathbb{R})$ (bounded C^∞ functions on $\mathbb{R}$) obviously satisfies the above conditions $||g||_\infty, ||g'||_\infty < \infty$, so we can say

$$\begin{aligned}|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &= \left| \int_0^\infty f'(x)\Big(1 - F_n(x) - (1 - F(x))\Big)\, dx - \int_{-\infty}^0 f'(x)\Big(F_n(x) - F(x)\Big)\, dx \right| \\ &= \left| \int_0^\infty f'(x)\Big(F(x) - F_n(x)\Big)\, dx \right| + \left| \int_{-\infty}^0 f'(x)\Big(F_n(x) - F(x)\Big)\, dx \right|\end{aligned}$$

We somehow need to make both these integrals arbitrarily small (e.g. by pulling the limit inside), but at this point, there's not really a way forward. However, as is often the case when dealing with infinite bounds, we should look to restrict ourselves to a compact set. A nice one to consider would be $[a, b]$ where $a, b \in C_F \cap \bigcap_{n=1}^\infty C_{F_n}$ (still $\mathbb{R}$ minus a null set), $F(a) < \epsilon$ and $F(b) > 1 - \epsilon$ (such points exist because $F(-\infty+) = 0$, $F(\infty-) = 1$, and F is increasing and thus is only discontinuous on a null set). Then,

$$\begin{aligned}|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &= \left| \int_\mathbb{R} f\, dF_n - \int_\mathbb{R} f\, dF \right| = \left| \int_\mathbb{R} f\, d(F_n - F) \right| \\ &= \left| \int_{(-\infty,a]} f\, d(F_n - F) + \int_{(a,b]} f\, d(F_n - F) + \int_{(b,\infty)} f\, d(F_n - F) \right| \\ &\le \left| ||f||_\infty \Big(\int_{(-\infty,a]} d(F_n - F) + \int_{(b,\infty)} d(F_n - F) \Big) + \int_{(a,b]} f\, d(F_n - F) \right| \\ &\le \left| ||f||_\infty \Big((F_n - F)(a) - (F_n - F)(b) \Big) + \int_{(a,b]} f\, d(F_n - F) \right|\end{aligned}$$

where

$$\begin{aligned}\int_{(a,b]} f\, d(F_n - F) &= \int_{(a,b]} f\, dF_n - \int_{(a,b]} f\, dF = \int_{X_n^{-1}(a,b]} f(X_n)\, dP - \int_{X^{-1}(a,b]} f(X)\, dP \\ &= \mathbb{E}\big[f(X_n) 1_{[a < X_n \le b]}\big] - \mathbb{E}\big[f(X) 1_{[a < X \le b]}\big] \\ &= f(a)\Big(P(X \le a) - P(X_n \le a)\Big) + f(b)\Big(P(X > b) - P(X_n > b)\Big) \\ &\quad + \int_0^b f'(x)\Big(P(X_n > x) - P(X > x)\Big) + \int_a^0 f'(x)\Big(P(X \le x) - P(X_n \le x)\Big) \\ &= f(a)\Big((F_n - F)(a)\Big) + f(b)\Big((F - F_n)(b)\Big) \\ &\quad + \int_0^b f'(x)\Big((F_n - F)(x)\Big)\, dx + \int_a^0 f'(x)\Big((F - F_n)(x)\Big)\, dx\end{aligned}$$

Now as $n \to \infty$, $(F_n - F)(a)$ and $(F - F_n)(b)$ both go to 0 because we chose a, b to be in the continuity sets of the distribution functions. The integral terms (i.e. $\int_0^b f'(x)(F_n - F)(x)\, dx + \int_a^0 f'(x)(F - F_n)(x)\, dx$) go to 0 by the DCT: $|f'(x)(F_n - F)(x)| \le 2||f'||_\infty$ (because $|F_n|, |F| \le 1$) which yields a finite integral over bounded sets (e.g. $[0, b]$ and $[a, 0]$). Thus, $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \to 0$, and so $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

($\Longleftarrow$) It would be very easy if we had access to $f_t(x) = 1_{(-\infty,t]}(x)$ because then we could just say $\mathbb{E}[f_t(X_n)] = P(X_n \leq t) \to \mathbb{E}[f_t(X)] = P(X \leq t)$. We can't do this, but we can do something very similar; consider the $C^\infty(\mathbb{R})$ function

$$\psi(x) = \frac{\int_x^1 e^{-\frac{1}{t(1-t)}}\, dt}{\int_0^1 e^{-\frac{1}{t(1-t)}}\, dt} \text{ for } 0 \leq x \leq 1, \quad = 1 \text{ for } x \leq 0, \text{ and } \; = 0 \text{ for } x \geq 1$$

Then, $\psi(\frac{x-t}{\epsilon})$ is very close to $f_t(x)$ in the sense that

$$f_t(x) \leq \psi\left(\frac{x-t}{\epsilon}\right) \leq f_t(x-\epsilon) = f_{t+\epsilon}(x)$$

Thus, defining $\psi_{t,\epsilon}(x) = \psi(\frac{x-t}{\epsilon})$, we have that

$$\limsup_{n\to\infty} F_n(t) = \limsup_{n\to\infty} \mathbb{E}[f_t(X_n)] \leq \limsup_{n\to\infty} \mathbb{E}[\psi_{t,\epsilon}(X_n)] = \mathbb{E}[\psi_{t,\epsilon}(X)] \leq \mathbb{E}[f_{t+\epsilon}(x)] = F(t+\epsilon)$$

and

$$\liminf_{n\to\infty} F_n(t) = \liminf_{n\to\infty} \mathbb{E}[f_t(X_n)] \geq \liminf_{n\to\infty} \mathbb{E}[\psi_{t-\epsilon,\epsilon}(X_n)] = \mathbb{E}[\psi_{t-\epsilon,\epsilon}(X)] \geq \mathbb{E}[f_{t-\epsilon}(x)] = F(t-\epsilon)$$

which when put together, becomes

$$F(t-\epsilon) \leq \liminf_{n\to\infty} F_n(t) \leq \limsup_{n\to\infty} F_n(t) \leq F(t+\epsilon)$$

For all $t \in C_F$, the left and right sides go to $F(t)$, and so for all $t \in C_F$, the $F_n(t)$ have a limit, and that limit equals $F(t)$.

Problem 2

Suppose that $\log X \sim$ Normal(0,1).

(a) We know that

$$P(X \leq x) = P(\log X \leq \log x) = \int_{-\infty}^{\log x} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt = \int_{-\infty}^{x} \frac{e^{-(\log u)^2/2}}{\sqrt{2\pi}} \frac{1}{u}\, du$$

(by making the substitution $u = e^t \iff t = \log u$), and so $f_X(u) = \frac{e^{-(\log u)^2/2}}{u\sqrt{2\pi}}$.

(b) Now if we have a random variable Y_a with density $f_a(y) = f_X(y)(1 + a\sin(2\pi \log y))$, we want to show that $\mathbb{E}\left[X^k\right] = \mathbb{E}\left[Y_a^k\right]$ for all integers $k \geq 1$ and $a \in [-1, 1]$ ($|a| \leq 1$ because densities are ≥ 0). In order for this to be true, we must have that

$$\int_0^\infty x^k f_X(x)\, dx = \int_0^\infty x^k f_X(x)(1 + a\sin(2\pi \log x))\, dx$$

This is equivalent to showing that

$$\int_0^\infty ax^k f_X(x)\sin(2\pi\log x)\,dx = \frac{a}{\sqrt{2\pi}}\int_0^\infty \frac{x^k}{x}e^{-(\log x)^2/2}\sin(2\pi\log x)\,dx$$
$$= \frac{a}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{ku}e^{-u^2/2}\sin(2\pi u)\,du = 0$$

which is trivial because $e^{ku-u^2/2}$ is even centered at k (because $ku - u^2/2$ is even centered at u), and $\sin(2\pi k)$ is odd centered at k, leaving us with an integrand which is is odd centered at k, yielding an integral of 0.

Problem 3

Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{W}$ be random vectors $\in \mathbb{R}^k$ (i.e. $\mathbf{X} = (X_1, \ldots, X_k)$) s.t. $\mathbf{X}$ and $\mathbf{Y}$ are independent and $\mathbf{X}$ and $\mathbf{W}$ are independent, and where $\mathbb{E}\big[|\mathbf{Y}|^3\big], \mathbb{E}\big[|\mathbf{W}|^3\big] < \infty$,

$$\mathbb{E}[\mathbf{Y}] = \begin{bmatrix}\mathbb{E}[Y_1]\\ \vdots \\ \mathbb{E}[Y_k]\end{bmatrix} = \begin{bmatrix}\mathbb{E}[W_1]\\ \vdots \\ \mathbb{E}[W_k]\end{bmatrix} = \mathbb{E}[\mathbf{W}],$$

and

$$\mathrm{Cov}[\mathbf{Y}] = \begin{bmatrix}\mathrm{Cov}[Y_1,Y_1] & \ldots & \mathrm{Cov}[Y_1,Y_k]\\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_k,Y_1] & \ldots & \mathrm{Cov}[Y_k,Y_k]\end{bmatrix} = \begin{bmatrix}\mathrm{Cov}[W_1,W_1] & \ldots & \mathrm{Cov}[W_1,W_k]\\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[W_k,W_1] & \ldots & \mathrm{Cov}[W_k,W_k]\end{bmatrix} = \mathrm{Cov}[\mathbf{W}]$$

which means that $\mathbb{E}[Y_iY_j] = \mathbb{E}[W_iW_j]$ for all $i, j \in \{1, \ldots, k\}$ because

$$\mathrm{Cov}[Y_i, Y_j] = \mathbb{E}[(Y_i - \mathbb{E}[Y_i])(Y_j - \mathbb{E}[Y_j])] = \mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\mathbb{E}[Y_j]$$

Now for any $f : \mathbb{R}^k \to \mathbb{R}$ in $C^3(\mathbb{R}^k)$ (i.e. f, its 1st, 2nd, and 3rd-order partials all exist and are continuous everywhere), Taylor's theorem says that

$$f(\mathbf{x}+\mathbf{y}) = \sum_{|\alpha|\le 2}\frac{[\partial_\alpha f](\mathbf{x})}{\alpha!}\mathbf{y}^\alpha + \sum_{|a|=3}\frac{[\partial_\alpha f](\mathbf{x}+c\mathbf{y})}{\alpha!}\mathbf{y}^\alpha$$

for some $c \in (0,1)$, where $\alpha = (a_1, \ldots, a_k)$ for $a_i \in \mathbb{Z}_{\ge 0}$, $|\alpha| = a_1 + \ldots + a_k$, $\alpha! = a_1! \cdots a_k!$, $\mathbf{y}^\alpha = y_1^{a_1}\cdots y_k^{a_k}$, and $\partial_\alpha f = \frac{\partial^{|\alpha|} f}{\partial_1^{a_1}\cdots\partial_k^{a_k}}$. Writing things out more explicitly,

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + \sum_{i=1}^k [\partial_i f](\mathbf{x})y_i + \sum_{i=1}^k\sum_{j=1}^k \frac{[\partial_{ij}f](\mathbf{x})}{2}y_iy_j + \sum_{i=1}^k\sum_{j=1}^k\sum_{l=1}^k\frac{[\partial_{ijl}f](\mathbf{x}+c\mathbf{y})}{6}y_iy_jy_l$$

But $|[\partial_{ijl}f](\mathbf{x}+c\mathbf{y})| \le \max_{(i,j,l)\in\{1,\ldots,k\}^3} \sup_{\mathbf{x}\in\mathbb{R}^k} |[\partial_{ijl}f](\mathbf{x})| =: C_0$ and $\sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k |y_i y_j y_l| = (|y_1|+\ldots+|y_k|)^3 = ||\mathbf{y}||_1^3$, so

$$\left|\mathbb{E}[f(\mathbf{X}+\mathbf{Y})] - \mathbb{E}[f(\mathbf{X})] + \mathbb{E}\left[\sum_{i=1}^k [\partial_i f](\mathbf{X}) Y_i\right] + \mathbb{E}\left[\sum_{i=1}^k \sum_{j=1}^k \frac{[\partial_{ij} f](\mathbf{X})}{2} Y_i Y_j\right]\right| \le \frac{C_0}{6}||\mathbf{Y}||_1^3$$

and so

$$\left|\mathbb{E}[f(\mathbf{X}+\mathbf{Y})] - \mathbb{E}[f(\mathbf{X})] + \sum_{i=1}^k \mathbb{E}[[\partial_i f](\mathbf{X})]\mathbb{E}[Y_i] + \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}\left[\frac{[\partial_{ij} f](\mathbf{X})}{2}\right]\mathbb{E}[Y_i Y_j]\right| \le \frac{C_0}{6}||\mathbf{Y}||_1^3$$

by independence of $\mathbf{X}$ and $\mathbf{Y}$. We can do the exact same argument for $\mathbf{X}$ and $\mathbf{W}$, and subtracting and using the triangle inequality, we get that

$$\begin{aligned}
&\Bigg|\mathbb{E}[f(\mathbf{X}+\mathbf{Y})] - \mathbb{E}[f(\mathbf{X})] + \sum_{i=1}^k \mathbb{E}[[\partial_i f](\mathbf{X})]\mathbb{E}[Y_i] + \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}\left[\frac{[\partial_{ij} f](\mathbf{X})}{2}\right]\mathbb{E}[Y_i Y_j] \\
&- \mathbb{E}[f(\mathbf{X}+\mathbf{W})] + \mathbb{E}[f(\mathbf{X})] - \sum_{i=1}^k \mathbb{E}[[\partial_i f](\mathbf{X})]\mathbb{E}[W_i] - \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}\left[\frac{[\partial_{ij} f](\mathbf{X})}{2}\right]\mathbb{E}[W_i W_j]\Bigg| \\
&= \big|\mathbb{E}[f(\mathbf{X}+\mathbf{Y})] - \mathbb{E}[f(\mathbf{X}+\mathbf{W})]\big| \le \frac{C_0}{6}\Big(||\mathbf{Y}||_1^3 + ||\mathbf{W}||_1^3\Big)
\end{aligned}$$

Finally, Cauchy-Schwarz tells us that

$$\begin{aligned}
||\mathbf{Y}||_1^2 &= (|Y_1| + \ldots + |Y_k|)^2 = (Y_1 \cdot (\pm 1) + \ldots + Y_k \cdot (\pm 1))^2 \\
&\le (Y_1^2 + \ldots + Y_k^2)((\pm 1)^2 + \ldots + (\pm 1)^2) = k||\mathbf{Y}||_2^2 = k|\mathbf{Y}|^2
\end{aligned}$$

and so

$$\big|\mathbb{E}[f(\mathbf{X}+\mathbf{Y})] - \mathbb{E}[f(\mathbf{X}+\mathbf{W})]\big| \le \frac{C_0}{6}k^{3/2}\Big(|\mathbf{Y}|^3 + |\mathbf{W}|^3\Big)$$

Problem 4

The classical multivariate CLT is as follows: for $\mathbf{X}_1, \ldots, \mathbf{X}_n$ i.i.d. random vectors in $\mathbb{R}^k$ (i.e. $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$) with $\mathbb{E}[\mathbf{X}_1] = \boldsymbol{\mu}$ and $\mathbb{E}\big[|\mathbf{X}_1|^2\big] < \infty$, and $\overline{\mathbf{X}}_n := \frac{1}{n}(\mathbf{X}_1 + \ldots + \mathbf{X}_n)$, we have that

$$\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) \to_d \mathbf{X} = (X_1, \ldots, X_k) \sim \text{Normal}_k(0, \text{Cov}[\mathbf{X}_1])$$

Cramér-Wold tells us that this happens if and only if $\mathbf{a} \cdot \sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) \to_d \mathbf{a} \cdot \mathbf{X}$ for all $\mathbf{a} \in \mathbb{R}^k$. Define $Y_i = \mathbf{a} \cdot (\mathbf{X}_i - \boldsymbol{\mu})$. Note that the Y_i are i.i.d., with variance

$$\text{Var}[Y_1] = \text{Var}\left[\sum_{i=1}^k a_i(X_{1i} - \mu_i)\right] = \sum_{i=1}^k \sum_{j=1}^k \mathbb{E}[a_i(X_{1i} - \mu_i)a_j(X_{1j} - \mu_j)] = \sum_{i=1}^k \sum_{j=1}^k a_i a_j \text{Cov}[X_{1i}, X_{1j}]$$

Similarly, the variance of $\mathbf{a} \cdot \mathbf{X} = a_1 X_1 + \ldots + a_k X_k$ is

$$\text{Var}\left[\sum_{i=1}^{k} a_i X_i\right] = \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \text{Cov}[X_i, X_j] = \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \text{Cov}[X_{1i}, X_{1j}]$$

because $\text{Cov}[\mathbf{X}] = \text{Cov}[\mathbf{X}_1]$ by definition of $\mathbf{X}$. Furthermore, $\mathbf{a} \cdot \mathbf{X}$ is normally distributed (uniquely determined by the values of the covariance matrix above), and so the single variable CLT gives that

$$\mathbf{a} \cdot \sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) = \sqrt{n}(\overline{Y}_n - 0) \to \mathbf{a} \cdot \mathbf{X} \sim \text{Normal}\left(0, \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \text{Cov}[X_{1i}, X_{1j}]\right)$$

Because $\mathbf{a} \in \mathbb{R}^k$ was arbitrarily chosen, we satisfied the conditions of Cramér-Wold, and so the multivariate CLT holds.

Problem 5

The Lévy metric for distributions F, G is defined to be

$$\lambda(F,G) = \inf\{\epsilon > 0 : F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon, \forall x \in \mathbb{R}\}$$

We first want to verify that this is indeed a metric:

- $\lambda(F,G) \geq 0$ obviously because we are only looking at $\epsilon > 0$.
- $\lambda(F,G) = \lambda(G,F)$: note that $G(x) \leq F(x+\epsilon)+\epsilon$ for all $x \in \mathbb{R} \iff G(x-\epsilon)-\epsilon \leq F(x)$ for all $x \in \mathbb{R}$ and similarly $F(x-\epsilon)-\epsilon \leq G(x)$ for all $x \in \mathbb{R} \iff F(x) \leq G(x+\epsilon)+\epsilon$ for all $x \in \mathbb{R}$ (just a variable change).
- $\lambda(F,G) = 0 \iff F = G$: ($\Longleftarrow$) is obvious because distribution functions are increasing, and so always $F(x-\epsilon) \leq F(x) \leq F(\epsilon)$. ($\Longrightarrow$) follows by right continuity and monotonicity of distribution functions (i.e. at any fixed $x \in \mathbb{R}$, $\forall \epsilon, \exists \delta$ s.t. $F(x+\delta) - F(x) < \epsilon$), which together with $G(x) \leq F(x+\delta)+\delta$ yields that $G(x) \leq F(x)+\epsilon+\delta$. ϵ is arbitrary and δ can be as small as we want it to be, so $G(x) \leq F(x)$ for all $x \in \mathbb{R}$. For the $\geq$ direction, just use symmetry (proven above) to switch F and G.
- $\lambda(F,G) \leq \lambda(F,H) + \lambda(H,G)$: for any ϵ_1 s.t. $F(x-\epsilon_1)-\epsilon_1 \leq H(x) \leq F(x+\epsilon_1)+\epsilon_1$ and ϵ_2 s.t. $H(x-\epsilon_2)-\epsilon_2 \leq G(x) \leq H(x+\epsilon_2)+\epsilon_2$, we have that

$$\begin{aligned} F(x-(\epsilon_1+\epsilon_2)) - (\epsilon_1+\epsilon_2) &\leq H(x-\epsilon_2)-\epsilon_2 \\ &\leq G(x) \\ &\leq H(x+\epsilon_2)+\epsilon_2 \leq F(x+(\epsilon_2+\epsilon_1)) + (\epsilon_1+\epsilon_2) \end{aligned}$$

  i.e. that $(\epsilon_1+\epsilon_2)$ works for F, G, $\forall \epsilon_1$ that work for F, H and $\forall \epsilon_2$ that work for H, G, implying that the infimum of the ϵ for F, G must be $\leq$ the infimum of the sum of the ϵ's for F, H and H, G.

Now that we know λ is a metric, we want to prove that going to 0 in the metric corresponds exactly to convergence in distribution, i.e. $\lambda(F_n, F) \to 0 \iff F_n \to_d F$.

($\Longrightarrow$) for $\epsilon > 0$ fixed and n sufficiently large, $\lambda(F_n, F) < \epsilon$. Similarly to our argument of $\lambda(F,G) = 0 \implies F = G$, we use right continuity of d.f.'s to get that $F_n(x) \leq F(x)+\epsilon+\delta$ (at some fixed x). Similarly, for the other direction, $F_n(x) \geq F(x) - \epsilon - \delta_n$ (this δ_n is based on the right continuity of the F_n, and so they might not be the same for all n). Thus, $|F_n(x) - F(x)| \leq \epsilon + \max\{\delta, \delta_n\} \leq \epsilon + \sup_{n\in\mathbb{N}}\{\delta_n\}$. The δ's can be made arbitrarily small (because if δ works, so do all $0 < \delta' < \delta$), and so $|F_n(x) - F(x)| \leq \epsilon$. Thus $F_n(x) \to F(x)$ at all continuity points, and so $F_n \to_d F$.

($\Longleftarrow$) fix any $\epsilon > 0$. Looking on any compact set $[a,b]$, find $x_1, \dots, x_N$ in the continuity set of F (which is almost every point) s.t. $x_{i+1} - x_i < \epsilon$. Now for all n large enough, $|F_n(x_i) - F(x_i)| < \epsilon$ for all $i \in \{1, \dots, N\}$. Thus for any $x \in [a,b]$, it is between some x_{i-1} and x_i and so $F_n(x) \leq F_n(x_i) \leq$

$F(x_i) + \epsilon \leq F(x + \epsilon) + \epsilon$ and similarly $F_n(x) \geq F_n(x_{i-1}) \geq F(x_{i-1}) - \epsilon \geq F(x - \epsilon) - \epsilon$. Doing this on all compact sets, we get that for n sufficiently large $\lambda(F_n, F) \leq \epsilon \implies \lambda(F_n, F) \to 0$.

