Homework Solutions

Lemma 3.4. Let $\mathcal{F}$ denote the intersection $\cap_\alpha \mathcal{F}_\alpha$. Since each $\mathcal{F}_\alpha$ is a σ-algebra, we have that $\varnothing \in \mathcal{F}_\alpha$ for all α. Thus, $\varnothing \in \mathcal{F}$ as well (because it is the intersection). Similarly, if $A \in \mathcal{F}$, then $A \in \mathcal{F}_\alpha$ for all α and $A^c \in \mathcal{F}_\alpha$ for all α and hence $A^c \in \mathcal{F}$. The same reasoning works for $\cup_n A_n$ where $A_n \in \mathcal{F}$ for all $n \geq 1$. Indeed, $A_n \in \mathcal{F}_\alpha$ for all α (and all $n \geq 1$) and hence $\cup_n A_n \in \mathcal{F}_\alpha$ for all α and consequently $\cup_n A_n \in \mathcal{F}$.

Lemma 3.11(i). If A_n are increasing then let $B_n = A_n \setminus A_{n-1}$. These are disjoint sets and $A_N = \cup_{n=1}^N B_n$. But then

$$\mu(A_N) = \mu(\cup_{n=1}^N B_n) = \sum_{n=1}^N \mu(B_n).$$

Also, note that $\cup_n A_n = \cup_n B_n$. Taking $n \to \infty$ and using σ-additivity shows

$$\lim_{N\to\infty} \mu(A_N) = \sum_{n=1}^\infty \mu(B_n) = \mu(\cup_n B_n) = \mu(\cup_n A_n).$$

Conversely, if say we did not know we have σ-additivity but knew we had additivity and continuity under increasing unions. Then consider B_n disjoint. Sets $A_N = \cup_{n=1}^N B_n$ are increasing to $\cup_n B_n$. We now have

$$\mu(\cup_n B_n) = \mu(\cup_n A_n) = \lim_{N\to\infty} \mu(A_N) = \lim_{N\to\infty} \sum_{n=1}^N \mu(B_n) = \sum_{n=1}^\infty \mu(B_n).$$

The second-last equality comes from the assumed additivity.

Lemma 3.11(ii). This follows from the above by using complements. However, one issue is that if $\mu(A_n) = \infty$ then we cannot deduce what $\mu(A_n^c)$ is. Hence the need for finiteness for this part of the argument. Namely, that if μ is finite and we have additivity and continuity under decreasing intersections, then we have σ-additivity.

Lemma 3.9. $\mu(A \cup B) = \mu(A) + \mu(B \setminus A)$ because of additivity and the two sets on the right-hand side being disjoint and their union making up the set on the left-hand side. Similarly, $\mu(B) = \mu(B \setminus A) + \mu(A \cap B) \geq \mu(B \setminus A)$. Together these give

$$\mu(A \cup B) \leq \mu(A) + \mu(B).$$

(In fact, from the above we get that if $\mu(A \cap B) < \infty$ then $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$. Of course, if it is infinite then $\mu(A \cup B) = \infty$ too.)

Induction gives

$$\mu(\cup_{n=1}^N A_n) \leq \sum_{n=1}^N \mu(A_n).$$

Since $\cup_{n=1}^N A_n$ increases to $\cup_n A_n$, a σ-additive μ would give

$$\mu(\cup_n A_n) = \lim_{n\to\infty} \mu(\cup_{n=1}^N A_n) \leq \lim_{n\to\infty} \sum_{n=1}^N \mu(A_n) = \sum_n \mu(A_n).$$

Borel σ-algebra is generated by intervals: Let $\mathcal{B}$ be the Borel σ-algebra. Let $\mathcal{F}$ be the σ-algebra generated by finite unions of intervals of the form $(a, b]$. Since finite unions of intervals are clearly $\mathcal{B}$-measurable we know that $\mathcal{B}$ must contain $\mathcal{F}$. On the other hand, any interval $(a, b]$ is an intersection of intervals $(a, b - 1/n)$, which are all open. Hence, intervals of the form (a, b) are all in $\mathcal{F}$. But any open set can be written as a (countable) union of such intervals. Hence, open sets are all in $\mathcal{F}$ and so is the σ-algebra generated by them. But this is $\mathcal{B}$. In other words we also have $\mathcal{B} \subset \mathcal{F}$.

3.2. Let $\Omega = (0, 1]$. For $n \geq 1$ let $\mathcal{F}_n$ be made up of $\varnothing$ and any unions of intervals of the form $(i/2^n, (i+1)/2^n]$, with $i \in \{0, 1, \ldots, 2^n - 1\}$. This is an algebra (CHECK!). But since it is made up of finitely many sets, it is also a σ-algebra. The union of $\mathcal{F}_n$ clearly does not have all open sets, since not all open sets are in SOME $\mathcal{F}_n$. On the other hand, any interval of the form (a, b) is a union of intervals $((i_n + 1)/2^n, j_n/2^n]$, where i_n is the largest integer such that $i/2^n < a \leq (i + 1)/2^n$ and j_n is the smallest integer such that $j_n/2^n < b \leq (j_n + 1)/2^n$. Note that $(i_n + 1)/2^n \leq 1/2_n$ and $b - j_n/2^n \leq 1/2_n$. Furthermore, i_n is nondecreasing and j_n nonincreasing (CHECK!). Therefore, interval $((i_n + 1)/2_n, j_n/2^n]$ (which is in $\mathcal{F}_n$) increases to (a, b). This proves that intervals (a, b) are all in the σ-algebra generated by $\cup_n \mathcal{F}_n$. But then so are open sets. Hence, the σ-algebra generated by $\cup_n \mathcal{F}_n$ is not $\cup_n \mathcal{F}_n$ itself. It is bigger. In fact, the proof shows that it is exactly the Borel σ-algebra on $(0, 1]$.

An easier example: Let $\Omega = \mathbb{N}$ and let $\mathcal{F}_n$ be the collection of subsets of $\{1, \ldots, n\}$ and their complements in $\mathbb{N}$ (i.e. subsets of the form $A \cup \{n +$

$1, n+2, \dots\}$ where A is a subset of $\{1, \dots, n\}$. Since $\mathcal{F}_n$ is an algebra (CHECK!) with finitely many sets in it, it is also a σ-algebra. This sequence is increasing, and the σ-algebra generated by the union must have in it the set of even numbers (because that is the countable union of sets $\{2n\} \in \mathcal{F}_{2n}$). But the set of all even numbers is not in any of $\mathcal{F}_n$. Hence the union is not the same as the σ-algebra generated by the union.

3.9. If $F(a) = \mu((-\infty, a])$ then $F(a)$ is nondecreasing because if $a < b$ then $(-\infty, a] \subset (-\infty, b]$ and thus the μ-measures are ordered the same way and as a result $F(a) \leq F(b)$. It also increases to 1 as $a \to \infty$ and decreases to 0 as $a \to -\infty$ and is right-continuous, all for the same reason: σ-additivity of μ implies its continuity under monotone limit (decreasing limits are OK because μ is a probability measure and hence is finite). Right-continuity comes from the fact that if a_n decreases to a then $(-\infty, a_n]$ decrease to $(-\infty, a]$ and hence the measures converge, in other words $F(a_n)$ converges to $F(a)$. Convergence to 1 and to 0 comes from the fact that $(-\infty, a]$ decrease to $\varnothing$ as $a \to -\infty$ and increase to all of $\mathbb{R}$ as $a \to \infty$.

Conversely, if $F(a)$ is a CDF, then we can define $\mu((a, b]) = F(b) - F(a)$. We can then define the measure of a finite intersection of such intervals. This defines an additive (probability) measure on an algebra. We can extend this measure uniquely to the σ-algebra generated by this algebra (which is nothing other than the Borel σ-algebra) by means of Carathéodory's theorem. But for this we need to show that μ is σ-additive on the algebra. This can be shown in exactly the same way as was shown for the Lebesgue measure, except that this time we use $F(b) - F(a)$ instead of $b - a$ to measure an interval. One thing that was needed in that construction is that if a is given, then we can find a $b > a$ close to a to make $F(b) - F(a)$ as small as we wish. This is where right-continuity comes in. The responsible student should fill in the details (e.g. adapting the proof of the Lebesgue measure case).

3.10. First, let us show that it makes sense to talk about $a + A$ and cA. Namely, if $a, c \in \mathbb{R}$ are numbers, then the functions $x \mapsto x - a$ and $x \mapsto x/c$ are continuous. We have seen in class (albeit in lectures pertaining to chapter 4!) that these functions are then measurable. In particular, if A is a Borel set, then $a + A$ (which is the inverse image of A under $x \mapsto x - a$) and cA (inverse image of $x \mapsto x/c$) are also Borel sets. To do this more directly one can consider $\mathcal{C} = \{A \in \mathcal{B} : a + A \in \mathcal{B}\}$ and observe that this clearly contains open sets and is a monotone class. Hence it contains $\mathcal{B}$ (the

σ-algebra generated by the open sets), but then this means it equals $\mathcal{B}$. In other words, for any $A \in \mathcal{B}$ we have $a + A \in \mathcal{B}$. The same can be done for cA.

