Ryan McCorvie

Exercises from Probability with Martingales

> **EG 1** Two points are chosen at random on a line $\overline{AB}$ independently according to the uniform distribution. If the line is cut at these two points, what's the probability the segments will form a triangle?

The sample space is the unit square $[0,1]^2$ with points corresponding to (x_1, x_2). The probability measure is the same as area in this square. A necessary and sufficient condition to form a triangle is that the segments $[0, x_1]$, $[x_1, x_2]$ and $[x_2, 1]$ satisfy the triangle inequalities. Assuming $x_1 < x_2$, these take the form

$$x_1 < (x_2 - x_1) + (1 - x_2) \qquad 1 - x_2 < x_1 + (x_2 - x_1) \qquad x_2 - x_1 < (1 - x_2) + x_1 \tag{1}$$

The first inequality says $x_1 < \frac{1}{2}$, the second that $x_2 > \frac{1}{2}$ and the last that $x_2 - x_1 < \frac{1}{2}$. These inequalities describe a triangle in the unit square with area $\frac{1}{8}$. Considering also the symmetrical when $x_1 > x_2$, the total probability is $\frac{2}{8}$ ∎

> **EG 2** Planet X is a ball with center O. Thre spaceships A, B and C land at random on its surface, their positions being independent and each uniformly distributed on the surface. Spaceships A and B can communicate directly by radio if $\angle AOB \leq 90°$. What's the probability all three can communicate? (For example, A can communicate to C via B if necessary).

A spaceship X may communicate with any spaceship in the points in the hemisphere centered at X. Denote this by $h(X)$. Without loss of generality, rotate the sphere so that A is the north pole, and B lies in a fixed plane through O and A. The three ships may communicate if:

1. B is in the northern hemisphere and C is in the northern hemisphere (the ships can communicate via A)

2. B is in the northern hemisphere and C is in the half-lune $h(B) \setminus h(A)$ (the ships can communicate via B)

3. B is in the southern hemisphere and C is in the half-lune $h(A) \cap h(B)$ (the ships can communicate via C)

These cases are disjoint, so we can consider them separately and sum their contributions. The area of a half lune is proportional to angle $\angle XOY$ where X and Y are the centers of the hemispheres. This is the same as the dihedral angle between the equitorial planes of the hemispheres. Thus the probability of being in a half lune of angle θ is $\theta/2\pi$.

So writing one integral for each case above, conditioning on the angle $\theta = \angle AOB$ and giving the probabity for C to be in the the configuration described by the case gives the equation

$$P = \int_0^{\pi/2} \frac{1}{2}\, p(d\theta) + \int_0^{\pi/2} \frac{\theta}{2\pi}\, p(d\theta) + \int_{\pi/2}^{\pi} \frac{(\pi - \theta)}{2\pi}\, p(d\theta) \tag{2}$$

Now, the probability density of $\theta = \angle AOB$ is given by $\frac{1}{2}\sin\theta\, d\theta$. Heuristically, this can be seen by consiering the infinitesimally thin strip on the surface of the sphere with constant latitue θ. This is like a cylinder (or frustrum of a cone) of side edge length $r\, d\theta$ and radius $r \sin\theta$, so the area is like $2\pi r^2 \sin\theta\, d\theta$. Normalizing by the total surface area $4\pi r^2$ gives the probability density.

Calculating each contribution

$$\begin{aligned}
\int_0^{\pi/2} \frac{1}{2} \cdot \frac{1}{2} \sin\theta\, d\theta &= \frac{1}{4} \\
\int_0^{\pi/2} \frac{\theta}{2\pi} \cdot \frac{1}{2} \sin\theta\, d\theta &= \frac{1}{4\pi} \\
\int_{\pi/2}^{\pi} \frac{(\pi - \theta)}{2\pi} \cdot \frac{1}{2} \sin\theta\, d\theta &= \frac{1}{4\pi}
\end{aligned} \tag{3}$$

Summing the contributions we find $P = \frac{2+\pi}{4\pi}$ ∎

EG.3 Let G be the free group with two generators a and b. Start at time 0 with the unit element 1, the empty word. At each second multiply the current word on the right by one of the four elements a, a^{-1}, b, b^{-1}, choosing each with probability $\frac{1}{4}$ (independently of previous choices). The choices

$$a, a, b, a^{-1}, a, b^{-1}, a^{-1}, a, b \tag{4}$$

at times 1 to 9 will produce the reduced word aab of length 3 at time 9. Prove that the probability that the reduced word 1 ever occurs at a positive time is 1/3, and explain why it is intuitively clear that (almost surely)

$$(\text{length of reduced word at time } n)/n \to \frac{1}{2}. \tag{5}$$

Multiplying by random elements in $\{a, a^{-1}, b, b^{-1}\}$ is the same as taking a random walk on the Cayley graph of the free group with two generators (see figure 1). This is the rooted 4-regular tree. The depth of a vertex is the length of the unique path to the root, which is the vertex corresponding to 1 in the free group element. For any vertex of depth $d > 0$, there are three edges leading to a node of depth $d + 1$ and one edge leading to a node of depth $d - 1$.

Let p_d be the probability of returning to 1 starting from a node of depth d. Its intuitively clear from the symmetry of the graph that the probability only depends on the depth. There is a graph isomorphism mapping any node of depth d to another node of

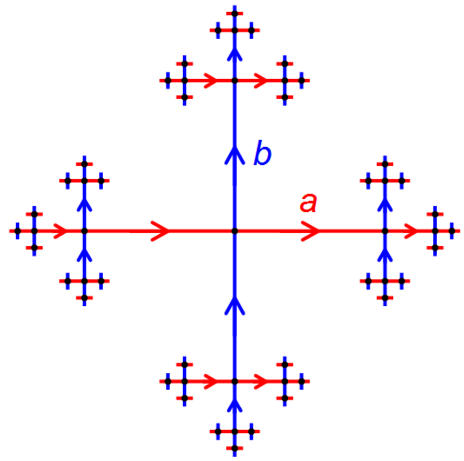

Figure 1: Cayley graph on the free group of 2 generators

depth d, so there is a one-to-one correspondence between paths which lead back to 1, and those paths have equal probability. The probability must satisfy the recurrance

$$p_n = \frac{3}{4} p_{n+1} + \frac{1}{4} p_{n-1} \tag{6}$$

This is because we can condition the return probability on the first node in the walk and use the fact that the return probability depends only the depth of the node. Fundamental solutions of recurrance equations have the form $p_n = \lambda^n$ and satisfy a characteristic polynomial equation which, for this recurrance is

$$3\lambda^2 - 4\lambda + 1 = 0 \qquad \Rightarrow \qquad \lambda = \frac{1}{3} \text{ or } \lambda = 1 \tag{7}$$

The only solution of the form $p_n = a \cdot \frac{1}{3}^n + b$ which also satisfies the boundary conditions $p_0 = 1$ and $p_n \to 0$ as $n \to \infty$ is $p_n = \frac{1}{3}^n$. In particular, a random walk starting at time 1 from a node of depth 1 returns to the origin with probability $\frac{1}{3}$. This is the same as the event that the reduced word is ever 1 at positive time, since every word sequence as one letter at time 1.

Let L_n be the random variable which represents word length at time n. Note that

$$\mathrm{E}[L_n \mid L_{n-1}] = \frac{3}{4}(L_{n-1} + 1) + \frac{1}{4}(L_{n-1} - 1) = L_{n-1} + \frac{1}{2} \tag{8}$$

since L_n is the same thing as vertex depth. Taking expectations and using the tower law

$$\mathrm{E}\, L_n = \mathrm{E}\, L_{n-1} + \frac{1}{2} \tag{9}$$

Hence the expectation satisfies a simple recurrance. Since $L_0 = 0$, this means $\mathrm{E}\, L_n = n/2$. More precisely, the quantity $L_n - L_{n-1}$ is an i.i.d. random variable with finite variance and mean $\frac{1}{2}$. Therefore by the SLLN, $L_n / n \to \frac{1}{2}$ almost surely. ∎

EG.4 Suppose now that the elements a, a^{-1}, b, b^{-1} are chosen instead with respective probabilities $\alpha, \alpha, \beta, \beta$, where $\alpha > 0, \beta > 0, \alpha + \beta = \frac{1}{2}$. Prove that the conditional probability that the reduced word 1 ever occurs at a positive time, *given that* the element a is chosen at time 1, is the unique root $x = (a)$ (say) in $(0,1)$ of the equation

$$3x^2 + (3 - 4\alpha^{-1})x^2 + x + 1 = 0 \tag{10}$$

As time goes on, (it is almost surely true that) more and more of the reduced word becomes fixed, so that a final word is built up. If in the final word, the symbols a and a^{-1} are both replaced by A and the symbols b and b^{-1} are both replaced by B, show that the sequence of A's and B's obtained is a Markov chain on $\{A, B\}$ with (for example)

$$p_{AA} = \frac{\alpha(1-x)}{\alpha(1-x) + 2\beta(1-y)} \tag{11}$$

where $y = r(\beta)$. What is the (almost sure) limiting proportion of occurrence of the symbol a in the final word? (Note. This result was used by Professor Lyons of Edinburgh to solve a long-standing problem in potential theory on Riemannian manifolds.)

■

Algebra's, etc.

1.1 Let $V \subset \mathbb{N}$, and say V has *Cesaro* density $\gamma(V)$ whenever

$$\gamma(V) = \lim_{n\to\infty} \frac{\sharp(V \cap \{1, 2, \ldots, n\})}{n} \tag{12}$$

exists. Let $\mathcal{C}$ be the set of all sets with Cesaro density. Give an example of $V_1, V_2 \in \mathcal{C}$ such that $V_1 \cap V_2 \notin \mathcal{C}$

Take V_1 to be 1 and $n > 10$ whose base 10 representation has an equal first and last digit. Take V_2 to be the natural numbers whose last digit in its base 10 representation is 1.

Then $0 \leq \sharp(V_i \cap \{1, 2, \ldots, n\}) - \lfloor n/10 \rfloor \leq 1$ since in every subset $\{q \cdot 10, q \cdot 10 + 1, q \cdot 10 + 2, \ldots, q \cdot 10 + 9\}$ each of V_1 and V_2 have exactly one representative. Therefore $\gamma(V_1) = \gamma(V_2) = 1$.

Note $V_1 \cap V_2$ consists of 1 and all numbers of the form $1b_2b_3\ldots b_{n-1}1$ where b_i can take any digit from 0 to 9. Thus between 10^k and 10^{k+1} there are 10^{k-1} members of $V_1 \cap V_2$. Therefore if $n = 10^k$,

$$\sharp(V_1 \cap V_2 \cap \{1, \ldots, n\}) = 1 + 1 + 10 + 100 + \cdots + 10^{k-2} = \frac{10^{k-1} - 1}{9} + 1 \tag{13}$$

and $\frac{\sharp(V_1 \cap V_2 \cap \{1, \ldots, n\})}{n} < \frac{1}{89}$. However if $n = 2 \cdot 10^k$ then

$$\sharp(V_1 \cap V_2 \cap \{1, \ldots, n\}) = 1 + 1 + 10 + 100 + \cdots + 10^{k-1} = \frac{10^k - 1}{9} + 1 \tag{14}$$

and $\frac{\sharp(V_1 \cap V_2 \cap \{1,\ldots,n\})}{n} > \frac{1}{18}$. Therefore the limit diverges since the lim sup is greater than the lim inf. ∎

Independence

4.1 Let $(\Omega, \mathcal{F}, \Pr)$ be a probability triple. Let $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ be three π-systems on Ω such that for $k = 1, 2, 3$

$$\mathcal{I}_k \subset \mathcal{F} \text{ and } \Omega \in \mathcal{I}_k \tag{15}$$

Prove that if

$$\Pr(I_1 \cap I_2 \cap I_3) = \Pr(I_1)\Pr(I_2)\Pr(I_3) \tag{16}$$

whenever $I_k \in \mathcal{I}_k$ then the $\sigma(\mathcal{I}_k)$ are independent. Why did we require $\Omega \in \mathcal{I}_k$?

Fix $I_2 \in \mathcal{I}_2$ and $I_3 \in \mathcal{I}_3$. Then consider the mappings for $I \subset \mathcal{I}_1$

$$\mu(I) \hookrightarrow \Pr(I \cap I_2 \cap I_3) \qquad \mu'(I) \hookrightarrow \Pr(I)\Pr(I_2)\Pr(I_3) \tag{17}$$

Each of μ and μ' are countably additive measures because the probability Pr is a countably additive measure on $\mathcal{F}$. By the hypothesis (16), the measures are equal on the π-system $\mathcal{I}_1$. By assumption $\Omega \in \mathcal{I}_1$ as well, so $\mu(\Omega) = \mu'(\Omega)$. If we dropped the assumption that $\Omega \in \mathcal{I}_1$, then the condition that $\mu(\Omega) = \mu'(\Omega)$ is the condition that $\Pr(I_2 \cap I_3) = \Pr(I_2)\Pr(I_3)$, which doesn't hold generically. We conclude from theorem 1.6 that $\mu(I) = \mu'(I)$ for $I \in \sigma(\mathcal{I}_1)$. Since I_2 and I_3 were arbitrary, equation (16) holds for all $I_1 \in \sigma(\mathcal{I}_1), I_2 \in \mathcal{I}_2, I_3 \in \mathcal{I}_3$.

Now we can repeat the above argument with the π-systems $\mathcal{I}'_1 = \mathcal{I}_2, \mathcal{I}'_2 = \mathcal{I}_3, \mathcal{I}'_3 = \sigma(\mathcal{I}_1)$, and observe (16) is invariant to permutations of $\{1, 2, 3\}$, to conclude that (16) holds for $I_1 \in \sigma(\mathcal{I}_1), I_2 \in \sigma(\mathcal{I}_2), I_3 \in \mathcal{I}_3$. Applying the arugment one more time for $\mathcal{I}'_1 = \mathcal{I}_3, \mathcal{I}'_2 = \sigma(\mathcal{I}_1), \mathcal{I}'_3 = \sigma(\mathcal{I}_2)$ gives the result we want, that (16) holds for $I_k \in \sigma(\mathcal{I}_k)$ ∎

4.2 For $s > 1$ define $\zeta(s) = \sum_{n \in \mathbb{N}} n^{-s}$ and consider the distribution on $\mathbb{N}$ given by $\Pr(X = n) = \zeta(s)^{-1} n^{-s}$. Let $E_k = \{n \text{ divisible by } k\}$. Show E_p and E_q are independent for primes $p \neq q$. Show Euler's formula $1/\zeta(s) = \prod_p (1 - 1/p^s)$. Show $\Pr(\text{X is square-free}) = 1/\zeta(2s)$. Let $H = \gcd(X, Y)$. Show $\Pr(H = n) = n^{-2s}/\zeta(2s)$

The elements of $m \in E_k$ are in one-to-one correspondence with elements of $n \in \mathbb{N}$ by the relationship $m = nk$. Therefore

$$\Pr(E_k) = \sum_{m \in E_k} \Pr(m) = \sum_{n \in \mathbb{N}} \Pr(nk) \tag{18}$$

Since

$$\Pr(nk) = k^{-s} n^{-s}/\zeta(s) = k^{-s}\Pr(n) \tag{19}$$

we have

$$\Pr(E_k) = k^{-s} \sum_{n \in \mathbb{N}} \Pr(n) = k^{-s} \tag{20}$$

Furthermore, the conditional probability satisfies

$$Pr(X = kn \mid X \in E_k) = \frac{\Pr(X = nk)}{\Pr(E_k)} = \frac{(nk)^{-s}/\zeta(s)}{k^{-s}} = n^{-s}/\zeta(s) \tag{21}$$

Thus the conditional distribution of X is given by the $\zeta(s)$ distribution on the relative divisor X/k.

Now, for m, n relatively prime, $E_m \cap E_n = E_{mn}$ by unique factorization. Therefore

$$\Pr(E_m \cap E_n) = \Pr(E_{mn}) = (mn)^{-s} = m^{-s}n^{-s} = \Pr(E_m)\Pr(E_n) \tag{22}$$

which proves independence.

Consider the probability X is not divisible by any of the first n primes $p_1, p_2, \ldots, p_n$. This is the event $E = E^c_{p_1} \cap \cdots \cap E^c_{p_n}$ and

$$\Pr(E) = \prod_{i=1}^{n}(1 - \Pr(E_{p_i})) = \prod_{i=1}^{n}(1 - p_i^{-s}) \tag{23}$$

As $n \to \infty$ the set E consists of 1 and really big numbers, certainly every element of E exceeds p_{n+1}, and $\Pr(X > n) \downarrow 0$ as $n \to \infty$. Thus $\Pr(E) \to \Pr(X = 1) = 1/\zeta(s)$. This shows

$$\zeta(s)^{-1} = \prod_p (1 - 1/p^s) \tag{24}$$

The set of integers which don't contain a square prime factor up to p_n is given by

$$F = E^c_{p_1^2} \cap \cdots \cap E^c_{p_n^2} \tag{25}$$

Each of the sets E_{p^2} are independent, because the numbers p^2 are relatively prime. Thus as $n \to \infty$, F converges to a set of square-free numbers up to p_n (at least) and a set of negligible probability, so

$$\Pr(F) \to \prod_p (1 - (p^2)^{-s}) = \prod_p (1 - 1/p^{2s}) = 1/\zeta(2s) \tag{26}$$

Now $n = \gcd(x, y)$ iff X and Y are divisible by n, and the relative divisors n_x, n_y such that $x = n \cdot n_x$ and $y = n \cdot n_y$. By independence the event that both $X \in E_n$ and $Y \in E_n$ satisfies

$$\Pr(X \in E_n, Y \in E_n) = \Pr(X \in E_n)\Pr(Y \in E_n) = n^{-2s} \tag{27}$$

Let $R = \{(a, b) \mid a, b \text{ relatively prime}\}$ where we take the convention $(1, 1) \notin R$. We can calulate $\Pr((X, Y) \in R)$ as follows. First consider the sets $F_p = \{X \in E_p\} \cap \{Y \in E_p\}$. By independence $\Pr(F_p) = \Pr(X \in E_p)\Pr(Y \in E_p) = p^{-2s}$. The F_p are independent of each other since the E_p are. Then F_p^c is the event that p is not a divisor of at least one of X and Y. Hence $F = \bigcap_p F_p^c$ is the event that X and Y have no common divisor.

$$\Pr(X, Y \text{ relatively prime}) = \Pr(F) = \prod_p (1 - p^{-2s}) = 1/\zeta(2s) \tag{28}$$

Pulling it together

$$\Pr(\gcd(X,Y)=n) = \Pr(A_n)\Pr(B_n \mid A_n) = n^{-2s}/\zeta(2s) \tag{29}$$

where $A_n = \{X \in E_n\} \cap \{Y \in E_n\}$ and $B_n = \{(X/n, Y/n) \in R\}$.

■

4.3 Let $X_1, X_2, \ldots$ be i.i.d. random variables from a continuous distribution. Let $E_1 = \Omega$ and for $n \geq 2$ let $E_n = \{X_n = \max(X_1, X_2, \ldots, X_n)\}$, that is that a "record" occurs at event n. Show that the events E_n are independent and that $\Pr(E_n) = 1/n$.

A continuous distribution has the property that singletons have measure zero $\mu(X = x) = 0$. Thus conditioning on the value of X_1, it follows that $\{X_1 = X_2\}$ has measure zero also, by Fubini's theorem $\Pr(X_1 = X_2) = \int_\Omega \Pr(X_2 = x)\,\mu(dx) = 0$. Since there are a finite number of pairs of variables, we may assume none of the X_i are coincident by excluding a set of measure zero.

Allow a permutation $\sigma \in S_n$ to act on the sample space by permuting the indices. That is $\sigma : (X_1, X_2, \ldots, X_n) \hookrightarrow (X_{\sigma 1}, X_{\sigma 2}, \ldots, X_{\sigma n})$. Clearly this also permutes the ranks so that $\rho_i(\sigma X) = \rho_{\sigma i}(X)$, or in vector form, $\rho(\sigma X) = \sigma\rho(X)$. Because the X_i are i.i.d., the permutation σ is measure preserving $\Pr(E) = \Pr(\sigma E)$ for any event $E \subset \Omega^n$. Consider the Weyl chamber $W = \{x_1 < x_2 < \cdots < x_n\}$. Clearly the disjoint σW partition Ω^n, excluding a set of measure zero of coincident points, since every point $x \in \Omega^n$ is in $\rho(x)^{-1}(W)$ where $\rho(x)$ is the permutation induced by mapping each component to its ordinal ranking from smallest to greatest.

We infer that $\Pr(W) = 1/n!$, since the disjoint copies of W have equal probability and comprise the entire space. The event E_n is the union of all the sets σW where $\sigma n = n$. There are $(n-1)!$ such permutations, so $\Pr E_n = (n-1)!/n! = 1/n$.

For any $\sigma W \subset E_n$ we can apply any permutation $\sigma' \in S_m \subset S_n$ which acts only on the first m letters and keeps the rest fixed. For any $\sigma' \in S_m$, $\sigma'\sigma W \subset E_n$ since the nth component is largest and unchanged by the action of σ'. Let us group the sets $\sigma W \subset E_n$ into the orbits of the action of S_m. From each orbit we may choose the representative with $\rho_1(x) < \rho_2(x) < \cdots < \rho_m(x)$ for each $x \in \sigma W$, which is formed by permuting the first m elements into sorted order. Let U be the union of all orbit representatives. For $\sigma' \in S_m$ the sets $\sigma' U$ are disjoint and partition E_n. There are $m!$ such sets, they all have equal probability, and their union has probability $1/n$, so $\Pr(U) = 1/(n \cdot m!)$. Of these exactly $(m-1)!$ are also in E_m. These correspond to $\sigma' \in S_m$ which satisfy $\sigma' m = m$. Therefore $\Pr(E_m \cap E_n) = 1/(mn) = \Pr(E_m)\Pr(E_n)$ and the sets are independent.