522 Homework 7

Daniel Rui - 3/4/20

Problem 1

Let $\mathbb{S}$ be a standard Brownian motion on $[0,\infty)$, and let $\tau_b := \inf\{t > 0 : \mathbb{S}(t) = b\}$ (for some fixed $b > 0$). τ_b is of course a stopping time, and we want to use optional sampling of the exponential martingale $Y_r(t) := e^{r\mathbb{S}(t) - r^2 t/2}$ to show that:

(a) $P(\tau_b < \infty) = 1$: $\{Y_n, \mathscr{A}_n\}_{n\geq 0}$ is a martingale, and $\tau_b \wedge m := \min\{\tau_b, m\}$ (for some fixed $m \in \mathbb{N}$) is a stopping time $\leq m$, and so by the simple optional sampling theorem, $\mathbb{E}(Y_r(\tau_b \wedge m)|\mathscr{A}_0) = Y_r(0)$ a.s. $\implies \mathbb{E}[Y_r(\tau_b \wedge m)] = \mathbb{E}[Y_r(0)] = \mathbb{E}\big[e^{r\mathbb{S}(0)}\big] = \mathbb{E}\big[e^0\big] = 1$ (because $\mathbb{S}(0) = 0$ by definition of Brownian motion). But we know $\tau_b \wedge m \leq \tau_b$ and $\tau_b = \inf\{t > 0 : \mathbb{S}(t) = b\}$, and so $\mathbb{S}(\tau_b \wedge m) \leq b$. Thus (for $r > 0$),

$$Y_r(\tau_b \wedge m) = \frac{e^{r\mathbb{S}(\tau_b\wedge m)}}{e^{r^2(\tau_b \wedge m)/2}} \leq \frac{e^{rb}}{e^{r^2(\tau_b\wedge m)/2}} \leq \frac{e^{rb}}{e^0} = e^{rb}$$

We've just bound a sequence (in m) of random variables by a constant e^{rb} (constant w.r.t. ω), so we can use the dominated convergence theorem, which tells us that

$$\mathbb{E}[Y_r(\tau_b)] = \mathbb{E}\Big[\lim_{m\to\infty} Y_r(\tau_b \wedge m)\Big] = \lim_{m\to\infty} \mathbb{E}[Y_r(\tau_b \wedge m)] = \lim_{m\to\infty} 1 = 1$$

This is very helpful, because now we can say that

$$\mathbb{E}[Y_r(\tau_b)] = \mathbb{E}\Big[e^{rb - r^2\tau_b/2}\Big] = 1 \implies \mathbb{E}\Big[e^{-r^2\tau_b/2}\Big] = e^{-rb}$$

Now as $r \searrow 0$, on $[\tau_b = \infty]$, $\frac{1}{e^{r^2\tau_b/2}} \to 0$ (obviously), but on $[\tau_b < \infty]$, $\frac{1}{e^{r^2\tau_b/2}} \nearrow 1$. On the other hand, $e^{-rb} \to 1$, and so heuristically $e^{-r^2\tau_b/2}$ must go to 1 a.s., i.e. $P([\tau_b < \infty]) = 1$. More rigorously, if $P([\tau_b = \infty]) = c > 0$, then $\mathbb{E}\Big[e^{-r^2\tau_b/2}\Big]$ is at most $P([\tau_b < \infty]) \cdot 1 = 1 - c$ (for all $r > 0$). This contradicts that $e^{-rb} \nearrow 1$, and so $P([\tau_b = \infty])$ must be 0.

(b) $\mathbb{E}[e^{-s\tau_b}] = e^{-b\sqrt{2s}}$ (for $s \geq 0$): defining $s = \frac{r^2}{2}$ ($\iff r = \sqrt{2s}$) and observing that the random variables $\mathbb{S}(\tau_b) = b$ and τ_b are independent (obviously because one is constant w.r.t. ω), we see that

$$1 = \mathbb{E}[Y_r(\tau_b)] = \mathbb{E}\left[\frac{e^{b\sqrt{2s}}}{e^{s\tau_b}}\right] = \mathbb{E}\Big[e^{b\sqrt{2s}}\Big]\mathbb{E}\big[e^{-s\tau_b}\big] \implies \mathbb{E}\big[e^{-s\tau_b}\big] = e^{-b\sqrt{2s}}$$

(c) $\mathbb{E}[\tau_b] = \infty$: with the moment generating function above, we just differentiate once (w.r.t. s) to see that $\mathbb{E}[\tau_b]$ is the negative of the function $-\frac{1}{2}b(2s)^{-1/2}e^{-b\sqrt{2s}}$ evaluated at $s = 0$, which is ∞.

(d) $\mathbb{E}[\tau_b^r] < \infty$ for $r \in (0, 1/2)$ and $= \infty$ for $r = 1/2$, using the fact that the density of τ_b is $f_{\tau_b}(t) = \frac{b}{t^{3/2}}\phi\Big(\frac{b}{\sqrt{t}}\Big) \cdot 1_{(0,\infty)}(t) = \frac{b}{\sqrt{2\pi t^3}}e^{-\frac{b^2}{2t}} \cdot 1_{(0,\infty)}(t)$:

$$\mathbb{E}[\tau_b^r] = \int_0^\infty t^r \frac{b}{\sqrt{2\pi t^3}}e^{-\frac{b^2}{2t}}\,dt \geq \int_0^1 \frac{be^{-b^2/2}}{\sqrt{2\pi}}t^{r-3/2}\,dt = \begin{cases} < \infty & \text{if } r < 1/2 \\ = \infty & \text{if } r \geq 1/2 \end{cases}$$

Problem 2

Because F and F_n are all continuous and monotonically increasing (0 at $-\infty$ and 1 at ∞), we can find (for any $N \in \mathbb{N}$) $N+1$ points $x_0, \ldots, x_N$ s.t. $F(x_k) = \frac{k}{N}$ (where $x_0 = -\infty$ and $x_N = \infty$ for convenience). Now for any $x \in \mathbb{R}$, there is exactly one interval $(x_i, x_{i+1}]$ that it lies in, and furthermore, because F is increasing,

$$F(x) - F_n(x) \leq F(x_{i+1}) - F_n(x_i) = F(x_i) + \tfrac{1}{N} - F_n(x_i)$$

and

$$F_n(x) - F(x) \leq F_n(x_{i+1}) - F(x_i) = F_n(x_{i+1}) - F(x_{i+1}) + \tfrac{1}{N}$$

and so $|F(x) - F_n(x)| \leq \max\{F_n(x_{i+1}) - F(x_{i+1}), F(x_i) - F_n(x_i)\} + \frac{1}{N}$. This bound is dependent on which interval the x is in, but we can easily tweak this to be a uniform bound for all $x \in \mathbb{R}$:

$$\sup_{x \in \mathbb{R}} |F(x) - F_n(x)| \leq \max_{i \in \{0, \ldots, N\}} |F(x_i) - F_n(x_i)| + \frac{1}{N}$$

Because $F_n \to_d F \iff F_n \to F$ pointwise everywhere (because the F_n are all continuous, and by the portmanteau theorem: $P_n \to_d P \iff \lim_{n\to\infty} P_n(B) = P(B)$ for all $P-$continuity sets $B \in \mathscr{B}_{[0,1]}$, namely $B = [0, x]$), the max over finitely many things goes to 0 for n large enough, and $\frac{1}{N}$ goes to 0 for N large enough. A rather easy consequence of this (i.e. uniform convergence and continuity) and the triangle inequality applied to $|F_n(x_n) - F(x)| = |F_n(x_n) - F(x_n) + F(x_n) - F(x)|$ is that $x_n \to x \implies F_n(x_n) \to F(x)$,

Problem 3

(b) $F_n \to F \implies \{F_n\}_{n \geq 1}$ is tight: fix $\epsilon > 0$. We know $F(\overline{\mathbb{R}}) = P([X \in \overline{\mathbb{R}}]) = 1$, and because $\mathbb{R} = \bigcup_{n=1}^{\infty} (-n, n)$ is a union of increasing sets, we can interchange measures and limits to get that $\lim_{n\to\infty} F((-n, n)) = 1$. Thus, there is some N s.t. $F((-N, N)) > 1 - \epsilon$.

The portmanteau theorem gives us that $F_n \to F \iff F_n \to_d F \iff \liminf_{n\to\infty} F_n(B) \geq F(B)$ for all open sets B, which in our case means that $\liminf_{n\to\infty} F_n((-N, N)) \geq F((-N, N)) > 1 - \epsilon$. The strict inequality here allows us to say that for some K, $F_k((-N, N)) > 1 - \epsilon$ for all $k > K$.

For $F_1, \ldots, F_K$, just use the same argument we used on F to find $N_1, \ldots, N_K$ s.t. $F_i((-N_i, N_i)) > 1 - \epsilon$. Taking $N_0 = \max\{N_1, \ldots, N_K, N\}$, we get that $\forall F_i$, $F_i([-N_0, N_0]) \geq F_i((-N_0, N_0)) > 1 - \epsilon$. I.e., for any fixed ϵ we've found K compact s.t. $F_i(K) > 1 - \epsilon$, $\forall F_i$; thus $\{F_n\}$ is tight.

(a) $\limsup_{n\to\infty} \mathbb{E}[|X_n|^r] = M < \infty \implies \{F_n\}$ is tight: to translate this assumption into a statement regarding probability measures, observe that

$$\begin{aligned}
\limsup_{n\to\infty} F_n((-\infty, -\lambda) \cup (\lambda, \infty)) &= \limsup_{n\to\infty} P(|X_n| > \lambda) \\
&= \frac{1}{\lambda^r} \limsup_{n\to\infty} \lambda^r P(|X_n| > \lambda) \leq \frac{1}{\lambda^r} \limsup_{n\to\infty} \mathbb{E}\left[|X_n|^r \cdot 1_{[|X_n| > \lambda]}\right] \\
&\leq \frac{1}{\lambda^r} \limsup_{n\to\infty} \mathbb{E}[|X_n|^r] = \frac{M}{\lambda^r}
\end{aligned}$$

and so $\limsup_{n\to\infty} F_n((-\infty,-\lambda)\cup(\lambda,\infty)) \to 0$ as $\lambda \to \infty$. Thus, there is some Λ s.t. $\limsup_{n\to\infty} F_n((-\infty,-\Lambda)\cup(\Lambda,\infty)) < \epsilon$ for any fixed $\epsilon > 0$. Then,

$$\begin{aligned}
\liminf_{n\to\infty} F_n([-\Lambda,\Lambda]) &= \liminf_{n\to\infty} \big(1 - F_n((-\infty,-\Lambda)\cup(\Lambda,\infty))\big) \\
&= 1 - \liminf_{n\to\infty} F_n((-\infty,-\Lambda)\cup(\Lambda,\infty)) \\
&\geq 1 - \limsup_{n\to\infty} F_n((-\infty,-\Lambda)\cup(\Lambda,\infty)) > 1-\epsilon
\end{aligned}$$

This means that for all large enough n, $F_n([-\Lambda,\Lambda]) > 1-\epsilon$, and so similarly as above in (b), we just find the corresponding Λ_i for all the F_i for i not "large enough", and then take the max to find a compact set that works for all the F_i.

Problem 4

Suppose that we have $Z \sim$ Normal(0,1) and $\mu_n \to \mu < \infty$ and $\sigma_n^2 \to \sigma^2 < \infty$. Then, defining $X_n =_d \mu_n + \sigma_n Z$ and $X =_d \mu + \sigma Z$, we get:

(b) $\big|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]\big| \leq ||f||_{BL}\mathbb{E}\big[\min\big\{1, |\mu_n-\mu| + |\sigma_n - \sigma|\cdot|Z|\big\}\big]$ (for any $f \in BL(\mathbb{R})$):

$$\begin{aligned}
\big|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]\big| &= \big|\mathbb{E}[f(X_n) - f(X)]\big| = \big|\mathbb{E}[f(\mu_n+\sigma_n Z) - f(\mu+\sigma Z)]\big| \\
&\leq \mathbb{E}[|f(\mu_n+\sigma_n Z) - f(\mu+\sigma Z)|] \\
&\leq \mathbb{E}\big[||f||_{BL}\min\{1, \big|(\mu+\sigma Z) - (\mu_n+\sigma_n Z)\big|\}\big] \\
&\leq ||f||_{BL}\mathbb{E}\big[\min\big\{1, |\mu_n-\mu| + |\sigma_n-\sigma|\cdot|Z|\big\}\big]
\end{aligned}$$

where we used the fact that $X =_d Y \implies \mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$ (could be proven with the law of the unconscious statistician), the BL-inequality, and the triangle inequality.

(a) $X_n \to_d X$: this is equivalent to showing $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for any $f \in BL(\mathbb{R})$. From (b), we know that

$$\begin{aligned}
\big|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]\big| &\leq ||f||_{BL}\mathbb{E}\big[\min\big\{1, |\mu_n-\mu| + |\sigma_n-\sigma|\cdot|Z|\big\}\big] \\
&\leq ||f||_{BL}\mathbb{E}[|\mu_n-\mu| + |\sigma_n-\sigma|\cdot|Z|] \\
&= ||f||_{BL}\Big(|\mu_n-\mu| + |\sigma_n-\sigma|\cdot\mathbb{E}[|Z|]\Big)
\end{aligned}$$

For large enough n, we can easily bound this by any $\epsilon > 0$, because $\mu_n \to \mu$ and $\sigma_n \to \sigma$ ($||f||_{BL}$ and $\mathbb{E}[|Z|]$ are just constants here). We've just found that for any $f \in BL(\mathbb{R})$, we can find N s.t. $n > N \implies \big|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]\big|$, and so $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all $f \in BL(\mathbb{R})$.

522 Homework 6

Daniel Rui - 2/26/20

Problem 1

Polyá's urn: initially at $t = 0$, we have 1 black and 1 white ball. At each time $t = 1, 2, \ldots$, a ball is chosen randomly and is replaced together with a new ball of the same color. Thus, immediately after $t = n$, there are $n + 2$ balls total and $B_n + 1$ black balls, where B_n is the number of black balls chosen by time n and the 1 is from the initial one. Define $M_n := \frac{B_n+1}{n+2}$, i.e. the fraction of the balls that are black immediately after $t = n$. Then, with $\mathscr{A}_n := \sigma[M_1, \ldots, M_n]$, $\{M_n, \mathscr{A}_n\}_{n\geq 0}$ is a martingale:

$$\begin{aligned}
\mathbb{E}(M_{n+1}|\mathscr{A}_n) &= \mathbb{E}\left(\frac{B_{n+1}+1}{n+3}\middle|\mathscr{A}_n\right) = \frac{1}{n+3}\mathbb{E}(B_{n+1}|B_n) + \frac{1}{n+3} \\
&= \frac{1}{n+3}\big((B_n+1)M_n + B_n(1-M_n)\big) + \frac{1}{n+3} \\
&= \frac{M_n + B_n + 1}{n+3} = \frac{\frac{B_n+1}{n+2} + B_n + 1}{n+3} = \frac{B_n+1}{n+2} = M_n
\end{aligned}$$

Furthermore, notice that $P(B_1 = k) = \frac{1}{2}$ for $k \in \{0, 1\}$. Assuming that $P(B_n = k) = \frac{1}{n+1}$ for all $k \in \{0, \ldots, n\}$, we inductively get that

$$\begin{aligned}
P(B_{n+1} = k) &= \mathbb{E}[P(B_{n+1} = k|B_n)] = \mathbb{E}\left[\sum_{i=0}^{\infty} P(B_{n+1} = k|B_n = i) \cdot 1_{[B_n=i]}\right] \\
&= \mathbb{E}\big[P(B_{n+1} = k|B_n = k) \cdot 1_{[B_n=k]} + P(B_{n+1} = k|B_n = k-1) \cdot 1_{[B_n=k-1]}\big] \\
&= P(B_{n+1} = k|B_n = k)P(B_n = k) + P(B_{n+1} = k|B_n = k-1)P(B_n = k-1) \\
&= \left(1 - \tfrac{k+1}{n+2}\right) \cdot P(B_n = k) + \tfrac{(k-1)+1}{n+2} \cdot P(B_n = k-1) \\
&= \tfrac{1}{n+2}\big((n-k+1)P(B_n = k) + kP(B_n = k-1)\big)
\end{aligned}$$

which equals $\frac{1}{n+2}$ in all three cases: $k = n + 1$, for which the first term is 0; $k = 0$, for which the second term is 0; and $k \in \{1, \ldots, n\}$, for which the terms cancel neatly. Thus $P(B_n = k) = P(M_n = \frac{k+1}{n+2}) = \frac{1}{n+1}$ for all $k \in \{0, \ldots, n\}$ (a discrete uniform distribution), and so $\Theta := \lim_{n\to\infty} M_n \sim \text{Unif}(0, 1)$.

Lastly, $\{N_n^\theta, \mathscr{A}_n\}_{n\geq 0}$, where $N_n^\theta := \frac{(n+1)!}{B_n!(n-B_n)!}\theta^{B_n}(1-\theta)^{n-B_n}$, is a martingale:

$$\begin{aligned}
\mathbb{E}(N_{n+1}^\theta|\mathscr{A}_n) &= \mathbb{E}\left(\frac{(n+2)!}{B_{n+1}!(n+1-B_{n+1})!}\theta^{B_{n+1}}(1-\theta)^{n+1-B_{n+1}}\middle|B_n\right) \\
&= \underbrace{\frac{(n+2)!\theta^{B_n+1}(1-\theta)^{n-B_n}}{(B_n+1)!(n-B_n)!} \cdot M_n}_{\text{picked black, prob.} M_n \implies B_{n+1}=B_n+1} + \underbrace{\frac{(n+2)!\theta^{B_n}(1-\theta)^{n+1-B_n}}{B_n!(n+1-B_n)!} \cdot (1-M_n)}_{\text{picked white, prob.} 1-M_n \implies B_{n+1}=B_n} \\
&= \frac{(n+1)!\theta^{B_n}(1-\theta)^{n-B_n}}{B_n!(n-B_n)!} \cdot \left(\frac{(n+2)\theta}{B_n+1} \cdot M_n + \frac{(n+2)(1-\theta)}{(n+1-B_n)} \cdot (1-M_n)\right) \\
&= N_n^\theta \cdot \left(\frac{\theta}{M_n} \cdot M_n + \frac{(1-\theta)}{(1-M_n)} \cdot (1-M_n)\right) = N_n^\theta
\end{aligned}$$

Problem 2

We want to prove Doob's $\mathscr{L}_r$-inequality: for a martingale $\{X_n, \mathscr{A}_n\}_{n\geq 0}$ and $r > 1$, the following are equivalent:

(i) $\{|X_n|^r\}$ is integrable

(ii) $X_n \to_{\mathscr{L}_r} X_\infty$

(iii) $\{X_n\}$ is uniformly integrable and $X_n \to_{\text{a.s.}} X_\infty \in \mathscr{L}_r$

(iv) $\{|X_n|^r\}$ is uniformly integrable

(v) $\{|X_n|^r, \mathscr{A}_n\}_{n\in[0,\infty]}$ is a submartingale and $\mathbb{E}[|X_n|^r] \nearrow \mathbb{E}[|X_\infty|^r] < \infty$

Just from the initial assumption (that $\{X_n, \mathscr{A}_n\}_{n\geq 0}$ is a martingale and $r > 1$), we get that $g(x) = |x|$ is convex $\implies$ $\{|X_n|, \mathscr{A}_n\}_{n\geq 0}$ is a submartingale. Furthermore $g(x) = |x|^r$ is also convex, so $\{|X_n|^r, \mathscr{A}_n\}_{n\geq 0}$ is also a submartingale.

Immediately, we see that $\{|X_n|, \mathscr{A}_n\}_{n\geq 0}$ is submg $\implies$ [(i) $\iff$ (iv)] (by HW#5P5), and $\{|X_n|^r, \mathscr{A}_n\}_{n\geq 0}$ is submg $\implies$ [(iv) $\iff$ (v)] (by part (B) of the (sub-)martingale convergence theorem).

(i) $\implies$ (iii): also, in HW#5P5, we proved as an intermediary step that $\{|X_n|^r\}$ integrable $\implies$ $\{|X_n|\}$ uniformly integrable, which of course is iff $\{X_n\}$ uniformly integrable. Part (A) of the s-mg convergence theorem gives us that $X_n \to_{\text{a.s.}} X_\infty \in \mathscr{L}_1$, and Fatou tells us that $\mathbb{E}[|X_\infty|^r] = \mathbb{E}\Big[\liminf_{n\to\infty} |X_n|^r\Big] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|^r] \leq \sup_{n\in} \mathbb{E}[|X_n|^r] < \infty$, and so $X_\infty \in \mathscr{L}_r$.

$\{|X_n|\}$ being uniformly integrable implies that $\{X_n\}$ is uniformly integrable, and because we are on a finite measure space that implies that $\{X_n^+\}$ is integrable. Hence, by part A of the (sub-)martingale convergence theorem, (iii) $\implies$ $X_n \to_{\text{a.s.}} X_\infty \in \mathscr{L}_1$ $\implies$ $X_n \to_p X_\infty$, which by Vitali's theorem, means that (iii) $\implies$ ((iv) $\iff$ (ii)). Because (iii) can be deduced from (iv), this proves (iv) $\implies$ (ii). The other direction (ii) $\implies$ (iv) is given by one of the big theorems from chapter 3 (Theorem 5.7).

Lastly, (iii) $\implies$ (i): part Paul Vondiziano B of the s-mg convergence theorem gives that $\{|X_n|, \mathscr{A}_n\}_{n\in[0,\infty]}$ is a submartingale, and so $|X_n| \leq \mathbb{E}\big(|X_\infty|\big|\mathscr{A}_n\big)$ $\implies$ $|X_n|^r \leq \mathbb{E}\big(|X_\infty|\big|\mathscr{A}_n^r\big) \leq \mathbb{E}(|X_\infty|^r|\mathscr{A}_n)$ $\implies$ $\mathbb{E}[|X_n|^r] \leq \mathbb{E}[|X_\infty|^r] < \infty$ $\implies$ $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^r] < \infty$ (because $X_\infty \in \mathscr{L}_r$).

Problem 3

!Warning!: I've switched the use of Y and X because I borrowed some notation from homework #4. Also, as written, the problem statement looks like it's saying that P and Q are measures on the real line, which would imply that $\frac{dQ}{dP}$ is a function of the real line, while $Y_1, \ldots$ are functions on $(\Omega, \mathscr{A})$. Immediately this seems incompatible, so I've tried my best to interpret the problem in such a way that the types match up.

Let $Y_1, Y_2, \ldots$ be independent on some probability space $(\Omega, \mathscr{A}, \mathsf{P})$, and suppose (Paul Vondiziano assume?) they have density p_k (w.r.t. some measure μ), i.e. $F_{Y_k}(B) = P_k(B) = \int_B p_k \, d\mu$, i.e. $p_k = \frac{dP_k}{d\mu}$, the Radon-Nikodym derivative. Now suppose that we have some other candidate density q_k (w.r.t the same measure μ) s.t. $Q_k \ll P_k \ll \mu$. Define $X_k = \frac{q_k(Y_k)}{p_k(Y_k)}$ (which is ≥ 0 because densities are ≥ 0).