Now that we have established that shifting and scaling leave the sets measurable we want to prove that the Lebesgue measure is invariant under shifts. For this, fix an $a \in \mathbb{R}$ and define the measure μ that assigns values $\mu(A) = m(a+A)$. Here, m is Lebesgue measure. Note that if A is an interval then $\mu(A)$ is in fact the same thing as $m(A)$ (both are equal to the length of the interval). So μ and m coincide on intervals and hence also on finite disjoint unions of intervals. Because both measures are clearly σ-finite, we deduce that they must match on the whole Borel σ-algebra (by the uniqueness part in Carathéodory's theorem). This says that $m(a + A) = m(A)$, which is what we wanted to prove.

For the scale invariance we do the same as above: define a measure $\nu(A) = \alpha m_1(\alpha^{-1}A)$ on the Borel σ-algebra of $[0, \alpha]$. We want to show that $\nu = m_\alpha$. But for this, it is enough to check that the two measures match on intervals. This in turn is pretty clear: $\nu((a, b]) = \alpha m_1((a/\alpha, b/\alpha]) = b - a = m_\alpha((a, b])$. By the uniqueness of the extension we deduce that the two measures actually match on the whole $\mathcal{B}$.

3.11. First, let us assume that $\mu((\alpha, \beta]) < \infty$ for some $\alpha < \beta$. By shift-invariance we have $\mu((\alpha, \beta]) = \mu((0, \beta - \alpha])$. So we can assume $\alpha = 0$ (and rename $\beta - \alpha > 0$ as $\beta > 0$). Abbreviate $f(t) = \mu((0, t])$. Notice that for $t, s > 0$

$$f(t+s) = \mu((0, t+s]) = \mu((0, t])+\mu((t, t+s]) = \mu((0, t])+\mu((0, s]) = f(t)+f(s).$$

The first equation came by additivity and the second by the assumed shift-invariance of μ. By induction we have that for any $n \geq 1$ and $s > 0$ we have $f(ns) = nf(s)$. One consequence of this is that $\mu((0, n\beta]) = n\mu((0, \beta]) < \infty$ for all n. Hence, for any $a < b$ we have $\mu((a, b]) = \mu((0, b-a]) \leq \mu((0, n\beta]) < \infty$, where we choose n large enough to have $b - a \leq n\beta$. As a result, we see that we could in fact choose $\beta = 1$.

Substituting s in the above display with s/n we get $f(s) = nf(s/n)$ or equivalently $f(s/n) = (1/n)f(s)$. These last two facts say that if $t = m/n$ with $m, n \in \mathbb{N}$, and if $s = 1$, then $f(t) = f(m/n) = mf(1/n) = mf(1)/n = tf(1)$. Furthermore, if $t > 0$ is irrational, then we can find a rational sequence

t_n decreasing to t. Then $(0, t_n]$ decreases to $(0, t]$ and since we have shown that $\mu((a, b]) < \infty$ for all $a < b$, we have that $\mu((0, t + 1]) < \infty$ and σ-additivity of μ implies that $\mu((0, t_n])$ decreases to $\mu((0, t])$. Thus

$$f(t) = \lim_{n\to\infty} f(t_n) = \lim_{n\to\infty} t_n f(1) = tf(1).$$

We have thus shown that

$$\mu((0, t]) = \mu((0, 1])t, \quad \text{for all } t > 0.$$

Using shift-invariance of μ and that $m((a, b]) = b - a$ we have

$$\mu((a, b]) = \mu((0, b - a]) = \mu((0, 1])(b - a) = \mu((0, 1])m((a, b]), \quad \text{for all } b > a.$$

This says that the measures μ and $\mu((0, 1])m$ match on semi-open intervals. Since they are both σ-additive and σ-finite, uniqueness in Carathéodory's theorem says they must match on the whole Borel σ-algebra. This answers the question with $c = 1/\mu((0, 1])$. Note that all this makes sense even if $\mu((0, 1]) = 0$. In this is the case, then $c = \infty$ and μ assigns 0 to everything.

It remains to dismiss the case when $\mu((a, b]) = \infty$ for all $a < b$. It may seem at first that σ-finiteness alone should suffice. However, consider the measure ν that assigns to a Borel set A the number of rationals in that set. You can quickly check that this is a σ-additive nonnegative measure. It is also σ-finite because we can number the rationals as say $\{q_n : n \geq 1\}$ and then consider the events $A_n = \Big((-n, n) \setminus \mathbb{Q}\Big) \cup \{q_1, \ldots, q_n\}$. It is clear that A_n increases to $\mathbb{R}$ and that $\nu(A_n) = n < \infty$. But even though measure ν is σ-additive, it assigns an infinite value to all intervals. It is not shift-invariant, though. So we must use shift-invariance crucially in our argument.

The actual argument requires knowing a bit more probability, namely knowing about product measures and the Radon-Nikodým derivative. We will thus revisit this exercise when we learn more about these.

3.13. Define $\mathrm{supp}(\mu)$, the support of μ, to be the intersection of all closed sets C such that $\mu(C^c) = 0$. Since the whole space Ω is one such set, the intersection is not empty. Since any intersection of closed sets is always a closed set, the support $\mathrm{supp}(\mu)$ is a closed set and is hence measurable. Next, note that $x \notin \mathrm{supp}(\mu)$ is equivalent to saying that x is in the union of all C^c

such that $\mu(C^c) = 0$. In other words, this is equivalent to the existence of an open set U such that $\mu(U) = 0$.

Comment: The problem in the textbook asks for a bit more than the above. It says that the support of μ is THE smallest closed set whose complement has measure 0. That is, it is asking us to also show that the measure of the complement of the support is 0. This is not a trivial matter. Let us denote the complement of the support by A. Then we know that A is the union of all open sets U with $\mu(U) = 0$. But since this union can be uncountable, in a general topological space, we cannot deduce that $\mu(A) = 0$. (And there are in fact counter-examples.) However, if the space is also second-countable (e.g. metric and separable), then there is a countable collection O_n of open sets such that each open set U can be written as a countable union of a subcollection of such sets. If U happens to also satisfy $\mu(U) = 0$, then every O_n in the subcollection will also have $\mu(O_n)$ (because O_n is inside of U). This shows that A can in fact be written as a countable union of of open sets O_n with $\mu(O_n) = 0$ for each n. Now this then guarantees that $\mu(A) = 0$.

3.14. Let $A_m \in \mathcal{M}$ be increasing. Since $A_m \in \mathcal{M}$ we have for each n a set $B_{m,n} \in \mathcal{A}$ such that $\mu(A_m \Delta B_{m,n}) \leq 1/(2^m n)$. For a given n let $B = \cup_m B_{m,n}$. Write

$$\begin{aligned}\mu(A\Delta B) &= \mu(\cup_m A_m \Delta \cup_m B_{m,n}) \leq \mu\big(\cup_m (A_m \Delta B_{m,n})\big)\\ &\leq \sum_{m\geq 1} \mu(A_m \Delta B_{m,n}) \leq \frac{1}{n}\sum_{m=1}^{\infty}\frac{1}{2^m} = \frac{1}{n}.\end{aligned}$$

It seems like we have thus shown that $A \in \mathcal{M}$. In other words, $\mathcal{M}$ is closed under increasing unions. One caveat, though: we have not shown that $B \in \mathcal{A}$. This would be the case if $\mathcal{A}$ were a σ-algebra. But $\mathcal{A}$ is a monotone class, and so it is only closed under increasing unions, not any arbitrary unions. So it is not clear that the way we chose B keeps it in $\mathcal{A}$. Also, we have not used the fact that A_m's are increasing.

So we will proceed a little differently from above. First, let $A = \cup_{m\geq 1} A_m$. We want to show that $A \in \mathcal{M}$. So fix any n. We want to find a $B \in \mathcal{A}$ such that $\mu(A\Delta B) \leq 1/n$. We have that $\mu(A_m)$ increases to $\mu(A)$. So there exists an m such that $\mu(A \setminus A_m) = \mu(A) - \mu(A_m) \leq 1/(2n)$. And since A_m itself is in $\mathcal{M}$, there exists a $B \in \mathcal{A}$ such that $\mu(A_m \Delta B) \leq 1/(2n)$. But now $A\Delta B \subset (A_m \Delta B) \cup (A \setminus A_m)$ (CHECK!). Hence

$$\mu(A\Delta B) \leq \mu(A_m \Delta B) + \mu(A \setminus A_m) \leq 1/(2n) + 1/(2n) = 1/n,$$

as desired. Closure under decreasing intersections works similarly (but the student should CHECK).