■

Borel-Cantelli Lemmas

4.4 Suppose a coin with probability p of heads is tossed repeatedly. Let A_k be the event that a sequence of k (or more) heads occurs among the tosses numbered $2^k, 2^k+1, 2^k+1, \ldots, 2^{k+1}-1$. Prove that

$$\Pr(A_k \text{ i.o}) = \begin{cases} 1 & \text{if } p \geq \frac{1}{2} \\ 0 & \text{if } p < \frac{1}{2} \end{cases} \tag{30}$$

Let $E_{k,i}$ be the event that there is a string of k heads at times j for

$$2^k + i \leq j < 2^k + i + k - 1 \tag{31}$$

Observe $\Pr(E_{k,i}) = p^k$. If $m_k = 2^k - k$ and $i \leq m_k$ then $E_{k,i} \subset A_k$. Furthmore $A_k \subset \bigcup_{i=0}^{m_k} E_{k,i}$ since if there's a string of k heads in this range, it has to start somewhere. Thus by the union bound

$$\Pr(A_k) \leq \sum_{i=0}^{m_k} E_{k,i} = m_k \cdot \Pr(E_{k,i}) \leq (2p)^k \tag{32}$$

If $p < \frac{1}{2}$ then $\sum_k \Pr(A_k) < \infty$. By Borel-Cantelli 1, we must have $\Pr(A_k \text{ i.o.}) = 0$.

Now consider $E_{k,0}, E_{k,k}, E_{k,2k}, \ldots$ so long as $ik \leq m_k$. Now the sets are independent and

$$A_k \supset \bigcup_{i=0}^{\lfloor m_k/k \rfloor} E_{k,ik} \tag{33}$$

Taking the inclusion-exclusion inequality

$$\Pr(A_k) \geq \sum_i \Pr(E_{k,ik}) - \sum_{i,j} \Pr(E_{k,ik} \cap E_{k,jk}) \tag{34}$$

All of the terms in the first sum are p^k and there are $\lfloor m_k/k \rfloor \geq 2^k/k - 2$ terms. All the terms in the second sum are p^{2k} since the $E_{k,ik}$ are independent and there are $\binom{\lfloor m_k/k \rfloor}{2}$ of them. Therefore when $p = \frac{1}{2}$

$$\Pr(A_k) \geq \left(\frac{2^k}{k} - 2\right) p^k - \left(\frac{2^{2k}}{2k^2}\right) p^{2k} = \frac{1}{k} - 2^{-(k-1)} - \frac{1}{2k^2} \tag{35}$$

Therefore $\sum \Pr(A_k) = \infty$, since its greater than sum of the harmonic series and a convergent series. Since the A_k are independent, the second Borel Cantelli lemma says that $\Pr(A)k$ i.o.$) = 1$.

As a function of p, the probability $\Pr(A_k\text{i.o.})$ must non-decreasing, since increasing the probability of heads increases the likelihood of the event. However we just showed this probability is 1 for $p = \frac{1}{2}$. Therefore the probability is 1 for any $p > \frac{1}{2}$. ■

4.5 Prove that if $G \sim \mathcal{N}(0,1)$ then for $x > 0$

$$\Pr(G > x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{36}$$

Let X_i $\mathcal{N}(0,1)$ be i.i.d.. Show that $L = 1$ almost surely where

$$L = \limsup(X_n/\sqrt{2\log n}) \tag{37}$$

Let $S_n = \sum_i X_i$. Recall $S_n/\sqrt{n} \sim \mathcal{N}(0,1)$. Prove that

$$|S_n| < 2\sqrt{n \log n}, \text{ eventually almost surely} \tag{38}$$

Let's do a calcualtion, noting that $u/x \geq 1$ in the range of the integral

$$\int_x^\infty \frac{e^{-\frac{1}{2}u^2}}{u^k}\,du \leq \frac{1}{x^{k+1}} \int_x^\infty u e^{-\frac{1}{2}u^2}\,du = \frac{e^{-\frac{1}{2}x^2}}{x^{k+1}} \tag{39}$$

Now integrating by parts with $v = xe^{-\frac{1}{2}x^2}$ and $u' = 1/x$ we get

$$\sqrt{2\pi}\,\Pr(G > x) = \int_x^\infty e^{-\frac{1}{2}u^2}\,du = \frac{e^{-\frac{1}{2}x^2}}{x} - \int_x^\infty \frac{e^{-\frac{1}{2}x^2}}{x^2}\,dx \tag{40}$$

Alternately dropping the last term or applying (39) with $k = 2$ we find

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-\frac{1}{2}x^2} \leq \Pr(G > x) \leq \frac{1}{\sqrt{2\pi}}\frac{1}{x} e^{-\frac{1}{2}x^2} \tag{41}$$

If $I(x) = x^{-1}e^{-\frac{1}{2}x^2}$ then this shows for any $\epsilon > 0$ of x is large enough then $\frac{1-\epsilon}{\sqrt{2\pi}} I(x) \leq P(G > x) \leq \frac{1}{\sqrt{2\pi}}$. Now

$$I(\sqrt{2\alpha \log n}) = \frac{1}{\sqrt{\alpha \log n}} e^{-\alpha \log n} = \frac{1}{x^\alpha \sqrt{\alpha \log n}} \tag{42}$$

Using the substitution the substitution $u = \sqrt{\log x}$ we can apply the integral test

$$\int_a^\infty I(\sqrt{2\alpha \log n})\,dx = \int_a^\infty \frac{dx}{x^\alpha \sqrt{\alpha \log x}} = \frac{2}{\sqrt{\alpha}} \int_a^\infty e^{(1-\alpha)u^2}\,du \tag{43}$$

Clearly the integral converges if $\alpha > 1$ and diverges if $\alpha \leq 1$, so the same is true of

$$\sum_n \Pr(X_n > \sqrt{2\alpha \log n}) = O(\sum_n I(\sqrt{2\alpha \log n})) \tag{44}$$

The X_n are independent so we may apply Borel-Cantelli and its converse to conclude for $\alpha > 1$

$$\limsup X_n / \sqrt{2\alpha \log n} \leq 1 \text{ a.s.} \qquad \liminf X_n / \sqrt{2 \log n} \geq 1 \text{ a.s.} \tag{45}$$

Let $\alpha \downarrow 1$ to conclude $\lim L = 1$ almost surely.

Note that we only used independence in the lower bound, the upper bound uses plain jane Borel-Cantelli which does not assume independence. Thus $Y_n = S_n/\sqrt{n}$ are a sequence of $\mathcal{N}(0,1)$ random variables. If we take $\alpha = \sqrt{2}$ then conclude

$$Y_n < 2\sqrt{\log n} \text{ eventually, a.s.} \tag{46}$$

The same is true for $-Y_n$ which has the same distribution, so eventually (taking the greater eventuality) $|Y_n| < 2\sqrt{\log n}$. This is the same as $|S_n| < 2\sqrt{n \log n}$ eventually almost surely. In particular this implies the strong law of large numbers for $X_n \sim \mathcal{N}(\mu, \sigma^2)$ since $|S_n/n - \mu| < 2\sigma\sqrt{\frac{\log n}{n}} \to 0$ eventually almost surely. ∎

4.6 This is a converse to the SLLN. Let X_n be a sequence of i.i.d. RV with $\mathrm{E}[|X_n|] = \infty$. Prove that

$$\sum_n \Pr(|X_n| > kn) = \infty \text{ for } \forall k \qquad \text{and} \qquad \limsup \frac{|X_n|}{n} = \infty \text{ a.s.} \tag{47}$$

For $S_n = \sum_{i=1}^n X_n$ conclude

$$\limsup \frac{|S_n|}{n} = \infty \text{ a.s.} \tag{48}$$

Then

$$\lfloor Z \rfloor = \sum_{n \in \mathbb{N}} I_{Z \geq n} \tag{49}$$

since each indicator with $n < Z$ adds 1 to the sum and all the other indicaters are 0, and there are $\lfloor Z \rfloor$ such indicators. Taking expectations of both sides and noting that $\lfloor Z \rfloor \leq Z \leq \lfloor Z \rfloor + 1$

$$E[Z] - 1 \leq \sum_{n \in \mathbb{N}} \Pr(Z \geq n) \leq E[Z] \tag{50}$$

FOr any $k \in \mathbb{N}$, take $Z = |X_1|/k$, since $E[Z] = \infty$ we get that

$$\sum_n \Pr(|X_1|/k > n) \geq E[Z] - 1 = \infty \tag{51}$$

For i.i.d. X_n since $\Pr(|X_n| > \alpha) = \Pr(|X_1| > \alpha)$ for any $\alpha \geq 0$ this implies

$$\sum_n \Pr(|X_n| > kn) = \infty \tag{52}$$

By the converse of Borel-Cantelli we must have $|X_n|/n > k$ infinitely often, and therefore $\limsup |X_n|/n \geq k$ for every $k \in \mathbb{N}$. We conclude $\limsup |X_n|/n = \infty$.

Suppose $|S_n|/n \leq k$ eventually for some $k > 0$. That is, there exists a $k > 0$ and $N > 0$ such that if $n \geq N$ then $|S_n| \leq kn$. We know that almost surely there exists an $n \geq N + 1$ such that $|X_n| \geq 2kn$. But then

$$|S_n| \geq |X_n| - |S_{n-1}| \geq 2kn - k(n-1) > kn \tag{53}$$

This is a contradiction. Hence, with probability 1, $|S_n|/n > k$ infinitely often for any $k > 0$ and therefore

$$\limsup \frac{|S_n|}{n} = \infty \text{ a.s.} \tag{54}$$

■

4.7 What's fair about a fair game? Let $X_1, X_2, \ldots$ be independent random variables such that

$$X_n = \begin{cases} n^2 - 1 & \text{with probability } n^{-2} \\ -1 & \text{with probability } 1 - n^{-2} \end{cases} \tag{55}$$

Prove that $\mathrm{E}\, X_n = 0$ for all n, but that if $S_n = X_1 + X_2 + \cdots + X_n$ then

$$\frac{S_n}{n} \to -1 \text{ a.s.} \tag{56}$$

An easy calculation shows

$$\mathrm{E}\,X_n = (n^2-1)\cdot \Pr(n^2-1) + -1\cdot \Pr(-1) \tag{57}$$
$$= (1-n^{-2}) + (n^{-2}-1) = 0 \tag{58}$$

Now let A_n be the event $X_n = -1$. Since

$$\sum_n \Pr(A_n^c) = \sum_n n^{-2} < \infty \tag{59}$$

By Borel-Cantelli, with probability 1, only finitely many of the A_n^c occur. Therefore, with probability 1, $X_n = -1$ with a finite number of exceptions. On any such sample, let N be the largest index such that $X_N = 1 - n^{-2}$. Then for $n > N$

$$\frac{S_n}{n} = \frac{S_N}{n} + \frac{-1\cdot(n-N)}{n} \tag{60}$$

The as $n \to \infty$, the first term tends to zero because S_N is fixed and the second term tends to -1. (Note the indepdendence of X_n was not needed in this argument) ∎

4.8 This exercise assumes that you are familiar with continuous-parameter Markov chains with two states.

For each $n \in \mathbb{N}$, let $X^{(n)} = \{X^{(n)}(t) : t \geq 0\}$ be a Markov chain with state-space the two-point set $\{0,1\}$ with Q-matrix

$$Q^{(n)} = \begin{pmatrix} -a_n & a_n \\ b_n & -b_n \end{pmatrix}, \qquad a_n, b_n > 0 \tag{61}$$

and transition function $P^{(n)}(t) = \exp(tQ^{(n)})$. Show that, for every t,

$$p_{00}^{(n)}(t) \geq b_n/(a_n+b_n), \qquad p_{01}^{(n)}(t) \leq a_n/(a_n+b_n) \tag{62}$$

The processes $(X^{(n)} : n \in \mathbb{N})$ are independent and $X^{(n)}(0) = 0$ for every n. Each $X^{(n)}$ has right-continuous paths.

Suppose that $\sum a_n = \infty$ *and* $\sum a_n/b_n < \infty$. Prove that if t is a fixed time then

$$\Pr(X^{(n)}(t) = 1 \text{ for infinitely many } n) = 0 \tag{63}$$

Use Weierstrass's M-test to show that $\sum_n \log p_{00}^{(n)}(t)$ is *uniformly convergent* on $[0,1]$, and deduce that

$$\Pr(X^{(n)}(t) = 0 \text{ for ALL } n) \to 1 \qquad \text{as } t \downarrow 0 \tag{64}$$

Prove that

$$\Pr(X^{(n)}(s) = 0, \forall s \leq t, \forall n) = 0 \text{ for every } t > 0 \tag{65}$$

and discuss with your tutor why it is almost surely true that within every non-empty time interval, infinitely many of the $X^{(n)}$) chains jump.

First let's calculate the matrix exponential

$$Q = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix} \qquad \text{we have} \qquad Q^2 = \begin{pmatrix} a^2+ab & -a^2-ab \\ -ab-b^2 & ab+b^2 \end{pmatrix} = -(a+b)Q \tag{66}$$

Thus, by induction, $Q^n = (-a-b)^{n-1}\, Q$, and therefore

$$\begin{aligned} \exp(tQ) &= I + \sum_{k\geq 1} \frac{t^k Q^k}{k!} = I - \frac{Q}{a+b} \sum_{k \geq 1} \frac{t^k(-a-b)^k}{k!} = I - \frac{e^{-(a+b)t}-1}{a+b} Q \\ &= \frac{1}{a+b} \begin{pmatrix} b + ae^{-t(a+b)} & a - ae^{-t(a+b)} \\ b - be^{-t(a+b)} & a + be^{-t(a+b)} \end{pmatrix} \end{aligned} \tag{67}$$

From this (62) is evident.

For fixed $t > 0$,

$$\sum_n \Pr(\{X^{(n)}(t) = 1\}) = \sum_n p_{01}^{(n)}(t) \leq \sum_n \frac{a_n}{a_n + b_n} \leq \sum_n \frac{a_n}{b_n} < \infty \tag{68}$$

Thus by Borel-Cantelli, almost surely, at any time t only finitely many of the $X^{(n)}(t) = 1$

Now consider the joint probability $\Pi(t)$ that every $X^{(n)}(t) = 0$ at time t. This satisfies the bound

$$-\log \Pi(t) = -\log \prod_n p_{00}^{(n)}(t) \leq \sum_n \log\left(\frac{a_n + b_n}{b_n}\right) \leq \sum_n \frac{a_n}{b_n} < \infty \tag{69}$$

This gives a uniform bound for the convergence of $-\log \Pi(t)$, implies the product converges uniformly on any compact subset of $\mathbb{R}_{\geq 0}$, such as $[0,1]$. In particular this means that

$$\lim_{t \downarrow 0} \Pi(t) = \Pi(0) = \prod_{n=1}^{\infty} 1 = 1 \tag{70}$$

TODO finish this ∎

Tail σ-algebras

4.9 Let $Y_0, Y_1, Y_2, \dots$ be independent random variables with

$$Y_n = \begin{cases} +1 & \text{probability } \frac{1}{2} \\ -1 & \text{probability } \frac{1}{2} \end{cases} \tag{71}$$

Define

$$X_n = Y_0 Y_1 \cdots Y_n \tag{72}$$

Prove the X_n are independent. Define

$$\mathcal{Y} = \sigma(Y_1, Y_2, \dots), \qquad \mathcal{T}_n = \sigma(X_r : r > n) \tag{73}$$

Prove

$$\mathcal{L} := \bigcap_n \sigma(\mathcal{Y}, \mathcal{T}_n) \neq \sigma\left(\mathcal{Y}, \bigcap_n \mathcal{T}_n\right) =: \mathcal{R} \tag{74}$$

First note $\Pr(X_n = 1) = \Pr(X_n = -1) = \frac{1}{2}$, so X_n has the same distribution as Y_n. If we let $X_0 = Y_0$, this follows by induction since

$$X_n = \begin{cases} X_{n-1} & \text{with probability } \frac{1}{2} \\ -X_{n-1} & \text{with probability } \frac{1}{2} \end{cases} \tag{75}$$

and so, for example,

$$\Pr(X_n = 1) = \frac{1}{2}\Pr(X_n = 1) + \frac{1}{2}\Pr(X_{n-1} = -1) = 2 \cdot \frac{1}{4} = \frac{1}{2} \tag{76}$$

For $m < n$ and $\epsilon_m, \epsilon_n \in \{-1, 1\}$, consider $\Pr(\{X_m = \epsilon_m\} \cap \{X_n = \epsilon_n\})$. Clearly this is the same as $\Pr(\{X_m = \epsilon_m\} \cap \{X_m \cdot X_n = \epsilon_m \epsilon_n\})$. Given X_m and X_n we can multiply them to generate $X_m \cdot X_n$. Conversely, given X_m and $X_m \cdot X_n$ we can multiply them to recover X_n. Note also that

$$X_m \cdot X_n = (Y_0 \cdot Y_1 \cdots Y_m)^2 \, Y_{m+1} \cdots Y_n = Y_{m+1} \cdots Y_n \tag{77}$$

Hence X_m and $X_m \cdot X_n$ are independent, since they are functions of independent random variables. Thus

$$\begin{aligned} \Pr(\{X_m = \epsilon_m\} \cap \{X_n = \epsilon_n\}) &= \Pr(\{X_m = \epsilon_m\} \cap \{X_m \cdot X_n = \epsilon_m \epsilon_n\}) \\ &= \Pr(\{X_m = \epsilon_m\}) \Pr(\{X_m \cdot X_n = \epsilon_m \epsilon_n\}) \\ &= \frac{1}{2} \cdot \frac{1}{2} = \Pr(X_m = \epsilon_m) \Pr(X_n = \epsilon_n) \end{aligned} \tag{78}$$

This argument generalizes to any finite set of $X_{n_1}, \ldots, X_{n_k}$ by considering the isometric transformation

$$\begin{pmatrix} X_{n_1} \\ X_{n_2} \\ \vdots \\ X_{n_k} \end{pmatrix} \to \begin{pmatrix} X_{n_1} \\ X_{n_1} X_{n_2} \\ \vdots \\ X_{n_1} X_{n_2} \cdots X_{n_k} \end{pmatrix} \tag{79}$$

and noting each of the variables on the right hand side are independent with the same distribution as X_n so the joint probability is 2^{-k} for any fixed values ϵ_{n_k}. But this is the same as $\prod_k \Pr(X_{n_k} = \epsilon_{n_k})$.

Furthermore X_n is independent of Y_m. If $m > n$ this is obvious because the definition of X_m involves terms independent of Y_n by definition. If $m \leq n$ then consider the isometric transformation $(Y_m, X_n) \to (Y_m, X_n \cdot Y_m)$. The second variable is the product $Y_0 \cdots Y_{m-1} \cdot Y_{m+1} \cdots Y_n$ which is clearly independent of Y_m

Suppose we know the values of Y_k for all $k \geq 1$ and also some X_r. Then we can immediately deduce the value of Y_0 since

$$Y_0 = Y_0 \cdot (Y_1 \cdots Y_r)^2 = X_r \cdot Y_1 \cdot Y_2 \cdots Y_r \tag{80}$$

This shows that for any n, Y_0 is measurable in the algebra $\sigma(Y_1, \ldots, Y_n, X_n) \subset \sigma(\mathcal{Y}, \mathcal{T}_{n-1})$. Therefore Y_0 is measurable in $\mathcal{L} = \bigcap_n \sigma(\mathcal{Y}, \mathcal{T}_n)$

On the other hand, Y_0 is independent of $\mathcal{R}$. Let $\mathcal{T} = \cap_n \mathcal{T}_n$ be the tail field. By the above argument Y_0 is independent of $\sigma(Y_1, \ldots, Y_m, \mathcal{T}_n)$ for any $n > m + 1$. Therefore Y_0 is independent of $\mathcal{R}_m = \sigma(Y_1, \ldots, T_m, \mathcal{T})$ since its a subset. But $\mathcal{R}_m \subset \mathcal{R}_{m+1}$ and so Y_0 is independent of $\cup_n \mathcal{R}_n = \mathcal{R}$ ∎

5.1 Let $S = [0,1]$, $\Sigma = \mathcal{B}(S)$, μ be the Lesbegue measure. Define $f_n = nI_{(0,1/n)}$. Prove that $f_n(s) \to 0$ for every $s \in S$ but that $\mu(f_n) = 1$ for every n. Draw a picture of $g = \sup_n |f_n|$ and show that $g \notin \mathcal{L}^1(S, \Sigma, \mu)$

If $n > 1/s$ then $f_n(s) = 0$ because the indicator $I_{(0,1/n)}$ doesn't put any mass at points as large as s. So $f_n(s)$ eventually for every s and $\lim f_n(s) \to 0$. On the other hand $\mu(f_n) = n\mu(0, 1/n) = \frac{n}{n} = 1$. The function $g = \sup_n |f_n|$ sort of looks like the function $1/x$ near 0, except it has stairsteps at each $1/n$. It takes the value n on the range $(1/(n+1), n]$. Now $g_n = \max(|f_1|, \ldots, |f_n|)$ is a simple function which satisfies

$$\mu g_n \geq \sum_{k=1}^{n} k \left(\frac{1}{k} - \frac{1}{k+1} \right) = \sum_{k=1}^{n} \frac{1}{k+1} \tag{81}$$

The sum diverges as $n \uparrow \infty$, which means μg also diverges since its greater than any μg_n. ∎

5.2 Prove inclusion-exclusion considering integrals of indicator functions

We will use the following property of indicators repeatedly: for events A and B,

$$I_A I_B = I_{A \cap B} \tag{82}$$

This can be verified by considering the four cases $x \in (A \cup B)^c$, $x \in A \setminus B$, $x \in B \setminus A$ and $x \in A \cap B$. These four cases partition Ω, and the values of the expressins on both sides of (82) are equal.