(a) Defining $M_n := \prod_{k=1}^n X_k$ and $\mathscr{A}_k = \sigma[Y_1, \ldots, Y_k]$, $\{M_n, \mathscr{A}_k\}_{k \geq 1}$ is essentially Kakutani's martingale. Note our usage of the law of the unconscious statistician and of the change of variables theorem:

$$\begin{aligned}
\mathbb{E}(M_{n+1}|\mathscr{A}_n) &= \mathbb{E}\left(\frac{q_{n+1}(Y_{n+1})}{p_{n+1}(Y_{n+1})} \cdot M_n \middle| \mathscr{A}_n\right) =_{\text{a.s.}} M_n \mathbb{E}\left[\frac{q_{n+1}(Y_{n+1})}{p_{n+1}(Y_{n+1})}\right] \\
&= M_n \int_\Omega \frac{q_{n+1}(Y_{n+1})}{p_{n+1}(Y_{n+1})} \, d\mathsf{P} = M_n \int_{\mathbb{R}} \frac{q_{n+1}(x)}{p_{n+1}(x)} \, dF_{Y_{n+1}}(x) \\
&= M_n \int_{\mathbb{R}} \frac{q_{n+1}(t)}{p_{n+1}(t)} \frac{dF_{Y_{n+1}}}{d\mu}(t) \, d\mu(t) = M_n \int_{\mathbb{R}} \frac{q_{n+1}(t)}{p_{n+1}(t)} p_{n+1}(t) \, d\mu(t) \\
&= M_n \int_{\mathbb{R}} q_{n+1}(t) \, d\mu(t) = M_n Q_{n+1}(\mathbb{R}) = M_n
\end{aligned}$$

Lastly, regarding the mean: $\mathbb{E}[M_n] = \mathbb{E}\left[\prod_{k=1}^n X_k\right] = \prod_{k=1}^n \mathbb{E}[X_k] = \prod_{k=1}^n \int_\Omega \frac{q_k(Y_k)}{p_k(Y_k)} \, d\mathbf{P} = \prod_{k=1}^n 1 = 1.$

Regarding the likelihood ratio interpretation: note that the P_k and Q_k are all measures on the real line; i.e. $P_k, Q_k : \mathscr{B} \to \mathbb{R}$. Define $\mathscr{A}_k = \sigma[Y_k] = Y_k^{-1}(\mathscr{B})$. Then for any $A_k \in \mathscr{A}_k$, there is some $B \in \mathscr{B}$ s.t. $Y_k^{-1}(B) = A_k$. Thus for every $A_k \in \mathscr{A}_k$ and its corresponding B, we can define $\widetilde{P}_k(A_k) = P_k(B)$, meaning that $\widetilde{P}_k : \mathscr{A}_k \to \mathbb{R}$. We can similarly do this to Q_k to get $\widetilde{Q}_k : \mathscr{A}_k \to \mathbb{R}$, $Q_k(B) = \widetilde{Q}_k(A_k)$.

Now because $Q_k \ll P_k$, there is some function $\frac{dQ_k}{dP_k}$ s.t. $Q_k(B) = \int_B \frac{dQ_k}{dP_k} \, dP_k$. The law of the unconscious statistician gives that this integral is $= \int_{Y_k^{-1}(B)} \frac{dQ_k}{dP_k}(Y_k) \, d\widetilde{P}_k = \int_{A_k} \frac{dQ_k}{dP_k}(Y_k) \, d\widetilde{P}_k$ (because $P_k(B) = \widetilde{P}_k(Y_k^{-1}(B))$ is in fact the induced measure). But $Q_k(B) = \widetilde{Q}_k(A_k)$ so $\widetilde{Q}_k(A_k) = \int_{A_k} \frac{dQ_k}{dP_k}(Y_k) \, d\widetilde{P}_k$!

Furthermore $\widetilde{Q}_k \ll \widetilde{P}_k$ (b/c $\widetilde{P}_k(A_k) = 0 \iff P_k(B) = 0 \implies Q_k(B) = 0 \iff \widetilde{Q}_n(A_k) = 0$), meaning that there is a unique (a.s.) $\frac{d\widetilde{Q}_k}{d\widetilde{P}_k}$ s.t. $\widetilde{Q}_k(A_k) = \int_{A_k} \frac{d\widetilde{Q}_k}{d\widetilde{P}_k} \, d\widetilde{P}_k$. Thus, $\frac{d\widetilde{Q}_k}{d\widetilde{P}_k} =_{\text{a.s.}} \frac{dQ_k}{dP_k}(Y_k)$. This relates to the X_k above because the chain rule for Radon-Nikodym derivatives tells us that,

$$q_k = \frac{dQ_k}{d\mu} =_{\text{a.s.}} \frac{dQ_k}{dP_k} \cdot \frac{dP_k}{d\mu} = \frac{dQ_k}{dP_k} \cdot p_k \implies \frac{dQ_k}{dP_k} =_{\text{a.s.}} \frac{q_k}{p_k}$$

and hence $\widetilde{Q}_k(A_k) = \int_{A_k} X_k \, d\widetilde{P}_k$, $\forall A_k \in \mathscr{A}_k$. We are now ready to define the product measure $\widetilde{\mathbf{P}}_\mathbf{n} = \prod_{k=1}^n \widetilde{P}_k$, where $\widetilde{\mathbf{P}}_\mathbf{n}(A_1 \times A_2 \times \ldots \times A_n) = \widetilde{P}_1(A_1) \cdot \widetilde{P}_2(A_2) \cdots \widetilde{P}_n(A_n)$ (similarly $\mathbf{Q}_n$). Because the Radon-Nikodym derivative of a product measure is the product of the R-N derivatives, we have that

$$\widetilde{\mathbf{Q}}_{\mathbf{n}}(\mathbf{A}) = \int_{\mathbf{A}} \prod_{k=1}^{n} X_k \, d\widetilde{\mathbf{P}}_{\mathbf{n}}$$

for any $\mathbf{A} \in \mathscr{A}_{\boldsymbol{n}} := \sigma[\{A_1 \times \ldots \times A_n : A_i \in \mathscr{A}_i\}]$. Thus, $M_n = \prod_{k=1}^{n} X_k$ is equal (a.s.) to $\frac{d\widetilde{\mathbf{Q}}_{\mathbf{n}}}{d\widetilde{\mathbf{P}}_{\mathbf{n}}}$. If we think of the infinite dimensional σ-algebra $\mathscr{A} := \sigma[\{A_1 \times A_2 \times \ldots : A_i \in \mathscr{A}_i\}]$ and the infinite product measures $\widetilde{\boldsymbol{P}} := \prod_{k=1}^{\infty} \widetilde{P}_k$ and $\widetilde{\boldsymbol{Q}} := \prod_{k=1}^{\infty} \widetilde{Q}_k$ (both measures : $\mathscr{A} \to \mathbb{R}$), then $\widetilde{\boldsymbol{P}}_{\boldsymbol{n}} = \widetilde{\boldsymbol{P}}|_{\mathscr{A}_{\boldsymbol{n}}}$ (restriction to a sub-σ-field $\mathscr{A}_{\boldsymbol{n}}$, where we tweak $\mathscr{A}_{\boldsymbol{n}}$ to be $:= \sigma[\{A_1 \times \ldots \times A_n \times \Omega \times \ldots : A_i \in \mathscr{A}_i\}]$ and so $\mathscr{A}_{\boldsymbol{n}} \subseteq \mathscr{A}_{\boldsymbol{n+1}}$), and same for $\widetilde{\boldsymbol{Q}}_{\boldsymbol{n}}$. Thus $\{M_n, \mathscr{A}_{\boldsymbol{n}}\}_{n \geq 1}$ fits the form of a likelihood ratio martingale: for any $\mathbf{A} \in \mathscr{A}_{\boldsymbol{m}}$ $(m < n)$,

$$\int_{\mathbf{A}} M_n \, d\widetilde{\boldsymbol{P}} = \int_{\mathbf{A}} M_n \, d\widetilde{\boldsymbol{P}}_{\boldsymbol{n}} = \widetilde{\boldsymbol{Q}}_{\boldsymbol{n}}(\mathbf{A}) = \widetilde{\boldsymbol{Q}}_{\boldsymbol{m}}(\mathbf{A}) = \int_{\mathbf{A}} M_m \, d\widetilde{\boldsymbol{P}}_{\boldsymbol{m}} = \int_{\mathbf{A}} M_m \, d\widetilde{\boldsymbol{P}}$$

implying that $\int_{\mathbf{A}} M_n - M_m \, d\widetilde{\boldsymbol{P}} = 0$ for all $\mathbf{A} \in \mathscr{A}_{\boldsymbol{m}}$.

(b) Not quite sure what's going on here...if we don't know that $Q \ll P$, how do we even define $\frac{dQ}{dP}$?

(c) From Kakutani's martingale theorem, $\{M_n\}$ uniformly integrable $\iff$ $\prod_1^{\infty} \mathbb{E}\left[\sqrt{X_n}\right] > 0$; thus we just need to prove the following equality:

$$\prod_{n=1}^{\infty} \mathbb{E}\left[\sqrt{X_n}\right] = \prod_{n=1}^{\infty} \int_{\mathbb{R}} \sqrt{p_n(x) q_n(x)} \, dx > 0$$

Under my interpretation, we would have that

$$\mathbb{E}\left[\sqrt{X_n}\right] = \int_{\Omega} \sqrt{\frac{q_n}{p_n}}(Y_n) \, d\mathsf{P} = \int_{\mathbb{R}} \sqrt{\frac{q_n}{p_n}}(x) \, dF_{Y_n}(x)$$

which doesn't get us to Lebesgue measure/regular integration on the real line...

(d) If $p_n = q_n$, then $\int_{\mathbb{R}} \sqrt{p_n q_n} \, dx = \int_{\mathbb{R}} p_n \, dx$ will be 1 (integrating a density function on $\mathbb{R}$ always yields 1 and so the product in (c) will be $1 > 0$. Now for example take p_n to be the density of Unif(0,1) and q_n to be the density of Unif(0,2). Then $\int_{\mathbb{R}} \sqrt{p_n q_n} \, dx = \int_{\mathbb{R}} \sqrt{1_{[0,1]} \cdot \frac{1}{2} \cdot 1_{[0,2]}} \, dx = \int_0^1 \sqrt{\frac{1}{2}} \, dx = \sqrt{\frac{1}{2}}$ which would go to 0 in the infinite product. Similarly if we take q_n to be the density of Unif(0,$\frac{1}{2}$), we would get $\int_{\mathbb{R}} \sqrt{p_n q_n} \, dx = \int_{\mathbb{R}} \sqrt{1_{[0,1]} \cdot 2 \cdot 1_{[0,\frac{1}{2}]}} \, dx = \int_0^{1/2} \sqrt{2} \, dx = \frac{\sqrt{2}}{2}$ which would also go to 0. Thus, the condition (c) is satisfied in the case of uniform r.v.'s when the distributions are exactly the same each other. In general, the statistical meaning when it holds will be if the distributions are similar, and the meaning when it fails will be if the distributions are different.

Problem 4

Let X and Y be random variables ≥ 0 where $\lambda P(X \geq \lambda) \leq \mathbb{E}\big[Y \cdot 1_{[X \geq \lambda]}\big]$ for every $\lambda > 0$. We know that for any $p > 1$,

$$\mathbb{E}[X^p] = \int_0^\infty P(X^p \geq x)\, dx = \int_0^\infty P(X \geq x^{1/p})\, dx = \int_0^\infty pu^{p-1} P(X \geq u)\, du$$

(where these equalities hold even if the integrals end up being ∞ — i.e. one diverges $\iff$ the other diverges, and one converges $\iff$ the other converges to the same value). The following (in)equalities use Fubini-Tonelli (because every function involved here is ≥ 0) and Hölder's inequality:

$$\begin{aligned}
\mathbb{E}[X^p] &= \int_0^\infty pu^{p-1} P(X \geq u)\, du \leq \int_0^\infty pu^{p-2} \cdot \mathbb{E}\big[Y \cdot 1_{[X \geq u]}\big]\, du \\
&= \int_0^\infty \mathbb{E}\big[pu^{p-2} \cdot Y \cdot 1_{[X \geq u]}\big]\, du = \mathbb{E}\left[\int_0^\infty pu^{p-2} \cdot Y(\omega) \cdot 1_{[X \geq u]}(\omega)\, du\right] \\
&= \mathbb{E}\left[pY(\omega) \int_0^\infty u^{p-2} \cdot 1_{[X \geq u]}(\omega)\, du\right] = \mathbb{E}\left[pY(\omega) \int_0^{X(\omega)} u^{p-2}\, du\right] \\
&= \mathbb{E}\left[pY \cdot \frac{1}{p-1} X^{p-1}\right] = \frac{p}{p-1} \mathbb{E}\big[Y \cdot X^{p-1}\big] \leq \frac{p}{p-1} (\mathbb{E}[Y^p])^{\frac{1}{p}} \left(\mathbb{E}\left[(X^{p-1})^{\frac{1}{1-\frac{1}{p}}}\right]\right)^{1-\frac{1}{p}} \\
&= \frac{p}{p-1} (\mathbb{E}[Y^p])^{\frac{1}{p}} (\mathbb{E}[X^p])^{\frac{p-1}{p}}
\end{aligned}$$

Rearranging yields that $\mathbb{E}[X^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[Y^p]$, as desired. Lastly note that $\mathbb{E}[X^p] = 0$ or $\mathbb{E}[Y^p] = \infty$ don't affect this inequality, because everything here is ≥ 0 and $\leq \infty$ respectively.

522 Homework 5

Daniel Rui - 2/19/20

Problem 1

We have a martingale $\{X_n, \mathscr{A}_n\}_{n\geq 0}$, a predictable process $\{H_n, \mathscr{A}_n\}_{n\geq 0}$ (i.e. H_n is $\mathscr{A}_{n-1}$-measurable, and H_0 is constant i.e. $\{0,\Omega\}$-measurable) and $W_n := (H\cdot X)_n := \sum_{k=1}^n H_k \Delta X_k := \sum_{k=1}^n H_k(X_k - X_{k-1})$.

(a) Assuming that the H_n are bounded (for ease I guess?) and keeping in mind that H_{n+1} is $\mathscr{A}_n$-measurable,

$$\begin{aligned}\mathbb{E}[W_{n+1} - W_n|\mathscr{A}_n] &= \mathbb{E}(H_{n+1}(X_{n+1}-X_n)|\mathscr{A}_n) =_{\text{a.s.}} H_{n+1}\mathbb{E}(X_{n+1}-X_n|\mathscr{A}_n) \\ &= H_{n+1}\big(\mathbb{E}(X_{n+1}|\mathscr{A}_n) - X_n\big) =_{\text{a.s.}} H_{n+1}\big(X_n - X_n\big) = 0\end{aligned}$$

and so clearly $\{W_n, \mathscr{A}_n\}_{n\geq 0}$ is a martingale.

(b) We want to show that $L_n = W_n^2 - \langle W\rangle_n$ is a 0-mean martingale w.r.t. $\mathscr{A}_n$ where $L_0 := H_0^2(X_0^2 - \mathbb{E}[X_0^2])$ and

$$\begin{aligned}\langle W\rangle_n := \langle H\cdot X\rangle_n &= \sum_{k=1}^n H_k^2 \mathbb{E}\big((\Delta X_k)^2|\mathscr{A}_{k-1}\big) + H_0^2\mathbb{E}\big[X_0^2\big] \\ &= \underbrace{H_n^2\mathbb{E}\big((X_n - X_{n-1})^2|\mathscr{A}_{n-1}\big)}_{\mathscr{A}_{n-1}\text{-msble}} + \dots\end{aligned}$$

First, we compute

$$\begin{aligned}\mathbb{E}\big(W_{n+1}^2|\mathscr{A}_n\big) &= \mathbb{E}\big((W_{n+1} - W_n + W_n)^2|\mathscr{A}_n\big) \\ &= \mathbb{E}\big((W_{n+1}-W_n)^2 + 2W_n(W_{n+1}-W_n) + (W_n)^2|\mathscr{A}_n\big) \\ &=_{\text{a.s.}} \mathbb{E}\Big(\big(H_{n+1}(X_{n+1}-X_n)\big)^2\Big|\mathscr{A}_n\Big) + 2W_n\mathbb{E}(W_{n+1}-W_n|\mathscr{A}_n) + (W_n)^2 \\ &=_{\text{a.s.}} H_{n+1}^2\mathbb{E}\Big(\big((X_{n+1}-X_n)\big)^2\Big|\mathscr{A}_n\Big) + 2W_n\big(W_n - W_n\big) + (W_n)^2 \\ &= H_{n+1}^2\mathbb{E}\big((\Delta X_{n+1})^2\big|\mathscr{A}_n\big) + (W_n)^2\end{aligned}$$

Then,

$$\begin{aligned}\mathbb{E}(L_{n+1} - L_n|\mathscr{A}_n) &= \mathbb{E}\big(W_{n+1}^2 - \langle W\rangle_{n+1} - W_n^2 + \langle W\rangle_n|\mathscr{A}_n\big) \\ &= \mathbb{E}\big(W_{n+1}^2 - W_n^2 - \langle W\rangle_{n+1} + \langle W\rangle_n|\mathscr{A}_n\big) \\ &= H_{n+1}^2\mathbb{E}\big((\Delta X_{n+1})^2\big|\mathscr{A}_n\big) + (W_n)^2 - (W_n)^2 - \langle W\rangle_{n+1} + \langle W\rangle_n \\ &= H_{n+1}^2\mathbb{E}\big((\Delta X_{n+1})^2\big|\mathscr{A}_n\big) + (W_n)^2 - (W_n)^2 - H_{n+1}^2\mathbb{E}\big((\Delta X_{n+1})^2\big|\mathscr{A}_n\big) = 0\end{aligned}$$

Because $\{L_n, \mathscr{A}_n\}_{n\geq 0}$ is now a martingale, $\mathbb{E}[L_{n+1}] = \mathbb{E}[L_n] = \ldots = \mathbb{E}[L_0] = \mathbb{E}\big[H_0^2(X_0^2 - \mathbb{E}\big[X_0^2\big])\big] = H_0^2\mathbb{E}\big[X_0^2 - \mathbb{E}\big[X_0^2\big]\big] = 0$ (recall that H_0 is constant). Thus $\{L_n, \mathscr{A}_n\}_{n\geq 0}$ is also 0-mean; this in combination with the predictable process $\langle W\rangle_n$ is Doob's decomposition of $\{W_n^2, \mathscr{A}_n\}_{n\geq 1}$

Problem 2

We suppose that $\{X_n, \mathscr{A}_n\}$ and $\{Y_n, \mathscr{A}_n\}$ are submartingales. Then by the definition of submartingale and that $X \le Y \implies \mathbb{E}(X|\mathscr{D}) \le_{\text{a.s.}} \mathbb{E}(Y|\mathscr{D})$, $X_n \le_{\text{a.s}} \mathbb{E}[X_{n+1}|\mathscr{A}_n] \le \mathbb{E}[\max\{X_{n+1}, Y_{n+1}\}|\mathscr{A}_n]$. Likewise, $Y_n \le_{\text{a.s}} \mathbb{E}[Y_{n+1}|\mathscr{A}_n] \le \mathbb{E}[\max\{X_{n+1}, Y_{n+1}\}|\mathscr{A}_n]$, and so $\max\{X_n, Y_n\} \le_{\text{a.s}} \mathbb{E}[Y_{n+1}|\mathscr{A}_n] \le \mathbb{E}[\max\{X_{n+1}, Y_{n+1}\}|\mathscr{A}_n]$, which of course means that $\{\max\{X_n, Y_n\}, \mathscr{A}_n\}$ is a sub-martingale.

Problem 3

Let $X, Y \in L_2(\Omega, \mathscr{A}, P)$ and $\mathscr{D}$ a sub-σ-field of $\mathscr{A}$. Then because $\mathbb{E}(Y|\mathscr{D})$ is $\mathscr{D}$-measurable, $\mathbb{E}[\mathbb{E}(Y|\mathscr{D})] = \mathbb{E}[Y]$, and $\mathbb{E}(XY|\mathscr{D}) = Y\mathbb{E}(X|\mathscr{D})$ when Y is $\mathscr{D}$-measurable,

$$\mathbb{E}[X\mathbb{E}(Y|\mathscr{D})] = \mathbb{E}\Big[\mathbb{E}\Big(X\mathbb{E}(Y|\mathscr{D})\Big|\mathscr{D}\Big)\Big] = \mathbb{E}[\mathbb{E}(X|\mathscr{D})\mathbb{E}(Y|\mathscr{D})]$$

We can switch the X and Y to get that $\mathbb{E}[X\mathbb{E}(Y|\mathscr{D})] = \mathbb{E}[\mathbb{E}(X|\mathscr{D})\mathbb{E}(Y|\mathscr{D})] = \mathbb{E}[Y\mathbb{E}(X|\mathscr{D})]$.

Problem 4

Let $T_1, T_2, \dots (\ge 0)$ be (extended) stopping times (w.r.t. a filtration $\mathscr{A}_t$ of $\mathscr{A}$, i.e. all $\mathscr{A}_t$ sub-σ-fields of $\mathscr{A}$ and $A_s \subseteq A_t$ for $s \le t$); i.e. $[T_i \le t] \in \mathscr{A}_t$ for all $t \ge 0$.

(a) If the $\mathscr{A}_t$ are right-continuous (i.e. $\bigcap_{j>i} \mathscr{A}_j = \mathscr{A}_i$ for all i or equivalently $\bigcap_{m=1}^{\infty} \mathscr{A}_{i+(1/m)} = \mathscr{A}_i$), then $T_1 + T_2$ is also a stopping time:

$$\begin{aligned}[T_1 + T_2 \le t] &= \{\omega \in \Omega : \forall m \in \mathbb{N}, \exists a \in \mathbb{Q} \text{ s.t. } T_1(\omega) \le a + \tfrac{1}{m} \text{ and } T_2(\omega) \le (t-a) + \tfrac{1}{m}\} \\ &= \bigcap_{m=1}^{\infty} \bigcup_{a \in \mathbb{Q}} [T_1 \le a + \tfrac{1}{m}] \cap [T_2 \le (t-a) + \tfrac{1}{m}]\end{aligned}$$

but $a < -\frac{1}{m}$ or $a > t + \frac{1}{m} \implies [T_1 \le a + \frac{1}{m}] \cap [T_2 \le (t-a) + \frac{1}{m}] = \varnothing$, so for any $a \in \mathbb{Q}$, $[T_1 \le a + \frac{1}{m}] \cap [T_2 \le (t-a) + \frac{1}{m}]$ is definitely in $\mathscr{A}_{t+(2/m)}$. Thus the sets that remain after intersecting over all $m \in \mathbb{N}$ must be in the $\mathscr{A}_{t+(2/m)}$ for all $m \in \mathbb{N}$, which by right-continuity means that $[T_1 + T_2 \le t] \in \mathscr{A}_t$.