Note that this time around, we did not use the fact that $\mathcal{A}$ is a monotone class! $\mathcal{A}$ could be any collection of measurable sets, and $\mathcal{M}$ would still be a monotone class. The point of the exercise is that if you have a collection of sets $\mathcal{A}_0$ with which you can approximate the measure of sets in an algebra $\mathcal{A}$, in the sense that for each $A \in \mathcal{A}$ and each $\varepsilon > 0$ there exists a $B \in \mathcal{A}_0$ such that $\mu(A\Delta B) < \varepsilon$, then you can also approximate the sets $A \in \sigma(\mathcal{A})$ by sets $B \in \mathcal{A}_0$, in that same sense.

3.17. (1) is straightforward. (2) is a direct use of the axiom of choice: you have a family of nonempty sets and you want to choose one representative from each set. For (3) say that the intersection is not empty. Then there exist $s, t \in (0, 1]$ such that $e^{2\pi is}, e^{2\pi it} \in \Lambda$, and $e^{2\pi i\alpha}e^{2\pi is} = e^{2\pi i\beta}e^{2\pi it}$. But then $e^{2\pi i(t-s)} = e^{2\pi i(\alpha-\beta)}$. Since $\alpha - \beta$ is rational there exists an integer n such that $n(\alpha - \beta)$ is a whole number. Then $1 = e^{2\pi in(\alpha-\beta)} = e^{2\pi in(t-s)}$. This implies that $n(t - s)$ is a whole number and $t - s$ is rational. But then $e^{2\pi it} \sim e^{2\pi is}$ and they belong to the same equivalence class. This forces $t = s$ for otherwise $e^{2\pi is}$ and $e^{2\pi it}$ cannot represent two distinct elements in Λ. But now if $t = s$, then we also have $\alpha = \beta$, which was assumed not to be the case.

(5) really follows from (4), for if Λ were measurable, then we would be able to compute $\mu(\Lambda)$. To show (4) let us first prove the statement in the hint: take element e^{it} with $t \in (0, 2\pi]$. Let $s \in \Lambda$ be the representative of the equivalence class of t. Set $\alpha = t - s$. Since $t \sim s$, we have $\alpha \in \mathbb{Q}$. But then $e^{it} = e^{i\alpha}e^{is}$, i.e. $e^{it} \in \Lambda_\alpha$, and the claim in the hint is proved. (One little detail: it could be that $t < s$ and thus α is negative. In this case, use $\alpha' = 2\pi + \alpha$, which is then in $(0, 2\pi]$ and we have $e^{it} = e^{i\alpha'}e^{is} \in \Lambda_{\alpha'}$.)

Next, we have just shown that Λ_α are all disjoint. Furthermore, we have already shown that Lebesgue measure is shift-invariant and since Λ_α is essentially a shift of Λ we have $\mu(\Lambda_\alpha) = \mu(\Lambda)$ for all α. (This actually needs proof, but the proof goes along the same lines as that of problem 3.10.) But then σ-additivity says that

$$2\pi = \mu(S^1) = \sum_{\alpha\in(0,2\pi]\cap\mathbb{Q}} \mu(\Lambda_\alpha) = \sum_{\alpha\in(0,2\pi]\cap\mathbb{Q}} \mu(\Lambda).$$

Either $\mu(\Lambda) = 0$ and then we would have $2\pi = 0$, or $\mu(\Lambda) > 0$ and then $2\pi = \infty$. Therefore, it must be the case that $\mu(\Lambda)$ is not defined, which means Λ is not measurable.

4.1 That the collection is a σ-algebra follows by checking all the necessary properties in the definition of a σ-algebra (and using the fact that $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$ and $X^{-1}(A^c) = (X^{-1}(A))^c$). That it is the smallest σ-algebra on which X is measurable is also clear: X is clearly measurable with respect to $\sigma(X)$ (because for every $A \in \mathcal{A}$ we have $B \in \sigma(X)$ by the definition of $\sigma(X)$!) and if X is measurable with respect to some σ-algebra $\mathcal{F}$, then for every $A \in \mathcal{A}$ we must have $X^{-1}(A) \in \mathcal{F}$, but this says that $\sigma(X) \subset \mathcal{F}$!

4.6. First observe that we can approximate $\mathbb{1}_{(a,b]}$ by a sequence of continuous functions, e.g. $f_n(x) = 0$ for $x \le a$ or $x \ge b + 1/n$, $f_n(x) = 1$ for $a + 1/n \le x \le b$, and interpolating linearly on $(a, a + 1/n)$ and $(b, b + 1/n)$. Next, we can approximate the indicator function of $A = \prod_{i=1}^d (a_i, b_i]$ by functions $f_n(x_1, \ldots, x_d) = \prod_{i=1}^d f_{n,i}(x_i)$, where $f_{n,i}$ approximates $(a_i, b_i]$. We can thus approximate indicators of finite disjoint unions of such boxes, by continuous functions. We have that

$$\begin{aligned}
\mu(|f_n - \mathbb{1}_A| > \varepsilon) \le \mu(f_n \ne \mathbb{1}_A) &\le \sum_{i=1}^d \mu(f_{n,i} \ne \mathbb{1}_{(a_i,b_i]}) \\
&\le \sum_{i=1}^d \big(\mu([a, a+1/n]) + \mu([b, b+1/n])\big) \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

(The last limit is due to σ-additivity.) We have that for each n, $\int f_n \, d\mu = \int f_n \, d\nu$. Take $n \to \infty$. The approximating functions are all bounded and so bounded convergence allows us to slip the limit under the integrals and we get that μ and ν match on finite unions of semi-open boxes. These form an algebra and since both measures are assumed to be σ-finite, the uniqueness in Carathéodory's theorem implies the two measures must be equal.

Counter-example to Fatou. Consider Lebesgue measure on $[0, 1]$. Let $f_n = -n\mathbb{1}_{(0,\frac{1}{n}]}$. Then $\underline{\lim}\, f_n(x) = 0$ for all x. But $\int f_n(x)\, dx = -1$ for all n. Fatou's lemma fails here, since it says that $\int 0 \, dx \le -1$, which is false.

4.16. Say $X_n \to X$ in probability. If we show that there exists a subsequence that converges almost surely, then we can repeat this argument starting with any subsequence, to show that that subsequence has a further subsequence converging almost surely.

For a given m we have that $P(|X_n - X| > 1/m) \to 0$ as $n \to \infty$. Thus, there exists an n_m such that $P(|X_{n_m} - X| > 1/m) < 1/2^m$. Now we claim that with probability one there exists an m_0 such that $|X_{n_m} - X| \le 1/m$ for all $m \ge m_0$. This then would imply that $X_{n_m} \to X$ almost surely. To prove our claim we instead show that the probability of its complement is zero. For this we write

$$
\begin{aligned}
&P\Big(\forall m_0 \ge 1\, \exists m \ge m_0 : |X_{n_m} - X| > 1/m\Big) = P\Big(\bigcap_{m_0 \ge 1} \bigcup_{m \ge m_0} \{|X_{n_m} - X| > 1/m\}\Big) \\
&\quad = \lim_{m_0 \to \infty} P\Big(\bigcup_{m \ge m_0} \{|X_{n_m} - X| > 1/m\}\Big) \le \lim_{m_0 \to \infty} \sum_{m \ge m_0} P(|X_{n_m} - X| > 1/m) \\
&\quad \le \lim_{m_0 \to \infty} \sum_{m \ge m_0} \frac{1}{2^m} = 0.
\end{aligned}
$$

4.18. There are several ways to do this. I will do it in several steps. First, observe that any function f can be written as $f = f^+ - f^-$ and so it is enough to approximate nonnegative functions. Second, notice that if $f \ge 0$ is in L^p and we define $f_n = f \wedge n$ then

$$\int |f - f_n|^p \, d\mu = \int (f-n)^p \mathbb{1}\{f \ge n\} \, d\mu \le \int f^p \mathbb{1}\{f \ge n\} \, d\mu.$$

The last integral converges to 0 by dominated convergence because $\mathbb{1}\{f \ge n\} \to 0$ and $f^p \mathbb{1}\{f \ge n\} \le f^p \in L^1$. The upshot is that we can approximate L^p functions by bounded ones and so our problem reduces to approximating (in L^p norm) nonnegative bounded functions by continuous ones. But now that the functions are bounded, we can in fact even do better.

To this end, note that we can approximate nonnegative bounded functions pointwise by simple functions. Indeed, if $f \le 0$ is bounded by M then for a given n and $0 \le i \le nM$ $A_i = f^{-1}([i/n, (i+1)/n))$. This is a measurable set and if we let

$$f_n = \sum_{0 \le i \le nM} \frac{i}{n} \mathbb{1}_{A_i}$$

then it is immediate that $0 \le f - f_n \le 1/n$ everywhere. In other words, $f_n \to f$ pointwise and f_n is a simple function.