Let $A_1, \ldots, A_n$ be a collection of events. To simplify notation I'm going to write $I(A)$ for indicators rather than I_A, but the indicator remains a function of $\omega \in \Omega$. Now by de Morgan's law $(A_1 \cup A_2 \cup \ldots A_n)^c = A_1^c \cap A_2^c \cap \cdots \cap A_n^c$.

$$\begin{aligned}
1 - I(A_1 \cup \cdots \cup A_n) &= I((A_1 \cup \ldots A_n)^c) \\
&= I(A_1^c \cap A_2^c \cap \cdots \cap A_n^c) \\
&= I(A_1^c) I(A_2^c) \ldots I(A_n^c) \\
&= (1 - I(A_1))(1 - I(A_2)) \ldots (1 - I(A_n)) \\
&= 1 - \sum_i I(A_i) + \sum_{i<j} I(A_i) I(A_j) - \sum_{i<j<k} I(A_i) I(A_j) I(A_k) \\
&\quad + \cdots + (-1)^{n-1} I(A_1) I(A_2) \ldots I(A_n) \\
&= 1 - \sum_i I(A_i) + \sum_{i<j} I(A_i \cap A_j) - \sum_{i<j<k} I(A_i \cap A_j \cap A_k) \\
&\quad + \cdots + (-1)^{n-1} I(A_1 \cap A_2 \cap \cdots \cap A_n)
\end{aligned} \tag{83}$$

Take expectations of each side yields inclusion exclusion

$$\begin{aligned}\Pr(A_1 \cup \cdots \cup A_n) = &\sum_i \Pr(A_i) + \sum_{i<j} \Pr(A_i \cap A_j) - \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) \\ &+ \cdots + (-1)^{n-1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n)\end{aligned} \tag{84}$$

Now we prove the inclusion-exclusion in equalities. The d-depth inequality on n terms on the indicator functions is given by

$$\begin{aligned}I(A_1 \cup \cdots \cup A_n) \ \lesseqgtr\ &\sum_i I(A_i) + \sum_{i<j} I(A_i \cap A_j) - \sum_{i<j<k} I(A_i \cap A_j \cap A_k) \\ &+ \cdots + (-1)^{d-1} \sum_{i_1<i_2<\cdots<i_d} I(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_d})\end{aligned} \tag{85}$$

where the inequality is $\leq$ if d is odd and $\geq$ if d is even. For each d we induct on n. Clearly its true for $d = 1$ and $n = 2$ since by inclusion-exclusion

$$I(A_1 \cup A_2) = I(A_1) + I(A_2) - I(A_1 \cap A_2) \leq I(A_1) + I(A_2) \tag{86}$$

Then its true inductively for any n and $d = 1$, assuming its true for $n - 1$

$$\begin{aligned}I(A_1 \cup \cdots \cup A_n) &\leq I(A_1 \cup \cdots \cup A_{n-1}) \cup I(A_n) \\ &\leq I(A_1) + \cdots + I(A_{n-1}) + I(A_n)\end{aligned} \tag{87}$$

For arbitrary d, the inequality is true for $n \leq k$, since in this case its just the inclusion-exclusion *equality*. No terms are truncated. So inductively if its true for $n - 1$, we can group together $A_1 \cup A_2$ and treat it as one set

$$\begin{aligned}I(A_1 \cup \cdots \cup A_n) \ \lesseqgtr\ &I(A_1 \cup A_2) + \sum_i I(A_i) \\ &- \sum_i I((A_1 \cup A_2) \cap A_i) - \sum_{i<j} I(A_i \cap A_j) \\ &+ \sum_{i,j} I((A_1 \cup A_2) \cap A_i \cap A_j) + \sum_{i<j<k} I(A_i \cap A_j \cap A_k) \\ &+ \ldots \\ &+ (-1)^{d-1} \sum_{i_1<i_2<\cdots<i_{d-1}} I((A_1 \cup A_2) \cap A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_{d-1}}) \\ &+ (-1)^{d-1} \sum_{i_1<i_2<\cdots<i_d} I(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_d})\end{aligned} \tag{88}$$

Here indices range from 3 to n since we've explicitly broken out 1 and 2. Using the distributive law, each indicator which includes $A_1 \cup A_2$ can be written as the indicator of a union of two sets $(A_1 \cap B) \cup (A_2 \cap B)$ where $B = A_{i_1} \cap \cdots \cap A_{i_m}$. For all such indicators of $m < d$ terms, use inclusion-exclusion for $n = 2$ to expand $I((A_1 \cup A_2) \cap B) \to I(A_1 \cap B) + I(A_2 \cap B) - I(A_1 \cap A_2 \cap B)$. After performing all these expansions, we see we recover most of the terms inclusion-exclusion formula up to level d. This is because the terms of level m expand to provide all the missing terms in level m which have exactly one of A_1 or A_2, and also to provide all the missing terms in level $m+1$ including both A_1 and A_2. Terms with neither A_1 or A_2 were already present before the expansion.

Finally we extend the inequality by expanding the terms in level d using the inequality $I((A_1 \cup A_2) \cap B) \leq I(A_1 \cap B) + I(A_2 \cap B)$. This provides the missing terms at level d. This also extends the inequality in the right direction, $\leq$ for odd d and $\geq$ for even d, owing to the factor $(-1)^{d-1}$ on these terms.

■

The Strong Law

7.1 Let f be a bounded continuous function on $[0, \infty)$. The Laplace transform of f is the function L on $(0, \infty)$ defined by

$$L(\lambda) = \int_0^\infty e^{-\lambda x} f(x)\, dx \tag{89}$$

Let $X_1, X_2, \ldots$ be i.i.d. random variables with the exponential distriution of rate λ, so $\Pr(X > x) = e^{-\lambda x}$, $\mathrm{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$. If $S_n = X_1 + X_2 + \cdots + X_n$ show that

$$\mathrm{E}\, f(S_n) = \frac{(-1)^{n-1}\lambda^n}{(n-1)!} L^{(n-1)}(\lambda) \tag{90}$$

Show that f may be recovered from L as follows: for $y > 0$

$$f(y) = \lim_{n \uparrow \infty} (-1)^{n-1} \frac{(n/y)^n L^{n-1}(n/y)}{(n-1)!} \tag{91}$$

We can write

$$\begin{aligned} \mathrm{E}\, f(S_n) &= \int_{\mathbb{R}_+^n} f(x_1 + x_2 + \cdots + x_n) e^{-\lambda x_1} e^{-\lambda x_2} \ldots e^{-\lambda x_n}\, dx_1 dx_2 \ldots dx_n \\ &= \int_0^\infty \int_{u_1}^\infty \cdots \int_{u_{n-1}}^\infty f(u_n) e^{-\lambda u_n}\, du_n \ldots du_1 \end{aligned} \tag{92}$$

where we performed a change of variables $u_1 = x_1, u_2 = x_1 + x_2, \ldots, u_n = x_1 + x_2 + \cdots + x_n$.

To simplify this, we're going to successively apply Fubini's theorem. To that end consider the calculation

$$\int_0^\infty u^k \int_u^\infty g(v)\, dv\, du = \int_0^\infty g(v) \int_0^v u^k\, du\, dv = \int_0^\infty \frac{v^{k+1}}{k+1} g(v)\, dv \tag{93}$$

So let $u = u_1, v = u_2, k = 0$ and $g(v) = \int_v^\infty \cdots \int_{u_{n-1}} f(u_n) e^{-\lambda u_n}$. We are left with

$$\int_0^\infty u_2 \int_{u_2}^\infty \cdots \int_{u_{n-1}}^\infty f(u_n) e^{-\lambda u_n}\, du_n \ldots du_1 \tag{94}$$

We may repeatedly interchange the order of integration, and then integrate out one variable until we are left with

$$\mathrm{E}\, f(S_n) = \int_0^\infty \frac{u_n^{n-1}}{(n-1)!} f(u_n) e^{-\lambda u_n}\, du_n \tag{95}$$

Now we need two fact from the theory of Laplace transforms

$$\begin{aligned}
\frac{d}{d\lambda} L[f] &= \int_0^\infty \frac{d}{d\lambda} f(x) e^{-\lambda x}\, dx = \int_0^\infty -x f(x) e^{-\lambda x}\, dx && (96)\\
&= -L[xf] && (97)\\
L[f'] &= \int_0^\infty f'(x) e^{-\lambda x}\, dx = f(x) e^{-\lambda x}\Big|_0^\infty + \lambda \int_0^\infty f(x) e^{-\lambda x}\, dx && (98)\\
&= \lambda f(0) + \lambda L[f] && (99)
\end{aligned}$$

Comparing this equation to (95) its clear that

$$\mathrm{E}\, f(S_n) = \frac{(-1)^{n-1}}{(n-1)!} L^{n-1}[f] \tag{100}$$

TODO where is the factor of λ^{n-1}???

Note $\mathrm{E}\, x^4 = \lambda^{-5}\Gamma(4) < \infty$. Therefore by the SLLN, $S_n/n \to \lambda$ almost surely. By continuity, $f(S_n/n) \to f(1/\lambda)$ almost surely. By bounded convergence theorem (since $f \leq B$ for some constant B and $\mathrm{E}\, B = B$), $\mathrm{E}\, f(S_n/n) \to f(1/\lambda)$. We connect this to our previous expression by first noting for $\alpha > 0$

$$L[f(\alpha x)](\lambda) = \int_0^\infty f(\alpha x) e^{-\lambda x}\, dx = \alpha^{-1} \int_0^\infty e^{-\lambda \alpha u} f(u)\, du = \alpha^{-1} L[f(x)](\alpha\lambda) \tag{101}$$

where the middle equality comes from the substitution $u = x/\alpha$. From this its clear that

$$\frac{d^{n-1}}{dx^n} L[f(\alpha x)](\lambda) = \alpha^{-n} \frac{d^{n-1}}{dx^n} L[f(x)](\alpha\lambda) \tag{102}$$

Taking $y = 1/\lambda$ and $\alpha = 1/n$ we get the desired result

$$f(y) = \mathrm{E}\, f(S_n/n) = \lim_{n\to\infty} (-1)^{n-1} \frac{(n/y)^n L^{(n-1)}(n/y)}{(n-1)!} \tag{103}$$

■

7.2 As usual, write $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$. There is a unique probability measure ν^{n-1} on $(S^{n-1}, \mathcal{B}(S^{n-1}))$ such that $\nu^{n-1}(A) = \nu^{n-1}(HA)$ for every orthogonal $n \times n$ matrix H and every A in $\mathcal{B}(S^{n-1})$. This is the same as the radial measure in polar coordinates, or the Haar measure under the action of orthogonal matricies.

(a) Prove that if $\boldsymbol{X}$ is a vector in $\mathbb{R}^n$, the components of which are independent $\mathcal{N}(0,1)$ variables, then for every orthogonal $n \times n$ matrix H the vector $H\boldsymbol{X}$ has the same property. Deduce that $\boldsymbol{X}/\|\boldsymbol{X}\|$ has law ν^{n-1}

(b) Let $Z_1, Z_2, \cdots \sim \mathcal{N}(0,1)$ and let

$$R_n = \|(Z_1, \ldots, Z_n)\| = (Z_1^2 + Z_2^2 + \cdots + Z_n^2)^{\frac{1}{2}} \tag{104}$$

Show $R_n/\sqrt{n} \to 1$ a.s.

(c) For each n, let $(Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)})$ be a point chosen on S^{n-1} according to the distribution ν^{n-1}. Then

$$\lim_{n\to\infty} \Pr(\sqrt{Y}_1^{(n)} \le x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \tag{105}$$

$$\lim_{n\to\infty} \Pr(\sqrt{Y}_1^{(n)} \le x_1; \sqrt{Y}_2^{(n)} \le x_2) = \Phi(x_1)\Phi(x_2) \tag{106}$$

(a) Let $\boldsymbol{Y} = H\boldsymbol{X}$. Each component is the linear combination of Gaussian random variables, so it too must be jointly Gaussian. Let's compute the moments of the components. Note $\mathrm{E}\,\boldsymbol{Y} = \mathrm{E}\,H\boldsymbol{X} = \mathrm{E}\,H\boldsymbol{0} = \boldsymbol{0}$. Therefore

$$\operatorname{Cov} \boldsymbol{Y} = \mathrm{E}\,\boldsymbol{Y}\boldsymbol{Y}^\intercal = \mathrm{E}(H\boldsymbol{X})(H\boldsymbol{X})^\intercal = \mathrm{E}\,H\boldsymbol{X}\boldsymbol{X}^\intercal H^\intercal = H(\mathrm{E}\,\boldsymbol{X}\boldsymbol{X}^\intercal)H^\intercal = HIH^\intercal = I \tag{107}$$

We've used the identity $HH^\intercal$ for orthogonal matricies. Thus each $Y_i \sim \mathbb{N}(0,1)$. Furthermore $\operatorname{Cov}(Y_i, Y_j) = 0$ if $i \neq j$. Since these are jointly Gaussian random variables, zero correlation is the same thing as independence. Now $\boldsymbol{X}/\|\boldsymbol{X}\| \in S^{n-1}$ and the law of $\boldsymbol{X}$, by the above calculation, is invariant under orthogonal transformations. That is, $H\boldsymbol{X}/\|H\boldsymbol{X}\|$ has the same distribution as $\boldsymbol{X}/\|\boldsymbol{X}\|$, which is true since $H\boldsymbol{X}$ has the same distribution as $\boldsymbol{X}$. We conclude that the law of $\boldsymbol{X}/\|\boldsymbol{X}\|$ is ν^{n-1} by the uniqueness of this measure.

(b) Note $\mathrm{E}\,Z_k^2 = 1$ since this is just the variance of Z_k. Also $\mathrm{E}\,Z_k^8 < \infty$ since a Gaussian has finite moments of all orders. By the SLLN applied to the random variables Z_k,

$$R_n^2/n = (\sum_{k=1}^{n} Z_k^2)/n \to 1 \text{ a.s.} \tag{108}$$

By continuity of square roots, it must also be the case that $R_n/\sqrt{n} \to 1$ almost surely.

(c) Note $\boldsymbol{Y}$ has the same law as $\boldsymbol{X}/\|\boldsymbol{X}\| = \boldsymbol{X}/R_n$. As $n \to \infty$, this tends to $\boldsymbol{X}/\sqrt{n}$ almost surely. Thus $\Pr(\sqrt{n}Y_1 < x) \to \Pr(X_1 < x) = \Phi(x)$. Similarly $\Pr(\sqrt{n}Y_1 < x_1; \sqrt{n}Y_2 < x_2) \to \Pr(X_1 < x_1; X_2 < x_2) = \Phi(x_1)\Phi(x_2)$.

■

Conditional Expectation

9.1 Prove that if $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$ and if $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \Pr)$ and if $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \Pr)$ and

$$\mathrm{E}(X; G) = E(Y; G) \tag{109}$$

for every G in a π-system which contains Ω and generates $\mathcal{G}$, then (109) holds for every $G \in \mathcal{G}$.

We make a standard monotone class argument. We will show that the collection of sets satisfying (109) is a d-system, so by Dynkin's lemma and the hypothesis the π-system $\mathcal{G}$ has property implies $\sigma(\mathcal{G})$ has the property.

(a) By assumption $\Omega \in \mathcal{G}$

(b) Suppose $A, B \in \mathcal{G}$. Then $L = A \cap B \in \mathcal{G}$ because $\mathcal{G}$ is a π-system. Now since $B = (B \setminus A) \coprod L$ is the union of these disjoint sets. Therefore $I_B = I_{B\setminus A} + I_L$. Thus by the linearity of expectations

$$\mathrm{E}(X; B \setminus A) = \mathrm{E}\, X I_{B\setminus A} = \mathrm{E}\, X(I_B - I_L) = \mathrm{E}(X; B) - \mathrm{E}(X; L) \tag{110}$$

Since the same is true for Y we get

$$\mathrm{E}(X; B \setminus A) = \mathrm{E}(X; B) - \mathrm{E}(X; L) = \mathrm{E}(Y; B) - \mathrm{E}(Y; L) = \mathrm{E}(Y; B \setminus A) \tag{111}$$

and the class of sets satisfying (109) is closed under taking differences.

(c) Suppose $A_n \uparrow A$. Then $X I_{A_n}$ is dominated by $|X I_A|$ and also $X I_{A_n} \to X I_A$. Thus by dominated convergence

$$\mathrm{E}(X; A_n) \to \mathrm{E}(X; A) \tag{112}$$

Mutatis mutandis, the same is true for Y hence

$$\mathrm{E}(X; A) = \lim_{n\to\infty} \mathrm{E}(X; A_n) = \lim_{n\to\infty} \mathrm{E}(Y; A_n) = \mathrm{E}(Y; A) \tag{113}$$

This shows that A is in the class of sets which satisfies (109)

Thus, the class of sets which satisfies (109) is a d-system as well as a π-system, so it includes $\sigma(\mathcal{G})$.

I guess another approach is to note that $G \to E[X; G]$ is a (signed) measure, so we can use something like theorem 1.6 on the uniqueness of measures defined on π-systems? ■

9.2 Suppose that $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \Pr)$ and that

$$\mathrm{E}(X|Y) = Y \text{ a.s.} \qquad \mathrm{E}(Y|X) = X \text{ a.s.} \tag{114}$$

Prove that $\Pr(X = Y) = 1$

Recall the defining property of of conditional expectation. First, $\mathrm{E}[Y \mid X]$ is $\sigma(X)$measurable, and second and for $G \in \sigma(X)$,

$$\mathrm{E}[\mathrm{E}[Y \mid X]; G] = \mathrm{E}[Y; G] \tag{115}$$

Let $A \in \sigma(X)$. From (114) and (115)

$$\mathrm{E}[X; A] = \mathrm{E}[E[Y \mid X]; A] = \mathrm{E}[Y; A] \tag{116}$$

Also, mutatis mutandis, $\mathrm{E}[Y; B] = \mathrm{E}[X; B]$ for any $B \in \sigma(Y)$. For $c \in \mathbb{R}$, taking $A = \{X \leq c\}$ and $B = \{Y \leq c\}$ gives

$$\begin{aligned}
&\mathrm{E}(X; X \leq c) = \mathrm{E}(Y; X \leq c) \Rightarrow \\
&\qquad \mathrm{E}(X - Y; X \leq c, Y > c) + \mathrm{E}(X - Y; X \leq c, Y \leq c) = 0 \\
&\mathrm{E}(X; Y \leq c) = \mathrm{E}(Y; Y \leq c) \Rightarrow \\
&\qquad \mathrm{E}(X - Y; X > c, Y \leq c) + \mathrm{E}(X - Y; X \leq c, Y \leq c) = 0
\end{aligned} \tag{117}$$

Therefore

$$\mathrm{E}(X - Y; X \leq c, Y > c) = \mathrm{E}(X - Y; X > c, Y \leq c) \tag{118}$$

since both equal $\mathrm{E}(Y - X; X \leq c, Y \leq c)$. Since the left hand side is non-positive, and the right hand side is non-negative, both sides must be zero.

I think this clinches it. Note $\{X > Y\} \subset \bigcup_{c \in \mathbb{Q}} \{X > c, Y \leq c\}$. However

$$\mathrm{E}(X - Y; X > Y) \leq \sum_{c \in \mathbb{Q}} \mathrm{E}(X - Y; X > c, Y \leq c) = 0 \tag{119}$$

However if $\Pr(X > Y) > 0$ then $\mathrm{E}(X - Y; X > Y) > 0$. Thus $X \leq Y$ almost surely. By symmetry, $Y \leq X$ almost surely, and the intersection of these almost sure events is almost sure. So $X = Y$ almost surely. ■

Martingales

10.1 Polya's urn At time 0 an urn contains 1 black ball and 1 white ball. At each time $t = 1, 2, \ldots$ a ball is chosen at random from the urn and is replaced, and another ball of the same color is also added. Just after time n there are therefore $n + 2$ balls in the urn, of which $B_n + 1$ is black, where B_n is the number of black balls chosen by time n.

Let $M_n = (B_n + 1)/(n + 2)$ be the proportion of blacks balls just after time n. Prove that M is a martingale.

Prove $\Pr(B_n = k) = \frac{1}{n+1}$ for $0 \leq k \leq n$. What's the distribution of $\Theta = \lim M_n$. Prove that for $0 < \theta < 1$,

$$N_n^\theta = \frac{(n+1)!}{B_n!(n - B_n)!} \theta^{B_n} (1 - \theta)^{n - B_n} \tag{120}$$

defines a martingale N^θ

We will consider M with respect to the filtration $\mathcal{F}_n = \sigma(B_1, \ldots, B_n)$. Thus, since $\sigma(B_{n-1}) \subset \mathcal{F}_{n-1}$

$$\begin{aligned}
\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] &= \Pr[B_n = B_{n+1}]\frac{B_{n-1}+1}{n+2} + \Pr[B_n = B_{n+1}+1]\frac{B_{n-1}+2}{n+2} \\
&= \frac{n-B_{n-1}}{n+1} \cdot \frac{B_{n-1}+1}{n+2} + \frac{B_{n-1}+1}{n+1} \cdot \frac{B_{n-1}+2}{n+2} \\
&= \frac{(n+2)(B_{n-1}+1)}{(n+2)(n+1)} = M_{n-1}
\end{aligned} \tag{121}$$

which verifies M_n is a martingale.