(b) Define $\mathscr{A}_{T_1} := \{A \in \mathscr{A} : A \cap [T_1 \le t] \in \mathscr{A}_t \text{ for all } t \ge 0\}$; similarly $\mathscr{A}_{T_2}$. Now if $A \in \mathscr{A}_{T_1}$, then $A \cap [T_1 \le T_2] \in \mathscr{A}_{T_2}$ because

$$[T_1 \le T_2 \le t] = \{\omega \in \Omega : \forall m \in \mathbb{N}, \exists a \in \mathbb{Q}_{[0,t-\frac{1}{m}]} \text{ s.t. } T_1 \le a + \tfrac{1}{m} \text{ and } a - \tfrac{1}{m} < T_2 \le t\}$$

where $\mathbb{Q}_m := \mathbb{Q}_{[0,t-\frac{1}{m}]} := \mathbb{Q} \cap [0, t - \frac{1}{m}) \cup \{t - \frac{1}{m}\}$, and so

$$A \cap [T_1 \le T_2] \cap [T_2 \le t] = \big(A \cap [T_1 \le t]\big) \cap \Big(\bigcap_{m=1}^{\infty} \bigcup_{a \in \mathbb{Q}_m} [T_1 \le a + \tfrac{1}{m}] \cap [a - \tfrac{1}{m} < T_2 \le t]\Big) \in \mathscr{A}_t$$

Taking $A = \Omega$, $[T_1 \le T_2] \in \mathscr{A}_{T_2} \implies [T_1 > T_2] \in \mathscr{A}_{T_2}$ and by symmetry $[T_2 \le T_1] \in \mathscr{A}_{T_1} \implies [T_2 > T_1] \in \mathscr{A}_{T_1}$. Furthermore, $[T_2 \le T_1] \in \mathscr{A}_{T_1} \implies [T_2 \le T_1] \cap [T_1 \le T_2] = [T_1 = T_2] \in \mathscr{A}_{T_2}$

and so it is clear that $[T_1 \le T_2], [T_1 = T_2]$, and $[T_1 \ge T_2]$ are all in both $\mathscr{A}_{T_1}$ and $\mathscr{A}_{T_2}$.

(c) If $T_1 \le T_2$ then $[T_1 \le T_2] = \Omega$ and so $A \in \mathscr{A}_{T_1} \implies A \in \mathscr{A}_{T_2}$, i.e. $\mathscr{A}_{T_1} \subseteq \mathscr{A}_{T_2}$. Therefore, $\mathscr{A}_{\min\{T_1,T_2\}} = \mathscr{A}_{T_1} \cap \mathscr{A}_{T_2}$ because $\min\{T_1, T_2\} \le T_1$ and $T_2 \implies \mathscr{A}_{\min\{T_1,T_2\}} \subseteq \mathscr{A}_{T_1}$ and $\mathscr{A}_{T_2}$, and because $A \cap [\min\{T_1, T_2\} \le t] = A \cap \big([T_1 \le t] \cup [T_2 \le t]\big) = \big(A \cap [T_1 \le t]\big) \cup \big(A \cap [T_2 \le t]\big) \implies \big(A \in T_1 \cap T_2 \implies A \in \mathscr{A}_{\min\{T_1,T_2\}}\big)$.

Additionally $A \in \mathscr{A}_{T_1} \implies A \cap [T_1 \le T_2] \in \mathscr{A}_{T_2}$, but also $[T_1 \le T_2] \in \mathscr{A}_{T_1}$ so $A \in \mathscr{A}_{T_1} \implies A \in [T_1 \le T_2] \in \mathscr{A}_{T_1} \cap \mathscr{A}_{T_2}$.

(d) If we have $\{T_n\}$ such that $T_n \ge T_{n+1} \ge \ldots \ge T$ and $\lim_{n\to\infty} T_n = T$, then we know that $\mathscr{A}_T \subseteq \ldots \subseteq \mathscr{A}_{T_{n+1}} \subseteq \mathscr{A}_{T_n}$ and so $\mathscr{A}_T \subseteq \bigcap_{i=n}^{\infty} \mathscr{A}_{T_i}$. For the other direction, it would be easy if $[T \le t] = \bigcap_{i=n}^{\infty} [T_i \le t]$; this of course does not work because all the T_n could be $> t$ while still having their limit be $T = t$. The fix is to write

$$\begin{aligned}[T \le t] &= \{\omega \in \Omega : \forall m \in \mathbb{N}, \exists N \in \mathbb{N} \text{ s.t. } \forall i \ge N,\ T_i \le t + \tfrac{1}{m}\} \\ &= \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{i=N}^{\infty} [T_i \le t + \tfrac{1}{m}] \in \bigcap_{m=1}^{\infty} \mathscr{A}_{t+(1/m)}\end{aligned}$$

We are allowed right-continuity, so $[T \le t] \in \mathscr{A}_t$. Adding back an $A \in \bigcap_{i=n}^{\infty} \mathscr{A}_{T_i}$,

$$\begin{aligned}A \cap [T \le t] &= A \cap \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{i=N}^{\infty} [T_i \le t + \tfrac{1}{m}] \\ &= \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{i=N}^{\infty} A \cap [T_i \le t + \tfrac{1}{m}] \in \bigcap_{m=1}^{\infty} \mathscr{A}_{t+(1/m)}\end{aligned}$$

and so $A \cap [T \le t] \in \mathscr{A}_t$ for all $t \ge 0$.

Problem 5

Let $\{X_n, \mathscr{A}_n\}_{n\ge0}$ be a submartingale with $X_n \ge 0$. Then for any $r > 1$, $\{X_n^r\}$ is uniformly integrable if and only if $\{X_n^r\}$ is integrable:

($\implies$) recall that being uniformly integrable on a finite measure space ($\mu(\Omega) < \infty$) implies being integrable:

$$\int_{\Omega} |f| \, d\mu = \int_{[|f|>\lambda]} |f| \, d\mu + \int_{[|f|\le\lambda]} |f| \, d\mu \le \epsilon + \lambda \cdot \mu(\Omega) < \infty$$

for a sufficiently large λ.

($\impliedby$) $g(x) = |x|^r$ is convex, and so $\{X_n^r, \mathscr{A}_n\}_{n\ge0}$ is also a submartingale. The (sub)-martingale convergence theorem tells us that $X_n^r \to_{\text{a.s.}} X_\infty^r \in \mathscr{L}_1$ (everything still ≥ 0), and that $\{X_n^r\}$ is uniformly integrable $\iff$ $\{X_n^r\}$ is integrable and $\limsup_{n\to\infty} \mathbb{E}[|X_n^r|] \le \mathbb{E}[|X_\infty^r|] < \infty$ and so we just need to show $\limsup_{n\to\infty} \mathbb{E}[X_n^r] \le \mathbb{E}[X_\infty^r] < \infty$.

First we show that $\{X_n\}$ is uniformly integrable (and hence that X_n converges a.s. to an $X_\infty \in \mathscr{L}_1$

that closes the martingale which would give that $\mathbb{E}(X_\infty|\mathscr{A}_n) \geq_{\text{a.s.}} X_n$, $\forall n \in \mathbb{N}$):

$$
\begin{aligned}
X_n \cdot 1_{[X_n>\lambda]} &= \frac{\lambda^{r-1}}{\lambda^{r-1}} X_n \cdot 1_{[X_n>\lambda]} < \frac{X_n^{r-1} \cdot X_n \cdot 1_{[X_n>\lambda]}}{\lambda^{r-1}} \\
\implies \mathbb{E}\big[X_n \cdot 1_{[X_n>\lambda]}\big] &\leq \frac{1}{\lambda^{r-1}} \mathbb{E}\big[X_n^r \cdot 1_{[X_n>\lambda]}\big] \leq \frac{1}{\lambda^{r-1}} \mathbb{E}[X_n^r] \leq \frac{1}{\lambda^{r-1}} \sup_{n\in\mathbb{N}} \mathbb{E}[X_n^r]
\end{aligned}
$$

which goes to 0 as $\lambda \to \infty$. Then as advertised $\mathbb{E}(X_\infty|\mathscr{A}_n) \geq_{\text{a.s.}} X_n \implies \mathbb{E}(X_\infty|\mathscr{A}_n)^r \geq_{\text{a.s.}} X_n^r \implies \mathbb{E}[X_n^r] \leq \mathbb{E}[\mathbb{E}(X_\infty|\mathscr{A}_n)^r] \leq \mathbb{E}[\mathbb{E}(X_\infty^r|\mathscr{A}_n)] = \mathbb{E}[X_\infty^r]$ for all n, and so clearly $\limsup_{n\to\infty} \mathbb{E}[X_n^r] \leq \mathbb{E}[X_\infty^r]$.

Finally, Fatou (which only requires $f \geq 0$) tells us that

$$
\mathbb{E}[X_\infty^r] = \mathbb{E}\Big[\lim_{n\to\infty} X_n^r\Big] = \mathbb{E}\Big[\liminf_{n\to\infty} X_n^r\Big] \leq \liminf_{n\to\infty} \mathbb{E}[X_n^r] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[X_n^r] < \infty
$$

and we are done.

Problem 6

Let's start at generation zero with a single individual, who then produces some individuals in the first generation, and so on. Let Z_n be the number of individuals from the n-th generation (so $Z_0 = 1$). Let X_{nj} denote the number of offspring of the j-th individual of the n-th generation (so $X_{01} = Z_1$). Then $Z_{n+1} = \sum_{j=1}^{Z_n} X_{nj}$ (for $n \geq 0$). Furthermore, all the X's are i.i.d., and $p_k = P(X = k)$ for $k \in \{0, 1, 2, \ldots\}$, and $m := \mathbb{E}[X] = \sum_{k=0}^{\infty} k p_k < \infty$ and $p_0 > 1$ and $p_0 + p_1 < 1$.

(a) Define $W_n := Z_n/m^n$ and $\mathscr{A}_n := \sigma[W_1, \ldots, W_n]$. Then $\{W_n, \mathscr{A}_n\}_{n\geq 0}$ is a martingale with $\mathbb{E}[W_n] = 1$. Furthermore, if $\sigma^2 := \text{Var}[X] < \infty$, we can calculate $\text{Var}[W_n]$:

- When $m = 1$, $W_n = Z_n$ and so $\text{Var}[W_n] = \text{Var}[Z_n]$ where

$$
\begin{aligned}
\text{Var}[Z_{n+1}] &= \mathbb{E}[\text{Var}(Z_{n+1}|Z_n)] + \text{Var}[\mathbb{E}(Z_{n+1}|Z_n)] \\
&= \mathbb{E}\left[\text{Var}\left(\sum_{j=1}^{Z_n} X_{nj} \middle| Z_n\right)\right] + \text{Var}\left[\mathbb{E}\left(\sum_{j=1}^{Z_n} X_{nj} \middle| Z_n\right)\right] \\
&= \mathbb{E}[Z_n \text{Var}[X_{nj}]] + \text{Var}[Z_n \mathbb{E}[X_{nj}]] = \sigma^2 \mathbb{E}[Z_n] + \text{Var}[Z_n]
\end{aligned}
$$

  Notice that in the computation above we said that $\mathbb{E}[Z_{n+1}] = \mathbb{E}[\mathbb{E}(Z_{n+1}|Z_n)] = \mathbb{E}[Z_n \mathbb{E}[X_n j]] = \mathbb{E}[Z_n]$, and so all the $\mathbb{E}[Z_n] = \mathbb{E}[Z_0] = 1$. The formula then becomes $\text{Var}[Z_{n+1}] = \sigma^2 + \text{Var}[Z_n]$, where $\text{Var}[Z_0] = 0$, and so $\text{Var}[Z_{n+1}] = (n+1)\sigma^2$. More beautifully that's $\text{Var}[W_n] = n\sigma^2$.

- For $m \neq 1$, note that $\mathbb{E}[Z_{n+1}] = \mathbb{E}[\mathbb{E}(Z_{n+1}|Z_n)] = \mathbb{E}[Z_n \mathbb{E}[X_n j]] = m\mathbb{E}[Z_n] \implies \mathbb{E}[Z_n] = m^n$. Then,

$$
\begin{aligned}
\text{Var}[Z_{n+1}] &= \mathbb{E}[\text{Var}(Z_{n+1}|Z_n)] + \text{Var}[\mathbb{E}(Z_{n+1}|Z_n)] \\
&= \mathbb{E}\left[\text{Var}\left(\sum_{j=1}^{Z_n} X_{nj} \middle| Z_n\right)\right] + \text{Var}\left[\mathbb{E}\left(\sum_{j=1}^{Z_n} X_{nj} \middle| Z_n\right)\right] \\
&= \mathbb{E}[Z_n \text{Var}[X_{nj}]] + \text{Var}[Z_n \mathbb{E}[X_{nj}]] = \sigma^2 \mathbb{E}[Z_n] + \text{Var}[m Z_n] \\
&= \sigma^2 m^n + m^2 \text{Var}[Z_n]
\end{aligned}
$$

Listing the first couple $\mathrm{Var}[Z_n]$ out we get that

n	$\mathrm{Var}[Z_n]$
0	0
1	$\sigma^2 m^0 + m^2 0 = \sigma^2$
2	$\sigma^2 m^1 + m^2(\sigma^2) = \sigma^2 m^1 + \sigma^2 m^2$
3	$\sigma^2 m^2 + m^2(\sigma^2 m^1 + \sigma^2 m^2) = \sigma^2 m^2 + \sigma^2 m^3 + \sigma^2 m^4$
4	$\sigma^2 m^3 + \sigma^2 m^4 + \sigma^2 m^5 + \sigma^2 m^6$
5	$\sigma^2 m^4 + \sigma^2 m^5 + \sigma^2 m^6 + \sigma^2 m^7 + \sigma^2 m^8$

and so it is clear that $\mathrm{Var}[Z_{n+1}] = \sigma^2 \sum_{i=n}^{2n} m^i = \sigma^2 m^n \sum_{i=0}^{n} m^i = \sigma^2 m^n \frac{1-m^{n+1}}{1-m}$. Thus,

$$\begin{aligned}\mathrm{Var}[W_{n+1}] = \mathrm{Var}\left[\frac{Z_{n+1}}{m^{n+1}}\right] &= \frac{1}{m^{2n+2}}\mathrm{Var}[Z_{n+1}] \\ &= \frac{\sigma^2 m^n(1-m^{n+1})}{(1-m)m^{2n+2}} = \sigma^2\frac{m^{n+1}-1}{m^{n+2}(m-1)} = \sigma^2\frac{1-m^{-(n+1)}}{m(m-1)}\end{aligned}$$

and so $\mathrm{Var}[W_n] = \sigma^2\frac{1-m^{-n}}{m(m-1)}$.

(b) If we define generating functions f and f_n of X and Z_n resp. by $f(s) := \sum_{k=0}^{\infty} s^k p_k$ and $f_n(s) := \sum_{k=0}^{\infty} s^k P(Z_n = k)$.

$$\begin{aligned}f(f_n(s)) &= \sum_{i=0}^{\infty}\left(\sum_{j=0}^{\infty} s^j P(Z_n = j)\right)^i p_i \\ &= \sum_{i=0}^{\infty}\left(\sum_{a_1+a_2+\ldots=i}\frac{i!}{a_1!\cdots a_j!}s^{a_1+2a_2+\ldots}P(Z_n=1)^{a_1}\cdot P(Z_n=2)^{a_2}\cdots\right)p_i\end{aligned}$$

The coefficients of s^k will be :(

522 Homework 4

Daniel Rui - 2/12/20

Problem 1

Let $X_1, X_2, \ldots$ be independent r.v.'s with $X_k \geq 0$ and $\mathbb{E}[X_k] = 1$, and let $M_n := \prod_{k=1}^n X_k$ with $M_0 := 1$. Then — recalling that $\mathbb{E}(XY|\mathscr{D}) =_{\text{a.s.}} Y\mathbb{E}(X|\mathscr{D})$ if Y is $\mathscr{D}$-measurable, and $\mathbb{E}(Y|\mathscr{D}) =_{\text{a.s.}} \mathbb{E}[Y]$ if $\mathscr{F}(Y)$ and $\mathscr{D}$ are independent — $\{M_n, \mathscr{A}_n\}_{n\geq 0}$ (where $\mathscr{A}_n := \sigma[X_1, \ldots, X_n]$) is a martingale with $\mathbb{E}[M_n] = 1$ for all M_n:

$$\mathbb{E}(M_{n+1}|\mathscr{A}_n) = \mathbb{E}(X_{n+1} \cdot M_n|\mathscr{A}_n) = M_n\mathbb{E}(X_{n+1}|\mathscr{A}_n) =_{\text{a.s.}} M_n \cdot \mathbb{E}[X_{n+1}] =_{\text{a.s.}} M_n$$

Problem 2

Recall from class that $\mathbb{M}_t := \mathbb{N}_t - \lambda t$ (for $\{\mathbb{N}_t\}$ a Poisson process with $\lambda > 0$) and $\mathbb{M}_t^2 - \lambda t$ are both martingales. We want to recover both of these (and more!) from the exponential martingale

$$\mathbb{Y}(c,t,\omega) := \frac{e^{c\mathbb{M}_t(\omega)}}{\mathbb{E}\big[e^{c\mathbb{M}_t(\omega)}\big]} = \frac{e^{c\mathbb{N}_t(\omega)}/e^{c\lambda t}}{\mathbb{E}\big[e^{c\mathbb{N}_t(\omega)}\big]/e^{c\lambda t}} = \frac{e^{c\mathbb{N}_t(\omega)}}{\mathbb{E}\big[e^{c\mathbb{N}_t(\omega)}\big]}$$

Note that $\mathbb{N}_t \sim \text{Poisson}(\lambda t)$, so the induced measure is $F_{\mathbb{N}_t}(A) = \sum_{k\in A\cap\mathbb{Z}_{\geq 0}} \frac{(\lambda t)^k e^{-\lambda t}}{k!}$. Thus by the law of the unconscious statistician,

$$\mathbb{E}\big[e^{c\mathbb{N}_t}\big] = \int_{\mathbb{R}} e^{cx}\, dF_{\mathbb{N}_t}(x) = \sum_{k=0}^{\infty} e^{ck}\frac{(\lambda t)^k e^{-\lambda t}}{k!} = e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(e^c\lambda t)^k}{k!} = e^{-\lambda t}e^{e^c\lambda t} = e^{(e^c-1)\lambda t}$$

Therefore, $\mathbb{Y}(c,t,\omega) = \dfrac{e^{c\mathbb{N}_t(\omega)}}{e^{(e^c-1)\lambda t}}$. To verify that these $\mathbb{Y}$ do indeed form a martingale, observe that

$$\begin{aligned}\mathbb{E}(\mathbb{Y}(c,t)|\mathscr{A}_s) &= \frac{1}{e^{(e^c-1)\lambda t}}\mathbb{E}\big(e^{c\mathbb{N}_t}|\mathscr{A}_s\big) = \frac{1}{e^{(e^c-1)\lambda t}}\mathbb{E}\big(e^{c\mathbb{N}_t - c\mathbb{N}_s}\cdot e^{c\mathbb{N}_s}|\mathscr{A}_s\big)\\ &=_{\text{a.s.}} \frac{e^{c\mathbb{N}_s}}{e^{(e^c-1)\lambda t}}\mathbb{E}\big(e^{c\mathbb{N}_t - c\mathbb{N}_s}|\mathscr{A}_s\big) =_{\text{a.s.}} \frac{e^{c\mathbb{N}_s}}{e^{(e^c-1)\lambda t}}\mathbb{E}\big[e^{c\mathbb{N}_t - c\mathbb{N}_s}\big]\end{aligned}$$

$\mathbb{N}_t - \mathbb{N}_s$ counts the number of things that happen between s and t, so this is also $\sim \text{Poiss}(\lambda(t-s))$, and so

$$\mathbb{E}(\mathbb{Y}(c,t)|\mathscr{A}_s) =_{\text{a.s.}} \frac{e^{c\mathbb{N}_s}}{e^{(e^c-1)\lambda t}}e^{(e^c-1)\lambda(t-s)} = \frac{e^{c\mathbb{N}_s}}{e^{(e^c-1)\lambda s}} = \mathbb{Y}(c,s)$$

Lastly, we just do some differentiation:

$$\begin{aligned}\frac{\partial}{\partial c}\mathbb{Y}(c,t) &= (\mathbb{N}_t - e^c\lambda t)e^{c\mathbb{N}_t - (e^c-1)\lambda t}\\ \frac{\partial^2}{\partial c^2}\mathbb{Y}(c,t) &= (\mathbb{N}_t - e^c\lambda t)^2 e^{c\mathbb{N}_t-(e^c-1)\lambda t} - e^c\lambda t\cdot e^{c\mathbb{N}_t-(e^c-1)\lambda t}\end{aligned}$$

and so $\big[\frac{\partial}{\partial c}\mathbb{Y}\big](0,t) = \mathbb{N}_t - \lambda t$ and $\implies \big[\frac{\partial^2}{\partial c^2}\mathbb{Y}\big](0,t) = (\mathbb{N}_t - \lambda t)^2 - \lambda t$. See Problem 6 (now commented out) for higher derivatives.

Problem 3

For $X_1, \ldots, X_n$ independent with mean 0, $S_k = X_1 + \ldots + X_k$, and $\mathscr{A}_k = \sigma[X_1, \ldots, X_k]$, observe that (because $\mathbb{E}(Y|\mathscr{D}) =_{\text{a.s.}} Y$ if Y is $\mathscr{D}$-measurable, $\mathbb{E}(XY|\mathscr{D}) =_{\text{a.s.}} Y\mathbb{E}(X|\mathscr{D})$ if Y is $\mathscr{D}$-measurable, $\mathbb{E}(Y|\mathscr{D}) =_{\text{a.s.}} \mathbb{E}[Y]$ if $\mathscr{F}(Y)$ and $\mathscr{D}$ are independent, and linearity):

$$\begin{aligned}
\mathbb{E}\big(S_k^2|\mathscr{A}_i\big) &= \mathbb{E}\big((S_k - S_i)^2 + 2S_i(S_k - S_i) + S_i^2|\mathscr{A}_i\big) \\
&= \mathbb{E}\big((S_k - S_i)^2|\mathscr{A}_i\big) + 2\mathbb{E}(S_i(S_k - S_i)|\mathscr{A}_i) + \mathbb{E}\big(S_i^2|\mathscr{A}_i\big) \\
&=_{\text{a.s.}} \mathbb{E}\big[(S_k - S_i)^2\big] + S_i\mathbb{E}[S_k - S_i] + S_i^2 \\
&= \mathbb{E}\big[S_k^2\big] - 2\mathbb{E}[S_kS_i] + \mathbb{E}\big[S_i^2\big] + 0 + S_i^2 \\
&= \mathbb{E}\big[S_k^2\big] - 2\mathbb{E}\big[(S_k - S_i)S_i + S_i^2\big] + \mathbb{E}\big[S_i^2\big] + 0 + S_i^2 \\
&= \mathbb{E}\big[S_k^2\big] - 2\mathbb{E}[S_k - S_i]\mathbb{E}[S_i] - 2\mathbb{E}\big[S_i^2\big] + \mathbb{E}\big[S_i^2\big] + S_i^2 \\
&= \mathbb{E}\big[S_k^2\big] + S_i^2 - \mathbb{E}\big[S_i^2\big]
\end{aligned}$$

and so $\mathbb{E}\big[S_k^2 - \mathbb{E}\big[S_k^2\big]|\mathscr{A}_i\big] = S_i^2 - \mathbb{E}\big[S_i^2\big]$, implying that $\{S_k^2 - \mathbb{E}\big[S_k^2\big], \mathscr{A}_k\}_{1\le k\le n}$ is a martingale.