Now, it suffices to approximate (in L^p norm) indicator functions $\mathbb{1}_A$ by continuous functions. (CHECK that this is enough to deduce that any simple function can be approximated by continuous functions.) If set A is a box of the form $\prod_{j=1}^d (a_j, b_j]$, then we have seen in Problem 4.6 how to approximate $\mathbb{1}_A$ pointwise with a continuous function and then dominated convergence implies that this approximation also works in L^p. This approximation also works for a disjoint union of such boxes, which forms an algebra. For a general measurable set A we want to avoid a messy construction of an approximation and so we rely on an abstract theorem! Let

$$\mathcal{C} = \{A \in \mathcal{B}([0,1]^d) : \exists f_n \text{ continuous, such that } \|f_n - \mathbb{1}_A\|_p \to 0\}.$$

We claim that this is a monotone class. If so, then we just argued that it contains an algebra that generates all of the Borel σ-algebra. It thus must equal the Borel σ-algebra, which means that we can approximate any indicator function by continuous functions.

To show that $\mathcal{C}$ is a monotone class consider an increasing sequence A_n of measurable sets from $\mathcal{C}$. Then for each n there exists a continuous function f_n such that $\|f_n - \mathbb{1}_{A_n}\|_p \leq 1/n$. Now let $A_n = \cup_n A$ and write

$$\|f_n - \mathbb{1}_A\|_p \leq \|f_n - \mathbb{1}_{A_n}\|_p + \|\mathbb{1}_{A_n} - \mathbb{1}_A\|_p \leq 1/n + \|\mathbb{1}_{A \setminus A_n}\|_p = 1/n + \mu(A \setminus A_n)^{1/p} \underset{n\to\infty}{\longrightarrow} 0.$$

This shows that the sequence of continuous functions f_n approximates $\mathbb{1}_A$ in L^p norm and thus $\mathcal{C}$ is closed under increasing unions. Closure under decreasing intersections comes similarly (pretty much word for word, but do CHECK!).

To recap: we have shown that indicator functions can be approximated by continuous functions, therefore simple functions have the same property. But bounded functions can be approximated by simple functions and thus by continuous functions. Nonnegative functions can be approximated by bounded ones and hence by continuous ones. And finally any function can be written as a combination of two nonnegative ones. This proves the claim of the problem.

4.28. One direction is easy:

$$\sup_{m \geq n} E[|X_m|; |X_m| \geq t] \leq \sup_m E[|X_m|; |X_m| \geq t]$$

and so

$$\overline{\lim_{n\to\infty}} E[|X_n|; |X_n| \geq t] \leq \sup_m E[|X_m|; |X_m| \geq t]$$

and if the right-hand side goes to 0 as $t \to \infty$ then so does the left-hand side.

For the other direction fix $\varepsilon > 0$. We are assuming that

$$\lim_{t\to\infty} \overline{\lim_{n\to\infty}} E[|X_n|; |X_n| \geq t] = 0.$$

This implies that we can find t_0 such that

$$\overline{\lim_{n\to\infty}} E[|X_n|; |X_n| \geq t_0] < \varepsilon.$$

This in turn implies that we can find n_0 such that

$$\sup_{m \geq n_0} E[|X_m|; |X_m| \geq t_0] < \varepsilon.$$

Now observe that if $t \geq t_0$ then $E[|X_m|; |X_m| \geq t] \leq E[|X_m|; |X_m| \geq t_0]$. Therefore, we actually have

$$\sup_{m \geq n_0} E[|X_m|; |X_m| \geq t] < \varepsilon$$

for all $t \geq t_0$. On the other hand, for any fixed m we have that

$$E[|X_m|; |X_m| \geq t] \to 0$$

as $t \to \infty$ by dominated convergence (since $X_m \in L^1$). Therefore, we can find t_1 such that for $t \geq t_1$ we get that $E[|X_m|; |X_m| \geq t] < \varepsilon$. We can do this for every $m < n_0$ (as long as n_0 is fixed and we do not have to do this infinitely many times!). This then implies that for $t \geq t_1$ we have

$$\sup_{m < n_0} E[|X_m|; |X_m| \geq t] < \varepsilon.$$

But now we have shown that the supremum for $m \geq n_0$ is below ε and the supremum for $m < n_0$ is below ε. This simply means that

$$\sup_m E[|X_m|; |X_m| \geq t] < \varepsilon.$$

And we have shown that this holds for any t that is larger than t_0 (for the supremum over $m \geq n_0$ to be below ε) and larger than t_1 (for the supremum over $m < n_0$ to be $< \varepsilon$). In other words, the above holds for t greater than both t_0 and t_1, and this for any $\varepsilon > 0$ we choose. This is exactly what it means for

$$\lim_{t\to\infty} \sup_m E[|X_m|; |X_m| \geq t] = 0.$$

We prove the rest of the claims. (1) is clear because $E[|X_n|; |X_n| \geq t] \leq E[|Y_n|; |Y_n| \geq t]$. For (2) note that

$$|X_n + Y_n| \mathbb{1}\{|X_n + Y_n| \geq t\} \leq 2|X_n| \mathbb{1}\{|X_n| \geq t/2\} + 2|Y_n| \mathbb{1}\{|Y_n| \geq t/2\}.$$

The inequality is trivial when $|X_n + Y_n| < t$. On the other hand, when $|X_n + Y_n| \geq t$ then we must have that $|X_n| \geq t/2$ or $|Y_n| \geq t/2$. We can thus look at the three cases $|X_n| \geq t/2 > |Y_n|$ (the inequality then comes from $|X_n + Y_n| \leq |X_n| + |Y_n| \leq 2|X_n|$), $|Y_n| \geq t/2 > |X_n|$ (this is the symmetric case), and both $|X_n|$ and $|Y_n|$ are $\geq t/2$ (the inequality now comes from $|X_n + Y_n| \leq |X_n| + |Y_n| \leq 2|X_n| + 2|Y_n|$). Now integrate the inequality to get

$$E[|X_n + Y_n|; |X_n + Y_n| \geq t] \leq 2E[|X_n|; |X_n| \geq t/2] + 2E[|Y_n|; |Y_n| \geq t/2].$$

Claim (2) follows. For (3) note that $\mathbb{1}\{|X_n| \geq t\} \leq |X_n|^{p-1}/t^{p-1}$ to deduce that

$$E[|X_n|; |X_n| \geq t] \leq E[|X_n|^p]/t^{p-1} \leq C/t^{p-1} \to 0 \quad \text{as } t \to \infty.$$

(4) is a special case of Problem 4.29, with $p = 1$.

4.29. First, observe that if $X \in L^p$ then $|X|^p$, viewed as a sequence (that does not depend on n), is uniformly integrable. This is because $E[|X|^p; |X|^p \geq t] \to 0$ as $t \to \infty$, e.g. by dominated convergence. Furthermore, note that

$$|X_n - X|^p \leq 2^{p-1}(|X_n|^p + |X|^p) \quad \text{and} \quad |X_n|^p \leq 2^{p-1}(|X|^p + |X_n - X|^p).$$

(For both inequalities apply $(a+b)^p \leq 2^{p-1}(a^p + b^p)$, which itself comes from Jensen's inequality and convexity of x^p.) The above shows that uniform integrability of $|X_n|^p$ is equivalent to uniform integrability of $|X_n - X|^p$. Now we go on with the proof of the exercise.

First, assume $X_n \to X$ in L^p. This implies that

$$E[|X_n - X|^p; |X_n - X|^p > t] \leq E[|X_n - X|^p] \longrightarrow 0.$$

Thus we see that $|X_n - X|^p$ is uniformly integrable and from the above we get that $|X_n|^p$ is uniformly integrable.

Next, we prove the other direction: uniform integrability of $|X_n|^p$ and convergence in probability of X_n to X imply together that $X_n \to X$ in L^p. By the above paragraph we know that $|X_n - X|^p$ is uniformly integrable. Fix $\varepsilon > 0$ and $t > \varepsilon$ and write

$$
\begin{aligned}
&E[|X_n - X|^p] \\
&= E[|X_n - X|^p; |X_n - X|^p > t] + E[|X_n - X|^p; \varepsilon^p < |X_n - X|^p \le t] \\
&\qquad + E[|X_n - X|^p; |X_n - X|^p \le \varepsilon^p] \\
&\le E[|X_n - X|^p; |X_n - X|^p > t] + tP(|X_n - X| > \varepsilon) + \varepsilon^p.
\end{aligned}
$$

Taking $n \to \infty$ and using convergence in probability of X_n to X we get

$$0 \le \varliminf_{n\to\infty} E[|X_n - X|^p] \le \varlimsup_{n\to\infty} E[|X_n - X|^p] \le \varlimsup_{n\to\infty} E[|X_n - X|^p; |X_n - X|^p > t] + \varepsilon^p.$$

Taking $t \to \infty$ and $\varepsilon \to 0$ makes the right-hand side go to 0. This proves that the limit of $E[|X_n - X|^p]$ exists and equals 0, i.e. that $X_n \to X$ in L^p.