Now let t_i specify on which draw the ith black ball selected, and let s_j specify which draw the jth white ball is selected are white. That is, if $n = 7$ and $k = 4$ and the balls drawn are *bwwbbwb* then $t_1 = 1, t_2 = 4, t_3 = 5, t_4 = 7$ and t_i is given by $s_1 = 2, s_2 = 3, s_3 = 6$. The length of each sequence is the total number of balls chosen of the respective color. Since every draw is either black or white, thus it must appear either in the black sequence or the white sequence. Therefore $\bigcup_i t_i \cup \bigcup_i s_i = \{1, 2, \ldots, n\}$. Let's compute the probability of a specific sequence

$$P = \prod_{i=1}^{k} \frac{i}{t_i+1} \cdot \prod_{j=1}^{n-k} \frac{j}{s_j+1} = \frac{k!(n-k)!}{n+1} = \frac{1}{(n+1)}\binom{n}{k}^{-1} \tag{122}$$

We've learned a remarkable fact: the probability doesn't depend on the order of the draws, just the total number which are black. The sequence is exchangable, any permutation of the outcomes results in an event with the same probability. There are $\binom{n}{k}$ sequences with k black balls and $n-k$ white ones.

$$\Pr(B_n = k) = \binom{n}{k}\frac{1}{(n+1)}\binom{n}{k}^{-1} = \frac{1}{n+1} \tag{123}$$

Thus we've learned another remarkable fact, the distribution of B_n is uniform on $1, 2, \ldots, n+1$. Thus $\Theta = \lim_{n\to\infty} M_n$ must converge in distribution to the uniform distribution $U(0,1)$.

Turning to N_n^θ, let B be the event that there's a black ball on the nth draw. That is that $B_n = B_{n-1}+1$. Let $W = B^c$, the probability there is a white ball.

$$\begin{aligned}
E[N_n^\theta \mid \mathcal{F}_{n-1}] =& \Pr(W)\frac{(n+1)!}{B_{n-1}!(n-B_{n-1})!}\theta^{B_{n-1}}(1-\theta)^{n-B_{n-1}} \\
&+ \Pr(B)\frac{(n+1)!}{(B_{n-1}+1)!(n-B_{n-1}-1)!}\theta^{B_{n-1}+1}(1-\theta)^{n-B_n-1} \\
=& \frac{n!}{B_{n-1}!(n-1-B_{n-1})!}\theta^{B_n-1}(1-\theta)^{n-1-B_n} \times \\
& \left[\frac{(1-\theta)(n+1)}{n-B_{n-1}}P(W) + \frac{\theta(n+1)}{B_{n-1}+1}P(B)\right]
\end{aligned} \tag{124}$$

The factor before the parantheses is N_{n-1}^θ. The expression in the parentheses is

$$(1-\theta)\frac{n+1}{n-B_{n+1}}\left(\frac{n-B_{n-1}}{n+1}\right) + \theta\frac{n+1}{B_{n-1}+1}\left(\frac{B_{n-1}+1}{n+1}\right) = 1-\theta+\theta = 1 \tag{125}$$

■

So the martingale property is proved

10.2 Belman's optimality principle Your winnings per unit stake on game n are ϵ_n where ϵ_n are i.i.d. RV with distribution

$$\epsilon_n = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } q \end{cases} \tag{126}$$

Your stake C_n on game n must lie between 0 and Z_{n-1}, where Z_{n-1} is your fortune at time $n-1$. Your object is to maximize the expected interest rate $\mathrm{E}\log(Z_N/Z_0)$ where N is a given integer representing the number of times you will play and Z_0, your initial wealth, is a given constant. Let $\mathcal{F}_n = \sigma(\epsilon_1, \ldots, \epsilon_n)$ be your history up to time n. Show that if C is any previsible strategy, then $\log Z_n - n\alpha$ is a supermartingale where α denotes the *entropy*

$$\alpha = p \log p + q \log q + \log 2 \tag{127}$$

so that $\mathrm{E}\log(Z_N/Z_0) \leq N\alpha$, but that, for a certain strategy, $\log Z_n - n\alpha$ is a *martingale*. What is the best strategy?

For a discrete distribution on $\{1, 2, \ldots, n\}$, lets maximize $E \log X$ for a random variable subject to the constraint $\sum_{i=1}^{n} X(i) = K$. Using the method of Lagrange multipliers we maximize the expression

$$F = \sum_{i=1}^{n} p_i \log x_i - \lambda(\sum_{i=1}^{n} x_i - K) \tag{128}$$

over the $n+1$ variables $x_1, x_2, \ldots, x_n, \lambda$. (Maximizing with respect to λ) is just the same as respecting the constraint. The first order conditions are

$$\frac{\partial F}{\partial x_i} = 0 \qquad \Rightarrow \qquad \frac{p_i}{x_i} = \lambda \tag{129}$$

Summing over the equations $1 = \sum_i p_i = \lambda \sum_i x_i = \lambda K$. Thus $x_i = Kp_i$ is the extremal solution. This is a maximum, which can be seen by considering the distribution where $p_1 = 1$, or the second order conditions $\frac{\partial^2 F}{\partial x_i^2} = -\frac{p_i}{x_i^2} = -\frac{1}{p_i}$. The maximum value is given by

$$\sum_i p_i \log x_i = \sum_i p_i(\log p_i + \log K) = -H(p) + \log K \tag{130}$$

where $H(p) = -\sum_i p_i \log p_i$ is the entropy of the distribution

Calculating

$$\log Z_n - \log Z_{n-1} = \log\left(\frac{Z_{n-1} + C_n \epsilon_n}{Z_n}\right) \tag{131}$$

Let $f_n = C_n/Z_{n-1}$. For any choice $0 \leq C_n \leq Z_{n-1}$, let $f_n = C_n/Z_{n-1}$ we get $0 \leq f \leq 1$.

$$\mathrm{E}[\log Z_n - \log Z_{n-1}] = p \log(1 + f_n) + q \log(1 - f_n) \leq p \log p + q \log q + \log 2 \tag{132}$$

where the last inequality follow from our previous calculation. Therefore $Z_n - n\alpha$ is a supermartingale for any previsible C_n and the optimal C_n is given by $C_n = (2p-1)Z_{n-1} = (p-q)Z_{n-1}$. The supermartingale property implies that $E[\log Z_N - N\alpha] \leq E[\log Z_0]$ and thus $E[\log(Z_N/Z_0)] \leq N\alpha$ ■

10.3 Stopping times. Suppose S and T are stopping times relative to $(\Omega, \mathcal{F}, \{\mathcal{F}_n\})$. Prove that $S \wedge T = \min(S,T)$ and $S \vee T = \max(S,T)$ and $S+T$ are all stopping times.

First note that for any constant n, $\{T = n\} \in \mathcal{F}_n$. This is $\{T \leq n\} \setminus \{T \leq n-1\}$ and the former is in $\mathcal{F}_n$ and the latter is in $\mathcal{F}_{n-1} \subset \mathcal{F}_n$. Therefore $\{T \geq n\} \in \mathcal{F}_n$ since $\{T \geq n\} = \{T = n\} \cup \{T > n\}$. The latter is just $\{T \leq n\}^c \in \mathcal{F}_n$.

Turning to $S \wedge T$, suppose $S \wedge T = n$. Without loss of generality, suppose $S = n$ and $T \geq n$. We've already shown that each of these sets $\{S = n\}$ and $\{T \geq n\}$ are in $\mathcal{F}_n$, therefore their intersection is as well. The function $S \vee T$ is even simpler, since if $S \vee T = n$, without loss of generality assume $S = n$ and $T \leq n$. These events are both in $\mathcal{F}_n$, so their intersection is as well.

For $S + T$, note that each of S and T are natural numbers. Therefore if $S + T = n$ for some $n \in \mathbb{N}$, there are a finite number of possibilities that $S = s$ and $T = t$ and $s + t = n$. (Namely $s \in \{0, 1, \ldots, n\}$ and $t = n - s$). So $\{T + S = n\} = \bigcup_{k=0}^{n} \{S = k\} \cap \{T = n - k\}$, and each of the sets in this expression are clearly in $\mathcal{F}_n$ ■

10.4 Let S and T be stopping times with $S \leq T$. Define the process $1_{(S,T]}$ with parameter set $\mathbb{N}$ via

$$1_{(S,T]}(n,\omega) = \begin{cases} 1 & \text{if } S(\omega) < n \leq T(\omega) \\ 0 & \text{otherwise} \end{cases} \tag{133}$$

Prove $1_{(S,T]}$ is previsible and deduce that if X is a supermartingale then

$$\mathrm{E}(X_{T\wedge n}) \leq \mathrm{E}(X_{S\wedge n}), \qquad \forall n \tag{134}$$

Note

$$\{1_{(S,T]}(n,\omega) = 1\} = \{S(\omega) \leq n-1\} \cap \{T(\omega) \geq n\} \tag{135}$$

The first event $\{S \leq n-1\} \in \mathcal{F}_{n-1}$ because S is a stopping time. Similarly the second event $\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$ because T is a stopping time. Since $1_{(S,T]}$ only takes the values $\{0,1\}$ this proves that $\omega \hookrightarrow 1_{(S,T]}(n,\omega)$ is $\mathcal{F}_{n-1}$ measurable. Hence the process $1_{(S,T)}$ is previsible.

Thus for any supermartingale X_n, the process $Y_n = 1_{(S,T]} \bullet X_n$ is also a supermartingale with value at time n given by

$$1_{(S,T]} \bullet X_n = X_{T\wedge n} - X_{S\wedge n} \tag{136}$$

Without loss of generality we can assume $Y_0 = 0$, reindexing time if necessary so that $S, T \geq 1$. Taking expectations of both sides gives the desired result

$$\mathrm{E}\, Y_n = \mathrm{E}\, X_{T\wedge n} - \mathrm{E}\, X_{S\wedge n} \leq Y_0 = 0 \tag{137}$$

■

10.5 Suppose that T is a stopping time such that for some $N \in \mathbb{N}$ and some $\epsilon > 0$ we have, for every n

$$\Pr(T \le n + N \mid \mathcal{F}_n) > \epsilon, \qquad \text{a.s.} \tag{138}$$

By induction using $\Pr(T > kN) = \Pr(T > kN; T > (k-1)N)$ that for $k = 1, 2, 3, \ldots$

$$\Pr(T > kN) \le (1-\epsilon)^k \tag{139}$$

Show that $\mathrm{E}(T) < \infty$

Let $G_n = \{T > n\}$. Taking complements, our hypothesis implies

$$\Pr(G_{n+N} \mid \mathcal{F}_n) < 1 - \epsilon, \qquad \text{a.s.} \tag{140}$$

Taking $n = 0$ this implies $\Pr(G_N) = \Pr(G_N \mid \mathcal{F}_n) < 1 - \epsilon$. By induction, using the relation above for $n = N(k-1)$ we get

$$\Pr(G_{Nk}) = \Pr(G_{N(k-1)+N} \mid G_{N(k-1)}) \Pr(G_{N(k-1)}) \le (1-\epsilon)(1-\epsilon)^{n-1} \tag{141}$$

so the induction holds.

Let's derive an alternate expression for expectation

$$\mathrm{E}\, T = \sum_{n=1}^{\infty} n \Pr(T = n) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \Pr(T = n) = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \Pr(T = n) = \sum_{k=1}^{\infty} \Pr(T \ge k) \tag{142}$$

Also if $m < n$ then $G_m \supset G_n$ we have the trivial inequality $\Pr(G_m) \ge \Pr(G_n)$. So using this inequality for $n = Nk + 1$ to $n = N(k+1) - 1$ compared with $m = Nk$ we get

$$\mathrm{E}\, T = \sum_{n=1}^{\infty} \Pr(T \ge n) \le \sum_{k=0}^{\infty} N \Pr(T \ge Nk) \le N \sum_{k=0}^{\infty} (1-\epsilon)^k = \frac{N}{\epsilon} \tag{143}$$

Clearly this is finite. Note that the inequalities used above are true almost surely. However, we used only a countable number of inequalities, and the countable intersection of almost sure events is still almost sure. Thus the inequality above also holds almost surely. The result may be summarized, "so long as there's a chance something happens, no matter how small (so long as the chance is bounded below by a positive number), then it will happen eventually" ∎

10.6 ABRACADABRA. At each of times $1, 2, 3, \dots$ a monkey types a capital letter at random, so the sequence typed is an i.i.d. sequence of RV's each chosed uniformly among the 26 possible capital letters. Just before each time $t = 1, 2, \dots$ a new gambler arrives. He bets \$1 that

$$\text{the } n\text{th letter will be } A \tag{144}$$

If he loses he leaves. If he wins he bets his fortune of \$26 that

$$\text{the } (n+1)\text{st letter will be } B \tag{145}$$

If he loses he leaves. If he wins he bets his fortune of \$$26^2$ that

$$\text{the } (n+2)\text{st letter will be } R \tag{146}$$

and so on through ABRACADABRA. Let T be the first time the monkey has produced the consecutive sequence. Show why

$$\mathrm{E}(T) = 26^{11} + 26^4 + 26 \tag{147}$$

Intuitively, the cumulative winnings of all the gamblers are a martingale, since its the sum of a bunch of martingales. The cumulative winnings are $-T$ for the initial stake placed at each time $1, \dots, T$ plus $26^{11} + 26^4 + 26$ payout for the winnings on the bets placed on the first A (winning for the whole word ABRACADABRA), the bet placed on the second to last A (winning for the terminal fragment ABRA) and the bet placed on the final A (winning for A). For this to be a martingale, expectated cumulative winnings must be 0 so $\mathrm{E}\,T = 26^{11} + 26^4 + 26$

More formally, let L_n be the nth randomly selected letter. The L_n are i.i.d., with uniform distribution over $\{A, B, \dots, Z\}$. For a sequence of letters $a_1 a_2 \dots a_n$ consider the martingale

$$M_n^{(j)} = \begin{cases} 1 & \text{if } n \le j \\ 26^k & \text{if } n = j + k \text{ and } L_j = a_1, L_{j+1} = a_2, \dots, L_{j+k-1} = a_k \\ 26^{11} & \text{if } n > j + 11 \text{ and } L_j = a_1, L_{j+1} = a_2, \dots, L_{j+10} = a_{11} \\ 0 & \text{otherwise} \end{cases} \tag{148}$$

This is bounded, and hence is in L^1. $M_n^{(j)}$ satisfies the martingale prorperty since, except for case 2, its constant and in that case

$$\mathrm{E}[M_{n+1}^{(j)} \mid \mathcal{F}_n] = \frac{1}{26} \cdot 26 M_n^{(j)} + \frac{25}{26} \cdot 0 = M_n^{(j)} \tag{149}$$

Now form the martingale $N^{(j)} = I_{n \ge j} \bullet M^{(j)}$ where $I_{n \ge j}$ is the indicator for $n \ge j$. Now form the martingale

$$B = \sum_{j \ge 1} N^{(j)} \tag{150}$$

At any time n only finitely many of the terms are non-zero. Each term is in L^1 , so B is also in L^1. The martingale property holds since B is the sum of martingales. For $n \in \mathbb{N}$

let S be the set of $j \le n$ such that $M_n^{(j)}$ is in case 2, matching some number of letters. Let S' be the set of $j \le n$ such that $M_n^{(j)}$ is in case 3, matching all the letters. The value of B is given by

$$B_n = -n + \sum_{j \in S} 26^{n-j} + \sum_{j \in S'} 26^{11} \tag{151}$$

Let T be the stopping time for the first time some $M^{(j)}$ enters case 3, that is $M^{(j)}$ matches all 11 letters in ABRACADABRA. It must be that $\mathrm{E}|T| < \infty$ because it statisfies the property in 10.5 with $N = 11$ and $\epsilon = 26^{-11}$– i.e., it stops if the next 26 letters are ABRACADABRA in order. Whenever the stopping time happens, B_n has the same value for the second and third terms. In this case $S = \{n-4, n-1\}$ since the terminal ABRA and A matched those $M^{(n-4)}$ and $M^{(n-1)}$ respectively, and S' has exactly one element, since this is the first time ABRACADABRA appeared. Hence by Doob's optional stopping theorem

$$0 = B_0 = \mathrm{E}\, B_T = -\mathrm{E}\, T + 26^4 + 26 + 26^{11} \tag{152}$$

This is the same as (147) ∎

> **10.7** Gambler's Ruin. Suppose that $X_1, X_2, \dots$ are i.i.d. RV's with
>
> $$\Pr[X = +1] = p, \qquad \Pr[X = -1] = q, \qquad \text{where } 0 < p = 1 - q < 1 \tag{153}$$
>
> and $p \neq q$. Suppose that a and b are integers with $0 < a < b$. Define
>
> $$S_n = a + X_1 + \cdots + X_n, \qquad T = \inf\{n : S_n = 0 \text{ or } S_n = b\} \tag{154}$$
>
> Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Explain why T satisfies the condition in 10.5. Prove that
>
> $$M_n = \left(\frac{q}{p}\right)^{S_n} \qquad \text{and} \qquad N_n = S_n - n(p-q) \tag{155}$$
>
> define martingales M and N. Deduce the values of $\Pr(S_T = 0)$ and $\mathrm{E}\, S_T$

Let's compute

$$\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] = p\left(\frac{q}{p}\right)^{S_{n-1}+1} + q\left(\frac{q}{p}\right)^{S_{n-1}-1} = (q+p)\left(\frac{q}{p}\right)^{S_{n-1}} = M_{n-1} \tag{156}$$

$$\mathrm{E}[N_n \mid \mathcal{F}_{n-1}] = p(S_{n-1} + 1 - n(p-q)) + q(S_{n-1} - 1 - n(p-q)) \tag{157}$$

$$= (p+q)S_{n-1} + p - q - n(p-q) = N_{n-1} \tag{158}$$

Now the condition described in 10.5 holds with $N = b - 1$. For any n suppose that $X_{n+1} = X_{n+2} = \cdots = X_{n+b} = +1$. Conditional on $S_n \in (0, b)$, then some $S_{n+b} = S_n + b > b$, so $S_{n+k} = b$ for some $k \in \{1, 2, \dots, b-1\}$. In particular, $T \le n + b - 1$. However $\Pr(X_{n+1} = X_{n+2} = \cdots = X_{n+b} = +1) = p^b > 0$. So

$$\Pr(T > N + b - 1 \mid \mathcal{F}_n) > p^b \tag{159}$$

and the condition is satisfied. We conclude that $\mathrm{E}\, T < \infty$.

Let $\pi = \Pr(S_T = 0) = 1 - \Pr(S_T = b)$. By Doob's optional-stopping theorem

$$\left(\frac{q}{p}\right)^a = M_0 = \mathrm{E}\, M_T = \pi \cdot 1 + (1-\pi) \cdot \left(\frac{q}{p}\right)^b \tag{160}$$

or

$$\Pr(S_T = 0) = \pi = \frac{(q/p)^a - (q/p)^b}{1 - (q/p)^b} \qquad \Pr(S_T = b) = 1 - \pi = \frac{1 - (q/p)^a}{1 - (q/p)^b} \tag{161}$$

Also

$$a = N_0 = \mathrm{E}\, N_T = \pi \cdot 0 + (1-\pi) \cdot b - T(p-q) \tag{162}$$

or

$$T = \frac{(1-\pi)b - a}{p - q} \tag{163}$$

■

10.8 A random number Θ is chosen uniformly in $[0,1]$ and a coin with probability Θ of heads is minted. The coin is tossed repeatedly. Let B_n be the number of heads in n tosses. Prove that B_n has exactly the same probabilistic structure as the (B_n) sequence in 10.1 on Polya's urn. Prove that N_n^θ is the regular conditional pdf of Θ given $B_1, \ldots, B_n$

Consider the beta function $B(m+1, n+1) = \int_0^1 u^m (1-u)^n \, du$. Note $B(m,n) = B(n,m)$ by a change of variables $v = 1 - u$. Thus, without loss of generality, assume $m \geq n$

$$\begin{aligned}
B(m+1, n+1) &= \int_0^1 u^m (1-u)^n \, du \\
&= \frac{1}{m+1} u^{m+1} u^n \Big|_0^1 + \frac{n}{m+1} \int_0^1 u^{m+1} (1-u)^{n-1} \, du \\
&= \frac{n}{m+1} B(m+2, n) = \left(\prod_{k=1}^{n} \frac{n+1-k}{m+k} \right) B(m+n+1, 1) \\
&= \frac{n! m!}{(m+n)!} \frac{1}{m+n+1} = \frac{n! m!}{(n+m+1)!}
\end{aligned} \tag{164}$$

Now conditional on $\Theta = \theta$, $B_n \sim \mathrm{Bin}(n, \theta)$. Thus we can marginalize over θ to find

$$\begin{aligned}
\Pr(B_n = k) &= \int_0^1 \Pr(B_n = k \mid \Theta = \theta) \, d\theta = \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} \, d\theta \\
&= \binom{n}{k} B(k+1, n-k+1) = \frac{1}{n+1}
\end{aligned} \tag{165}$$

Compare this with (123). Now

$$\begin{aligned} \Pr(B_n = k; B_{n-1} = k-1) &= \int_0^1 \Pr(B_n = k; B_{n-1} = k-1 \mid \Theta = \theta)\, d\theta \\ &= \int_0^1 \theta \cdot \binom{n-1}{k-1} \theta^{k-1} (1-\theta)^{n-k}\, d\theta \\ &= \binom{n-1}{k-1} B(k+1, n-k+1) = \frac{k}{n(n+1)} \end{aligned} \tag{166}$$

We've used the fact that $\Pr(B_n = k \mid B_{n-1} = k-1, \Theta = \theta) = \theta$, since conditional on Θ, the trials are independent with probability θ. Therefore using Baye's rule

$$\Pr(B_n = k \mid B_{n-1} = k-1) = \frac{\Pr(B_n = k; B_{n-1} = k-1)}{\Pr(B_{n-1} = k-1)} = \frac{k}{n} \tag{167}$$

This is eactly the rule for for Polya's urn.