Problem 4

Defining $T_k := \frac{S_n}{b_n} + \sum_{n+1}^k \frac{X_i}{b_i}$, where $0 < b_1 \le \cdots \le b_N$, $\{S_k, \mathscr{A}_k\}_{1\le k\le N}$ is a martingale, $\mathbb{E}[S_k] = 0$, and $X_k := S_k - S_{k-1}$ (and $X_1 := S_1$), we have that for any j s.t. $n \le j < k$,

$$\begin{aligned}
\mathbb{E}(T_k|\mathscr{A}_j) = \mathbb{E}(T_k - T_j + T_j|\mathscr{A}_j) &= \mathbb{E}(T_k - T_j|\mathscr{A}_j) + \mathbb{E}(T_j|\mathscr{A}_j) =_{\text{a.s.}} \mathbb{E}\left(\sum_{j+1}^k \frac{X_i}{b_i}\middle|\mathscr{A}_j\right) + T_j \\
= \sum_{j+1}^k \frac{\mathbb{E}(X_i|\mathscr{A}_j)}{b_i} + T_j &= \sum_{j+1}^k \frac{\mathbb{E}(S_i|\mathscr{A}_j) - \mathbb{E}(S_{i-1}|\mathscr{A}_j)}{b_i} + T_j =_{\text{a.s.}} \sum_{j+1}^k \frac{S_j - S_j}{b_i} + T_j = T_j
\end{aligned}$$

and so $\{T_k, \mathscr{A}_k\}_{n\le k\le N}$ is a martingale. For the variance, we can just go back to the basic definition (let's overwrite $b_1, \ldots, b_{n-1}$ to all just be b_n; it's not like they show up anyways), keeping in mind that S_k is A_k-measureable, the $\mathscr{A}_k$ are increasing (hence X_k is also A_k-measureable), and that $\mathbb{E}(XY|\mathscr{D}) =_{\text{a.s.}} Y\mathbb{E}(X|\mathscr{D})$ if Y is $\mathscr{D}$-measurable:

$$\begin{aligned}
\mathrm{Var}[T_N] &= \mathbb{E}\big[T_N^2\big] - \mathbb{E}[T_N]^2 = \mathbb{E}\big[T_N^2\big] = \mathbb{E}\left[\left(\frac{S_n}{b_n} + \sum_{n+1}^N \frac{X_i}{b_i}\right)^2\right] = \mathbb{E}\left[\left(\frac{\sum_1^n X_i}{b_n} + \sum_{n+1}^N \frac{X_i}{b_i}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_1^n \frac{X_i}{b_n} + \sum_{n+1}^N \frac{X_i}{b_i}\right)^2\right] = \mathbb{E}\left[\left(\sum_1^N \frac{X_i}{b_i}\right)^2\right] = \mathbb{E}\left[\sum_1^N \frac{X_i^2}{b_i^2} + \sum_{1\le i\ne j\le N} \frac{X_iX_j}{b_ib_j}\right] \\
&= \sum_1^N \frac{\mathbb{E}\big[X_i^2\big]}{b_i^2} + 2\sum_{1\le i<j\le N} \frac{\mathbb{E}[X_iX_j]}{b_ib_j} = \sum_1^N \frac{\sigma_i^2}{b_i^2} + 2\sum_{1\le i<j\le N} \frac{\mathbb{E}\Big[\mathbb{E}\Big(X_iX_j\Big|\mathscr{A}_{j-1}\Big)\Big]}{b_ib_j} \\
&= \sum_1^N \frac{\sigma_i^2}{b_i^2} + 2\sum_{1\le i<j\le N} \frac{\mathbb{E}\Big[X_i\mathbb{E}\Big(X_j\Big|\mathscr{A}_{j-1}\Big)\Big]}{b_ib_j} = \sum_1^N \frac{\sigma_i^2}{b_i^2} + 2\sum_{1\le i<j\le N} \frac{\mathbb{E}[X_i(S_{j-1} - S_{j-1})]}{b_ib_j} = \sum_1^N \frac{\sigma_i^2}{b_i^2}
\end{aligned}$$

which when we un-overwrite the $b_1, \ldots, b_{n-1}$, becomes $\sum_1^n \frac{\sigma_i^2}{b_n^2} + \sum_{n+1}^N \frac{\sigma_i^2}{b_i^2}$

Problem 5

Let $Y_1, Y_2, \ldots$ be independent on some probability space $(\Omega, \mathscr{A}, \mathsf{P})$, and suppose (assume?) they have density p_k (w.r.t. some measure μ), i.e. $F_{Y_k}(B) = P_k(B) = \int_B p_k \, d\mu$, i.e. $p_k = \frac{dP_k}{d\mu}$, the Radon-Nikodym derivative. Now suppose that we have some other candidate density q_k (w.r.t the same measure μ) s.t. $Q_k \ll P_k \ll \mu$. Define $X_k = \frac{q_k(Y_k)}{p_k(Y_k)}$ (which is ≥ 0 because densities are ≥ 0). Then:

(a) defining $M_n := \prod_{k=1}^n X_k$ and $\mathscr{A}_k = \sigma[Y_1, \ldots, Y_k]$, $\{M_n, \mathscr{A}_k\}_{k \geq 1}$ is essentially Kakutani's martingale. Note our usage of the law of the unconscious statistician and of the change of variables theorem:

$$\begin{aligned}
\mathbb{E}(M_{n+1}|\mathscr{A}_n) &= \mathbb{E}\left(\frac{q_{n+1}(Y_{n+1})}{p_{n+1}(Y_{n+1})} \cdot M_n \middle| \mathscr{A}_n\right) =_{\text{a.s.}} M_n \mathbb{E}\left[\frac{q_{n+1}(Y_{n+1})}{p_{n+1}(Y_{n+1})}\right] \\
&= M_n \int_\Omega \frac{q_{n+1}(Y_{n+1})}{p_{n+1}(Y_{n+1})} \, d\mathsf{P} = M_n \int_\mathbb{R} \frac{q_{n+1}(x)}{p_{n+1}(x)} \, dF_{Y_{n+1}}(x) \\
&= M_n \int_\mathbb{R} \frac{q_{n+1}(t)}{p_{n+1}(t)} \frac{dF_{Y_{n+1}}}{d\mu}(t) \, d\mu(t) = M_n \int_\mathbb{R} \frac{q_{n+1}(t)}{p_{n+1}(t)} p_{n+1}(t) \, d\mu(t) \\
&= M_n \int_\mathbb{R} q_{n+1}(t) \, d\mu(t) = M_n Q_{n+1}(\mathbb{R}) = M_n
\end{aligned}$$

Lastly, regarding the mean: $\mathbb{E}[M_n] = \mathbb{E}\left[\prod_{k=1}^n X_k\right] = \prod_{k=1}^n \mathbb{E}[X_k] = \prod_{k=1}^n \int_\Omega \frac{q_k(Y_k)}{p_k(Y_k)} \, d\mathbf{P} = \prod_{k=1}^n 1 = 1.$

(b) note that the P_k and Q_k are all measures on the real line; i.e. $P_k, Q_k : \mathscr{B} \to \mathbb{R}$. Define $\mathscr{A}_k = \sigma[Y_k] = Y_k^{-1}(\mathscr{B})$. Then for any $A_k \in \mathscr{A}_k$, there is some $B \in \mathscr{B}$ s.t. $Y_k^{-1}(B) = A_k$. Thus for every $A_k \in \mathscr{A}_k$ and its corresponding B, we can define $\widetilde{P}_k(A_k) = P_k(B)$, meaning that $\widetilde{P}_k : \mathscr{A}_k \to \mathbb{R}$. We can similarly do this to Q_k to get $\widetilde{Q}_k : \mathscr{A}_k \to \mathbb{R}$, $Q_k(B) = \widetilde{Q}_k(A_k)$.

Now because $Q_k \ll P_k$, there is some function $\frac{dQ_k}{dP_k}$ s.t. $Q_k(B) = \int_B \frac{dQ_k}{dP_k} \, dP_k$. The law of the unconscious statistician gives that this integral is $= \int_{Y_k^{-1}(B)} \frac{dQ_k}{dP_k}(Y_k) \, d\widetilde{P}_k = \int_{A_k} \frac{dQ_k}{dP_k}(Y_k) \, d\widetilde{P}_k$ (because $P_k(B) = \widetilde{P}_k(Y_k^{-1}(B))$ is in fact the induced measure). But $Q_k(B) = \widetilde{Q}_k(A_k)$ so $\widetilde{Q}_k(A_k) = \int_{A_k} \frac{dQ_k}{dP_k}(Y_k) \, d\widetilde{P}_k$!

Furthermore $\widetilde{Q}_k \ll \widetilde{P}_k$ (b/c $\widetilde{P}_k(A_k) = 0 \iff P_k(B) = 0 \implies Q_k(B) = 0 \iff \widetilde{Q}_n(A_k) = 0$), meaning that there is a unique (a.s.) $\frac{d\widetilde{Q}_k}{d\widetilde{P}_k}$ s.t. $\widetilde{Q}_k(A_k) = \int_{A_k} \frac{d\widetilde{Q}_k}{d\widetilde{P}_k} \, d\widetilde{P}_k$. Thus, $\frac{d\widetilde{Q}_k}{d\widetilde{P}_k} =_{\text{a.s.}} \frac{dQ_k}{dP_k}(Y_k)$. This relates to the X_k above because the chain rule for Radon-Nikodym derivatives tells us that,

$$q_k = \frac{dQ_k}{d\mu} =_{\text{a.s.}} \frac{dQ_k}{dP_k} \cdot \frac{dP_k}{d\mu} = \frac{dQ_k}{dP_k} \cdot p_k \implies \frac{dQ_k}{dP_k} =_{\text{a.s.}} \frac{q_k}{p_k}$$

and hence $\widetilde{Q}_k(A_k) = \int_{A_k} X_k \, d\widetilde{P}_k$, $\forall A_k \in \mathscr{A}_k$. We are now ready to define the product measure $\widetilde{\mathbf{P}}_\mathbf{n} = \prod_{k=1}^n \widetilde{P}_k$, where $\widetilde{\mathbf{P}}_\mathbf{n}(A_1 \times A_2 \times \ldots \times A_n) = \widetilde{P}_1(A_1) \cdot \widetilde{P}_2(A_2) \cdots \widetilde{P}_n(A_n)$ (similarly $\mathbf{Q}_n$). Because the Radon-Nikodym derivative of a product measure is the product of the R-N derivatives, we have that

$$\widetilde{\mathbf{Q}}_\mathbf{n}(\mathbf{A}) = \int_\mathbf{A} \prod_{k=1}^n X_k \, d\widetilde{\mathbf{P}}_\mathbf{n}$$

for any $\mathbf{A} \in \mathscr{A}_{\boldsymbol{n}} := \sigma[\{A_1 \times \ldots \times A_n : A_i \in \mathscr{A}_i\}]$. Thus, $M_n = \prod_{k=1}^{n} X_k$ is equal (a.s.) to $\frac{d\widetilde{\mathbf{Q}}_{\mathbf{n}}}{d\widetilde{\mathbf{P}}_{\mathbf{n}}}$. If we think of the infinite dimensional σ-algebra $\mathscr{A} := \sigma[\{A_1 \times A_2 \times \ldots : A_i \in \mathscr{A}_i\}]$ and the infinite product measures $\widetilde{\boldsymbol{P}} := \prod_{k=1}^{\infty} \widetilde{P}_k$ and $\widetilde{\boldsymbol{Q}} := \prod_{k=1}^{\infty} \widetilde{Q}_k$ (both measures : $\mathscr{A} \to \mathbb{R}$), then $\widetilde{\boldsymbol{P}}_{\boldsymbol{n}} = \widetilde{\boldsymbol{P}}|_{\mathscr{A}_{\boldsymbol{n}}}$ (restriction to a sub-σ-field $\mathscr{A}_{\boldsymbol{n}}$, where we tweak $\mathscr{A}_{\boldsymbol{n}}$ to be $:= \sigma[\{A_1 \times \ldots \times A_n \times \Omega \times \ldots : A_i \in \mathscr{A}_i\}]$ and so $\mathscr{A}_{\boldsymbol{n}} \subseteq \mathscr{A}_{\boldsymbol{n+1}}$), and same for $\widetilde{\boldsymbol{Q}}_{\boldsymbol{n}}$. Thus $\{M_n, \mathscr{A}_{\boldsymbol{n}}\}_{n \geq 1}$ fits the form of a likelihood ratio martingale: for any $\mathbf{A} \in \mathscr{A}_{\boldsymbol{m}}$ $(m < n)$,

$$\int_{\mathbf{A}} M_n \,\mathrm{d}\widetilde{\boldsymbol{P}} = \int_{\mathbf{A}} M_n \,\mathrm{d}\widetilde{\boldsymbol{P}}_{\boldsymbol{n}} = \widetilde{\boldsymbol{Q}}_{\boldsymbol{n}}(\mathbf{A}) = \widetilde{\boldsymbol{Q}}_{\boldsymbol{m}}(\mathbf{A}) = \int_{\mathbf{A}} M_m \,\mathrm{d}\widetilde{\boldsymbol{P}}_{\boldsymbol{m}} = \int_{\mathbf{A}} M_m \,\mathrm{d}\widetilde{\boldsymbol{P}}$$

implying that $\int_{\mathbf{A}} M_n - M_m \,\mathrm{d}\widetilde{\boldsymbol{P}} = 0$ for all $\mathbf{A} \in \mathscr{A}_{\boldsymbol{m}}$.

522 Homework 3

Daniel Rui - 2/5/20

Problem 1

We have X, Y i.i.d. with continuous d.f. F, and $M = \max\{X, Y\}$. We want to verify that

$$P(X \leq x|M) = 1_{[M \leq x]} + \tfrac{1}{2}\tfrac{F(x)}{F(M)} \cdot 1_{[M>x]}$$

From definition, $P(X \leq x|M) := P(X \leq x|\mathscr{F}(M)) := \mathbb{E}\big(1_{[X \leq x]}|\mathscr{F}(M)\big)$, so the following must hold:

$$\int_D \mathbb{E}\big(1_{[X\leq x]}|\mathscr{F}(M)\big)\, dP = \int_D 1_{[X \leq x]}\, dP, \qquad \forall D \in \mathscr{F}(M)$$

It suffices to show the equality for all $D \in \{[M \leq m] : m \in \mathbb{R}\}$ (cf. Axler MIRA). Note that we are integrating w.r.t. $\omega \in \Omega$ here, not x, and so $F(x)$ can be taken out of the integral as a constant.

$$\begin{aligned}\int_{[M\leq m]} \mathbb{E}\big(1_{[X\leq x]}|\mathscr{F}(M)\big)\, dP &= \int_{[M\leq m]} 1_{[M\leq x]} + \tfrac{1}{2}\frac{F(x)}{F(M)} \cdot 1_{[M>x]}\, dP \\ =&P([M \leq m] \cap [M \leq x]) + \tfrac{1}{2}F(x)\int_{[x<M\leq m]} \frac{1}{F(M)}\, dP\end{aligned}$$

If $x \geq m$, then the 2nd term is 0 and the 1st term is $P([M \leq m])$. Furthermore, $[M \leq m] \subseteq [X \leq x]$ (b/c for any ω, $M(\omega) \leq m \implies X(\omega) \leq m \leq x$). Thus, $P([M \leq m]) = P([M \leq m] \cap [X \leq x])$.
Now for $x < m$: note that $F_M(x) = P([M \leq x]) = P([X \leq x] \cap [Y \leq x]) = F^2(x)$. Then, by successive applications of the law of the unconscious statistician, we get that

$$\int_{[x<M\leq m]} \frac{1}{\sqrt{F_M(M)}}\, dP = \int_x^m \frac{1}{\sqrt{F_M(t)}}\, dF_M(t) = \int_{F_M(x)}^{F_M(m)} \frac{1}{\sqrt{y}}\, d\lambda(y)$$

and so the expression just becomes $P([M \leq x]) + \frac{1}{2}F(x) \cdot 2(\sqrt{F_M(m)} - \sqrt{F_M(x)}) = F^2(x) + P([X \leq x])(F(m) - F(x)) = P([X \leq x])P([Y \leq m]) = P([X \leq x] \cap [Y \leq m]) = P([M \leq m] \cap [X \leq x])$.

Problem 2

(a) Defining conditional variance as $\text{Var}[Y|X] = \mathbb{E}\Big(\big(Y - \mathbb{E}(Y|X)\big)^2\Big|X\Big)$, denoting $Y_X := \mathbb{E}(Y|X)$, and using the properties $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y|\mathscr{D})]$, $\mathbb{E}(Y|\mathscr{D})$ is $\mathscr{D}$-measurable, $\mathbb{E}(Y_X|D) =_{\text{a.s.}} Y_X$ if Y_X is $\mathscr{D}$-msble, $\mathbb{E}(XY|\mathscr{D}) =_{\text{a.s.}} Y\mathbb{E}(X|\mathscr{D})$ if Y is $\mathscr{D}$-measurable, and linearity, we see that

$$\begin{aligned}\text{Var}[Y] &= \mathbb{E}\big[(Y - \mathbb{E}[Y])^2\big] = \mathbb{E}\big[(Y - Y_X + Y_X - \mathbb{E}[Y])^2\big] \\ &= \mathbb{E}\big[(Y - Y_X)^2\big] + 2\mathbb{E}\Big[(Y - Y_X)(Y_X - \mathbb{E}[Y])\Big] + \mathbb{E}\big[(Y_X - \mathbb{E}[Y])^2\big] \\ &= \mathbb{E}\Big[\mathbb{E}\big((Y - Y_X)^2|X\big)\Big] + 2\mathbb{E}\Big[\mathbb{E}\Big((Y - Y_X)(Y_X - \mathbb{E}[Y])\Big|X\Big)\Big] + \mathbb{E}\Big[\big(Y_X - \mathbb{E}[Y_X]\big)^2\Big] \\ &= \mathbb{E}\big[\text{Var}[Y|X]\big] + 2\mathbb{E}\Big[(Y_X - \mathbb{E}[Y])\mathbb{E}\Big((Y - Y_X)\Big|X\Big)\Big] + \text{Var}[Y_X] \\ &= \mathbb{E}\big[\text{Var}[Y|X]\big] + 2\mathbb{E}\Big[(Y_X - \mathbb{E}[Y])\Big(Y_X - Y_X\Big)\Big] + \text{Var}[Y_X] = \mathbb{E}\big[\text{Var}[Y|X]\big] + \text{Var}[Y_X]\end{aligned}$$

(b) We want to show that $Z := \mathbb{E}(Y|\mathscr{D})$ minimizes $\mathbb{E}\big[(Y-Z)^2\big]$ over all $\mathscr{D}$-measurable random variables in L^2. Let Z' be some arbitrary L^2 $\mathscr{D}$-measurable random variable. Using the same properties as above, we see that

$$
\begin{aligned}
\mathbb{E}\big[(Y-Z')^2\big] &= \mathbb{E}\big[(Y-Z+Z-Z')^2\big] \\
&= \mathbb{E}\big[(Y-Z)^2\big] + 2\mathbb{E}[(Y-Z)(Z-Z')] + \mathbb{E}\big[(Z-Z')^2\big] \\
&= \mathbb{E}\big[(Y-Z)^2\big] + 2\mathbb{E}\big[\mathbb{E}\big((Y-Z)(Z-Z')|\mathscr{D}\big)\big] + \mathbb{E}\big[(Z-Z')^2\big] \\
&= \mathbb{E}\big[(Y-Z)^2\big] + 2\mathbb{E}\big[(Z-Z')\mathbb{E}\big((Y-Z)|\mathscr{D}\big)\big] + \mathbb{E}\big[(Z-Z')^2\big] \\
&= \mathbb{E}\big[(Y-Z)^2\big] + 2\mathbb{E}\big[(Z-Z')(\mathbb{E}(Y|\mathscr{D})-Z)\big] + \mathbb{E}\big[(Z-Z')^2\big] \\
&= \mathbb{E}\big[(Y-Z)^2\big] + 0 + \mathbb{E}\big[(Z-Z')^2\big]
\end{aligned}
$$

which means that $\mathbb{E}\big[(Y-Z')^2\big] \geq \mathbb{E}\big[(Y-Z)^2\big]$, with equality only when $Z' = Z$ (a.s.).