That $X_n \to X$ in L^p implies $E[|X_n|^p] \to E[|X|^p]$ comes easily from the fact that $|\|X_n\|_p - \|X\|_p| \le \|X_n - X\|_p$ (which itself comes from the triangle inequality).

It remains to show that if $X_n \to X$ in probability and $E[|X_n|^p] \to E[|X|^p]$ then $X_n \to X$ in L^p. For this we will need a small lemma.

Lemma 0.1. *If $Y \in L^1$ and A_n is a sequence of events such that $P(A_n) \to 0$ then $E[Y; A_n] \to 0$.*

Proof. Take $a_n = 1/\sqrt{P(A_n)}$. Then $a_n \to \infty$ but $a_n P(A_n) \to 0$. Now write

$$E[Y; A_n] = E[Y; A_n \cap \{Y \le a_n\}] + E[Y; A_n \cap \{Y > a_n\}] \le a_n P(A_n) + E[Y; Y > a_n].$$

The first term goes to 0 by the choice of a_n while the second one goes to 0 by dominated convergence. □

Now fix $\varepsilon > 0$ and write

$$
\begin{aligned}
&E[|X_n - X|^p] \\
&= E[|X_n - X|^p; |X_n - X| > \varepsilon] + E[|X_n - X|^p; |X_n - X| \le \varepsilon] \\
&\le 2^{p-1}E[|X_n|^p; |X_n - X| > \varepsilon] + 2^{p-1}E[|X|^p; |X_n - X| > \varepsilon] + \varepsilon^p \\
&= 2^{p-1}E[|X_n|^p] - 2^{p-1}E[|X_n|^p; |X_n - X| \le \varepsilon] + 2^{p-1}E[|X|^p; |X_n - X| > \varepsilon] + \varepsilon^p.
\end{aligned}
$$

By the mean value theorem we have that if $0 < a < b$ then

$$b^p = a^p + p\theta^{p-1}(b-a)$$

for some $\theta \in (a,b)$. Thus, $b^p \le a^p + pb^{p-1}|b-a|$. If, on the other hand $0 < b < a$ then we trivially have $b^p \le a^p + pb^{p-1}|b-a|$. This implies that if $|X_n - X| \le \varepsilon$ then $|X_n|^p \ge |X|^p - p|X|^{p-1}\varepsilon$. Then we can continue bounding the above by

$$\begin{aligned}
&2^{p-1}E[|X_n|^p] - 2^{p-1}E[|X|^p; |X_n - X| \le \varepsilon] + 2^{p-1}pE[|X|^{p-1}]\varepsilon \\
&\qquad + 2^{p-1}E[|X|^p; |X_n - X| > \varepsilon] + \varepsilon^p \\
&= 2^{p-1}E[|X_n|^p] - 2^{p-1}E[|X|^p] + 2^{p-1}E[|X|^p; |X_n - X| > \varepsilon] \\
&\qquad + 2^{p-1}pE[|X|^{p-1}]\varepsilon + 2^{p-1}E[|X|^p; |X_n - X| > \varepsilon] + \varepsilon^p.
\end{aligned}$$

Taking $n \to \infty$ makes $E[|X_n|^p] \to E[|X|^p]$, by assumption. Also taking $A_n = \{|X_n - X| > \varepsilon\}$ and $Y = |X|^p$ in the above lemma shows that

$$E[|X|^p; |X_n - X| > \varepsilon] \to 0.$$

All this gives

$$0 \le \varliminf_{n\to\infty} E[|X_n - X|^p] \le \varlimsup_{n\to\infty} E[|X_n - X|^p] \le 2^{p-1}pE[|X|^{p-1}]\varepsilon + \varepsilon^p.$$

Note that Jensen's inequality gives us that if $E[|X|^p] < \infty$ then $E[|X|^{p-1}] < \infty$. Now taking $\varepsilon \to 0$ proves the L^p convergence.

5.2. Note that for a fixed $y \in \mathbb{R}$ we have

$$D_y = \{x : (x,y) \in D\} = \{x : x = y\} = \{y\}.$$

Furthermore, $\mu(D_y) = \mu(\{y\}) = 0$. Hence

$$\mu \times \mu(D) = \int \mu(D_y)\, \mu(dy) = 0.$$

5.3. Define the measure $\lambda(C) = P\{(X,Y) \in C\}$, for $C \in \mathcal{B}(\mathbb{R}^2)$ (i.e. the distribution of (X,Y)). Note that if $C = A \times B$, with $A, B \in \mathcal{B}(\mathbb{R})$, then

$$\lambda(C) = P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} = \mu(A)\nu(B) = \mu \times \nu(C).$$

Since we have probability measures, the uniqueness of the Carathéodory extension implies that it must be that $\lambda = \mu \times \nu$.

Now write as in the previous problem

$$P(X = Y) = \lambda(D) = \mu \times \nu(D) = \int \mu(D_y)\,\nu(dy).$$

But $\mu(D_y) = P(X \in D_y) = P(X = y) = 0$ by the assumption of the problem. Hence the above integral vanishes and $P(X = Y) = 0$.

5.4. Let $\Omega = \mathbb{N}$, with the σ-algebra of all subsets of $\mathbb{N}$. Let μ be the counting measure: $\mu(A) = |A|$, the cardinality of A. (CHECK this is a σ-finite measure.) Note that if we think of a sequence $(a_i)_{i\geq 1}$ as a function $i \in \mathbb{N} \mapsto a_i \in \mathbb{R}$, then a sequence is a random variable! Note that $\int \mathbb{1}_A\,d\mu = \mu(A) = \sum_{i\geq 1} \mathbb{1}_A(i)$. By the standard machinery, this extends to $\int f\,d\mu = \sum_{i\geq 1} f(i)$. In other words, sums of series are just integrals against the counting measure. Now the claim of the problem is exactly what Fubini's theorem says.

5.7. This is a calculus exercise. Two helpful observations: $(x^2-y^2)/(x^2+y^2)^2$ is the derivative in x of $-x/(x^2+y^2)$ and the derivative in y of $y/(x^2+y^2)$. So

$$\int_0^1 \Big(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dx\Big)\,dy = \int_0^1 \frac{-1}{1+y^2}\,dy = -\tan^{-1} 1 + \tan^{-1} 0 = -\frac{\pi}{4}$$

and

$$\int_0^1 \Big(\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\Big)\,dx = \int_0^1 \frac{1}{x^2+1}\,dy = \tan^{-1} 1 - \tan^{-1} 0 = \frac{\pi}{4}\,.$$

Fubini's theorem failed because the double integral of the quantity with absolute values on it is infinite:

$$\int_0^1 \Big(\int_0^1 \frac{|x^2-y^2|}{(x^2+y^2)^2}\,dy\Big)\,dx \geq \int_0^1 \Big(\int_0^x \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\Big)\,dx = \int_0^1 \frac{x}{2x^2}\,dx = \int_0^1 \frac{1}{2x}\,dx = \infty.$$

5.13. Let X and Y be two random variables such that (X, Y) has distribution $\mu \times \mu$. This means that the two random variables are independent and have the same distribution μ. Now note that

$$\int fg\,d\mu = E[f(X)g(X)] = E[f(Y)g(Y)].$$

Also,

$$\int f\,d\mu = E[f(X)] = E[f(Y)] \quad \text{and} \quad \int g\,d\mu = E[g(X)] = E[g(Y)].$$

Since f and g are non-decreasing, we have (CHECK!)

$$(f(X) - f(Y))(g(X) - g(Y)) \geq 0.$$

Expand and take expected values to get

$$E[f(X)g(X)] - E[f(Y)g(X)] - E[f(X)g(Y)] + E[f(Y)g(Y)] \geq 0.$$

Use independence to write $E[f(Y)g(X)] = E[f(Y)]E[g(X)]$ and $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$. Then use the above equalities to get

$$2\int fg\,d\mu - 2\int f\,d\mu \int g\,d\mu \geq 0.$$

Divide by 2 to get the FKG inequality. To get the claim about the sums let $\mu = n^{-1}\sum_{i=1}^{n}\delta_i$, i.e. $\mu(A) = n^{-1}|A \cap \{1,\dots,n\}|$. Let $f(x) = a_x$ if $x \in \{1,\dots,n\}$ and 0 otherwise. Define g similarly, with a replaced by b. Check that the claim now follows directly from the FKG inequality.