Let's compute the regular conditinal pdf for Θ. Let $B^{(n)} = (B_1, B_2, \ldots, B_n)$ be the vector of values of B_k for $k = 1, \ldots, n$ and let $b = (b_1, \ldots, b_n)$ be a realization. Now

$$\Pr(B^{(n)} = b \mid \Theta = \theta) = \prod_{k=1}^{n} \theta^{I(b_k = b_{k-1}+1)} (1-\theta)^{I(b_k = b_{k-1})} \tag{168}$$

That is, we pick up a factor of θ whenever b_k increases (we flipped a head) or $1-\theta$ whenever b_k stays the same (we flipped a tails). Clearly this can be simplified to

$$\Pr(B^{(n)} = b \mid \Theta = \theta) = \theta^{B_n} (1-\theta)^{n-B_n} \tag{169}$$

Marginalizing over Θ, we compute

$$\Pr(B^{(n)} = b) = \int_0^1 \theta^{B_n} (1-\theta)^{n-B_n}\, d\theta = \frac{B_n!(n-B_n)!}{(n+1)!} \tag{170}$$

Now by Bayes rule

$$\begin{aligned} \Pr(\Theta = \theta \mid B^{(n)} = b) &= \frac{\Pr(B^{(n)} = b \mid \Theta = \theta) \Pr(\Theta = \theta)}{\Pr(B^{(n)} = b)} \\ &= \frac{(n+1)!}{B_n!(n-B_n)!} \theta^{B_n} (1-\theta)^{n-B_n}\, d\theta \\ &= N_n^{\theta}\, d\theta \end{aligned} \tag{171}$$

This proves the last assertion in the problem.

The jailhouse problem.

$$
\begin{aligned}
\Gamma(x)\Gamma(y) &= \int_0^\infty s^{x-1}e^{-x}\,ds \int_0^\infty t^{y-1}e^{-y}\,dt \\
&= \int_0^\infty \int_0^\infty s^{x-1}t^{y-1}e^{-x-y}\,ds\,dt \\
&= \int_0^\infty \int_0^1 (uv)^{x-1}((1-u)v)^{y-1}e^{-x-y}\,v\,du\,dv \\
&= \int_0^1 u^{x-1}(1-u)^{y-1}\,du \int_0^\infty v^{x+y-1}e^{-x-y}\,dv \\
&= B(x,y)\Gamma(x+y)
\end{aligned} \tag{172}
$$

where we made the substitution $u = \frac{s}{s+t}, v = s+t$ with Jacobian

$$
\frac{\partial(u,v)}{\partial(s,t)} = \begin{vmatrix} t/(s+t) & -s/(s+t) \\ 1 & 1 \end{vmatrix} = \frac{1}{s+t} = \frac{1}{v} \tag{173}
$$

When $x, y \in \mathbb{N}$ we recover our previous formula (164) since $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$ ∎

10.9 Show that if X is a non-negative supermartingale and T is a stopping time then

$$
\mathrm{E}(X_T; T < \infty) \le \mathrm{E}\,X_0 \tag{174}
$$

Deduce that

$$
\Pr(\sup X_n \ge c) \le \frac{\mathrm{E}\,X_0}{c} \tag{175}
$$

Let $Z_n = X_n I_{T<\infty}$. Then as $n \to \infty$, $Z_{n\wedge T} \to X_T I_{T<\infty}$ since if $T(\omega) = N$ then $Z_n = X_T$ for all $n > N$ so $Z_n \to X_T$ and if $T(\omega) = \infty$ then $Z_n = 0$ for all n. Since $X_n \ge 0$, by Fatou's lemma

$$
\mathrm{E}[X_T; T < \infty] = \mathrm{E}\lim Z_{T\wedge n} = \mathrm{E}\liminf Z_{T\wedge n} \le \liminf \mathrm{E}\,Z_{T\wedge n} \tag{176}
$$

so, because $X_n \ge 0$ and because of the supermartingale property

$$
\begin{aligned}
\mathrm{E}[X_T; T < \infty] - \mathrm{E}[X_0] &\le \liminf \mathrm{E}\,Z_{T\wedge n} - \mathrm{E}\,X_0 \\
&\le \liminf_{n\to\infty} \mathrm{E}\,X_{T\wedge n} - \mathrm{E}\,X_0 \\
&\le 0
\end{aligned} \tag{177}
$$

This gives us (174).

Now consider the stopping time $T = \inf\{X_n \ge c\}$. From the definition of T, if $n = T$ then $X_n \ge c$.

$$
c\Pr(\max X_n \ge c) = c\Pr(T < \infty) \le \mathrm{E}[X_T I_{\{T<\infty\}}] \le \mathrm{E}\,X_0 \tag{178}
$$

This gives us (175). Note this is a maximal version of Markov's inequality. ∎

10.10 The "Starship Enterprise" problem. The control system on the star-ship Enterprise has gone wonky. All that one can do is to set a distance to be travelled. The spaceship will then move that distance in a randomly chosen direction, then stop. The object is to get into the Solar System, a ball of radius ρ. Initially, the Enterprise is at a distance $R_0 > \rho$ from the Sun.
Show that for all ways to set the jump lengths r_n

$$\Pr(\text{ the Enterprise returns to the solar system}) \leq \frac{\rho}{R_0} \tag{179}$$

and find a strategy of jump lengths r_n so that

$$\Pr(\text{the Enterprise returns to the solar system}) \geq \frac{\rho}{R_0} - \epsilon \tag{180}$$

Let R_n be the distance after n space-hops. Now $f(\boldsymbol{x}) = \frac{1}{|\boldsymbol{x}|}$ is a harmonic function satisfying $\nabla^2 f = \sum_i \frac{\partial^2 f}{\partial x_i^2} = 0$ when $x \neq 0$. Let $r = |x|$, let $S(\boldsymbol{x}, r)$ be the 2-sphere of radius r centered at $\boldsymbol{x}$. Then let $S(r) = S(0, r)$, and let $d\sigma$ be the surface area element. Similarly let $B(\boldsymbol{x}, r)$ be the ball of radius r centered at x.

$$\frac{\partial f}{\partial x_i} = -\frac{x_i}{r^3} \qquad \frac{\partial^2 f}{\partial x_i^2} = -\frac{1}{r^3} + \frac{3x_i^2}{r^5} \qquad \nabla^2 f = -\frac{3}{r^3} + \frac{3r^2}{r^5} = 0 \tag{181}$$

Any harmonic function satisfies the mean-value property. Let

$$A(\boldsymbol{x}, r) = \frac{1}{4\pi r^2} \iint_{S(r)} f \, d\sigma = \frac{1}{4\pi} \iint_{S(1)} f(\boldsymbol{x} + r\boldsymbol{n}) \, d\sigma_1 \tag{182}$$

In the above we applied a dilation to integrate over $S(1)$ with surface element $d\sigma_1 = r^2 d\omega$. Then differentiating under the integral sign with respect to r

$$\begin{aligned} \frac{d}{dr} A(\boldsymbol{x}, r) &= \frac{1}{4\pi} \iint_{S(1)} \nabla f(\boldsymbol{x} + r\boldsymbol{n}) \cdot n \, d\sigma_1 = \frac{1}{4\pi r^2} \iint_{S(r)} \nabla f \cdot n \, d\sigma \\ &= \frac{1}{4\pi r^2} \iiint_{B(r)} \nabla^2 f \, dV = 0 \end{aligned} \tag{183}$$

In the last equation we used Gauss' law. By continuity $A(x, r) \to f(x)$ as $r \to 0$ so $A(x, r) = f(x)$ for all r.

The probability law for the starship enterprise for a jump of size r is given by $\frac{d\omega}{4\pi r^2}$ on the ball $S(\boldsymbol{x}, r)$. so $\mathrm{E}\, 1/R_{n+1} = A(\boldsymbol{x}_n, r)$ for $f(\boldsymbol{x}) = |x|^{-1}$. Hence, if $r < R_n$, by the above calculation

$$\mathrm{E}\, \frac{1}{R_{n+1}} = \frac{1}{R_n} \tag{184}$$

since $|x|^{-1}$ is harmonic inside of $B(\boldsymbol{x}, r)$, and therefore $1/R_n$ is a martingale.

Before considering the case $r > R_n$ we need to perform a calculation. For any $\epsilon > 0$ and $f(\boldsymbol{x}) = |\boldsymbol{x}|^{-1}$

$$\iint_{S(\epsilon)} \nabla f \cdot \boldsymbol{n}\, d\sigma = \iint_{S(\epsilon)} -\frac{\boldsymbol{x}\cdot\boldsymbol{x}}{r^4}\, d\sigma = -\frac{4\pi\epsilon^2}{\epsilon^2} = -4\pi \tag{185}$$

since $\nabla|\boldsymbol{x}|^{-1} = -\boldsymbol{x}/r^3$ and $\boldsymbol{n} = \boldsymbol{x}/r$. Remarkably, the integral does not depend on ϵ. Thus for $r > R_n$, we can modify (183) to integrate over $B(\boldsymbol{x}, r) \setminus B(0, \epsilon)$, and then take $\epsilon \to 0$. This gives the formula

$$\frac{d}{dr}A(\boldsymbol{x}, r) = \begin{cases} 0 & \text{if } r < |x| \\ -\frac{1}{r^2} & \text{if } r \geq |x| \end{cases} \tag{186}$$

Thus, this more complete analysis shows that $1/R_n$ is a supermartingale everywhere, and

$$\mathrm{E}\,\frac{1}{R_{n+1}} = \frac{1}{\max(r, R_n)} \leq \frac{1}{R_n} \tag{187}$$

By the positive supermartingale maximal inequality (175)

$$\Pr(\min_n R_n \leq \rho) = \Pr\left(\max_n \frac{1}{R_n} \geq \frac{1}{\rho}\right) \leq \frac{\rho}{R_0} \tag{188}$$

The probability on the left is the probability that the enterprise ever gets into the solar system, and it applies to any strategy r_n of jump-lengths.

To get a lower bound, consider the process with constant jump sizes $r_n = \epsilon$ for any $0 < \epsilon < \rho$, and also choose $\rho' > R_0$. Let $\tau = \inf\{R_n \leq \rho\}$ and $\tau' = \inf\{R_n \geq \rho'\}$ and let $t = \tau \wedge \tau'$. First we show $\Pr(t < \infty) = 1$, that is we hit one barrier or the other eventually. Note the x coordinate of the Enterprise's location follows a random walk since the x projection of each jump of length ϵ is i.i.d.. Let X_n be the x-coordinate at time n. By the symmetry of each spherical jump, X_n is a martingale, so $\mathrm{E}\,X_n = X_0$. Also if $J_n = X_n - X_{n-1}$ note $\mathrm{Var}(J_n) = \epsilon^2\sigma_0^2$ for some σ_0 since, by dilation symmetry of the spherical jump, the variance scales as ϵ^2.

$$\Pr(R_n \geq \rho') \geq \Pr(|X_n| \geq \rho') \geq \Pr(|X_n - X_0| > 2\rho') \tag{189}$$

However by the central limit theorem

$$\Pr(|X_n - X_0| > 2\rho') \to 2\Phi\left(-\frac{2\rho'}{\epsilon\sigma_0\sqrt{n}}\right) \to 1 \qquad \text{as } n \to \infty \tag{190}$$

We conclude

$$\Pr(t = \infty) \leq \Pr(\tau' = \infty) = \Pr(\max_n R_n \leq \rho') \leq \inf_n \Pr(R_n \leq \rho') = 0 \tag{191}$$

Now let's consider the stopped martingale $Z_n = 1/R_{t\wedge n}$. This is a martingale since our chose of $\epsilon < \rho$ implies $r_n = \epsilon < R_n$ for all n. We can apply Doob's optional stopping theorem since

$$\frac{1}{\rho' + \epsilon} \leq Z_n \leq \frac{1}{\rho - \epsilon} \tag{192}$$

and therefore Z_n is a bounded, positive martingale. We can break the expectation up into two parts, when R_n hits the upper and lower barriers ρ' and ρ.

$$\frac{1}{R_0} = \mathrm{E}[Z_t] = s\,\mathrm{E}[Z_t; t = \tau] + \mathrm{E}[Z_t; t = \tau'] \leq \frac{\Pr(t = \tau)}{\rho - \epsilon} + \frac{\Pr(t = \tau')}{\rho'} \tag{193}$$

where we've estimated the expectations by the simple upper bounds implied by

$$\begin{aligned} \rho - \epsilon \leq & R_t \leq \rho & \text{if } t = \tau \\ \rho' \leq & R_t \leq \rho' + \epsilon & \text{if } t = \tau' \end{aligned} \tag{194}$$

This gives the lower bound

$$\Pr(t = \tau) \geq \frac{\frac{1}{R_0} - \frac{1}{\rho'}}{\frac{1}{\rho - \epsilon} - \frac{1}{\rho'}} \tag{195}$$

Now as $\rho' \to \infty$, the event $\{t = \tau\}$ is essentially the event that the Enterprise returns to the solar system. It says that the distance $R_n \leq \rho$ occurs before $R_n \to \infty$. So we can interpret this equation in the limit that $\rho' \to \infty$ to say

$$\Pr(\text{Enterprise returns}) \geq \frac{\rho - \epsilon}{R_0} \tag{196}$$

■

10.11 Star Trek 2, Captain's Log
Mr Spock and Chief Engineer Scott have modified the control system so that the Enterprise is confined to move for ever in a fixed plane passing through the Sun. However, the next 'hop-length' is now automatically set to be the current distance to the Sun ('next' and 'current' being updated in the obvious way). Spock is muttering something about logarithms and random walks, but I wonder whether it is (almost) certain that we will get into the Solar System sometime. " '

Let $f(\boldsymbol{x}) = \log|\boldsymbol{x}| = \frac{1}{2}\log(x_1^2 + x_2^2)$.

$$\frac{\partial f}{\partial x_i} = -\frac{x_i}{r^2} \qquad \frac{\partial^2 f}{\partial x_i^2} = -\frac{1}{r^2} + \frac{2x_i^2}{r^4} \qquad \nabla^2 f = -\frac{2}{r^2} + \frac{2r^2}{r^4} = 0 \tag{197}$$

so this function is harmonic for $x \neq 0$. Thus $\log R_n$ is a martingale so $X_n = \log R_n - \log R_{n-1}$ has mean zero. Note the law of R_n is unique determined by the value of R_{n-1}. However, the law of X_n is independent of the value of R_{n-1} because

$$X_n = \log R_n - \log R_{n-1} = \log \alpha R_n - \log \alpha R_{n-1} \qquad \text{for any } \alpha > 0 \tag{198}$$

We can choose $\alpha = 1/R_{n-1}$. Therefore we conclude the sequence $X_1, X_2, \ldots$ is i.i.d. Furthermore, as we'll show later, $\mathrm{E}\,X_n^2 < \infty$, so the random variables have finite variance σ^2. Consider

$$S_n = X_1 + X_2 + \cdots + X_n \tag{199}$$

For any real number $\alpha > 0$,

$$\Pr(\inf_n S_n = -\infty) \geq \Pr(S_n \leq -\alpha\sigma\sqrt{n}, \text{i.o}) \tag{200}$$

since if the event on the right hand side happens, the event on the left hand side certainly does, since S_n takes on values which are arbitrarily negative if its less than $-\alpha\sigma\sqrt{n}$ infinitely often.

Let E_k be a sequence of events. Then the event that E_k occur infinitely often is given by

$$E_{\text{i.o.}} = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k \tag{201}$$

This implies that

$$\Pr(E_{\text{i.o}}) = \lim_{n\to\infty} \Pr(\bigcup_{k\geq n} E_k) = \limsup_{n\to\infty} \Pr(\bigcup_{k\geq n} E_k) \tag{202}$$

The limit exists because the terms are non-increasing, and it equals the lim sup. Since $\Pr(\bigcup_{k\geq n} E_k) \geq \Pr(E_n)$, a lower bound is given by

$$\Pr(E_{\text{i.o}}) \geq \limsup_{n\to\infty} \Pr(E_n) \tag{203}$$

By the central limit theorem

$$\lim_{n\to\infty} \Pr(S_n/\sqrt{n} \leq -\alpha\sigma) = \Phi(-\alpha) > 0 \tag{204}$$

and therefore $\Pr(\inf_n S_n = -\infty) > 0$. This implies that $\Pr(\inf_n S_n = -\infty) = 1$ by Kolmolgorov's 0-1 law, since we will show this event is in the tail field.

There are two loose ends to tie up. First we will demonstrate that the event $E = \{\inf S_n = -\infty\}$ is in the tail field. Select k such that $S_{n_k} \leq -k$. Now replace $X_1, \ldots, X_m$ by $X'_1, \ldots, X'_m$, and let

$$S'_n = \begin{cases} X'_1 + X'_2 + \cdots + X'_n & \text{if } n \leq m \\ X'_1 + X'_2 + \cdots + X'_m + X_{m+1} + \cdots + X_n & \text{if } n > m \end{cases} \tag{205}$$

If $D = \sum_{i=1}^m X'_i - X_i$, then for $n > m$, $S'_n = S_n + D$. From this its obvious that, for large enough k, $S'_{n_k} \leq -k + M \to -\infty$. Thus $\inf S'_n = -\infty$ as well. Therefore E is in $\sigma(X_{m+1}, X_{m+2}, \ldots)$ since its independent of the values of $X_1, \ldots, X_m$. Since this is true for any m, E is in the tail field.

Finally we show that $\sigma^2 = \text{Var}(X_i) < \infty$. We show $\text{E}\, X_i^2 < \infty$. Without loss of generality, assume $R_{n-1} = 1$, and we'll compute the pdf of $R = R_n$. Let Θ be the direction the Enterprise travels, with $\Theta = 0$ corresponding to the direction toward the center of the solar system. By hypothesis $\Theta \sim U(0, 2\pi)$. Now R is the base of an isoceles triangle with edges 1 and opposite angle Θ so

$$R = 2\sin\frac{\Theta}{2} \qquad \Rightarrow \qquad p_R(r) = \frac{4}{\sqrt{4-r^2}} \tag{206}$$

where $p_R(r)$ is the pdf of R. Thus we want to compute

$$\mathrm{E}(logR)^2 = \int_0^2 \frac{4(\log r)^2}{\sqrt{4-r^2}}\,dr = \int_{-\infty}^{\log 2} \frac{4u^2e^u}{\sqrt{4-e^u}}\,du \tag{207}$$

The left tail when r is close to zero equivalent to the transformed integral when u is close to $-\infty$. This integral resembles the tail of $\Gamma(3) = \int_0^\infty u^2e^{-u}$, and is therefore finite. On the right tail when $r \to 2$ the integrand resembles $4(\log 2)^2/\sqrt{4-r^2}$ which has antiderivative $A \arctan r/2$ for some constant A, and therefore finite integral on $[2-\epsilon, 2]$. On the intermediate range $[\epsilon, 2-\epsilon]$, the integrand is continuous and hence bounded, so the integral is bounded there too. Thus $\mathrm{E}(\log R)^2 < \infty$ ∎

12.1 Branching process. A branching process $Z = \{Z_n : n \geq 0\}$ is constructed in the usual way. Thus, a family $\{X_k^{(n)} : n, k \geq 0\}$ of i.i.d. $\mathbb{Z}_+$-valued random variables is supposed given. We define $Z_0 = 1$ and recursively

$$Z_{n+1} = X_1^{(n+1)} + \cdots + X_{Z_n}^{(n+1)} \tag{208}$$

Assume that if X denotes any of the $X_k^{(n)}$ then

$$\mu = \mathrm{E}\,X < \infty \qquad \text{and} \qquad 0 < \sigma^2 = \mathrm{Var}(X) < \infty \tag{209}$$

Provet that $M_n = Z_n/\mu^n$ defintes a martingale M relative to the filtration $\mathcal{F}_n = \sigma(Z_0, Z_1, \ldots, Z_n)$. Show that

$$\mathrm{E}(Z_{n+1}^2 \mid \mathcal{F}_n) = \mu^2 Z_n^2 + \sigma^2 Z_n \tag{210}$$

and deduce that M is bounded in L^2 iff $\mu > 1$. Show that when $\mu > 1$

$$\mathrm{Var}(M_\infty) = \frac{\sigma^2}{\mu(\mu-1)} \tag{211}$$

Note that since $\mathrm{E}(X_k^{(n+1)} \mid \mathcal{F}_n) = \mathrm{E}\,X_k^{(n+1)} = \mu$ and since Z_n is $\mathcal{F}_n$-measurable

$$\mathrm{E}(Z_{n+1} \mid \mathcal{F}_n) = \sum_{k=1}^{Z_n} \mathrm{E}(X_k^{(n+1)} \mid \mathcal{F}_n) = Z_n\mu \tag{212}$$