Problem 3

If we assume that $\Omega = \bigsqcup_{i\in I} D_i$ for finite or countable I, and let $\mathscr{D} := \sigma[\{D_1, D_2, \ldots\}]$, then

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

We verify this by showing that $\int_{D_i} P(A|\mathscr{D})\,dP = \int_{D_i} \mathbb{E}(1_A|\mathscr{D})\,dP = \int_{D_i} 1_A\,dP$ for all $D_i \in \mathscr{D}$ (because $\{D_i\}$ is a π-system, Dynkin's $\pi-\lambda$ theorem (+ corollaries) says that two measures that agree on $\{D_i\}$ agree on $\mathscr{D}$). Below we probably have to use Fubini-Tonelli (but it's not that bad because the things we are integrating are ≥ 0):

$$
\int_{D_j} \sum_{i\in I} \frac{P(A\cap D_i)}{P(D_i)} 1_{D_i}\,dP = \sum_{i\in I} \frac{P(A\cap D_i)}{P(D_i)} \int_{D_j} 1_{D_i}\,dP
$$
$$
= \sum_{i\in I} \frac{P(A\cap D_i)}{P(D_i)} P(D_j\cap D_i) = \frac{P(A\cap D_j)}{P(D_j)} P(D_j) = P(A\cap D_j)
$$

where the sum collapses because the D_i are disjoint $\implies P(D_i\cap D_j) = 0$. Now for general $Y \in L^1$, we want to verify that $\int_{D_i} \mathbb{E}(Y|\mathscr{D})\,dP = \int_{D_i} Y\,dP$ where $\mathbb{E}(Y|\mathscr{D})$ is defined as

$$
\mathbb{E}(Y|\mathscr{D}) = \sum_{i\in I} \left(\frac{1}{P(D_i)} \int_{D_i} Y\,dP\right) \cdot 1_{D_i}.
$$

We proceed similar to above:

$$
\int_{D_j} \sum_{i\in I} \left(\frac{1}{P(D_i)} \int_{D_i} Y\,dP\right) \cdot 1_{D_i} = \sum_{i\in I} \left(\frac{1}{P(D_i)} \int_{D_i} Y\,dP\right) \int_{D_j} 1_{D_i}\,dP
$$
$$
= \sum_{i\in I} \left(\frac{1}{P(D_i)} \int_{D_i} Y\,dP\right) P(D_j\cap D_i) = \frac{1}{P(D_j)} \int_{D_j} Y\,dP \cdot P(D_j) = \int_{D_j} Y\,dP
$$

Problem 4

Given independent X, Y and $r \geq 1$, we want to show that $\mathbb{E}[|X+\mu_Y|^r] \leq \mathbb{E}[|X+Y|^r]$. Recall the properties $\mathbb{E}(Y|\mathscr{D}) =_{\text{a.s.}} \mathbb{E}[Y]$ if $\mathscr{F}(Y)$ and $\mathscr{D}$ are independent, and $g(\mathbb{E}(Y|\mathscr{D})) \leq_{\text{a.s.}} \mathbb{E}(g(Y)|\mathscr{D})$ for convex g (for this problem we'll take $g(x) = |x|^r$). Then,

$$|X+\mu_Y|^r = |X+\mathbb{E}(Y|X)|^r = |\mathbb{E}((X+Y)|X)|^r \leq \mathbb{E}\big(|X+Y|^r \big| X\big).$$

Finally, by monotonicity of expectation and by $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y|\mathscr{D})]$, we conclude with

$$\mathbb{E}[|X+\mu_Y|^r] \leq \mathbb{E}\big[\mathbb{E}\big(|X+Y|^r\big|X\big)\big] = \mathbb{E}[|X+Y|^r].$$

Problem 5

C_r inequality: $\mathbb{E}[|X+Y|^r] \leq C_r\mathbb{E}[|X^r|] + C_r\mathbb{E}[|Y^r|]$, where $r > 0$, $C_r = 2^{\max\{r,1\}-1}$.
Proof: First recall from the proof of the regular C_r inequality that $|X+Y|^r \leq C_r|X|^r + C_r|Y|^r$. From monotonicity and linearity of conditional expectation, we can place $\mathbb{E}(\cdot|\mathscr{D})$ around this equality to get

$$\mathbb{E}\big(\mathbb{E}[|X+Y|^r]\big|\mathscr{D}\big) \leq_{\text{a.s.}} \mathbb{E}\big(C_r|X|^r + C_r|Y|^r\big|\mathscr{D}\big) = C_r\mathbb{E}\big(|X|^r\big|\mathscr{D}\big) + C_r\mathbb{E}\big(|Y|^r\big|\mathscr{D}\big)$$

Hölder: $\mathbb{E}[|XY|] \leq \big(\mathbb{E}[|X|^r]\big)^{1/r}\big(\mathbb{E}[|X|^s]\big)^{1/s}$ where $r > 1$, $1/r + 1/s = 1$.
Proof: Like in the proof of the regular Hölder's inequality, we use Young's inequality (inheriting our conditions on r and s):

$$|ab| \leq \frac{|a|^r}{r} + \frac{|b|^s}{s} \quad \text{with equality exactly when } |a|^r = |b|^s \iff |b| = |a|^{r-1}$$

Now we let $a = \frac{|X|}{\left(\mathbb{E}\left(|X|^r\middle|\mathscr{D}\right)\right)^{1/r}}$ and $b = \frac{|Y|}{\left(\mathbb{E}\left(|Y|^s\middle|\mathscr{D}\right)\right)^{1/s}}$ and then take conditional expectations:

$$\mathbb{E}\left(\frac{|X|}{\big(\mathbb{E}\big(|X|^r\big|\mathscr{D}\big)\big)^{1/r}}\frac{|Y|}{\big(\mathbb{E}\big(|Y|^s\big|\mathscr{D}\big)\big)^{1/s}}\middle|\mathscr{D}\right) \leq \mathbb{E}\left(\frac{|X|^r}{r\mathbb{E}\big(|X|^r\big|\mathscr{D}\big)}\middle|\mathscr{D}\right) + \mathbb{E}\left(\frac{|Y|^s}{s\mathbb{E}\big(|Y|^s\big|\mathscr{D}\big)}\middle|\mathscr{D}\right)$$

which simplifies down to

$$\frac{\mathbb{E}\big(|XY|\big|\mathscr{D}\big)}{\big(\mathbb{E}\big(|X|^r\big|\mathscr{D}\big)\big)^{1/r}\big(\mathbb{E}\big(|Y|^s\big|\mathscr{D}\big)\big)^{1/s}} \leq \frac{\mathbb{E}(|X|^r|\mathscr{D})}{r\mathbb{E}\big(|X|^r\big|\mathscr{D}\big)} + \frac{\mathbb{E}(|Y|^s|\mathscr{D})}{s\mathbb{E}\big(|Y|^s\big|\mathscr{D}\big)} = \frac{1}{r} + \frac{1}{s} = 1$$

Rearranging yields the desired result.

Liapunov: $\big(\mathbb{E}[|X|^r]\big)^{1/r}$ is increasing in r for $r \geq 0$.
Proof: Let $0 < a \leq b$. By the conditional version of Hölder with parameters $r = \frac{b}{a}$ and $s = \frac{b}{b-a}$ and r.v.'s $|X|^a$ and 1, we get that

$$\mathbb{E}\big(|X|^a \cdot 1\big|\mathscr{D}\big) \leq \Big(\mathbb{E}\Big((|X|^a)^{\frac{b}{a}}\big|\mathscr{D}\Big)\Big)^{\frac{a}{b}}\Big(\mathbb{E}\Big(1^{\frac{b}{b-a}}\big|\mathscr{D}\Big)\Big)^{\frac{b-a}{b}} \implies \mathbb{E}\big(|X|^a\big|\mathscr{D}\big)^{1/a} \leq \big(\mathbb{E}\big(|X|^b\big|\mathscr{D}\big)\big)^{1/b}$$

Minkowski: $\big(\mathbb{E}[|X+Y|^r]\big)^{1/r} \le \big(\mathbb{E}[|X|^r]\big)^{1/r} + \big(\mathbb{E}[|Y|^r]\big)^{1/r}$ for $r \ge 1$.
Proof: Just like in the proof of regular Minkowski, we start by noting that inequality is trivial for $r=1$, and that for $r>1$ (and $s = \frac{r}{r-1}$),

$$\begin{aligned}
\mathbb{E}\big(|X+Y|^r \big| \mathscr{D}\big) &\le \mathbb{E}\big((|X|+|Y|)|X+Y|^{r-1} \big| \mathscr{D}\big) \\
&= \mathbb{E}\big(|X| \cdot |X+Y|^{r-1} \big| \mathscr{D}\big) + \mathbb{E}\big((|Y| \cdot |X+Y|^{r-1} \big| \mathscr{D}\big) \\
&\le \big(\mathbb{E}\big(|X|^r \big| \mathscr{D}\big)\big)^{\frac{1}{r}} \big(\mathbb{E}\big((|X+Y|^{r-1})^s \big| \mathscr{D}\big)\big)^{\frac{1}{s}} + \big(\mathbb{E}\big(|Y|^r \big| \mathscr{D}\big)\big)^{\frac{1}{r}} \big(\mathbb{E}\big((|X+Y|^{r-1})^s \big| \mathscr{D}\big)\big)^{\frac{1}{s}} \\
&= \big(\mathbb{E}\big(|X|^r \big| \mathscr{D}\big)\big)^{1/r} \big(\mathbb{E}\big(|X+Y|^r \big| \mathscr{D}\big)\big)^{1/s} + \big(\mathbb{E}\big(|Y|^r \big| \mathscr{D}\big)\big)^{1/r} \big(\mathbb{E}\big(|X+Y|^r \big| \mathscr{D}\big)\big)^{1/s} \\
&= \Big(\big(\mathbb{E}\big(|X|^r \big| \mathscr{D}\big)\big)^{1/r} + \big(\mathbb{E}\big(|Y|^r \big| \mathscr{D}\big)\big)^{1/r}\Big) \big(\mathbb{E}\big(|X+Y|^r \big| \mathscr{D}\big)\big)^{1-(1/r)}
\end{aligned}$$

If $\mathbb{E}\big(|X+Y|^r \big| \mathscr{D}\big) = 0$, the result is trivial, and otherwise we divide to get an exponent of $(1/r)$ on the left hand side, which is the result.

Jensen: $g(\mathbb{E}(Y|\mathscr{D})) \le_{\text{a.s.}} \mathbb{E}(g(Y)|\mathscr{D})$.
Proof: Tweaking the proof of the regular Jensen inequality a bit (consider a set of lines instead of just one to account for the fact that we're dealing with inequalities on r.v.'s and not numbers), observe that $g(x) = \sup\{l(x) : l \in L\}$ where L is the set of lines $l(x)$ s.t. $l(x) \le g(x)$ for all $x \in \mathbb{R}$. Then, picking any line l from L, $l(X) \le g(X) \implies \mathbb{E}(l(X)|\mathscr{D}) \le_{\text{a.s.}} \mathbb{E}(g(X)|\mathscr{D})$. But by linearity of conditional expectation, $\mathbb{E}(l(X)|\mathscr{D}) = l(\mathbb{E}(X|\mathscr{D}) \le_{\text{a.s.}} \mathbb{E}(g(X)|\mathscr{D})$. This is true for any line l, and so we can take the supremum to get that $g(\mathbb{E}(X|\mathscr{D})) \le_{\text{a.s.}} \mathbb{E}(g(X)|\mathscr{D})$.

A random variable $h_{\mathscr{D}}(\omega) =_{\text{a.s.}} \mathbb{E}(Y|\mathscr{D}) \iff \mathbb{E}[XY] = \mathbb{E}[Xh_{\mathscr{D}}]$ for all $\mathscr{D}$-measurable r.v.'s X (where $\mathbb{E}[|XY|] < \infty$):

($\implies$) For all $D \in \mathscr{D}$, we know that $\int_D h_{\mathscr{D}}\, dP = \int_D Y\, dP$. Thus, we know that for all $X = 1_D$, the claim is true. We proceed via the standard machine: for any simple function $X := \sum_{i=1}^N c_i 1_{D_i}$,

$$\int_\Omega \sum_{i=1}^N c_i Y \cdot 1_{D_i} = \sum_{i=1}^N c_i \int_{D_i} Y\, dP = \sum_{i=1}^N c_i \int_{D_i} h_{\mathscr{D}}\, dP = \int_\Omega \sum_{i=1}^N c_i h_{\mathscr{D}} \cdot 1_{D_i}$$

Now supposing $X \ge 0$, we have some sequence X_n of simple functions that converge monotonically to X, and so using the monotone convergence theorem,

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}\big[XY^+\big] - \mathbb{E}\big[XY^-\big] = \lim_{n\to\infty} \mathbb{E}\big[X_n Y^+\big] - \lim_{n\to\infty} \mathbb{E}\big[X_n Y^-\big] \\
&= \lim_{n\to\infty} \mathbb{E}\big[X_n h_{\mathscr{D}}^+\big] - \lim_{n\to\infty} \mathbb{E}\big[X_n h_{\mathscr{D}}^-\big] = \mathbb{E}\big[X_n h_{\mathscr{D}}^+\big] - \mathbb{E}\big[X_n h_{\mathscr{D}}^-\big] = \mathbb{E}[X_n h_{\mathscr{D}}]
\end{aligned}$$

Finally, for arbitrary $\mathscr{D}$-measurable X,

$$\mathbb{E}[XY] = \mathbb{E}\big[X^+Y\big] - \mathbb{E}\big[X^-Y\big] = \mathbb{E}\big[X^+h_{\mathscr{D}}\big] - \mathbb{E}\big[X^-h_{\mathscr{D}}\big] = \mathbb{E}[Xh_{\mathscr{D}}]$$

($\Longleftarrow$) Take $X = 1_D$ for any D and we get that $\int_\Omega 1_D \cdot Y\,dP = \int_\Omega 1_D \cdot h_{\mathscr{D}}\,dP \implies \int_D Y\,dP = \int_D h_{\mathscr{D}}\,dP \implies h_{\mathscr{D}}(\omega) =_{\text{a.s.}} \mathbb{E}(Y|\mathscr{D})$.

Problem 6

We have a jar of six identical balls labelled 1, 2, 2, 3, 3, 3. The r.v.'s X_1 and X_2 represent sampling twice (WITHOUT replacement). The joint probability distribution for (X_1, X_2) is

			X_1		
		1	2	3	
	1	0	2/30	3/30	5/30
X_2	2	2/30	2/30	6/30	10/30
	3	3/30	6/30	6/30	15/30
		5/30	10/30	15/30	1

Now letting Ω be the 3×3 grid above, $Y = X_2$, $S = X_1 + X_2$, and $\mathscr{D} = S^{-1}(\mathscr{B})$, we want to explicitly compute the values of the r.v.'s $P(Y = 1|\mathscr{D}), P(Y = 2|\mathscr{D}), P(Y = 3|\mathscr{D})$, and $\mathbb{E}(Y|\mathscr{D})$.

Values of S

2	3	4
3	4	5
4	5	6

Associated probabilities

s	2	3	4	5	6
$P(S = s)$	0	4/30	8/30	12/30	6/30

Using the formula from problem #3, $P(Y = i|\mathscr{D}) = \displaystyle\sum_{j=2}^{6} \frac{P(A \cap D_i)}{P(D_i)} 1_{D_j}$, we get the following:

$P(Y = 1|\mathscr{D})(\omega)$

	2/4	3/8
2/4	3/8	0
3/8	0	0

$P(Y = 2|\mathscr{D})(\omega)$

	2/4	2/8
2/4	2/8	6/12
2/8	6/12	0

$P(Y = 3|\mathscr{D})(\omega)$

	0	3/8
0	3/8	6/12
3/8	6/12	1

Finally, $\mathbb{E}(Y|\mathscr{D}) = \mathbb{E}\big(1_{[Y=1]} + 2 \cdot 1_{[Y=2]} + 3 \cdot 1_{[Y=3]}|\mathscr{D}\big) = P(Y = 1|\mathscr{D}) + 2 \cdot P(Y = 2|\mathscr{D}) + 3 \cdot P(Y = 3|\mathscr{D})$:

$\mathbb{E}(Y|\mathscr{D})(\omega)$

	6/4	16/8
6/4	16/8	30/12
16/8	30/12	3

Problem 7

Supposing that $X, Y \in L^1(\Omega, \mathscr{F}, P)$ and $\mathbb{E}(Y|X) = X$ a.s. and $\mathbb{E}(X|Y) = Y$ a.s., we want to prove that $P([X = Y]) = 1$. I actually went and hunted down an online pdf of Williams's *Probability* *with Martingales*, and so we proceed with his hint (along with the properties $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}(Y|\mathscr{D})]$ and $\mathbb{E}(XY|\mathscr{D}) =_{\text{a.s.}} Y\mathbb{E}(X|\mathscr{D})$ if Y is $\mathscr{D}$-measurable):

$$\begin{aligned}
&\int_{[X>c]\cap[Y\le c]} (X-Y)\,dP + \int_{[X\le c]\cap[Y\le c]} (X-Y)\,dP \\
=&\int_{[Y\le c]} (X-Y)\,dP = \int_{[Y\le c]} X\,dP - \int_{[Y\le c]} Y\,dP \\
=&\mathbb{E}\big[1_{[Y\le c]}X\big] - \mathbb{E}\big[1_{[Y\le c]}Y\big] = \mathbb{E}\big[\mathbb{E}\big(1_{[Y\le c]}X|Y\big)\big] - \mathbb{E}\big[1_{[Y\le c]}Y\big] \\
=&\mathbb{E}\big[1_{[Y\le c]}\mathbb{E}(X|Y)\big] - \mathbb{E}\big[1_{[Y\le c]}Y\big] = \int_{[Y\le c]} \mathbb{E}(X|Y)\,dP - \int_{[Y\le c]} Y\,dP = 0
\end{aligned}$$

Of course, we can flip the letters around to get: $\int_{[Y>c]\cap[X\le c]}(Y-X)\,dP + \int_{[Y\le c]\cap[X\le c]}(Y-X)\,dP = 0$. Thus, when we sum the two formulas together we get that

$$\begin{aligned}
&\int_{[X>c]\cap[Y\le c]} (X-Y)\,dP + \int_{[X\le c]\cap[Y\le c]} (X-Y)\,dP \\
+&\int_{[Y>c]\cap[X\le c]} (Y-X)\,dP + \int_{[Y\le c]\cap[X\le c]} (Y-X)\,dP \\
=&\int_{[X>c]\cap[Y\le c]} (X-Y)\,dP + \int_{[Y>c]\cap[X\le c]} (Y-X)\,dP \\
=&\int_{[Y\le c<X]} (X-Y)\,dP + \int_{[X\le c<Y]} (Y-X)\,dP = 0
\end{aligned}$$

We are almost at the point where we are integrating $(X-Y)$ over $[X>Y]$ and $(Y-X)$ over $[Y>X]$; note that $[X>Y] = \bigcup_{c\in\mathbb{Q}}[X>c\ge Y]$ and $[Y>X] = \bigcup_{c\in\mathbb{Q}}[Y>c\ge X]$ (where we use $\mathbb{Q}$ because $\mathbb{R}$ is uncountable and we don't want to deal with uncountable sums). Therefore, we can say that

$$\begin{aligned}
0 &\le \int_{[X>Y]} (X-Y)\,dP + \int_{[Y>X]} (Y-X)\,dP \\
&\le \sum_{c\in\mathbb{Q}} \left(\int_{[Y\le c<X]} (X-Y)\,dP + \int_{[X\le c<Y]} (Y-X)\,dP \right) = \sum_{c\in\mathbb{Q}} 0 = 0
\end{aligned}$$

This implies that $\int_\Omega |X-Y|\,dP = 0$, which is only possible if $P([|X-Y|=0]) = 1 \implies X =_{\text{a.s.}} Y$.

522 Homework 2

Daniel Rui - 1/22/20

Problem 1

We want to prove the equivalence of the following three statements:

$$\text{(a)} \sum_{k=1}^{\infty} X_k < \infty \text{ a.s.} \quad \text{(b)} \sum_{k=1}^{\infty} \Big[P(X_k > 1) + \mathbb{E}\big[X_k \cdot 1_{[X_k \le 1]}\big] \Big] < \infty \quad \text{(c)} \sum_{k=1}^{\infty} \mathbb{E}\Big[\tfrac{X_k}{1+X_k}\Big] < \infty$$

We set out to prove that (c) implies (a) using the 3-series theorem:

(1) Note that $X_k > 1 \iff 1 + X_k > 2 \iff \frac{1}{2} < \frac{1}{1+X_k} \iff \frac{1}{2} < \frac{X_k}{1+X_k} \iff 1 < \frac{2X_k}{1+X_k}$. Thus,

$$\sum_{k=1}^{\infty} P(X_k > 1) = \sum_{k=1}^{\infty} \mathbb{E}\big[1 \cdot 1_{[X_k > 1]}\big] < \sum_{k=1}^{\infty} \mathbb{E}\Big[\tfrac{2X_k}{1+X_k} \cdot 1_{[X_k > 1]}\Big] \le \sum_{k=1}^{\infty} 2\mathbb{E}\Big[\tfrac{X_k}{1+X_k}\Big] < \infty.$$

(2) For the next two, note that $0 \le X_k \le 1 \iff \frac{1}{2} \le \frac{1}{1+X_k} \le 1 \iff X_k \le \frac{2X_k}{1+X_k} \le 2X_k$, and so

$$\sum_{k=1}^{\infty} \mathbb{E}\big[X_k \cdot 1_{[X_k \le 1]}\big] \le \sum_{k=1}^{\infty} \mathbb{E}\Big[\tfrac{2X_k}{1+X_k} \cdot 1_{[X_k \le 1]}\Big] \le \sum_{k=1}^{\infty} 2\mathbb{E}\Big[\tfrac{X_k}{1+X_k}\Big] < \infty.$$

(3) Expanding upon the above inequalities, note that $X_k \le \frac{2X_k}{1+X_k}$ together with $X_k \le 1$ imply that $X_k^2 \le X_k \le \frac{2X_k}{1+X_k}$, and recall also that $\mathrm{Var}\big[X_k \cdot 1_{[X_k \le 1]}\big] = \mathbb{E}\big[X_k^2 \cdot 1_{[X_k \le 1]}\big] - \mathbb{E}\big[X_k \cdot 1_{[X_k \le 1]}\big]^2$ is always ≥ 0. Thus,

$$\sum_{k=1}^{\infty} \mathbb{E}\big[X_k^2 \cdot 1_{[X_k \le 1]}\big] - \mathbb{E}\big[X_k \cdot 1_{[X_k \le 1]}\big]^2 \le \sum_{k=1}^{\infty} \mathbb{E}\big[X_k^2 \cdot 1_{[X_k \le 1]}\big] \le \sum_{k=1}^{\infty} \mathbb{E}\big[X_k \cdot 1_{[X_k \le 1]}\big] < \infty$$

Note that our proof in (3) also shows that the two conditions given by (b) imply the variance condition (because we simply bounded the variance series by the expectation series), and so assuming (b) we can still fulfill the three conditions set by the 3-series theorem. The 3-series theorem now tells us that $\sum_{k=1}^{n} X_k \to_{\text{a.s.}} S$, which trivially gives that $\sum_{k=1}^{\infty} X_k < \infty$.

Summarizing, we've shown that (c) implies (a) (and hence obviously (b) as well), and also that (b) implies (a), so all we need to do now is show (a) $\implies$ (c). We are given that $\sum_{k=1}^{\infty} X_k$ is finite a.s., so let's say that it is less than M a.s. for some M. Then,

$$\sum_{k=1}^{\infty} \mathbb{E}\Big[\tfrac{X_k}{1+X_k}\Big] \le \sum_{k=1}^{\infty} \mathbb{E}[X_k] \le \mathbb{E}\left[\sum_{k=1}^{\infty} X_k\right] \le \mathbb{E}[M] = M < \infty$$

Problem 2

Let $Y_1, Y_2, \dots$ be i.i.d. Cauchy(0,1). From Wikipedia, we know that the characteristic function of the Cauchy(0,1) distribution is

$$\varphi_{Y_k}(t) = \mathbb{E}\big[e^{itY_k}\big] = e^{-|t|}$$

Furthermore, Lévy's continuity theorem gives that $\varphi_{S_n}(t) \to \varphi_S(t)$ pointwise for all $t \in \mathbb{R} \implies [\varphi_S(t)$ is continuous $\iff$ $S_n \to_d S$ for some $S \iff$ S has characteristic function $\varphi_S]$. Let us define $S_n = \sum_{k=1}^{n} a_k Y_k$ for some sequence $\{a_k\}$. Then,

$$\varphi_{S_n}(t) = \mathbb{E}\big[e^{itS_n}\big] = \mathbb{E}\big[e^{itY_1} \cdots e^{itY_n}\big]$$

but because the Y_k are independent $\implies$ e^{itY_k} are independent, expectation is multiplicative, so

$$\varphi_{S_n}(t) = \mathbb{E}\big[e^{itS_n}\big] = \mathbb{E}\big[e^{ita_1Y_1}\big] \cdots \mathbb{E}\big[e^{ita_nY_n}\big] = e^{-|a_1t|} \cdots e^{-|a_nt|} = e^{-|t| \sum_{k=1}^{n} |a_k|}$$

If $\sum_{k=1}^{\infty} |a_k| = \infty$, then for all $t \neq 0$, $\varphi_{S_n}(t) \to 0$; however, for all $n \in \mathbb{N}$, $\varphi_{S_n}(0) = 1$, so $\varphi_{S_n}(0) \to 1$. This of course means that $\varphi_S(t)$ is not continuous, and so Lévy tells us that the $S_n \not\to_d$ to any r.v., and furthermore, theorem 8.1 from PfS says that $\to_d \iff \to_{\text{a.s.}}$, and so $S_n \not\to_{\text{a.s.}} S$ for any S.