5.17. Write

$$\sum_{n=1}^{\infty} \alpha^n H_n = \sum_{n=1}^{\infty} \alpha^n \sum_{j=1}^{n} \frac{1}{j} = \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} \frac{\alpha^n \mathbb{1}\{j \leq n\}}{j}.$$

Since the summands are nonnegative we can use Fubini's theorem to switch the order of sums and the above equals

$$\sum_{j=1}^{\infty}\sum_{n=1}^{\infty} \frac{\alpha^n \mathbb{1}\{j \leq n\}}{j} = \sum_{j=1}^{\infty} \frac{1}{j} \sum_{n=j}^{\infty} \alpha^n = \frac{1}{1-\alpha}\sum_{j=1}^{\infty} \frac{\alpha^j}{j}.$$

Now let us compute $\sum_{j=1}^{\infty} \frac{\alpha^j}{j}$. For this we use the hint and another application of Fubini's theorem to get

$$\sum_{j=1}^{\infty} \frac{\alpha^j}{j} = \sum_{j=1}^{\infty} \int_0^{\alpha} x^{j-1}\,dx = \int_0^{\alpha} \sum_{j=1}^{\infty} x^{j-1}\,dx = \int_0^{\alpha} \frac{1}{1-x}\,dx = -\log(1-\alpha).$$

Consequently,

$$\sum_{n=1}^{\infty} \alpha^n H_n = \frac{-\log(1-\alpha)}{1-\alpha}\,.$$

Revisiting 3.11. Earlier, we solved problem 3.11 but provided we can find some interval that gets assigned a finite measure. What we did not prove, however, is the existence of such an interval. Now that we know about the Radon-Nikodým derivative and Fubini's theorems, we can solve the exercise completely and even in a more elegant way.

We start by showing that $\mu \ll m$. Namely, that if A is a Borel set with $m(A) = 0$, then $\mu(A) = 0$. To this end, consider the set $B = \{(x, y) \in \mathbb{R}^2 : x + y \in A\}$. This is a measurable set in $\mathbb{R}^2$. For a fixed $x \in \mathbb{R}$ the section $B_x = \{y \in \mathbb{R} : x + y \in A\} = A - x$ (where a set minus a number means we subtract that number from each element of the set, i.e. we shift the set by that number). Since m is translation-invariant we have $m(A - x) = m(A)$, which is assumed to be 0. Hence, we have

$$\mu\times m(B) = \int m(B_x)\,\mu(dx) = \int m(A-x)\,\mu(dx) = \int m(A)\,\mu(dx) = \int 0\,\mu(dx) = 0.$$

On the other hand, noticing that $B^y = \{x \in \mathbb{R} : x + y \in A\} = A - y$ and recalling that μ is also assumed translation invariant, we now have

$$0 = \mu \times m(B) = \int \mu(B^y)\,m(dy) = \int \mu(A - y)\,m(dy) = \int \mu(A)\,m(dy).$$

Hence, $\mu(A) = 0$.

Since $\mu \ll m$ and both measures are σ-finite, the Radon-Nikodým derivative theorem implies that there exists a unique function $f \geq 0$ such that

$$\mu(A) = \int \mathbb{1}_A(x)\,f(x)\,m(dx),$$

for all measurable sets A. (In class, we proved the theorem assuming the measures are finite. But if μ is σ-finite then we can split the space $\mathbb{R}$ into a countable union of disjoint measurable sets I_n, $n \in \mathbb{N}$, each of which having $\mu(I_n) < \infty$. And then we can consider the sets $I_n \cap (-m, m]$, $n \in \mathbb{N}$, $m \in \mathbb{Z}$. This gives another countable collection of disjoint measurable sets, the union of which is all of $\mathbb{R}$ and whose μ and ν measures are both finite. Now, apply the theorem to each set separately. This gives a bunch of Radon-Nikodým derivatives, one for each set. Since the sets are disjoint and cover all of $\mathbb{R}$, this defines a measurable function $f \geq 0$ on all of $\mathbb{R}$. You can then check that the above display holds, by breaking A into the union of its intersections with the sets partitioning sets and using σ-additivity of μ and linearity of the integral [along with e.g. monotone convergence to handle the sum $\sum_{m,n} \mathbb{1}_{A \cap I_n \cap (-m,m]}$ that may potentially have infinitely many terms].)

Now fix a $y \in \mathbb{R}$ and write

$$\int \mathbb{1}_A(x)\, f(x)\, m(dx) = \mu(A) = \mu(A-y) = \int \mathbb{1}_{A-y}(x)\, f(x)\, m(dx)$$
$$= \int \mathbb{1}_A(x+y)\, f(x)\, m(dx) = \int \mathbb{1}_A(x)\, f(x-y)\, m(dx).$$

Explanation: second equality is translation invariance of μ, the third comes from the definition of f, the fourth is just saying that $x \in A - y$ is the same thing as $x + y \in A$, and the fifth comes from translation invariance of m (use the standard argument to go from $m(A-y) = m(A)$ to $\int g(x+y)\, m(dx) = \int g(x)\, m(dx)$).

But the uniqueness of the Radon-Nikodým derivative then implies that $f(x) = f(x-y)$, m-almost surely, meaning that $m\{x : f(x) \neq f(x-y)\} = 0$ (for a fixed y). We are almost done. If this claim were true for every x, rather than m-almost every x, then we could say that $f(x-y) = f(x)$ for all x, y and so $f(t) = f(x)$ for all t, x, meaning that f is a constant, say c. Then we conclude that $\mu(A) = \int c\, \mathbb{1}_A(x)\, m(dx) = c\, m(A)$, as desired. The problem is that $f(x) = f(x-y)$ holds for every y and m-almost every x, not every x. We can go around this issue with another application of Fubini's theorem! Here is how.

Consider the set $B = \{(x,y) : f(x) \neq f(x-y)\}$. We know that for every fixed y, $B^y = \{x : f(x) \neq f(x-y)\}$ satisfies $m(B^y) = 0$. Then, similar to the argument we used earlier:

$$0 = \int m(B^y)\, m(dy) = m \times m(B) = \int m(B_x)\, m(dx).$$

Here, $B_x = \{y : f(x) \neq f(x-y)\}$. The above says that

$$m\{y : f(x) \neq f(x-y)\} = 0,$$

for m-almost every x. In particular, it implies that there exists at least one x_0 such that $m\{y : f(x_0) \neq f(x_0 - y)\} = 0$. (Otherwise, if this were > 0 for every choice of x, the last integral in the above display would be positive, violating the equalities in the display.) The upshot is that for m-almost every y we have $f(x_0 - y) = f(x_0)$. Since m is translation invariant, this says that for m-almost every y we have $f(-y) = f(x_0)$. One can also show that m is reflection-invariant, i.e. that $m(A) = m(-A)$ (CHECK!). Thus, we have shown that $f(y) = f(x_0)$, for m-almost every y. In other words, f equals a constant, $c = f(x_0)$, m-almost surely. Now we get the claim of the problem:

$$\mu(A) = \int c\,\mathbb{1}_A(x)\,m(dx) = c\,m(A).$$

6.2. One direction is clear: if the random variables are independent, then $\mathbb{1}\{X \le x\}$ and $\mathbb{1}\{Y \le y\}$ are independent events and the claimed equality (6.75) follows. We now prove the other direction. That is, assume (6.75). Now compute

$$
\begin{aligned}
&P(a < X \le b, c < Y \le d) \\
&= P(X \le b, Y \le d) - P(X \le a, Y \le d) - P(X \le b, Y \le c) + P(X \le a, Y \le c) \quad \text{(CHECK!)} \\
&= P(X \le b)P(Y \le d) - P(X \le a)P(Y \le d) - P(X \le b)P(Y \le c) + P(X \le a)P(Y \le c) \\
&= \big[P(X \le b) - P(X \le a)\big]\,\big[P(Y \le d) - P(Y \le c)\big] \\
&= P(a < X \le b)P(c < Y \le d).
\end{aligned}
$$

This says that $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all semi-open intervals A and B.

Let μ be the distribution of (X, Y) under P, i.e. $\mu(C) = P\{(X, Y) \in C\}$. Let ν be the product measure $P_X \times P_Y$. In particular, for any two Borel sets A and B we have $\nu(A \times B) = P(X \in A)P(X \in B)$. What we proved above is that $\mu(C) = \nu(C)$ for all sets C of the form $(a, b] \times (c, d]$ for some $a, b, c, d \in \mathbb{R}$. As we have seen several times already, this equality extends to finite unions of such sets, which forms an algebra and then the uniqueness of the extension in Carathéodory's theorem implies that in fact $\mu = \nu$. In particular, for any two Borel sets A and B

$$P(X \in A, Y \in B) = \mu(A \times B) = \nu(A \times B) = P(X \in A)P(Y \in B).$$

Consequently, X and Y are independent.

6.4. (1) is a direct result of the Radon-Nikodým theorem. Let m_d denote Lebesgue measure on $\mathbb{R}^d$. If $P_{X,Y} \ll m_2$ then $P_X \ll m_1$ and $P_Y \ll m_1$. Indeed, if $m_1(A) = 0$, then $m_2(A \times \mathbb{R}) = m_2(\mathbb{R} \times A) = m_1(A) = 0$ (since $m_2 = m_1 \times m_1$). But then $P_{X,Y}(A \times \mathbb{R}) = P_{X,Y}(\mathbb{R} \times A) = 0$. This says $P(X \in A) = P(Y \in A) = 0$ and the absolute continuity claim follows. Now (2) comes again from the Radon-Nikodým theorem.