From this it immediately follows that M_n is a martingale

$$\mathrm{E}(M_{n+1} \mid \mathcal{F}_n) = \mathrm{E}(Z_{n+1}/\mu^{n+1} \mid \mathcal{F}_n) = Z_n/\mu^n = M_n \tag{213}$$

Similarly since $\mathrm{Var}(X_k^{(n+1)} \mid \mathcal{F}_n) = \sigma^2$,

$$\mathrm{Var}(Z_{n+1} \mid \mathcal{F}_n) = \sum_{k=1}^{Z_n} \mathrm{Var}(X_k^{(n+1)} \mid \mathcal{F}_n) = Z_n\sigma^2 \tag{214}$$

But by definition, $\mathrm{Var}(Z_{n+1} \mid \mathcal{F}_n) = \mathrm{E}(Z_n^2 \mid \mathcal{F}_n) - (\mathrm{E}(Z_{n+1} \mid \mathcal{F}_n))^2$, so

$$\mathrm{E}(Z_n^2 \mid \mathcal{F}_n) = (\mathrm{E}(Z_{n+1} \mid \mathcal{F}_n))^2 + Z_n\sigma^2 = Z_n^2\mu^2 + Z_n\sigma^2 \tag{215}$$

From this we get

$$\mathrm{E}(M_{n+1} \mid \mathcal{F}_n) = \frac{\mu^2 Z_n^2}{\mu^{2n+2}} + \frac{Z_n\sigma^2}{\mu^{2n+2}} = M_n^2 + M_n\frac{\sigma^2}{\mu^{n+2}} \tag{216}$$

Thus

$$b_{n+1} = \mathrm{E}(M_{n+1}^2 - M_n^2) = \frac{\sigma^2}{\mu^{n+2}}\,\mathrm{E}\,M_n = \frac{\sigma^2}{\mu^{n+2}} \tag{217}$$

Note b_n is a geometric series and therefore $\sum_n b_n < \infty$ iff $\mu > 1$. By the theorem in section 12.1, M is bounded in L^2 iff $\sum_n b_n < \infty$, and in that case it converges almost surely to some random variable M_∞ with $\mathrm{E}\,M_\infty^2 = \sum_n b_n + M_0^2$. When $\mu > 1$ we can sum the series

$$\mathrm{Var}(M_\infty) = \mathrm{E}\,M_\infty^2 - (\mathrm{E}\,M_\infty)^2 = \frac{\sigma^2/\mu^2}{1-\mu^{-1}} = \frac{\sigma^2}{\mu(\mu-1)} \tag{218}$$

since $\mathrm{E}\,M_\infty = M_0$ ■

12.2 Let $E_1, E_2, \ldots$ be independent events with $\Pr(E_n) = 1/n$. Let $Y_k = I_{E_k}$. Prove that $\sum(Y_k - \frac{1}{k})/\log k$ converges a.s. and use Kronecker's Lemma to deduce that

$$\frac{N_n}{\log n} \to 1 \qquad \text{a.s.} \tag{219}$$

where $N_n = Y_1 + \cdots + Y_n$

First note $\mathrm{E}\,Y_k = \Pr E_k = \frac{1}{k}$ and

$$\mathrm{Var}\,Y_k \le \mathrm{E}\,Y_k^2 = \mathrm{E}\,I_{E_k}^2 = \mathrm{E}\,I_{E_k} = \Pr(E_k) = \frac{1}{k} \tag{220}$$

Consider the random variable $Z_k = \frac{Y_k - k^{-1}}{\log k}$. Note that $\mathrm{E}\,Z_k = 0$ and

$$\sigma_k^2 = \mathrm{Var}\,Z_k = \frac{\mathrm{Var}(Y_k)}{(\log k)^2} \le \frac{1}{k(\log k)^2} \tag{221}$$

Now $\int \frac{dx}{x(\log x)^2} = -\frac{1}{\log x}$ so by the integral test, $\sum_k \sigma_k^2 < \infty$. Also the Z_k are independent, so by theorem 12.2,

$$\sum_k Z_k = \sum_k \left(Y_k - \frac{1}{k}\right) / \log k \tag{222}$$

converges almost surely. By Kronecker's lemma, this means

$$\frac{\sum_{k=1}^n Y_k - k^{-1}}{\log n} = \frac{N_n - H_n}{\log n} \to 0 \qquad \text{a.s.} \tag{223}$$

where $N_n = Y_1 + \cdots + Y_n$ and $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is the harmonic series. Now $H_n \to \log n + \gamma$ where γ is the Euler-Mascheroni constant. In particular $H_n / \log n \to 1$. Therefore

$$\frac{N_n}{\log n} \to 1 \qquad \text{a.s.} \tag{224}$$

Note that the events in 4.3 for $X_1, X_2, \ldots$ drawn from i.i.d. continuous distribution that there is a "record" at time k satisfies the hypothesis of this problem. In 4.3 we show the events are independent, and that they have probability $\frac{1}{k}$. ■

12.3 Star Trek 3. Prove that if the strategy in 10.11 is (in the obvious sense) employed – and for ever – in $\mathbb{R}^3$ rather than in $\mathbb{R}^2$, then

$$\sum R_n^{-2} < \infty \qquad \text{a.s.} \tag{225}$$

where R_n is the distance from the Enterprise to the Sun at time n.

Let E_n be the event that the Enterprise returns to the solar system for the first time at time n. The area of a cap on a sphere with central angle θ is $2\pi R^2(1 - \cos\theta)$ and therefore the probability of a jump landing in that cap is $\frac{1}{2}(1 - \cos\theta)$. The area formula can be seen by integrating the surface area element $R^2 \sin\theta\, d\theta d\phi$ or by using Archemedes observation that the projection of an sphere onto an enclosing cylendar preserves surface area. For the jumps strategy described in this problem, the radius of the jump sphere for jump n is R_{n-1}, the distance from the Sun. The probabilty we enter the solar system corresponds to the cap whose edge is distance r from the sun. Thus there is an isoceles triangle with equal sides R and base r, and the opposite angle of the base is θ. By the law of cosines, $\rho^2 = 2R^2 - 2R^2\cos\theta$. Let T be the stopping time that the Enterprise returns to the solar system $T = \inf\{R_n \le r\}$. This implies

$$\xi_k = \Pr(E_k \mid \mathcal{F}_{k-1}) = \begin{cases} \frac{\rho^2}{4R_{k-1}^2} & k \le T \\ 0 & \text{otherwise} \end{cases} \tag{226}$$

Using the notation of theorem 12.15, $Z_n = \sum_{k\le n} I_{E_n} \le 1$ for all n. Let $Y_\infty = \sum_{k=1}^{\infty} \xi_k = \sum_{k=1}^{T} \frac{\rho^2}{4R_{k-1}^2}$. If $Y_\infty = \infty$ then 12.15b implies that $\lim_{n\to\infty} Z_n/Y_n = 1$, but this is a contradiction because Z_n is bounded. Therefore $Y_\infty < \infty$ almost surely.

TODO: to clinch the argument I want to take $\rho \to 0$ and $T \to \infty$, but the sum also gets bigger in this limit since it includes the terms where R_n is really close to 0.

■

13.1 Prove that a class $\mathcal{C}$ of RVs is UI if and only if both of the following conditions hold

(i) $\mathcal{C}$ is bounded in L^1, so that $A = \sup\{E(|X|) : X \in \mathcal{C}\} < \infty$

(ii) for every $\epsilon > 0, \exists\delta > 0$ such that if $F \in \mathcal{F}, \Pr(F) < \delta$ and $X \in \mathcal{C}$ then $\mathrm{E}(|X|; F) < \epsilon$

Assume the conditions are true. Let

$$B_K = \sup_{X \in \mathcal{C}} \{\Pr(|X| > K)\} \tag{227}$$

Now for any $X \in \mathcal{C}$, Markov's inequality implies

$$\Pr(|X| > K) \leq \frac{\mathrm{E}|X|}{K} \leq \frac{A}{K} \tag{228}$$

Taking the supremum over all $X \in \mathcal{C}$, this shows $B_K \leq A/K$. Therefore $B_K \to 0$ as $K \to \infty$. For any $\epsilon > 0$, there is a δ satisfying condition (ii). Take K large enough that $B_K < \delta$. From the definition of B_K, all of the sets $F = \{|X| > K\}$ satisfy $\Pr(F) < \delta$, and hence condition (ii) implies $\mathrm{E}(|X|; |X| > K) < \epsilon$ for each $X \in \mathcal{C}$. This verifies that $\mathcal{C}$ is UI.

Conversely assume $\mathcal{C}$ is UI. For any $K > 0$ and $F \in \mathcal{F}$, note

$$I_F \leq I_F(I_{|X|>K} + I_{|X|\leq K}) \leq I_{|X|>K} + I_{|X|\leq K} I_F \tag{229}$$

Multiplying both sides by $|X|$ and taking expectatins gives

$$\begin{aligned} \mathrm{E}(|X|; F) &\leq \mathrm{E}(|X|; |X| > K) + \mathrm{E}(|X|; F \cap \{|X| \leq K\}) \\ &\leq \mathrm{E}(|X|; |X| > K) + K \Pr(F) \end{aligned} \tag{230}$$

Hence, given $\epsilon > 0$ choose K such that $\mathrm{E}(|X|; |X| > K) < \frac{\epsilon}{2}$ for all X. Then choose $\delta = \frac{\epsilon}{2K}$. Equation (230) shows that for this choice of δ, condition (ii) is satisfied. Furthermore with $F = \Omega$ and K chosen so $\mathrm{E}(|X|; |X| > K) \leq 1$, (230) gives a uniform bound of $1 + K$ for $\mathrm{E}(|X|)$, so condition (i) is satisfied. ∎

13.2 Prove that if $\mathcal{C}$ and $\mathcal{D}$ are UI classes of RVs, and if we define

$$\mathcal{C} + \mathcal{D} = \{X + Y : X \in \mathcal{C}, Y \in \mathcal{D}\} \tag{231}$$

then $\mathcal{C} + \mathcal{D}$ is UI. Hint.

We'll verify the conditions in 13.1. First note that for $Z_0 \in \mathcal{C} + \mathcal{D}$ with $Z_0 = X_0 + Y_0$

$$\mathrm{E}|Z_0| \leq \mathrm{E}|X_0| + \mathrm{E}|Y_0| \leq \sup_{X \in \mathcal{C}} \mathrm{E}|X| + \sup_{Y \in \mathcal{D}} \mathrm{E}|Y| \tag{232}$$

The bound on the right hand side is independent of Z, so its a bound for $\sup_{Z \in \mathcal{C}+\mathcal{D}} \mathrm{E}|Z|$, and condition (i) is satisfied.

For $\epsilon > 0$, choose δ such that for any $X \in \mathcal{C}$ and $Y \in \mathcal{D}$ and $F \in \mathcal{F}$ with $\Pr(F) < \delta$, then $\mathrm{E}(|X|; F) < \frac{\epsilon}{2}$ and $\mathrm{E}(|Y|; F) < \frac{\epsilon}{2}$ (Take δ to be the minimum of the δ needed for $\mathcal{C}$ or $\mathcal{D}$ alone). Then for $Z = X + Y$

$$\mathrm{E}(|Z|; F) \leq \mathrm{E}(|X|; F) + \mathrm{E}(|Y|; F) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \tag{233}$$

Thus condition (ii) is satisfied for $\mathcal{C} + \mathcal{D}$ ∎

13.3 Let $\mathcal{C}$ be a UI family of RVs. Say that $Y \in \mathcal{D}$ if for some $X \in \mathcal{C}$ and some sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, we have $Y = E(X \mid \mathcal{G})$, a.s. Prove that $\mathcal{D}$ is UI.

Let $G \in \mathcal{G}$ and $Y = \mathrm{E}(X \mid \mathcal{G})$. Then

$$\begin{aligned} \mathrm{E}(|Y|; G) &= \mathrm{E}(|\mathrm{E}(X \mid \mathcal{G})| \, I_G) = \mathrm{E}(|\mathrm{E}(X I_G \mid \mathcal{G})|) \\ &\leq \mathrm{E}(\mathrm{E}(|X I_G| \mid \mathcal{G})) = \mathrm{E}(|X| I_G) = \mathrm{E}(|X|; G) \end{aligned} \tag{234}$$

The inequality in the middle is the conditional Jensen's inequality for the convex function $|X|$. When $G = \Omega$, this gives a bound $\mathrm{E}(|Y|) \leq \mathrm{E}(|X|)$. So $A = \sup_{X \in \mathcal{C}} \mathrm{E}(|X|)$ is a bound for $\sup_{Y \in \mathcal{D}} \mathrm{E}(|Y|)$, and condition (i) of 13.1 is satisfied. Given $\epsilon > 0$, choose $\delta > 0$ to satisfy condition (ii) of 13.1 for $\mathcal{C}$. Then for any $G \in \mathcal{G} \subset \mathcal{F}$ with $\Pr(G) < \delta$ have for any $Y \in \mathcal{D}$

$$\mathrm{E}(|Y|; G) \leq \mathrm{E}(|X|; G) < \epsilon \tag{235}$$

Thus $\mathcal{D}$ also satisfies condition (ii) of 13.1. Satisfying both conditions, it is UI. ∎

14.1 Hunt's lemma. Suppose that (X_n) is a sequence of RV's such that $X = \lim X_n$ exists a.s. and that (X_n) is dominated by Y in $(L^1)^+$.

$$|X_n(\omega)| \leq Y(\omega), \qquad \forall (n, \omega), \qquad \text{and} \qquad \mathrm{E}(Y) < \infty \tag{236}$$

Let $\{\mathcal{F}_n\}$ be any filtration. Prove that

$$\mathrm{E}(X_n \mid \mathcal{F}_n) \to \mathrm{E}(X \mid \mathcal{F}_\infty) \qquad \text{a.s.} \tag{237}$$

Let $Z_m = \sup_{r \geq m} |X_r - X|$. This tends to 0 because

$$\lim_n X_n = X \qquad \Rightarrow \qquad \lim_n |X_n - X| = 0 \qquad \Rightarrow \qquad \limsup_n |X_n - X| = 0 \tag{238}$$

The last limit is the same as $\lim_m Z_m = 0$. Furthermore $Z_m \to 0$ in L^1 by dominated convergence. Note $|X| = \lim_n |X_n| \leq Y$, so $|Z_m| \leq \sup_{r \geq m} |X_r - X| \leq 2Y$ and is dominated by an L^1 function. Therefore $\mathrm{E}\, Z_m \to 0$.

Let's turn to $\mathrm{E}(X_n \mid \mathcal{F}_n)$. Note for any $m \leq n$,

$$\begin{aligned} |\mathrm{E}(X_n \mid \mathcal{F}_n) - \mathrm{E}(X \mid \mathcal{F}_\infty)| &\leq |\mathrm{E}(X \mid \mathcal{F}_n) - \mathrm{E}(X \mid \mathcal{F}_\infty)| + |\mathrm{E}(X_n \mid \mathcal{F}_n) - \mathrm{E}(X \mid \mathcal{F}_n)| \\ &\leq |\mathrm{E}(X \mid \mathcal{F}_n) - \mathrm{E}(X \mid \mathcal{F}_\infty)| + \mathrm{E}(|X_n - X| \mid \mathcal{F}_n) \\ &\leq |\mathrm{E}(X \mid \mathcal{F}_n) - \mathrm{E}(X \mid \mathcal{F}_\infty)| + \mathrm{E}(Z_m \mid \mathcal{F}_n) \\ &\leq |\mathrm{E}(X \mid \mathcal{F}_n) - \mathrm{E}(X \mid \mathcal{F}_\infty)| \\ &\qquad + |\mathrm{E}(Z_m \mid \mathcal{F}_n) - \mathrm{E}(Z_m \mid \mathcal{F}_\infty)| + \mathrm{E}(Z_m \mid \mathcal{F}_\infty) \end{aligned} \tag{239}$$

By Doob's upward convergence theorem, $\mathrm{E}(X \mid \mathcal{F}_n) \to \mathrm{E}(X \mid \mathcal{F}_\infty)$ and $\mathrm{E}(Z_m \mid \mathcal{F}_n) \to \mathrm{E}(Z_m \mid \mathcal{F}_\infty)$ in L^1 and almost surely. By the tower property and positivity,

$$\mathrm{E}|\mathrm{E}(Z_m \mid \mathcal{F}_\infty)| = \mathrm{E}\, Z_m \to 0 \tag{240}$$

so the last term converges to 0 in L^1. Furthermore by conditional dominated convergence, $\mathrm{E}(Z_m \mid \mathcal{F}_\infty) \to \mathrm{E}(0 \mid \mathcal{F}_\infty) = 0$, so the last term converges to 0 almost surely. Therefore $\mathrm{E}(X_n \mid \mathcal{F}_n) \to \mathrm{E}(X \mid \mathcal{F}_\infty)$ in L^1 and almost surely. ∎

14.2 Azuma-Hoeffding

(a) Show that if Y is a RV with values in $[-c, c]$ and $\mathrm{E}\, Y = 0$ then for $\theta \in \mathbb{R}$

$$\mathrm{E}\, e^{\theta Y} \leq \cosh \theta c \leq \exp\left(\frac{1}{2}\theta^2 c^2\right) \tag{241}$$

(b) Prove that if M is a positive martingale null at 0 such that for some sequence $(c_n : n \in \mathbb{N})$ of positive constants

$$|M_n - M_{n-1}| \leq c_n, \qquad \forall n \tag{242}$$

then, for $x > 0$,

$$\Pr\left(\sup_{k \leq n} M_k \geq x\right) \leq \exp\left(\frac{1}{2}x^2 \Big/ \sum_{k=1}^{n} c_k^2\right) \tag{243}$$

(a) Define α such that Y is the convex combination of c and $-c$

$$\alpha = \frac{Y + c}{2c} \qquad \Rightarrow \qquad Y = c\alpha - c(1 - \alpha) \tag{244}$$

Since $|Y| \leq c$ we have $\alpha \in [0, 1]$. By Jensen's inequality since $\mathrm{E}\, \alpha = \frac{1}{2}$

$$\mathrm{E}\, e^{\theta Y} = \mathrm{E}\, e^{\theta c\alpha - \theta c(1-\alpha)} \leq e^{\theta c}\, \mathrm{E}\, \alpha + e^{-\theta c}\, \mathrm{E}(1 - \alpha) = \frac{1}{2}(e^{\theta c} + e^{-\theta c}) = \cosh \theta c \tag{245}$$

Now compare the Taylor series of $\cosh x$ vs $\exp(\frac{1}{2}x^2)$

$$\cosh x = \sum_k \frac{x^{2k}}{(2k)!} \qquad \exp\left(\frac{1}{2}x^2\right) = \sum_k \frac{x^{2k}}{2^k k!} \tag{246}$$

For $x \geq 0$, we see $\exp(\frac{1}{2}x^2)$ dominates $\cosh x$ term by term since the product on the RHS below contains only the even terms

$$(2k)! = 2k \cdot (2k-1) \cdots 2 \cdot 1 \geq 2k \cdot (2k-2) \cdots 2 = 2^k k! \tag{247}$$

Thus we get the second inequality $\cosh \theta c \leq e^{\frac{1}{2}\theta^2 c^2}$.