If $\sum_{k=1}^{\infty} |a_k|$ converges (i.e. is finite) to say a, then $\varphi_{S_n}(t) \to \varphi_S(t) := \exp(-a|t|)$ pointwise for all t, so Lévy tells us that $S_n \to_d S$ where the characteristic function of S is $\varphi_S(t)$. This of course implies that S is distributed Cauchy(0,a). Theorem 8.1 again says that convergence in distribution iff convergence a.s., and so $S_n \to_{\text{a.s.}} S$. In the particular case that $a_k = \frac{1}{2^k}$, $S_n \to_{\text{a.s.}} S$ where $S \sim$ Cauchy(0,1).

Problem 3

Let a be some fixed value > 0; $\mathbb{V}, \mathbb{U}^{(1)}, \mathbb{U}^{(2)}$ be independent Brownian bridge processes; and Z be Normal(0,1) and independent of $\mathbb{V}$. Let us add to the conditions of Brownian bridge processes that $\mathbb{V}(0) = \mathbb{V}(1) = 0$. We want to verify that ...

(a) $\mathbb{B}(t) = \mathbb{V}(t) + tZ$ is a Brownian motion for $t \in [0,1]$: (1) $\mathbb{B}(0) = \mathbb{V}(0) + 0 \cdot Z = 0 + 0 = 0$; (2) $\mathbb{E}[\mathbb{B}(t)] = \mathbb{E}[\mathbb{V}(t)] + t\mathbb{E}[Z] = 0 + 0 = 0$; and (3) $\mathbb{E}[\mathbb{B}(s)\mathbb{B}(t)] = \mathbb{E}[(\mathbb{V}(s) + sZ)(\mathbb{V}(t) + tZ)] = \mathbb{E}\big[\mathbb{V}(s)\mathbb{V}(t) + sZ\mathbb{V}(t) + tZ\mathbb{V}(s) + stZ^2\big] = \mathbb{E}[\mathbb{V}(s)\mathbb{V}(t)] + s\mathbb{E}[Z]\mathbb{V}(t) + t\mathbb{E}[Z]\mathbb{V}(s) + st\mathbb{E}\big[Z^2\big] = \min\{s,t\} - st + 0 + 0 + st \cdot 1 = \min\{s,t\}$.

(b) $\mathbb{B}^{(1)}(t) = \mathbb{B}(at)/\sqrt{a}$ is a Brownian motion for $t \in [0, 1/a)$: (1) $\mathbb{B}^{(1)}(0) = \mathbb{B}(0)/\sqrt{a} = 0$; (2) $\mathbb{E}\big[\mathbb{B}^{(1)}(t)\big] = \mathbb{E}[\mathbb{B}(at)]/\sqrt{a} = 0$ for $t \in [0, 1/a)$; and (3) $\mathbb{E}\big[\mathbb{B}^{(1)}(s)\mathbb{B}^{(1)}(t)\big] = \frac{1}{a}\mathbb{E}[\mathbb{B}(as)\mathbb{B}(at)] = \frac{1}{a}\min\{as, at\} = \min\{s,t\}$ for $s,t \in [0, 1/a)$.

(c) $\mathbb{B}^{(2)}(t) = \mathbb{B}(a+t) - \mathbb{B}(a)$ is a Brownian motion for $t \in [-a, 1-a], a \in [0,1]$: (1) $\mathbb{B}^{(2)}(0) = \mathbb{B}(a) - \mathbb{B}(a) = 0$; (2) $\mathbb{E}\big[\mathbb{B}^{(2)}(t)\big] = \mathbb{E}[\mathbb{B}(a+t)] - \mathbb{E}[\mathbb{B}(a)] = 0$; and (3) $\mathbb{E}\big[\mathbb{B}^{(2)}(s)\mathbb{B}^{(2)}(t)\big] = \mathbb{E}[\mathbb{B}(a+s)\mathbb{B}(a+t)] - \mathbb{E}[\mathbb{B}(a)\mathbb{B}(a+t)] - \mathbb{E}[\mathbb{B}(a+s)\mathbb{B}(a)] + \mathbb{E}[\mathbb{B}(a)\mathbb{B}(a)] = \min\{a+s, a+t\} - \min\{a, a+t\} - \min\{a+s, a\} + \min\{a,a\} = \min\{a+s, a+t\} - a - a + a = \min\{s,t\}$.

(d) $\mathbb{U}^{(3)} = \sqrt{1-a}\mathbb{U}^{(1)} \pm \sqrt{a}\mathbb{U}^{(2)}$ is a Brownian bridge, $t \in [0,1]$: (1) $\sqrt{1-a}\mathbb{U}^{(1)}(0) \pm \sqrt{a}\mathbb{U}^{(2)}(0) = 0$; $\sqrt{1-a}\mathbb{U}^{(1)}(1) \pm \sqrt{a}\mathbb{U}^{(2)}(1) = 0$; (2) $\mathbb{E}\big[\sqrt{1-a}\mathbb{U}^{(1)} \pm \sqrt{a}\mathbb{U}^{(2)}\big] = \sqrt{1-a}\mathbb{E}\big[\mathbb{U}^{(1)}\big] \pm \sqrt{a}\mathbb{E}\big[\mathbb{U}^{(2)}\big] = 0$; and (3) $\mathbb{E}\big[\mathbb{U}^{(3)}(s)\mathbb{U}^{(3)}(t)\big] = \mathbb{E}\big[\big(\sqrt{1-a}\mathbb{U}^{(1)}(s) \pm \sqrt{a}\mathbb{U}^{(2)}(s)\big)\big(\sqrt{1-a}\mathbb{U}^{(1)}(t) \pm \sqrt{a}\mathbb{U}^{(2)}(t)\big)\big] = (1-a)\mathbb{E}\big[\mathbb{U}^{(1)}(s)\mathbb{U}^{(1)}(t)\big] \pm \sqrt{a(1-a)}\mathbb{E}\big[\mathbb{U}^{(1)}(s)\big]\mathbb{E}\big[\mathbb{U}^{(2)}(t)\big] \pm \sqrt{a(1-a)}\mathbb{E}\big[\mathbb{U}^{(1)}(t)\big]\mathbb{E}\big[\mathbb{U}^{(2)}(s)\big] + a\mathbb{E}\big[\mathbb{U}^{(2)}(s)\mathbb{U}^{(2)}(t)\big] = (1-a)(\min\{s,t\} - st) \pm 0 \pm 0 + a(\min\{s,t\} - st) = \min\{s,t\} - st$.

(e) $\mathbb{Z}(t) = [\mathbb{U}^{(1)}(t) + \mathbb{U}^{(2)}(1-t)]/\sqrt{2}$ is a Brownian bridge, $t \in [0,1]$: (1) $\mathbb{Z}(0) = \mathbb{Z}(1) = [\mathbb{U}^{(1)}(0) + \mathbb{U}^{(2)}(1)]/\sqrt{2} = (0+0)/\sqrt{2} = 0$; (2) $\mathbb{E}[\mathbb{Z}(t)] = \big(\mathbb{E}\big[\mathbb{U}^{(1)}(t)\big] + \mathbb{E}\big[\mathbb{U}^{(2)}(1-t)\big]\big)/\sqrt{2} = 0$; and (3) $\mathbb{E}[\mathbb{Z}(s)\mathbb{Z}(t)] = \frac{1}{2}\mathbb{E}\big[\big(\mathbb{U}^{(1)}(s) + \mathbb{U}^{(2)}(1-s)\big)\big(\mathbb{U}^{(1)}(t) + \mathbb{U}^{(2)}(1-t)\big)\big] = \frac{1}{2}\big(\mathbb{E}\big[\mathbb{U}^{(1)}(s)\mathbb{U}^{(1)}(t)\big] + \mathbb{E}\big[\mathbb{U}^{(2)}(1-s)\big]\mathbb{E}\big[\mathbb{U}^{(1)}(t)\big] + \mathbb{E}\big[\mathbb{U}^{(2)}(1-t)\big]\mathbb{E}\big[\mathbb{U}^{(1)}(s)\big] + \mathbb{E}\big[\mathbb{U}^{(2)}(1-s)\mathbb{U}^{(2)}(1-t)\big]\big) = \frac{1}{2}\big(\min\{s,t\} - st + 0 + 0 + \min\{1-s, 1-t\} - (1-s)(1-t)\big) = \frac{1}{2}\big(\min\{s,t\} - 2st + \min\{1-s,1-t\} - 1 + s + t\big) = \frac{1}{2}\big(\min\{s,t\} - 2st + \min\{t,s\}\big) = \min\{s,t\} - st$.

Problem 4

(a) Given any two Haar functions g_{nj} and g_{mk}, we want to show that $\int_0^1 g_{nj}(t)g_{mk}(t)\,d\lambda = 0$. First, let us what the set $[g_{nj} \neq 0]$ is. Well, $g_{nj}(t) = 2^{n/2}\big[1_{[0,1/2]}(2^n t - j) - 1_{(1/2,1]}(2^n t - j)\big]$ so it's non-zero only when $0 \le 2^n t - j \le 1 \iff t \in \big[\frac{j}{2^n}, \frac{j+1}{2^n}\big]$. If $n = m$ and w.l.o.g. $j \le k$; then k could either be $= j$, in which case the integral is $\int_{[\frac{j}{2^n}, \frac{j+1}{2^n}]} 1\,d\lambda = \frac{1}{2^n}$ and so $\int_0^1 g_n^2(t)\,d\lambda = \sum_0^{2^n-1} \frac{1}{2^n} = 1$; or k could be $= j+1$, in which case the integral would be $\int_{\{\frac{j+1}{2^n}\}} -1\,d\lambda = 0$; or k could be $> j+1$, in which case the integral would obviously just be 0.

Now w.l.o.g. $n < m$. Similar to above, the intersection of two intervals $\big[\frac{j}{2^n}, \frac{j+1}{2^n}\big]$, $\big[\frac{k}{2^m}, \frac{k+1}{2^m}\big]$ can only result in $\varnothing$, a singleton, or the smaller of the two intervals, i.e. $\big[\frac{k}{2^m}, \frac{k+1}{2^m}\big]$. If the intersection is $\varnothing$ or a singleton, the integral is clearly 0. Note that g_{nj} is 1 on $\big[\frac{2j}{2^{n+1}}, \frac{2j+1}{2^{n+1}}\big]$ and -1 on $\big(\frac{2j+1}{2^{n+1}}, \frac{2j+2}{2^{n+1}}\big]$, and furthermore, that $\big[\frac{k}{2^m}, \frac{k+1}{2^m}\big]$ lies cleanly inside exactly one of these two. We will say that it's 1 w.l.o.g.; then the integral becomes $\int_{[\frac{k}{2^m}, \frac{k+1}{2^m}]} g_{mk}(t)\,d\lambda$, which is just 0 (because of the equal amount of area above and below the x-axis).

(b) Recall from class that we proved that

$$|\mathbb{V}_n(t,\omega)| \le \tfrac{1}{2} \cdot 2^{-n/2} \max_{1 \le k \le 2^n - 1} |X_{nk}(\omega)| \le 2^{-n/2}\sqrt{n} \text{ almost surely}$$

for all $n \ge N$ for some N. We want to rigorously justify the following interchange (by thinking about sums as integrals on $\mathbb{N}$ w.r.t. counting measure, where counting measure is sigma finite):

$$\begin{aligned} &\int_\Omega \left[\sum_{n=0}^\infty \left[\sum_{m=0}^\infty \mathbb{V}_n(s,\omega)\mathbb{V}_m(t,\omega)\,\text{“}dm\text{”}\right] \text{“}dn\text{”}\right] dP \\ &= \sum_{n=0}^\infty \left[\int_\Omega \left[\sum_{m=0}^\infty \mathbb{V}_n(s,\omega)\mathbb{V}_m(t,\omega)\,\text{“}dm\text{”}\right] dP\right] \text{“}dn\text{”} \\ &= \sum_{n=0}^\infty \left[\sum_{m=0}^\infty \left[\int_\Omega \mathbb{V}_n(s,\omega)\mathbb{V}_m(t,\omega)\,dP\right] \text{“}dm\text{”}\right] \text{“}dn\text{”} \end{aligned}$$

which we can do by showing that $\int_\Omega \left[\sum_{n=0}^\infty \left[\left|\sum_{m=0}^\infty \mathbb{V}_n(s,\omega)\mathbb{V}_m(t,\omega)\,\text{“}dm\text{”}\right|\right] \text{“}dn\text{”}\right] dP < \infty$ and that

$\int_\Omega \left[\sum_{m=0}^\infty \Big| \mathbb{V}_n(s,\omega)\mathbb{V}_m(t,\omega) \Big| \text{“}dm\text{”} \right] dP < \infty$ (for any fixed n). For the first one, note that

$$\begin{aligned}
&\int_\Omega \left[\sum_{n=0}^\infty \left[\left| \sum_{m=0}^\infty \mathbb{V}_n(s,\omega)\mathbb{V}_m(t,\omega) \text{“}dm\text{”} \right| \right] \text{“}dn\text{”} \right] dP \\
\leq &\int_\Omega \left[\sum_{n=0}^\infty \left[\sum_{m=0}^\infty \Big| \mathbb{V}_n(s,\omega) \Big| \Big| \mathbb{V}_m(t,\omega) \Big| \text{“}dm\text{”} \right] \text{“}dn\text{”} \right] dP \\
= &\int_\Omega \left[\sum_{n=0}^\infty \Big| \mathbb{V}_n(s,\omega) \Big| \left[\sum_{m=0}^{N-1} \Big| \mathbb{V}_m(t,\omega) \Big| \text{“}dm\text{”} + \sum_{m=N}^\infty \Big| \mathbb{V}_m(t,\omega) \Big| \text{“}dm\text{”} \right] \text{“}dn\text{”} \right] dP \\
\leq &\int_\Omega \left[\sum_{n=0}^\infty \Big| \mathbb{V}_n(s,\omega) \Big| \left[\sum_{m=0}^{N-1} \Big| \mathbb{V}_m(t,\omega) \Big| \text{“}dm\text{”} + \sum_{m=N}^\infty \frac{\sqrt{m}}{2^{m/2}} \text{“}dm\text{”} \right] \text{“}dn\text{”} \right] dP
\end{aligned}$$

But $Y(t,\omega) := \sum_{m=0}^{N-1} \big| \mathbb{V}_m(t,\omega) \big|$ has

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_{m=0}^{N-1} \mathbb{E}\big[\big| \mathbb{V}_m(t,\omega) \big| \big] = \sum_{m=0}^{N-1} \sum_{j=0}^{2^m-1} |h_{mj}(t)| \mathbb{E}[|X_{mj}(\omega)|] \\
&= \underbrace{\sum_{m=0}^{N-1} \sum_{j=0}^{2^m-1} |h_{mj}(t)| \cdot \sqrt{\tfrac{2}{\pi}} \leq \sqrt{\tfrac{2}{\pi}} \sum_{m=0}^{N-1} 2^{-m/2}}_{\text{for any given } t, \text{ at most one } h_{nj} \text{ is } \neq 0} =: C_1 < \infty
\end{aligned}$$

(using formulas from Wikipedia for expectations of folded normals) and

$$\begin{aligned}
\operatorname{Var}[Y] &= \sum_{m=0}^{N-1} \operatorname{Var}\big[\big| \mathbb{V}_m(t,\omega) \big| \big] = \sum_{m=0}^{N-1} \sum_{j=0}^{2^m-1} (h_{mj}(t))^2 \operatorname{Var}[|X_{mj}(\omega)|] \\
&= \sum_{m=0}^{N-1} \sum_{j=0}^{2^m-1} h_{mj}^2(t) \Big(\mathbb{E}\big[X_{mj}^2(\omega)\big] - \mathbb{E}[|X_{mj}(\omega)|]^2 \Big) \leq \sum_{m=0}^{N-1} 2^{-m} \big(1 - \tfrac{2}{\pi}\big) =: C_2 < \infty.
\end{aligned}$$

$C_0 := \sum_{m=N}^\infty \frac{\sqrt{m}}{2^{m/2}}$ is clearly finite as well ($\sqrt{m} < 1.1^m$ for $m > M$, so the sum is $<$ finite number of terms $+ \sum_{m=M}^\infty (\frac{1.1}{\sqrt{2}})^m$, which is finite). We then continue the above chain of inequalities:

$$\begin{aligned}
\leq &\int_\Omega \left[\sum_{n=0}^\infty \Big| \mathbb{V}_n(s,\omega) \Big| [Y(t,\omega) + C_0] \text{“}dn\text{”} \right] dP \\
= &\int_\Omega [Y(t,\omega) + C_0] \left[\sum_{n=0}^{N-1} \Big| \mathbb{V}_n(s,\omega) \Big| \text{“}dn\text{”} + \sum_{n=N}^\infty \Big| \mathbb{V}_n(s,\omega) \Big| \text{“}dn\text{”} \right] dP \\
\leq &\int_\Omega [Y(t,\omega) + C_0][Y(t,\omega) + C_0] \, dP = \int_\Omega Y^2(t,\omega) + 2C_0 Y(t,\omega) + C_0^2 \, dP \\
= &\mathbb{E}\big[Y^2\big] + 2C_0 \mathbb{E}[Y] + C_0^2 = \operatorname{Var}[Y] + \mathbb{E}[Y]^2 + 2C_0 C_1 + C_0^2 \\
= &C_2 + C_1^2 + 2C_0 C_1 + C_0^2 < \infty
\end{aligned}$$

The second interchange is now trivial with the above machinery. The rest of the argument given in the handout is just interchanging integrals and finite sums, and Parseval's identity.

Problem 5

Let $\mu_0 := \mu|_{\mathscr{A}_0}$, i.e. for any $A_0 \in \mathscr{A}_0$, we have that $\mu_0(A_0) = \mu(A_0)$. Now for any indicator function, say on some A_0,

$$\int_\Omega 1_{A_0}\, d\mu = \mu(A_0) = \mu_0(A_0) = \int_\Omega 1_{A_0}\, d\mu_0.$$

Bumping up to any simple function $\sum_{k=1}^{n} c_k 1_{A_{0,k}}$ for real numbers c_k and disjoint $A_{0,k}$,

$$\int_\Omega \sum_{k=1}^{n} c_k 1_{A_{0,k}}\, d\mu = \sum_{k=1}^{n} c_k \mu(A_{0,k}) = \sum_{k=1}^{n} c_k \mu_0(A_{0,k}) = \int_\Omega \sum_{k=1}^{n} c_k 1_{A_{0,k}}\, d\mu_0$$

Then for any $X^+ \geq 0$, we can construct simple functions X_n that monotonically increase to X^+ (reference pg. 26 of PfS), and so with the monotone convergence theorem, we have that

$$\int_\Omega X^+\, d\mu = \lim_{n\to\infty} \int_\Omega X_n\, d\mu = \lim_{n\to\infty} \int_\Omega X_n\, d\mu_0 = \int_\Omega X^+\, d\mu_0$$

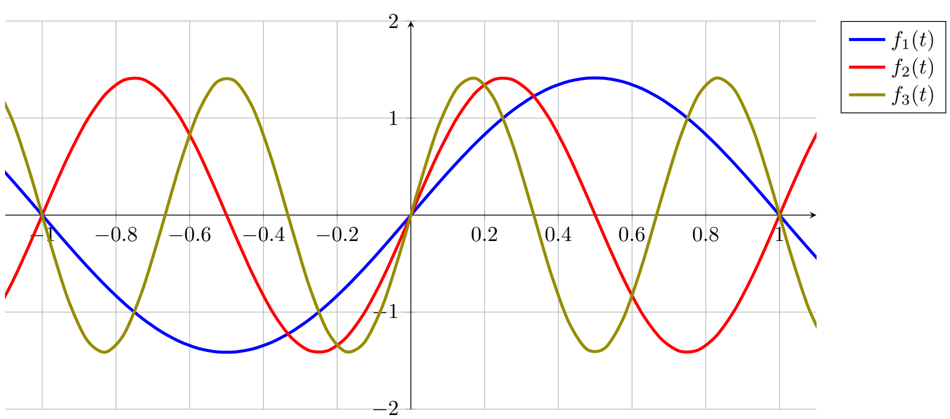
And so finally, for arbitrary $X = X^+ - X^-$,

$$\int_\Omega X\, d\mu = \int_\Omega X^+\, d\mu - \int_\Omega X^-\, d\mu = \int_\Omega X^+\, d\mu_0 - \int_\Omega X^-\, d\mu_0 = \int_\Omega X\, d\mu_0$$

Problem 6

Let $Z_0, Z_1, \ldots$ be i.i.d Normal(0,1), and let $f_k(t) := \sqrt{2}\sin(k\pi t)$, $k \in \mathbb{N}$.

(a) Here's a plot for $k = 1, 2, 3$:



Orthonomality:

$$\int_0^1 f_k^2(t)\, dt = 2\int_0^1 \sin^2(k\pi t)\, dt = 2\int_0^1 \frac{1-\cos(2k\pi t)}{2}\, dt = 1 - \left(\frac{\sin(2k\pi t)}{2k\pi}\right)\Bigg|_0^1 = 1$$

and (using WolframAlpha to evaluate the nastier integrals),

$$\int_0^1 f_j(t) f_k(t)\, dt = 2\int_0^1 \sin(j\pi t)\sin(k\pi t)\, dt$$
$$= 2\left(\frac{k\sin(j\pi t)\cos(k\pi t) - j\cos(j\pi t)\sin(k\pi t)}{\pi(j^2-k^2)}\right)\Bigg|_0^1 = 0$$

(b) Defining $\mathbb{U}(t,\omega) := \sum_{k=1}^{\infty} \frac{Z_k(\omega) f_k(t)}{k\pi}$ for $t \in [0,1]$, we want to verify that it is in fact a Brownian bridge:

(1) $\mathbb{U}(0) = \sum_{k=1}^{\infty} \sqrt{2}\sin(0)\cdot \frac{Z_k(\omega)}{k\pi} = 0$; and $\mathbb{U}(1) = \sum_{k=1}^{\infty} \sqrt{2}\sin(k\pi)\cdot \frac{Z_k(\omega)}{k\pi} = 0$

(2) $\mathbb{E}[\mathbb{U}(t)] = \sum_{k=1}^{\infty} \frac{f_k(t)}{k\pi}\mathbb{E}[Z_k] = \sum_{k=1}^{\infty} \frac{f_k(t)}{k\pi}\cdot 0 = 0$, where we would have to verify (for Tonelli) that

$$\mathbb{E}\left[\sum_{k=1}^{\infty}\left|\frac{Z_k f_k(t)}{k\pi}\right|\right] < \infty$$

which seems like it wouldn't converge because we just have $1/k$, which diverges....uh oh....