Now, if $f(x,y) = f_X(x)f_Y(y)$ for m_2-almost every (x,y), then

$$\begin{aligned}
P(X \in A, Y \in B) &= \int \mathbb{1}_{A\times B}(x,y) f(x,y)\, m_2(dx,dy) \\
&= \int \mathbb{1}_A(x)\mathbb{1}_B(y) f_X(x) f_Y(y)\, m_1(dx) m_1(dy) \\
&= \int \mathbb{1}_A(x) f_X(x)\, m_1(dx) \int \mathbb{1}_B(y) f_Y(y)\, m_1(dy) \\
&= P(X \in A) P(Y \in B).
\end{aligned}$$

The second-last equality is by Fubini's theorem. This computation shows that X and Y are independent.

On the other hand, assume X and Y are independent. Then we can run the same computation as above, but in a different order:

$$\begin{aligned}
\int \mathbb{1}_{A\times B}(x,y) f(x,y)\, m_2(dx,dy) &= P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \\
&= \int \mathbb{1}_A(x) f_X(x)\, m_1(dx) \int \mathbb{1}_B(y) f_Y(y)\, m_1(dy) \\
&= \int \mathbb{1}_A(x)\mathbb{1}_B(y) f_X(x) f_Y(y)\, m_1(dx) m_1(dy) \\
&= \int \mathbb{1}_{A\times B}(x,y) f_X(x) f_Y(y)\, m_2(dx,dy).
\end{aligned}$$

But this implies that $f_X(x)f_Y(y)$ is a Radon-Nikodým derivative of the distribution of $P_{X,Y}$ relative to m_2. Since this is supposed to be unique, it must be equal to $f(x,y)$, m_2-almost surely.

6.5. Let X_1 and X_2 be two independent Bernoulli(1/2) random variables and let $X_3 = X_1 + X_2 \bmod 2$. Note that X_3 is again a Bernoulli(1/2) random variable: $P(X_3 = 0) = P(X_1 = X_2) = 1/2$. Because of symmetry, we just need to check that X_1 and X_3 are independent and that the three are not jointly independent. For the first claim we see that

$$P(X_1 = 0, X_3 = 0) = P(X_1 = 0, X_2 = 0) = 1/4 = P(X_1 = 0)P(X_3 = 0).$$

(This is enough, as $\{X_i = 1\}$ is the complement of $\{X_i = 0\}$. But do sort this out, if it is not absolutely clear!)

Now compute

$$P(X_1 = X_2 = X_3 = 0) = P(X_1 = X_2 = 0) = 1/4 \neq 1/8 = P(X_1 = 0)P(X_2 = 0)P(X_3 = 0).$$

A similar example: two coin tosses and $X_i = \mathbb{1}\{\text{coin } i \text{ is H}\}$, $i \in \{1, 2\}$, and $X_3 = \mathbb{1}\{\text{coin } 1 = \text{coin } 2\}$. Work out the rest.

6.7. Let X and Z be two independent random variables such that

$$P(X = 0) = P(X = 1) = P(Z = 1) = P(Z = -1) = 1/2.$$

Let $Y = XZ$. Then $E[Y] = E[XZ] = E[X]E[Z] = 1/2 \times 0 = 0$ and

$$E[XY] = E[X^2 Z] = E[X^2]E[Z] = 0 = E[X]E[Y].$$

In other words, the two are uncorrelated. However,

$$P(X = 0, Y = 0) = P(X = 0) = 1$$

while $P(X = 0) = 1/2$ and $P(Y = 0) = P(X = 0) = 1/2$ and $1/2 \times 1/2 \neq 1$. So the two random variables are not independent. (In short, if we know $X = 0$ then we know for sure $Y = 0$, so Y does depend on X.)

6.9. Here is the quick version: Let Z be a random variable that is independent of X and that has a continuous distribution, i.e. $P(Z = z) = 0$ for all $z \in \mathbb{R}$ (no jumps in the CDF), and that takes all values on the real line, i.e. $P(Z \leq z)$ is strictly increasing in z (CDF has no constant portions). For example, Z can be a standard normal.

Let $X_n = X + Z/n$. Clearly, $X_n \to X$ almost surely. Fix an $x \in \mathbb{R}$. Applying Problem 5.3, with Z playing the role of X there and $n(x - X)$ playing the role of Y, we see that $P(X_n = x) = P(Z = n(x - X)) = 0$. This says that X_n has a continuous CDF. If the CDF were not strictly increasing, there would be two values $x < y$ such that $P(X_n \leq x) = P(X_n \leq y)$. But this says that $P(x < X + Z/n \leq y) = 0$. A reasoning similar to Problem 5.3 (i.e. thinking of the product measure that is the distribution of (X, Z) and using Fubini's theorem) one sees that since Z has a positive probability to be in any nonempty interval and since X is independent of Z we have that given X the probability that Z falls between $(n(x - X), n(y - X)]$ is always positive and so must be $P(x < X + Z/n \leq y)$. This shows that the CDF of X_n cannot have constant segments and it is thus strictly increasing.

Now, the longer more detailed and more direct version. To fix the notation say $X : \Omega \to \mathbb{R}$ with Ω equipped with some σ-algebra $\mathcal{F}$ and a probability measure P. Let $Z : \mathbb{R} \to \mathbb{R}$ be the identity function and equip $\mathbb{R}$ with the Borel σ-algebra $\mathcal{B}$ and the probability measure defined by $Q\{(-\infty, s]\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-z^2/2}\, dz$ for all $s \in \mathbb{R}$ (this defines a CDF and we have shown in an earlier homework problem that this identifies a unique probability measure). In other words, Q is the distribution of a standard normal random variable.

Now, consider the product space $\Omega \times \mathbb{R}$ with the product σ-algebra $\mathcal{F} \times \mathcal{B}$ and the product measure $P \times Q$. Let $Y : \Omega \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ be defined so that $Y(\omega, s) = (X(\omega), Z(s))$. In other words, we have extended the space Ω to the bigger space $\Omega \times \mathbb{R}$ and added the independent random variable Z.

Let $X_n = X + Z/n$. This is a random variable on the bigger product space $\Omega \times \mathbb{R}$. We verify that it has all the required properties. Indeed, it is clear that we have almost surely $X_n \to X$ as $n \to \infty$ (since $Z < \infty$ almost surely). Also, X_n has distribution function

$$
\begin{aligned}
P \times Q(X_n \le x) &= P \times Q(\{(\omega, s) : X(\omega) + Z(s)/n \le x\}) \\
&= \int Q\big(\{s : Z(s) \le n(x - X(\omega))\}\big)\, P(d\omega) \\
&= \frac{1}{\sqrt{2\pi}} \int \Big(\int_{-\infty}^{n(x-X(\omega))} e^{-z^2/2}\, dz \Big)\, P(d\omega) \\
&= \frac{n}{\sqrt{2\pi}} \int \Big(\int_{-\infty}^{x} e^{-n^2(y-X(\omega))^2/2}\, dy \Big)\, P(d\omega) \\
&= \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{x} \Big(\int e^{-n^2(y-X(\omega))^2/2}\, P(d\omega) \Big)\, dy \\
&= \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{x} E[e^{-n^2(y-X)^2/2}]\, dy.
\end{aligned}
$$

In the second-last equality we used Fubini's theorem because the integrand is nonnegative.

Now, the expected value in the last integral is that of a positive random variable (for all y). Thus, the integral is strictly increasing as a function of x. Also, the above shows that X_n is absolutely continuous relative to Lebesgue measure, with the Radon-Nikodým derivative (or pdf) being given by $y \mapsto E[e^{-n^2(y-X)^2/2}]$. As such, its CDF is a continuous function. (Alternatively, the above formula shows that $P \times Q(X_n = x)$ is the integral over the set $\{x\}$ and is thus 0.)

Or yet another way to see continuity: Note that for a fixed ω, $e^{-n^2(y-X(\omega))^2/2}$ is continuous in y. It is also bounded by 1. Bounded convergence then implies that the aforementioned pdf $E[e^{-n^2(y-X)^2/2}]$ is continuous in y (not just L^1). One consequence of this is that the last integral in the above computation is also a Riemann integral and the continuity in x is then the familiar calculus fact about antiderivatives.

6.10. First observe that the claim that (2) and (3) are equivalent does not really have to do with probability theory. The equivalence holds for any sequence of real numbers (X_n). Indeed, clearly (3) implies (2) since (3) has the larger quantity. And if, on the other hand, (2) holds then for any $\varepsilon > 0$ we can find n_0 such that for $n \geq n_0$ we have $|X_n| \leq \varepsilon n^{1/p}$. But then, having fixed n_0, we can find n_1 (which we can choose to be even larger than n_0) such that for $n \geq n_1$ we have $|X_m| \leq \varepsilon n^{1/p}$ for all $m \leq n_0$. Now for $n \geq n_1$

$$\max_{m \leq n} |X_m| = \max\Big(\max_{m \leq n_0} |X_m|, \max_{n_0 \leq m \leq n} |X_m|\Big) \leq \max\Big(\varepsilon n^{1/p}, \max_{n_0 \leq m \leq n} \varepsilon m^{1/p}\Big) = \varepsilon n^{1/p}.$$

So (3) holds.