(b) Now by the conditional Jensen inequality, e^{M_n} is a submartingale. By Doob's submartingale inequality

$$\Pr\left(\sup_{k \leq n} e^{\theta M_k} \geq e^{\theta x}\right) \leq e^{-\theta x}\, \mathrm{E}\, e^{\theta M_n} \tag{248}$$

Using the tower property, the inequality from (a)

$$\mathrm{E}\, e^{\theta M_n} = \mathrm{E}(\mathrm{E}(e^{M_n - M_{n-1}} e^{M_{n-1}} \mid \mathcal{F}_{n-1})) = e^{-\frac{1}{2}\theta^2 c_n^2}\, \mathrm{E}\, e^{\theta M_{n-1}} \tag{249}$$

Thus by induction $\mathrm{E}\, e^{\theta M_n} = e^{\frac{1}{2}\theta^2(c_n^2+c_{n-1}^2+\cdots+c_1^2)}\,\mathrm{E}\, e^{M_0} = \exp(\frac{1}{2}\theta^2\sum_k c_k^2)$. If we choose θ to minimize $\exp(-\theta x + \frac{1}{2}\theta^2\sum_k c_k^2)$ we get $\theta = x/(\sum_k c_k^2)$. Plugging this into (248), where we can rewrite the LHS because $e^{\theta x}$ is monotonically increasing

$$\Pr\left(\sup_{k\le n} M_k \ge x\right) \le \exp\left(-\frac{1}{2}x\Big/\sum_k c_k^2\right) \tag{250}$$

■

16.1 Prove that

$$\lim_{T\uparrow\infty}\int_0^T x^{-1}\sin x\,dx = \frac{\pi}{2} \tag{251}$$

Define

$$f(\lambda) = L[x^{-1}\sin x] = \int_0^\infty \frac{e^{-\lambda x}\sin x}{x}\,dx \tag{252}$$

Then by (97) $f'(\alpha) = -L[\sin x]$. We can compute this handily

$$\begin{aligned} f'(\alpha) &= -\int_0^\infty e^{-\lambda x}\sin x\,dx && (253)\\ &= \cos x e^{-\lambda x}\Big|_0^\infty + \lambda\int_0^\infty e^{-\lambda x}\cos x\,dx && (254)\\ &= 1+\lambda\left(\sin x e^{-\lambda x}\Big|_0^\infty + \lambda\int_0^\infty \sin x e^{-\lambda x}\right) && (255)\\ &= (1+\lambda^2)f'(\alpha) && (256)\end{aligned}$$

From dominated convergence theorem, $f(\infty) = 0$. Thus we integrate from ∞ to find

$$f(\lambda) = \int_\infty^\alpha \frac{d\lambda}{1+\lambda^2} = \arctan(\lambda) + \frac{\pi}{2} \tag{257}$$

Set $\lambda = 0$ to find $\int_0^\infty \frac{\sin x}{x}\,dx = \frac{\pi}{2}$

A second solution is to use complex analysis. Consider the contour γ in the complex plane along the real line from $-T$ to $-\epsilon$, then a semicircle centered at 0 from $-\epsilon$ to ϵ, then along the real line from ϵ to T, then a semicircle centered at 0 from T to $-T$. By Cauchy's theorem

$$\lim_{\epsilon\to 0}\int_\gamma \frac{e^{iz}}{z}\,dz = i\pi \tag{258}$$

since the integrand is holomorphic inside of the region bound by γ, but has a residue at 0, and the arc of radius ϵ is π radians. On the semicircle of radius T, writing $z = Te^{i\theta}$ we have $dz = iTe^{i\theta}d\theta$.

$$\left|\int_0^\pi \frac{\exp(iTe^{i\theta})}{Te^{i\theta}}\,iTe^{i\theta}d\theta\right| \le \int_0^\pi \left|e^{iT(\cos\theta+i\sin\theta))}\right|\,d\theta = \int_0^\pi e^{-T\sin\theta}\,d\theta \tag{259}$$

The integrand is dominated by 1, so by the dominated convergence theorem the integral converges to 0 as $T \to \infty$. Therefore as $T \to \infty$ and $\epsilon \to 0$ we have

$$\lim_{T\to\infty,\epsilon\to 0} \int_\gamma \frac{e^{iz}}{z}\, dz = \int_{-\infty}^{\infty} \frac{e^{ix}}{x}\, dx \tag{260}$$

and we conclude this integral equals $i\pi$. Making the substitution $x \to -x$ we conclude

$$\int_{-\infty}^{\infty} \frac{e^{-ix}}{x}\, dx = -i\pi \tag{261}$$

Using the fact $\sin x = \frac{e^{ix}-e^{-ix}}{2i}$, this implies

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\, dx = \pi \qquad \Rightarrow \qquad \int_0^{\infty} \frac{\sin x}{x}\, dx = \frac{\pi}{2} \tag{262}$$

Since the integrand on the left is even. ∎

16.2 Prove that if Z has $U[-1,1]$ distribution, then

$$\phi_Z(\theta) = \sin\theta/\theta \tag{263}$$

Show there do not exist i.i.d. RV X and Y such that

$$X - Y \sim U[-1,1] \tag{264}$$

Computing

$$\phi_Z(\theta) = \mathrm{E}\, e^{i\theta Z} = \int_{-1}^{1} e^{i\theta x}\frac{1}{2}\, dx = \left.\frac{e^{i\theta x}}{2i\theta}\right|_{-1}^{1} = \frac{\sin\theta}{\theta} \tag{265}$$

For X and Y i.i.d. and $Z = X - Y$

$$\phi_Z(\theta) = \mathrm{E}\, e^{i\theta(X-Y)} = \mathrm{E}\, e^{i\theta X}e^{-i\theta Y} = \mathrm{E}\, e^{i\theta X}\,\mathrm{E}\, e^{-i\theta Y} = \phi_X(\theta)\phi_X(-\theta) \tag{266}$$

where we used the property that $\mathrm{E}\, f(X)g(Y) = \mathrm{E}\, f(X)\,\mathrm{E}\, g(Y)$ for independent X and Y. However since $\phi_X(-\theta) = \overline{\phi_X(\theta)}$, in order for $Z = X - Y$ to satisfy $Z \sim U[-1,1]$ it must be that

$$\|\phi_X(\theta)\|^2 = \frac{\sin\theta}{\theta} \tag{267}$$

But this is impossible since the left hand side is strictly non-negative whereas the right hand side is sometimes negative. ∎

16.3 Calculate $\phi_X(\theta)$ where X has the Cauchy distribution, which has pdf $\frac{1}{1+x^2}$. Show that the Cauchy distribution is stable, i.e., that $(X_1 + X_2 + \cdots + X_n)/n \sim X$ for i.i.d. Cauchy X_i.

First assume $\theta > 0$. Consider the contour γ which goes from $-R$ to R along the real line, and then along a semicircle centered at 0 from R to $-R$. By Cauchy's theorem

$$\int_\gamma \frac{e^{\theta iz}}{1+z^2}\,dz = 2\pi i \frac{e^{-\theta}}{2i} = \pi e^{-\theta} \tag{268}$$

since there is exactly one residue at $z = i$ (the denominator can be written $(i+z)(i-z)$) and the residue there is $e^{-\theta}/2$. Along the semicircle we can use the elementary bound

$$\left|\int_{\gamma'} \frac{e^{i\theta z}}{1+z^2}\,dz\right| \leq \frac{\pi R}{R^2-1} \to 0 \tag{269}$$

since for $z = Re^{i\psi}$ the magnitude $\left|\exp(i\theta R e^{i\psi})\right| = e^{-R\theta \sin\psi} \leq 1$ (this is where we use the hypothesis $\theta > 0$ and that the path is a semicircle in the upper half plane). Also $|1+z^2| \geq |z|^2 - 1 = R^2 - 1$. Thus the integrand is bound by $\frac{1}{R^2-1}$ and the length of the path is πR. Thus in the limit only the segment along the real line makes a contribution to the path integral and we conclude for $\theta > 0$

$$\phi_X(\theta) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{i\theta x}}{1+x^2}\,dx = e^{-\theta} \qquad \text{if } \theta \geq 0 \tag{270}$$

For $\theta < 0$ we can modify the above argument, using a semicircle in the lower half plane from R to $-R$ instead. Then in Cauchy's theorem we pick up the residue at $-i$ (instead of i) going along a clockwise path (instead of counterclockwise). Thus the integral equals πe^{θ}. We get the same bounds on the integral which show the contribution of the segment along the semicircle tends to 0. Thus

$$\phi_X(\theta) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{i\theta x}}{1+x^2}\,dx = e^{\theta} \qquad \text{if } \theta \leq 0 \tag{271}$$

Both cases may be summarized by the formula $\phi_X(\theta) = e^{-|\theta|}$. From this it immediately follows that X is stable since

$$\phi_{S_n/n}(\theta) = \phi_X(\theta/n)^n = e^{-n|\theta/n|} = e^{-|\theta|} \tag{272}$$

By the uniqueness of characteristic functions, S_n/n is distributed like a standard Cauchy random variable.

All that remains is to justify $\phi_{S_n/n}(\theta) = \phi_X(\theta/n)^n$. This follows from the following calculation for i.i.d. X_k:

$$\phi_{S_n/n}(\theta) = \mathrm{E}\, e^{i\theta(X_1+\cdots+X_n)/n} = (\mathrm{E}\, e^{i\theta X_1/n})(\mathrm{E}\, e^{i\theta X_2/n})\cdots(\mathrm{E}\, e^{i\theta X_n/n}) = \phi_X(\theta/n)^n \tag{273}$$

■

16.4 Suppose that $X \sim \mathcal{N}(0,1)$. Show that $\phi_X(\theta) = \exp(-\frac{1}{2}\theta^2)$

Note that

$$\phi_X(\theta) = \mathrm{E}\, e^{i\theta X} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\theta x} e^{-\frac{1}{2}x^2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-i\theta))^2 - \frac{1}{2}\theta^2} = e^{-\frac{1}{2}\theta^2} I(i\theta) \quad (274)$$

where $I(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\alpha)^2}\, dx$. For a real parameter α, $I(\alpha) = I(0) = 1$, since we can just perform a change of variables $u = x - \alpha$, and the range of integration is unchanged.

In the complex case, consider a path integral in $\mathbb{C}$ along the rectangle with corners at $\pm R$ and $\pm R - i\theta$ of the function $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Since $f(z)$ is entire, the path integral is 0. As $R \to \infty$, the integral along the bottom side from $-R - i\theta$ to $R - i\theta$ tends to $I(i\theta)$. The integral along the top side from R to $-R$ tends to $-I(0)$ (since the limits of the integral are reversed). Along the sides of the rectangle, $z = \pm R - ix$ for $x \in [0, \theta]$, so the magnitude of the integrand satisfies

$$\left| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\pm R - ix)^2} \right| = \frac{1}{\sqrt{2\pi}} |e^{-\frac{1}{2}(R^2 - x^2 + 2Rxi)}| \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(R^2 - \theta^2)} \to 0 \quad (275)$$

The path along this side is constant length θ as $R \to \infty$. Therefore the contribution to the path integral from the left and right sides of the rectangle is negligible as $R \to \infty$, by the simple bound of the path length times the maximum integrand magnitude. We conclude

$$0 = \int_\gamma f(z)\, dz \to I(i\theta) - I(0) \quad (276)$$

Thus $I(i\theta) = I(0) = 1$ and $\phi_X(\theta) = e^{-\frac{1}{2}\theta^2}$. ∎

16.5 Prove that if ϕ is the characteristic function of a RV X then ϕ is *non-negative definite* in that for $c_1, c_2, \ldots, c_n \in \mathbb{C}$ and $\theta_1, \theta_2, \ldots, \theta_n \in \mathbb{R}$

$$\sum_{j,k} c_j \bar{c}_k \phi(\theta_j - \theta_k) \geq 0 \quad (277)$$

Consider the RV $Z = \sum_k c_k e^{i\theta_k X}$. Now by the positivity of expectations

$$\mathrm{E}|Z|^2 \geq 0 \quad (278)$$

Since

$$|Z|^2 = Z\overline{Z} = \left(\sum_j c_j e^{i\theta_j X} \right) \left(\sum_k \bar{c}_k e^{-i\theta_k X} \right) = \sum_{j,k} c_j \bar{c}_k e^{i(\theta_j - \theta_k)X} \quad (279)$$

Taking expectations of both sides and using linearity yields (277) ∎

16.6

(a) Let $(\Omega, \mathcal{F}, \Pr) = ([0,1], \mathcal{B}[0,1], \text{Leb})$. What is the distribution of the random variable $Z = 2\omega - 1$? Let $\omega = \sum 2^{-n} R_n(\omega)$ be the binary expansion of ω. Let

$$U(\omega) = \sum_{\text{odd } n} 2^{-n} Q_n(\omega) \qquad \text{where} \qquad Q_n(\omega) = 2R_n(\omega) - 1 \tag{280}$$

Find a random variable V independent of U such that U and V are identically distributed and $U + \frac{1}{2}V$ is uniformly distributed on $[-1,1]$

(b) Now suppose that (on some probability triple) X and Y are i.i.d. RV such that

$$X + \frac{1}{2}Y \qquad \text{is uniformly distributed on } [-1,1] \tag{281}$$

Let ϕ be the characteristic function of X. Calculate $\phi(\theta)/\phi(\frac{1}{4}\theta)$. Show that the distribution of X must be the same as that of U in part (a) and deduce that there exists a set $F \in \mathcal{B}[-1,1]$ such that $\text{Leb}(F) = 0$ and that $\Pr(X \in F) = 1$

(a) The Lebesgue measure gives the uniform distribution on $[0,1]$. Under the linear transform $Z = 2\omega - 1$ the distribution remains uniform, but over $[-1,1]$. To see this note for any $z \in [-1,1]$

$$\Pr(Z < z) = \Pr\left(\omega < \frac{1}{2}(z+1)\right) = \text{Leb}\left(0, \frac{1}{2}(z+1)\right) = \frac{1}{2}(z - (-1)) \tag{282}$$

which is exactly the distribution for $U[-1,1]$. Let V be given by a similar expression to U

$$V(\omega) = \sum_{\text{odd } n} 2^{-n}(2R_{n+1}(\omega) - 1) = \sum_{\text{even } n} 2^{-n+1}(2R_n(\omega) - 1) \tag{283}$$

the difference being that the nth term uses the $n+1$st Rademacher function R_{n+1} rather than R_n. Its a standard result that the Rademacher functions $R_n(\omega)$ are i.i.d.. The variables U and V are functions of disjoint collections of R_n, so they are independent. Because their expressions as functions of identically distributed Rademacher functions are the same, the variables have the same distribution. Its also clear that

$$\begin{aligned} U + \frac{1}{2}V &= \sum_{\text{odd } n} 2^{-n}(2R_n(\omega) - 1) + \sum_{\text{even } n} 2^{-n}(2R_n(\omega) - 1) \\ &= 2\left(\sum_{n \in \mathbb{N}} 2^{-n} R_n(\omega)\right) - 1 = 2\omega - 1 \end{aligned} \tag{284}$$

Thus $U + \frac{1}{2}V \sim U[-1,1]$

(b) From the assumption $Z = X + \frac{1}{2}Y$ has $Z \sim U[-1,1]$ we get

$$\phi_X(\theta)\,\phi_X(\frac{1}{2}\theta) = \int_{-1}^{1} e^{i\theta u} \cdot \frac{du}{2} = \frac{\sin\theta}{\theta} \tag{285}$$

From this we deduce $\phi_X(\frac{1}{2}\theta)\phi_X(\frac{1}{4}\theta) = \frac{\sin(\frac{1}{2}\theta)}{\frac{1}{2}\theta}$ and hence

$$\frac{\phi_X(\theta)}{\phi_X(\frac{1}{4}\theta)} = \frac{\phi_X(\theta)\,\phi_X(\frac{1}{2}\theta)}{\phi_X(\frac{1}{2}\theta)\,\phi_X(\frac{1}{4}\theta)} = \frac{\sin\theta}{2\sin(\frac{1}{2}\theta)} = \cos(\frac{1}{2}\theta) \tag{286}$$

Hence, by induction

$$\frac{\phi_X(\theta)}{\phi_X(4^{-k}\theta)} = \frac{\phi_X(\theta)}{\phi_X(\frac{1}{4}\theta)}\frac{\phi_X(\frac{1}{4}\theta)}{\phi_X(\frac{1}{16}\theta)}\cdots\frac{\phi(4^{-k+1}\theta)}{\phi(4^{-k}\theta)} = \cos(\frac{1}{2}\theta)\cos(\frac{1}{8}\theta)\cdots\cos(2^{-2k+1}\theta) \tag{287}$$

Since ϕ is uniformly continuous and $\phi(0) = 0$ we conclude by taking the limit $k \to \infty$

$$\phi(\theta) = \prod_{\text{odd } k} \cos(2^{-k}\theta) \tag{288}$$

On the other hand

$$\phi_{Q_n}(\theta) = \mathrm{E}\, e^{i\theta Q_n} = \frac{1}{2}e^{i\theta} + \frac{1}{2}e^{-i\theta} = \cos(\theta) \tag{289}$$

Therefore since $U = \sum_{\text{odd } k} 2^{-k} Q_n$ we have

$$\phi_U(\theta) = \prod_{\text{odd } k} \cos(2^{-k}\theta) = \phi_X(\theta) \tag{290}$$

So it must be that U and X have the same distribution. TODO prove the measure 0 thing. Its not hard to show that the values of X lie in a cantor-like set with middle thirds excluded. However this approach doesn't use the characteristic function at all... is there some clever thing with Fourier analysis?

■

18.1

(a) Supose that $\lambda > 0$ and that (for $n > \lambda$), F_n is the DF associated with the Binomial distribution $\mathrm{Bin}(n, \lambda/n)$. Prove using CF's that F_n converges weakly to F where F is the DF of the Poisson distribution with parameter λ.

(b) Suppose that $X_1, X_2, \ldots$ are i.i.d. RV's each with the density function $(1 - \cos x)/\pi x^2$ on $\mathbb{R}$. Prove that for $x \in \mathbb{R}$

$$\lim_{n\to\infty} \Pr\left(\frac{X_1 + X_2 + \cdots + X_n}{n} \le x\right) = \frac{1}{2} + \pi^{-1}\arctan x \tag{291}$$

where $\arctan \in (-\frac{\pi}{2}, \frac{\pi}{2})$

Consider the Bernoulli random variable Z with parameter p

$$\phi_Z(\theta) = \mathrm{E}\, e^{i\theta Z} = pe^{i\theta\cdot 1} + qe^{i\theta\cdot 0} = p(e^{i\theta} - 1) + 1 \tag{292}$$

(a) Since $B(n,p)$ is the sum of independent Bernoulli random variables, if $F_n \sim B(n,p) = B(n,\lambda/n)$ then

$$\phi_{F_n}(\theta) = \left(pe^{i\theta} + q\right)^n = \left(1 + \frac{\lambda}{n}(e^{i\theta} - 1)\right)^n \to \exp(\lambda(e^{i\theta} - 1)) \tag{293}$$

On the other if F has a Poisson distribution with parameter λ, its CF is given by

$$\phi_F(\theta) = \mathrm{E}\, e^{i\theta F} = \sum_{k=0}^{\infty} e^{i\theta k} e^{-\lambda} \frac{\lambda^k}{k!} = \exp(\lambda e^{i\theta} - \lambda) \tag{294}$$

which is exactly the same as $\lim_{n\to\infty} \phi_{F_n}(\theta)$. Thus we conclude that $F_n \to F$ in weakly in distribution.

(b) Let's compute

$$\phi_X(\theta) = \mathrm{E}\, e^{i\theta X} = \int_{\mathbb{R}} e^{i\theta x} \frac{1 - \cos x}{\pi x^2}\, dx \tag{295}$$

To this end first let's compute for $k \in \mathbb{N}$ and $\alpha > 0$

$$I_k(\alpha) = \int_{\mathbb{R}} x^{-k} e^{i\alpha x}\, dx \tag{296}$$

(Aside: there's something subtle going on here, because, like the integral in 16.1, this is not integrable in the L^1 sense). As in problem 16.1, take a contour along the real axis from $-R$ to $-\epsilon$, a semicircle in the upper half plane from $-\epsilon$ to ϵ, a line from ϵ to R, and then a semicircle in the upper half plane back to $-R$. We use a variation of the residue theorem $\int_{\gamma'} \frac{f(z)}{(z-z_0)^k}\, dz = \frac{i\Delta\theta}{(k-1)!} f^{(k-1)}(z_0)$, and $\Delta\theta$ is the radians in the infinitesimal path.

$$\oint_\gamma \frac{e^{i\alpha z}}{z^k}\, dz = \pi i \frac{(i\alpha)^{k-1}}{(k-1)!} \tag{297}$$

Using the technique of 16.1, we get an estimate of the integral along the big semicircle

$$\int_{\gamma''} \frac{e^{i\alpha z}}{z^k}\, dz = \int_0^\pi \frac{e^{-R\alpha \sin\theta}}{R^{k-1}}\, d\theta \to 0 \qquad \text{as } R \to \infty \tag{298}$$

Since $|x^{-k}| \le |x^{-1}|$ on the big semicircle, the estimate for the contribution of this segment holds. Therefore as $R \to \infty$ and $\epsilon \to 0$ we get

$$\oint_\gamma \frac{e^{i\alpha z}}{z^k}\, dz \to I_k(\alpha) \tag{299}$$

When $\alpha < 0$ we can modify the argument by taking semicircles in the lower half plane. This allows us to conclude the integral vanishes along the big semicircle, but now the sense of integration along the small semicircle is reversed to clockwise from counterclockwise. We can summarize both cases as

$$I_k(\alpha) = \frac{\pi (i\alpha)^k}{(k-1)!|\alpha|} \tag{300}$$

This gives $I_1(\alpha) = \pi i \operatorname{sgn}(x)$ and $I_2(\alpha) = -\pi|\alpha|$.

Expressing $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$, our CF can be written

$$\begin{aligned}\phi_X(\theta) &= \frac{1}{\pi}\left(I_2(\theta) - \frac{1}{2}(I_2(\theta+1) + I_2(\theta-1))\right) \\ &= \frac{1}{2}\left(|\theta+1| + |\theta-1|\right) - |\theta| \\ &= \begin{cases} 1-|\theta| & |\theta| \le 1 \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{301}$$

Thus ϕ_X is a kind of a witch-hat centered at 0.

Like in 16.3, for $S_n = X_1 + \cdots + X_n$,

$$\phi_{S_n/n}(\theta) = \phi_X(\theta/n)^n = \begin{cases} \left(1 - \frac{|\theta|}{n}\right)^n & \text{if } \frac{|\theta|}{n} \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{302}$$

For any fixed θ, for large enough n, $|\theta|/n \le 1$ so the first case applies eventually. Taking the limit as $n \to \infty$ this implies

$$\phi_{S_n/n}(\theta) \to e^{-|\theta|} \tag{303}$$

This is the CF of the Cauchy distribution, S_n/n converges in distribution to the Cauchy distribution. We conclude

$$P(S_n/n \le x) \to \int_{-\infty}^{x} \frac{dx}{\pi(1+x^2)} = \frac{1}{2} + \pi^{-1}\arctan(x) \tag{304}$$

■

18.2 Prove the weak law of large numbers in the following form. Suppose $X_1, X_2, \ldots$ are i.i.d. RV each with the same distribution as X. Suppose $X \in \mathcal{L}^1$ and that $\mathrm{E}\, X = \mu$. Prove by use of CF's that the distribution of

$$S_n = n^{-1}(X_1 + \cdots + X_n) \tag{305}$$

converges weakly to the unit mass at μ. Deduce that

$$S_n \to \mu \text{ in probability} \tag{306}$$

Of course, SLLN implies this weak law.