(3) $\mathbb{E}[\mathbb{U}(s)\mathbb{U}(t)] = \mathbb{E}\left[\sum_{j=1}^{\infty}\frac{Z_j f_j(s)}{j\pi}\sum_{k=1}^{\infty}\frac{Z_k f_k(t)}{k\pi}\right] = \mathbb{E}\left[\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\frac{Z_j f_j(s)}{j\pi}\frac{Z_k f_k(t)}{k\pi}\right]$ and of course again we would have to verify (for Tonelli) that

$$\mathbb{E}\left[\sum_{j=1}^{\infty}\left[\left|\sum_{k=1}^{\infty}\frac{Z_j f_j(s)}{j\pi}\frac{Z_k f_k(t)}{k\pi}\right|\right]\right] < \infty \quad \text{and} \quad \mathbb{E}\left[\left|\frac{Z_j f_j(s)}{j\pi}\right|\sum_{k=1}^{\infty}\left|\frac{Z_k f_k(t)}{k\pi}\right|\right] < \infty$$

(c)

522 Homework 1

Daniel Rui - 1/15/20

Problem 1

We have $X_1, X_2, \dots$ i.i.d with $P(X_k = 0) = P(X_k = 2) = \frac{1}{2}$, so $\mathbb{E}[X_k] = \mathrm{Var}[X_k] = 1$. Defining

$$S_n := \sum_{k=1}^{n} \frac{X_k}{3^k} \implies \mathbb{E}[S_n] = \sum_{k=1}^{n} \mathbb{E}\left[\frac{X_k}{3^k}\right] = \sum_{k=1}^{n} \frac{1}{3^k}, \mathrm{Var}[S_n] = \sum_{k=1}^{n} \mathrm{Var}\left[\frac{X_k}{3^k}\right] = \sum_{k=1}^{n} \frac{1}{9^k}$$

we see that $\mathbb{E}[S_n] \to \frac{1}{2}$ and $\mathrm{Var}[S_n] \to \frac{1}{8}$, so by the 2-series theorem, there is some r.v. S s.t. $S_n \to_{\text{a.s.}} S$ where $\mathbb{E}[S] = \frac{1}{2}$ and $\mathrm{Var}[S] = \frac{1}{8}$ Furthermore, we know that convergence a.s. implies convergence in distribution, so to find F_S, we will find F_{S_n}. Observe that (using the law of total probability) for $n > 1$

$$P(S_n \leq x) = P(S_{n-1} \leq x) \cdot P(X_n = 0) + P\left(S_{n-1} \leq x - \frac{2}{3^n}\right) \cdot P(X_n = 2)$$

or in other words

$$F_{S_n}(x) = \frac{1}{2}\left[F_{S_{n-1}}(x) + F_{S_{n-1}}\left(x - \frac{2}{3^n}\right)\right]$$

Drawing out F_{S_1} and subsequently F_{S_2}, we immediately jump to the conclusion that F_S is the Cantor function. We will prove this rigorously by induction. First note that we can "extend" the F_{S_n} backwards to F_{S_0} which is 0 on $(-\infty, 0)$ and 1 on $[0, \infty)$. In general, F_{S_n} satisfies a few properties:

- F_{S_n} is increasing.

  Proof: F_{S_0} is obviously increasing. Assuming that $F_{S_{n-1}}(x)$ is increasing, we know that the function $F_{S_{n-1}}\big(x - \frac{2}{3^n}\big)$ is also increasing. The sum of two increasing functions is increasing, and dividing by 2 doesn't change that, so F_{S_n} is increasing.

- $F_{S_n} \leq F_{S_{n-1}}$ (for all x)

  Proof: $F_{S_{n-1}}$ is increasing, so we know that $F_{S_{n-1}}\big(x - \frac{2}{3^n}\big) \leq F_{S_{n-1}}(x) F_{S_{n-1}}\big(x - \frac{2}{3^n}\big)$ for all x, implying that $F_{S_{n-1}}\big(x - \frac{2}{3^n}\big) + F_{S_{n-1}}(x) \leq 2F_{S_{n-1}}(x)$, yielding the result upon rearrangement.

- F_{S_n} only attains finitely many values.

  Proof: Base case: F_{S_1} attains 3 values. Assume that $F_{S_{n-1}}$ satisfies the proposition. Let $\#_{n-1}$ denote the number of values that $F_{S_{n-1}}$ attains. F_{S_n} can only have values of that are averages of the values from $F_{S_{n-1}}$, and there definitely $\leq (\#_{n-1})^2$ possible averages, which is finite.

- For any value v that F_{S_n} attains, $F_{S_n}^{-1}(v)$ is an interval I (a set I is an interval if and only if $\forall x, y \in I$, $x < z < y \implies z \in I$).

  Proof: Assume not; then there is $x, y \in I$ s.t. there is $z \in (x, y)$ s.t. $F_{S_n}(z) \neq v$. Because F_{S_n} is increasing, $x < z < y \implies F_{S_n}(x) = v \leq F_{S_n}(z) \leq F_{S_n}(y) = v \implies F_{S_n}(z) = v$; contradiction.

- I_0 for all F_{S_n} is $(-\infty, 0)$.

  Proof: Base case: trivially true for F_{S_0}. Assuming $F_{S_{n-1}}$ satisfies this, this would tell us that $F_{S_{n-1}} > 0$ on $[0, \infty)$. Thus, for any $x \in [0, \infty)$, $F_{S_n}(x)$ is either the average of 2 positive quantities, of the average of one positive quantity and 0, which is still positive. This implies that $F_{S_n} > 0$ on $[0, \infty)$. Finally, for all $x \in (-\infty, 0)$, we see that $F_{S_n}(x)$ is 0, so the result is true.

- Let $\#_n$ denote the number of values attained by F_{S_n}. Then there are $\#_n$ intervals $I_0, \ldots, I_{\#_n - 1}$ s.t. for any i, $F_{S_n}(x) = F_{S_n}(y)$ for all $x, y \in I_i$ (define this value to be v_i — note that $i < j \iff v_i < v_j$). Also denote $F_{S_n}(I_i) = v_i$. (We can do all of this is from the above properties). **Claim:** $\lambda(I_i) \geq \frac{2}{3^n}$, and all the I_i ($i \geq 1$) are of the form $[x, y)$. Note that $\lambda(I_0)$ always $= \infty$ from above.

  Proof: Base case: obviously true for F_{S_0}. Now assume that $F_{S_{n-1}}$ satisfies the proposition. Let $I_i = [x_i, y_i)$ (for $i \geq 1$); observe that $x_i = y_{i-1}$ (letting $y_0 = 0$). Then, all $x \in J_i := \left[x_i + \frac{2}{3^n}, y_i\right)$ satisfy $F_{S_n}(x) = F_{S_{n-1}}(x) = v_i$, and all $x \in J_{i'} := \left[x_i, x_i + \frac{2}{3^n}\right)$ satisfy $F_{S_n}(x) = \frac{v_{i-1} + v_i}{2}$. Call J_i an old interval because $F_{S_n}(J_i) = v_i$ is an old value; analogously call $J_{i'}$ a new interval. Because $\lambda([x_i, y_i)) \geq \frac{2}{3^{n-1}}$, we know that $\lambda(J_i) \geq \frac{2}{3^{n-1}} - \frac{2}{3^n} = \frac{4}{3^n} \geq \frac{2}{3^n}$, and $\lambda(J_{i'}) = \frac{2}{3^n}$. Because this holds for all i ($\neq 0$), the claim holds. Furthermore, note that when we go from $F_{S_{n-1}}$ to F_{S_n}, the left bounds of ALL old intervals are moved right by $\frac{2}{3^n}$, and ALL new intervals have length $\frac{2}{3^n}$.

- $\#_n = 2^n + 1$ for $n \geq 1$

  Proof: Base case: F_{S_1} has 3 pieces. Assuming that $\#_{n-1} = 2^{n-1} + 1$, we know from the above proof that every interval I_i, $i \in \{1, \ldots, 2^{n-1}\}$ splits into 2 pieces (while I_0 remains $(-\infty, 0)$).

- We know that new intervals created by F_{S_n}, $n \geq 1$, (i.e. intervals corresponding to values that F_{S_n} is the first to attain) initially have length $\frac{2}{3^n}$. **Claim:** Let $I = \left[x, x + \frac{2}{3^n}\right)$ be any such interval, with $v := F_{S_n}(I)$. As we iterate, the intervals corresponding to v have lengths $\to \frac{1}{3^n}$.

  Proof: Above, we showed that ALL old intervals get their left bound moved by $\frac{2}{3^n}$, so when we get to $F_{S_{n+1}}, F_{S_{n+2}}, \ldots$, I's left bound gets moved right by $\frac{2}{3^{n+1}}, \frac{2}{3^{n+2}}, \ldots$. Thus, in F_{S_N} ($N > n$), the left bound is moved right by $\sum_{k=n+1}^{N} \frac{2}{3^k}$, so as $N \to \infty$, $F_{S_N}^{-1}(v) \to \left[x + \frac{1}{3^n}, x + \frac{2}{3^n}\right)$. Furthermore, denote I_{-1} to be the last interval (i.e. the one with value 1, first attained by F_{S_0} on $[0, \infty)$). Although I_{-1} is not of the form covered above, we still know that its left bound gets moved right with every iteration. Thus, the I_{-1} approach $\left[\sum_{n=1}^{\infty} \frac{2}{3^n}, \infty\right) = [1, \infty)$

- For F_{S_n}, y is the right bound of one of the intervals $I_0, \ldots, I_{-2}$ (where I_{-2} denotes the second to last interval) $\iff y = \sum_{k=1}^{n} d_k \frac{2}{3^k}$ where d_k is 0 or 1. Furthermore, if y is the right bound of the interval I_i, then $F_{S_n}(I_i) = \sum_{k=1}^{n} d_k \frac{1}{2^k}$

  Proof: Base case: F_{S_1} has $I_0 = (-\infty, 0), I_1 = [0, \frac{2}{3}), I_2 = [\frac{2}{3}, \infty)$, so $y_0 = 0, y_1 = \frac{2}{3}$, and $F_{S_1}(I_0) = 0, F_{S_1}(I_1) = \frac{1}{2}$ as desired. Now assume $F_{S_{n-1}}$ satisfies the proposition; let the intervals of $F_{S_{n-1}}$ be denoted $I_{(n-1,0)}, \ldots, I_{(n-1,2^{n-1})}$. We know from the proofs above that any interval from any interval $I_{(n-1,i)}$, $1 \leq i \leq 2^{n-1}$ is split into two intervals $I_{(n,2i-1)}, I_{(n,2i)}$, where $y_{(n,2i)} = y_{(n-1,i)}$ and $y_{(n,2i-1)} = y_{(n-1,i-1)} + \frac{2}{3^n}$. This easily proves ($\implies$) in the first part; ($\impliedby$) follows because of the induction hypothesis and the fact that each interval of $F_{S_{n-1}}$ splits into 2, one where $d_n = 0$ and one where $d_n = 1$.

Finally, the intervals with $d_n = 0$ are of the form $I_{(n,2i)}$, and we know that $y_{(n,2i)} = y_{(n-1,i)}$ and $v_{(n,2i)} = v_{(n-1,i)}$. The intervals with $d_n = 1$ are of the form $I_{(n,2i-1)}$, and $y_{(n,2i-1)} = y_{(n-1,i-1)} + \frac{2}{3^n}$ and $v_{(n,2i-1)} = \frac{1}{2}(v_{(n-1,i-1)} + v_{(n-1,i)})$; but we know that $v_{(n-1,i)} = v_{(n-1,i-1)} + \frac{1}{2^{n-1}}$ (because the first part implies that $F_{S_{n-1}}$ attains all $\frac{a}{2^{n-1}}$, $0 \leq a \leq 2^{n-1}$, and because F_{S_n} is increasing, so if $v_{(n-1,i)} = \frac{a}{2^{n-1}}$, then $v_{(n-1,i-1)}$ must be $\frac{a-1}{2^{n-1}}$). Therefore, $v_{(n,2i-1)} = v_{(n-1,i-1)} + \frac{1}{2^n}$. Because the formulas for $v_{(n,2i-1)}$ and $y_{(n,2i-1)}$ "match", the second part of the proposition holds.

With all that, we see that F_S maps intervals of the form

$$\left[\left(\sum_{k=1}^{n} d_k \frac{2}{3^k}\right) - \frac{1}{3^n}, \sum_{k=1}^{n} d_k \frac{2}{3^k}\right) \mapsto \sum_{k=1}^{n} d_k \frac{1}{2^k}$$

(for all $n \in \mathbb{N}$) and of course $(-\infty, 0) \mapsto 0$ and $[1, \infty) \mapsto 1$. This, of course, describes the Cantor function (or the Cantor uniform distribution). QED!

Problem 2

We have independent $\{X_k\}_{k \geq 1}$ with $X_k \sim \text{Unif}(-k, k)$, so $\mathbb{E}[X_k] = 0$ and $\text{Var}[X_k] = \frac{k^2}{3}$ (formulas from Wikipedia). Define $S_n = \sum_{k=1}^{n} a^k X_k$ for some $a \in (0, 1)$. Observe that

$$\begin{aligned}
\mathbb{E}[S_n] &= \sum_{k=1}^{n} \mathbb{E}\left[a^k X_k\right] = \sum_{k=1}^{n} a^k \mathbb{E}[X_k] = 0 \to 0 < \infty \\
\text{Var}[S_n] &= \sum_{k=1}^{n} \text{Var}\left[a^k X_k\right] = \sum_{k=1}^{n} (a^k)^2 \text{Var}[X_k] = \frac{1}{3} \sum_{k=1}^{n} k^2 a^{2k}
\end{aligned}$$

Letting $V_n = \sum_{k=1}^{n} k^2 a^{2k}$, we can find the closed form by doing the following manipulations:

$$\begin{aligned}
V_n &= 1a^2 + 4a^4 + 9a^6 + \ldots + n^2 a^{2n} \\
a^2 V_n &= 1a^4 + 4a^6 + 9a^8 + \ldots + n^2 a^{2n+2} \\
(1 - a^2) V_n &= 1a^2 + 3a^4 + 5a^6 + \ldots + (2n+1) a^{2n} - n^2 a^{2n+2} \\
a^2 (1 - a^2) V_n &= 1a^4 + 3a^6 + 5a^8 + \ldots + (2n+1) a^{2n+2} - n^2 a^{2n+4} \\
(1 - a^2)^2 V_n &= 1a^2 + 2a^4 + 2a^6 + \ldots + 2a^{2n} - (n+1)^2 a^{2n+2} + n^2 a^{2n+4} \\
(1 - a^2)^2 V_n + a^2 &= 2 \sum_{k=1}^{n} a^{2k} - (n+1)^2 a^{2n+2} + n^2 a^{2n+4} \\
&= 2 \frac{a^2 (1 - a^{2n+2})}{1 - a^2} - (n+1)^2 a^{2n+2} + n^2 a^{2n+4}
\end{aligned}$$

Taking the limit as $n \to \infty$, we get

$$\begin{aligned}
(1 - a^2)^2 V + a^2 = \frac{2a^2}{1 - a^2} &\implies V = \frac{2a^2 - a^2(1 - a^2)}{(1 - a^2)^3} = \frac{a^2(1 + a^2)}{(1 - a^2)^3} \\
&\implies \text{Var}[S_n] \to \frac{1}{3} \cdot \frac{a^2(1 + a^2)}{(1 - a^2)^3} < \infty
\end{aligned}$$

By the 2-series theorem, $S_n \to_{\text{a.s.}} S$ where $\mathbb{E}[S]$ and $\text{Var}[S]$ equal the limiting values calculated above.

Problem 3

Let $\{X_k\}$ be a sequence of arbitrary random variables ≥ 0 with $\sum_{k=1}^{\infty} \mathbb{E}[\min(X_k, c)] < \infty$ $(\forall c > 0)$. We can split this expression into

$$\sum_{k=1}^{\infty} \mathbb{E}\big[X_k \cdot 1_{[X_k \leq c]} + c \cdot 1_{[X_k > c]}\big] = \sum_{k=1}^{\infty} \big(\mathbb{E}\big[X_i \cdot 1_{[X_k \leq c]}\big] + cP([X_k > c])\big) < \infty$$

which implies that $\sum_{k=1}^{\infty} P([X_k > c]) < \infty$. Borel-Cantelli tells us that $P([X_k > c] \text{ i.o.}) = 0$, which means that there is some N_c s.t. $\forall k > N_c$, $X_k \leq c$ almost surely. Then,

$$\sum_{k=N_c}^{\infty} \mathbb{E}[X_k] = \sum_{k=N_c}^{\infty} \big(\mathbb{E}\big[X_i \cdot 1_{[X_k \leq c]}\big] + cP([X_k > c])\big) = \sum_{k=N_c}^{\infty} \mathbb{E}[\min(X_k, c)] < \infty$$

Let $S_{N_c,n} = \sum_{k=N_c}^{n} X_k$ and let $a_n = \mathbb{E}[S_{N_c,n}]$. The a_n are increasing (because $X_k \geq 0$) and bounded, so $a_n \to a$. Markov's inequality gives that

$$P([S_{N_c,n} \geq \lambda]) \leq \frac{a_n}{\lambda} \leq \frac{a}{\lambda}$$

For any $\omega \in \Omega$, $S_{N_c,n}(\omega)$ is increasing (because $X_k \geq 0$). If it's unbounded, set $S_{(N_c)}(\omega) = \infty$; otherwise, we know the limit exists (by the monotone sequence theorem) so set $S_{(N_c)}(\omega) = \lim_{n\to\infty} S_{N_c,n}(\omega)$. In other words

$$S_{(N_c)}(\omega) = \begin{cases} \lim_{n\to\infty} S_{N_c,n}(\omega) & \text{if the limit exists} \\ \infty & \text{otherwise} \end{cases}$$

Now observe that for all appropriate n,

$$[S_{N_c,n} \geq \lambda] \subseteq [S_{N_c,n+1} \geq \lambda]$$

so by the limit-measure commutativity theorem for monotone sets,

$$P([S_{(N_c)} \geq \lambda]) \leq P\left(\bigcup_{N_c}^{\infty} [S_{N_c,n} \geq \lambda - \epsilon]\right) = \lim_{n\to\infty} P([S_{N_c,n} \geq \lambda - \epsilon]) \leq \frac{a}{\lambda - \epsilon}$$

Thus taking $\lambda \to \infty$, we see that $P([S_{(N_c)} = \infty]) = 0$. Which means that almost surely, $S_{N_c,n}(\omega)$ converge to $S_{(N_c)}(\omega)$. Finally, for $n \geq N_c$,

$$S_n = \sum_{k=1}^{N_c-1} X_k + S_{N_c,n} \quad \text{and} \quad S = \sum_{k=1}^{N_c-1} X_k + S_{(N_c)}$$

which inherits the result from above; mainly that $S_n \to_{\text{a.s.}} S$.

Problem 4

We have i.i.d. $Z_1, Z_2, \ldots \sim N(0,1)$. Define $W_n^2 = \sum\limits_{k=1}^{\infty} \frac{Z_k^2}{(\pi k)^2}$. Using formulas from Wikipedia about the moments of the normal distribution and known values of $\zeta(s)$,

$$\mathbb{E}\big[W_n^2\big] = \sum_{k=1}^{n} \mathbb{E}\left[\frac{Z_k^2}{(\pi k)^2}\right] = \frac{1}{\pi^2} \sum_{k=1}^{n} \frac{1}{k^2} \mathbb{E}\big[Z_k^2\big] \to \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{i^2} = \frac{1}{\pi^2} \cdot \frac{\pi^2}{6} = \frac{1}{6} < \infty$$

and

$$\begin{aligned}\mathrm{Var}\big[W_n^2\big] &= \sum_{k=1}^{n} \mathrm{Var}\left[\frac{Z_k^2}{(\pi k)^2}\right] = \frac{1}{\pi^4} \sum_{k=1}^{n} \frac{1}{k^4} \mathrm{Var}\big[Z_k^2\big] = \frac{1}{\pi^4} \sum_{k=1}^{n} \frac{1}{k^4} \big(\mathbb{E}\big[Z_k^4\big] - \mathbb{E}\big[Z_k^2\big]\big) \\ &= \frac{1}{\pi^4} \sum_{k=1}^{n} \frac{1}{k^4} \big(3\mathrm{Var}[Z_k] - \mathrm{Var}[Z_k]\big) \to \frac{1}{\pi^4} \sum_{k=1}^{\infty} \frac{2}{k^4} = \frac{1}{\pi^4} \cdot \frac{2\pi^4}{90} = \frac{1}{45} < \infty\end{aligned}$$

By the 2-series theorem, $W_n^2 \to_{\text{a.s.}} W^2$, where $E[W^2] = \dfrac{1}{6}$.

Problem 5

We have $Y_1, Y_2, \ldots$ i.i.d. Exponential(1), so $\mathbb{E}[Y_k] = 1$ and $\mathrm{Var}[Y_k] = 1$,

$$\mathbb{E}[S_n] = \sum_{k=1}^{n} \mathbb{E}\left[\frac{Y_k - 1}{\lambda_k}\right] = \sum_{k=1}^{n} \frac{1}{\lambda_k} \mathbb{E}[Y_k - 1] = 0 \to 0 < \infty$$

Also,

$$\mathrm{Var}[S_n] = \sum_{k=1}^{n} \mathrm{Var}\left[\frac{Y_k - 1}{\lambda_k}\right] = \sum_{k=1}^{\infty} \frac{1}{\lambda_k^2} \mathrm{Var}[Y_k - 1] = \sum_{k=1}^{n} \frac{1}{\lambda_k^2} \mathrm{Var}[Y_k] \to \sum_{k=1}^{\infty} \frac{1}{\lambda_k^2} < \infty$$

as from the given. Thus, by the 2-series theorem, $S_n \to_{\text{a.s.}} S$, where $\mathbb{E}[S] = 0$ and $\mathrm{Var}[S] = \dfrac{1}{\lambda^2}$.