Let $Z \sim \delta_\mu$ be equal to μ almost surely. Then

$$\phi_Z(\theta) = \mathrm{E}\, e^{i\theta Z} = e^{i\theta\mu} \tag{307}$$

Now consider the Taylor expansion of $\phi_X(\theta)$. Note that

$$\phi_X(0) = 1 \qquad \phi'_X(0) = \mathrm{E}\, iXe^{i\theta X}\Big|_{\theta=0} = i\mu \tag{308}$$

Therefore

$$\log \phi_X(\theta) = 1 + i\mu\theta + o(\theta) \tag{309}$$

So the characteristic function satisfies

$$\phi_{S_n}(\theta) = \phi_X(\theta/n)^n = \left(1 + \frac{i\mu\theta}{n} + o\left(\frac{\theta}{n}\right)\right)^n \to \exp(i\theta\mu) = \phi_Z(\theta) \tag{310}$$

This implies that $S_n \to \delta_\mu$ in distribution. However, for a unit mass, convergence in distribution is the same as convergence in probability since for any $\epsilon > 0$

$$\Pr(S_n < \mu - \epsilon) \to \Pr(Z < \mu - \epsilon) = 0 \qquad \Pr(S_n > \mu + \epsilon) \to \Pr(Z > \mu + \epsilon) = 0 \tag{311}$$

and hence $\Pr(|S_n - \mu| > \epsilon) \to 0$ for any $\epsilon > 0$. This proves the weak law of large numbers. ■

Weak Convergence for $[0,1]$

18.3 Let X and Y be RV's taking values in $[0,1]$. Suppose that

$$\mathrm{E}\, X^k = \mathrm{E}\, Y^k \qquad k = 0, 1, 2, \ldots \tag{312}$$

Prove that

(i) $\mathrm{E}\, p(X) = \mathrm{E}\, p(Y)$ for every polynomial p

(ii) $\mathrm{E}\, f(X) = \mathrm{E}\, f(Y)$ for every continuous function f on $[0,1]$

(iii) $\Pr(X \le \alpha) = \Pr(Y \le \alpha)$ for every $\alpha \in [0,1]$.

Item (i) follows from linearity of expectations. If $p(x) = \sum_{k=0}^n c_k x^k$ then

$$\mathrm{E}\, p(X) = \sum_{k=0}^n c_k \,\mathrm{E}\, X^k = \sum_{k=0}^n c_k \,\mathrm{E}\, Y^k = \mathrm{E}\, p(Y) \tag{313}$$

Item (ii) follows from the Weierstrass theorem. On the compact set $[0,1]$, every continuous function f can be approximated uniformly by a sequence of polynomials p_n. Choose N large enough so that for all $n > N$, $|f(x) - p_n(x)| < \epsilon$ for all x. In this case

$$|\mathrm{E}\, f(X) - \mathrm{E}\, p_n(X)| \le \mathrm{E}|f(X) - p_n(X)| \le \epsilon \tag{314}$$

Therefore

$$\begin{aligned} |\mathrm{E}\, f(X) - \mathrm{E}\, f(Y)| &\le |\mathrm{E}\, f(X) - \mathrm{E}\, p_n(X)| + |\mathrm{E}\, p_n(X) - \mathrm{E}\, p_n(Y)| + |\mathrm{E}\, p_n(X) - \mathrm{E}\, f(X)| \\ &\le \epsilon + 0 + \epsilon = 2\epsilon \end{aligned} \tag{315}$$

Since ϵ is arbitrary, this shows $\mathrm{E}\, f(X) = \mathrm{E}\, f(Y)$.

Approximate $I_{[0,\alpha]}$ by the continuous peicewise linear function

$$f_n(x) = \begin{cases} 1 & x \in [0,\alpha) \\ n\,(\alpha - x) & x \in [\alpha, \alpha + \frac{1}{n}) \\ 0 & x \in [\alpha + \frac{1}{n}, 1] \end{cases} \tag{316}$$

Using the dominated convergence theorem

$$\lim_{n\to\infty} \mathrm{E}\, f_n(X) = \mathrm{E}\, I_{[0,\alpha]}(X) = \Pr\left(X \le \alpha\right) \tag{317}$$

The same holds for Y. However, $\mathrm{E}\, f_n(X) = \mathrm{E}\, f_n(Y)$ for all n by part (ii). Therefore the limits are equal and $\Pr(X \le \alpha) = \Pr(Y \le \alpha)$. Thus the moments unique determine the distribution on $[0,1]$. ∎

18.4 Suppose that (F_n) is a sequence of DFs with

$$F_n(x) = 0 \text{ for } x < 0, \qquad F_n(1) = 1, \qquad \text{for every } n \tag{318}$$

Suppose that

$$m_k = \lim_n \int_{[0,1]} x^k dF_n \text{ exists for } k = 0, 1, 2, \ldots \tag{319}$$

Use the Helly-Bray Lemma and 18.3 to show that $F_n \xrightarrow{w} F$, where F is characterized by $\int_{[0,1]} x^k\, dF = m_k, \forall k$

Let F_n be associated with the measure μ_n. Suppose a subsequence $F_{n_i} \xrightarrow{w} F$ converges in distribution. By Helly-Bray there is a subsequence n_i such that $F_{n_i} \xrightarrow{w} F$, and call the associated measure μ. By the definition of weak convergence since x^k is a continuous function on $[0,1]$, and since n_i is a subsequence of n

$$m_k = \lim_n \mu_n(x^k) = \lim_{n_i} \mu_{n_i}(x^k) = \mu(x^k) = \int_{[0,1]} x^k\, dF \tag{320}$$

If another subsequence n_i' satisfies $F_{n_i'} \xrightarrow{w} F'$ with associated measure μ', then 10.3 implies that $F' = F$ since the above argument shows $\mu'(x^k) = m_k$ as well, so $\mu'(x^k) = \mu(x^k)$ for all k.

This implies that $F_n \xrightarrow{w} F$. For suppose there is some continuous $g : [0,1] \to \mathbb{R}$ such that $\limsup \mu_n(g) \ne \mu(g)$ or $\liminf \mu_n(g) \ne \mu(g)$. Then choose a subsequence n_i so that $\mu_{n_i}(g)$ converges to a value other than $\mu(g)$. By Helly-Bray we can take a further subsequence converging to a distribution F'. This distribution must satisfy $\mu'(g) \ne \mu(g)$, but this contradicts what we've previously shown, that every convergent subsequence has the same limit. Therefore $\lim_n \mu_n(g)$ exists for all $g \in C[0,1]$ and $\lim \mu_n(g) = \mu(g)$. ∎

18.5 Improving on 18.3, a moment inversion formula. Let F be a distribution with $F(0-) = 0$ and $F(1) = 1$. Let μ be the associated law, and define

$$m_k = \int_{[0,1]} x^k \, dF(x) \tag{321}$$

Define

$$\begin{aligned} &\Omega = [0,1] \times [0,1]^{\mathbb{N}}, \ \ = \mathcal{B} \times \mathcal{B}^{\mathbb{N}}, \ \ \Pr = \mu \times \text{Leb}^{\mathbb{N}} \\ &\Theta(\omega) = \omega_0, \ \ H_k(\omega) = I_{[0,\omega_0]}(\omega_k) \end{aligned} \tag{322}$$

This models the situation in which Θ is chosen with law μ, a coin with probability Θ is minted and tossed at times $1, 2, \ldots$. See E10.8. The RV H_k is 1 if the kth toss produces heads, 0 otherwise. Define

$$S_n = H_1 + H_2 + \cdots + H_n \tag{323}$$

By the strong law of large numbers and fubini's theorem

$$S_n/n \to \Theta, \qquad \text{a.s.} \tag{324}$$

Define a map D on the space of real sequences $(a_n : n \in \mathbb{Z}_+)$ by setting

$$Da = (a_n - a_{n+1} : n \in \mathbb{Z}_+) \tag{325}$$

Prove that

$$F_n(x) = \sum \binom{n}{i} (D^{n-i} m)_i \to F(x) \tag{326}$$

at every point of continuity of F

By the SLLN, we know that

$$F(x) = \Pr(\Theta \le x) = \lim_{n\to\infty} \Pr(S_n/n \le x) \tag{327}$$

Conditional on Θ, $S_n \sim \text{Bin}(n, \Theta)$, so marginalizing over Θ and using Fubini's theorem we find

$$\Pr(S_n \le nx) = \int_{[0,1]} \sum_{k \le nx} \binom{n}{k} \theta^{n-k} (1-\theta)^k \, \mu(d\theta) = \sum_{k \le nx} \binom{n}{k} \int_{[0,1]} \theta^{n-k} (1-\theta)^k \, \mu(d\theta) \tag{328}$$

But we can write the integral in terms of the m_k. By induction I claim

$$(D^n m)_k = \int_{[0,1]} x^k (1-x)^n \, \mu(dx) \tag{329}$$

Clearly this is true for $n = 0$, for $n \geq 1$

$$\begin{aligned}(D^n m)_k &= (D \cdot (D^{n-1} m))_k = (D^{n-1} m)_k - (D^{n-1} m)_{k+1} \\ &= \int_{[0,1]} x^k (1-x)^{n-1} \, \mu(dx) - \int_{[0,1]} x^{k+1} (1-x)^{n-1} \, \mu(dx) \\ &= \int_{[0,1]} x^k (1-x)^{n-1} (1-x) \, \mu(dx)\end{aligned} \tag{330}$$

Pulling together the parts we get (326) ∎

18.6 Moment Problem. Prove that if $(m_k : k \in \mathbb{Z}_+)$ is a sequence of numbers in $[0,1]$, then there exiss a RV X with values in $[0,1]$ such that $\mathrm{E}(X^k)$

∎

Weak Convergence for $[0, \infty)$

18.7 Using Laplace transforms instead of CF's. Suppose that F and G are DF's on $\mathbb{R}_+$ such that $F(0-) = G(0-)$ and

$$\int_{[0,\infty)} e^{-\lambda x} \, dF(x) = \int_{[0,\infty)} e^{-\lambda x} \, dG(x), \qquad \forall x \geq 0 \tag{331}$$

Note that the integral on the LHS has a contribution $F(0)$ from $\{0\}$. Prove that $F = G$. Suppose that (F_n) is a sequence of distribution functions on $\mathbb{R}$ each with $F_n(0-) = 0$ and such that

$$L(\lambda) = \lim_n \int e^{-\lambda x} \, dF_n(x) \tag{332}$$

exists for $\lambda \geq 0$ and that L is continuous at 0. Prove that F_n is tight and that

$$F_n \xrightarrow{w} F \text{ where } \int e^{-\lambda x} \, dF(x) = L(\lambda), \qquad \forall \lambda \geq 0 \tag{333}$$

Let $U = e^{-X}$ and $V = e^{-Y}$. Since the range of X, Y is $[0, \infty)$ the range of U, V is $[0,1]$. From (331)

$$\mathrm{E}\, U^k = \mathrm{E}\, e^{-kX} = \mathrm{E}\, e^{-kY} = \mathrm{E}\, V^k \tag{334}$$

Hence by 10.3 the distributions of U and V are equal. However, since e^x is monotonic and invertible, this means that the distributions of X and Y are equal. More concretely, the relationship $\Pr(U \geq c) = \Pr(X \leq -\log c)$ provides a dictionary to translate between the distributions of X and U, and similarly for Y and V.

Given $\epsilon > 0$ choose λ small enough so that $L(\lambda) > 1 - \epsilon/4$. This is possible because L is continuous at 0. Thus, because L is the limit of the Laplace transforms of the F_n, for all but finitely many n we must have

$$\int_{[0,\infty)} (1 - e^{-\lambda x}) \, dF_n(x) < \epsilon/2 \tag{335}$$

Thus for any K, for all but finitely many n

$$(1 - F_n(K))(1 - e^{-\lambda K}) \leq \int_{[0,\infty)} (1 - e^{-\lambda x})\, dF_n(x) < \epsilon/2 \tag{336}$$

Choose K large enough so that $1 - e^{-\lambda K} > \frac{1}{2}$. Thus we get, for all but finitely many n

$$F_n(K) > 1 - \epsilon \tag{337}$$

which shows that F_n is tight.

The rest of the proof is basically the same as 18.4. Having proved tightness, Hally-Bray implies that some subsequence converges in distribution to some distribution, say F. Any distribution which is the limit of a subsequence of F_n has the Laplace transform L (since the limit of a subsequences are the same as the limit of the sequence). By the argument above, there is a unique distribution with a given Laplace transform, Thus, the limit of a subsequence of F_n, when it exists, converges to a unique distribution regardless of the subsequence. This implies that $F_n \xrightarrow{w} F$ since, for any continuous function g, we must have $\limsup \int g\, dF_n = \liminf \int g\, dF_n = \int g\, dF$, since otherwise we could use Helly-Bray to find a subsequence which converged to a distribution other than F.

■

A13.1 Modes of convergence.

(a) Show that $(X_n \to X,\ \text{a.s.}) \Rightarrow (X_n \to X \text{ in prob})$

(b) Prove that $(X_n \to X \text{ in prob}) \not\Rightarrow (X_n \to X,\ \text{a.s.})$

(c) Prove that if $\sum \Pr(|X_n - X| > \epsilon) < \infty, \forall \epsilon > 0$ then $X_n \to X$, a.s.

(d) Suppose that $X_n \to X$ in probability. Prove that a subsequence (X_{n_k}) of (X_n) converges $X_{n_k} \to X$, a.s.

(e) Deduce from (a) and (d) that $X_n \to X$ in probability if and only if every subsequence of (X_n) contains a further subsequence which converges a.s. to X

(a) Suppose there is an event $E \subset \Omega$ with $\Pr(E) > 0$ and for some $\epsilon > 0$, $|X_n - X| > \epsilon$ infinitely often for $\omega \in \Omega$. Then for every $\omega \in E$, it can't be that $X_n \to X$ since either $X_n > X + \epsilon$ infinitely often or $\limsup_n X_n < X - \epsilon$ infinitely often, and hence either the lim inf or lim sup is different from X. Thus $\Pr(X_n \to X) \leq 1 - \Pr(E) < 1$.

(b) Let $\Omega = [0, 1]$, taking Pr as the Lesbegue measure on the Borel sets. Let

$$X_{n,k}(x) = \begin{cases} 1 & \text{if } \frac{k}{n} \leq x < \frac{k+1}{n} \\ 0 & \text{otherwise} \end{cases} \tag{338}$$

Then take the $X_{n,k}$ as a sequence taking the indices in lexicographic order. That is, take the n's in order, and for a given n take the k's in order. Now for $n \geq N$, and $\epsilon \in (0, 1)$

$$\Pr(|X_{n,k} - 0| > \epsilon) = \Pr(X_{n,k} = 1) = \frac{1}{n} \leq \frac{1}{N} \tag{339}$$

Thus $X_{n,k} \to 0$ in probability. On the other hand, $\liminf X_{n,k}(x) = 1$ for any $x \in [0,1)$. That's because for any n, let $k_n(x) = \lfloor n/x \rfloor$. Then $X_{n,k_n(x)}(x) = 1$, and $X_{n,k}(x) = 1$ infinitel often. Therefore it can't be that $X_{n,k} \to X$ almost sure (in fact, almost surely, the sequence doesn't converge).

A second solution is given by considering I_{E_n} for independent E_n where $\Pr(E_n) = p_n$. Note

$$\Pr(|I_{E_m} - I_{E_n}| > \epsilon) = \Pr(E_n \Delta E_m) = (1 - p_n)p_m + p_n(1 - p_m) \tag{340}$$

which are independent. TODO finish second solution

(c) By Borel-Cantelli, $|X_n - X| \leq \epsilon$ eventually almost surely, because only finitely of the events $\{|X_n - X| > \epsilon\}$ hold. By hypothesis, this is true for each $\epsilon = 1/k$ for $k \in \mathbb{N}$, so take the intersection of these countably many almost sure sets E_k to get an almost sure set $E = \bigcap_k E_k$. For samples $\omega \in E$, for each $k \in \mathbb{N}$, there exists an $N_k(\omega) \in \mathbb{N}$ such that if $n > N_k(\omega)$ then $|X_n(\omega) - X(\omega)| \leq 1/k$. For arbitrary $\epsilon > 0$, choose $1/k < \epsilon$ and $N_\epsilon(\omega) = N_k(\omega)$ to get a corresponding statement for ϵ. This is the definition of almost sure convergence– for each ϵ almost surely there exists an N such that $|X_n - X| < \epsilon$ for $n > N$.

(d) Start with $m = 1$. Since $X_n \to X$ in probability, we can choose a subsequence $n_{1,k}$ such that $\Pr(|X_{n_k} - X| > 1) < 2^{-k}$ for $k \in \mathbb{N}$. This subsequence evidently satisfies the property $\sum \Pr(|X_{n_{1,k}} - X| > 1) = 1 < \infty$.

Now recursively choose a subsequence $n_{m,k}$ of $n_{m-1,k}$ which satisfies $\Pr(|X_{n_{m,k}} - X| > 1/m) < 2^{-k}$. This subsequence satisfies the property $\sum \Pr(|X_{n_{1,k}} - X| > 1/m) = 1 < \infty$. Therefore, for any $m' \in \mathbb{N}$ the diagonal subsquence $n_m = n_{m,m}$ is entirely within the subsequence $n_{m',k}$ except for at most finitely many terms when $m < m'$. Thus $\sum \Pr(|X_{n_m} - X| > 1/m') < \infty$ for all m'. This shows that the subsequence satisfies property (c), and therefore $X_{n_m} \to X$ almost surely.

(e) If $X_n \to X$ in probability, then for any $\epsilon > 0$, $p_n = \Pr(|X_n - X|) \to 0$. Since the sequence p_n has a limit, any subsequence p_{n_k} converges to the same limit, 0. Thus the subsequence X_{n_k} also converges in probability to X. By (d) this subsequence has a further subsequence which converges almost surely to X.

Conversely, assume the property, and choose $\epsilon > 0$. We'll show $p_n = \Pr(|X_n - X| > \epsilon)$ has the property $p_n \to 0$. Choose $\delta > 0$, $n_1 = 1$, and recursively find a subsequence n_k by choosing $p_{n_k} + \delta > \sup_{m > n_{k-1}} p_m$. By the property, p_{n_k} has a subsequence n_l for which X_{n_l} converges almost surely to X. By (a), X_{n_l} converges in in probability, so $p_{n_l} \to 0$. Since n_l is a subsequence of n_k is has the property that $p_{n_l} + \delta > p_m$ for any $m > n_l$. Taking the limit $l \to \infty$, this shows that $\lim p_m < \delta$. Since δ is arbitrary, this shows $p_m \to 0$. Since ϵ is arbitrary, this shows $X_n \to X$ in probability.

∎

A13.2 If $\xi \sim \mathcal{N}(0,1)$, show $\mathrm{E}\, e^{\lambda\xi} = \exp(-\frac{1}{2}\lambda^2)$. Suppose $\xi_1, \xi_2, \ldots$ are i.i.d. RV's each with a $\mathcal{N}(0,1)$ distribution. Let $S_n = \sum_{i=1}^{n} \xi_i$ and let $a, b \in \mathbb{R}$ and define

$$X_n = \exp(aS_n - bn) \tag{341}$$

Show

$$(X_n \to 0 \text{ a.s.}) \Leftrightarrow (b > 0) \tag{342}$$

but that for $r \geq 1$

$$(X_n \to 0 \text{ in } \mathcal{L}^r) \Leftrightarrow (r < 2b/a^2) \tag{343}$$

Calculating

$$\begin{aligned}\mathrm{E}\, e^{\lambda\xi} &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\lambda\xi} e^{-\frac{1}{2}\xi^2}\, d\xi = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(\xi-\lambda)^2 - \frac{1}{2}\lambda^2}\, d\xi = e^{-\frac{1}{2}\lambda^2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}\zeta^2}\, d\zeta \\ &= e^{-\frac{1}{2}\lambda^2}\end{aligned} \tag{344}$$

where we performed the substitution $\zeta = \xi - \lambda$ ■

Suppose $b < 0$. By the SLLN, $S_n/n \to 0$ as $n \to \infty$ so, almost surely. Thus for each element of the sample space ω there is an $N(\omega)$ such that $\frac{S_n}{n} - b < \frac{1}{2}b$ for all $n > N$. In this case $\exp(S_n - bn) < \exp(-\frac{1}{2}bn) \to 0$. Now suppose $b > 0$. The similarly there is an $N(\omega)$ such that $\frac{S_n}{n} > \frac{1}{2}b$ and therefore $X_n > \exp(\frac{1}{2}bn) \to \infty$. For $b = 0$, the law of the iterated logarithm says that $S_n > n \log\log n$ infinitely often. Therefore $e^{aS_n} > (\log n)^{an}$ infinitely often, and hence does not converge to 0. Since $-S_n$ is also the sum of i.i.d. $\mathcal{N}(0,1)$ RV's, $e^{aS_n} < \frac{1}{(\log n)^{an}}$ infinitely often. Thus the limit of e^{aS_n} doesn't exist, it oscillates infinitely often reaching arbitrarily high and low positive values.

On the other hand

$$\mathrm{E}\, X_n^r = \mathrm{E}\, e^{r(aS_n - bn)} = e^{-rbn} (\mathrm{E}\, e^{ra\xi})^n = e^{(\frac{1}{2}(ra)^2 - br)n} \tag{345}$$

If $r < 2b/a^2$ then $\|X_n\|_r \to 0$, but if $r \geq 2b/a^2$ its not true. If $2b/a^2$ then $\|X_n\|_r \to \infty$ and if $r = 2b/a^2$ then $\|X_n\|_r \to 1$