Next, we show that (1) and (2) are equivalent. For this start with (2) and write

$$\begin{aligned}
P\Big(\lim \frac{X_n}{n^{1/p}} = 0\Big) = 1 \Leftrightarrow P\Big(\overline{\lim}\, \frac{|X_n|}{n^{1/p}} = 0\Big) = 1& \\
\Leftrightarrow \exists k_0 : \forall k \geq k_0,\ P\Big(\overline{\lim}\, \frac{|X_n|}{n^{1/p}} \leq \frac{1}{k}\Big) = 1.&
\end{aligned}$$

One direction in the last equivalence is trivial (because $0 \leq 1/k$) and the other one comes by continuity of probability under decreasing limits.

The last statement in the above display is equivalent to

$$\begin{aligned}
&\exists k_0 : \forall k \geq k_0,\ P\Big(\overline{\lim}\, \frac{|X_n|}{n^{1/p}} > \frac{1}{k}\Big) = 0 \\
&\Leftrightarrow \exists k_0 : \forall k \geq k_0,\ P\big(|X_n| > n^{1/p}/k \text{ infinitely often}\big) = 0.
\end{aligned}$$

There is a small subtlety in the last equivalence. One direction is again obvious: if the limsup is $> 1/k$ then there is a subsequence staying $> 1/k$.

The other direction has a small problem: for a fixed k, if the subsequence stays $> 1/k$ then this only says the limsup is $\geq k$, not $> k$. But 1. this actually is not a big deal for the purpose of the proof we are ultimately after and 2. in fact, it is not a problem at all because the claim is that this holds for all $k \geq k_0$, and so we get that the limsup is $\geq 1/k > 1/(k+1)$ for all $k \geq k_0$. So renaming k (so that it plays the role of $k+1$) we have the desired $> 1/k$ for all $k \geq k_0 + 1$ (and now $k_0 + 1$ also plays the role of the k_0). (WRITE THIS UP CLEANLY!)

Since the random variables are assumed independent, the events in the very last claim in the above display are independent. Hence, by Borel-Cantelli, the above is equivalent to

$$\exists k_0 : \forall k \geq k_0,\ \sum_{n=1}^{\infty} P\big(|X_n| > n^{1/p}/k\big) < \infty$$

which is the same as

$$\exists k_0 : \forall k \geq k_0,\ \sum_{n=1}^{\infty} P\big(k^p|X_1|^p > n\big) < \infty.$$

The X_n was replaced by X_1 because the random variables are assumed to have the same distribution. Now, we invoke a lemma (that we sort of saw in class). From this lemma we get that the above is equivalent to $k^p E[|X_1|^p] < \infty$ for all k large enough. But this k now plays no role! This is in fact equivalent to simply $X_1 \in L^p$, which is (1).

Lemma 0.2. *Say Y is a nonnegative random variable. Then*

$$\sum_{n=1}^{\infty} P(Y > n) < \infty \text{ if and only if } E[Y] < \infty.$$

Proof. This is just an application of Fubini's theorem:

$$\sum_{n=1}^{\infty} P(Y > n) = \sum_{n=1}^{\infty} E[\mathbb{1}\{Y > n\}] = E\Big[\sum_{n=1}^{\infty}\{Y > n\}] = E[f(Y)],$$

where $f(y) = y-1$ if $y \in \mathbb{N}$ and $[y]$ (the integral part of y, or the round-down of y) if $y \in \mathbb{R}_+ \setminus \mathbb{N}$. But note that $y - 1 \leq f(y) \leq y$ for all $y \geq 0$. Hence $E[f(Y)] < \infty$ if and only if $E[Y] < \infty$. $\square$

6.14. Fix an $\varepsilon > 0$. First, use Chebyshev's inequality to write

$$P(|n^{-1}S_n - \mu| > \varepsilon) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2 n^2} = o(n^{-\delta}).$$

So we see that $S_n/n \to \mu$ in probability. But the question is about almost sure convergence.

If $\delta \in (1, 2)$, then the above is summable and Borel-Cantelli tells us that

$$P(|n^{-1}S_n - \mu| > \varepsilon \text{ infinitely often}) = 0.$$

This means that for every $\varepsilon > 0$ we have

$$P\big(\exists n_0 : n \geq n_0 \Rightarrow |n^{-1}S_n - \mu| \leq \varepsilon\big) = 1.$$

So for every $k \geq 1$ we have

$$P\big(\exists n_0 : n \geq n_0 \Rightarrow |n^{-1}S_n - \mu| \leq 1/k\big) = 1.$$

Since the intersection of countably many full-measure events is a full measure event (CHECK!) we have that

$$P\big(\forall k \geq 1, \exists n_0 : n \geq n_0 \Rightarrow |n^{-1}S_n - \mu| \leq 1/k\big) = 1.$$

But the event in the above probability implies that $n^{-1}S_n \to \mu$. Hence we have established the almost sure convergence.

If $\delta \in (0, 1)$ then the above does not work. At least not right away. However, if we pick $r > 0$ large enough so that $r\delta > 1$, then $n^{-r\delta}$ is summable and the same argument as above shows that S_{n^r}/n^r converges to μ almost surely. Now, we use the nonnegativity of the random variables to conclude.

To this end, let $m_n = \lfloor n^r \rfloor$ (the round-down of n^r). Then $m_n \leq n^r < m_n + 1$. That is, $m_n^r \leq n < (m_n + 1)^r$. Therefore, $S_{m_n^r} \leq S_n \leq S_{(m_n+1)^r}$ (this is where the nonnegativity is used). Thus,

$$\frac{m_n^r}{n} \times \frac{S_{m_n^r}}{m_n^r} = \frac{S_{m_n^r}}{n} \leq \frac{S_n}{n} \leq \frac{S_{(m_n+1)^r}}{n} = \frac{S_{(m_n+1)^r}}{(m_n + 1)^r} \times \frac{(m_n + 1)^r}{n}.$$

Taking $n \to \infty$ and using the fact that $m_n \to \infty$ and $m_n^r/n \to 1$ as $n \to \infty$ and that we have shown that $S_{m^r}/m^r \to \mu$ as $m \to \infty$ we see that the two extremes of the above display converge to μ. This proves the claim of the problem.

6.19. Fix $\varepsilon > 0$. Observe that $\overline{\lim}(X_n/\log n) > \lambda^{-1} + \varepsilon$ implies that $X_n > (\lambda^{-1} + \varepsilon)\log n$ for infinitely many n. But

$$P\{X_n > (\lambda^{-1} + \varepsilon)\log n\} = e^{-\lambda(\lambda^{-1}+\varepsilon)\log n)} = n^{-1-\lambda\varepsilon},$$

which is summable. Borel-Cantelli then tells us that there is zero probability the above events happen infinitely often. In other words, we have $\overline{\lim}(X_n/\log n) \le \lambda^{-1} + \varepsilon$ with probability one. This is true for $\varepsilon = 1/k$, for all $k \ge 1$. Hence, by continuity of probability under monotone events, we can take $k \to \infty$ and conclude that $\overline{\lim}(X_n/\log n) \le \lambda^{-1}$ has probability one. Now for the other bound.

Since

$$P\{X_n \ge \lambda^{-1}\log n\} = e^{-\lambda\lambda^{-1}\log n} = n^{-1},$$

is not summable, and since the random variables are independent (and hence events $\{X_n \ge \lambda^{-1}\log n\}$ are independent) the Borel-Cantelli lemma tells us that there is full probability that $X_n \ge \lambda^{-1}\log n$ happens for infinitely many n. But then this implies that $\overline{\lim}(X_n/\log n) \ge \lambda^{-1}$ has probability one. Putting the two facts together, we get the first claim of the problem.

The second claim comes similarly, so we go through the proof a bit faster. First, observe that $P\{\log X_n \le -\log n\} = 1 - e^{-\lambda/n}$. Using Calculus we can show that $e^{-x} \le 1 - x/2$ for x small (CHECK). This implies that the probability is $\ge \lambda/(2n)$, which is not summable. Hence with probability one the events in question happen infinitely often. This implies that $\underline{\lim}(\log X_n/\log n) \le -1$ almost surely. On the other hand, for any $\varepsilon > 0$ we have $P\{\log X_n \le -(1+\varepsilon)\log n\} = 1 - e^{-\lambda/n^{1+\varepsilon}}$. Since for $x \ge 0$ we have $e^{-x} \ge 1 - x$ (CHECK!), we can bound the probability from above by $\lambda/n^{1+\varepsilon}$, which is summable. Thus, with probability one $\log X_n \le -(1+\varepsilon)\log n$ happens only finitely often and for n large enough we have $\log X_n \ge -(1+\varepsilon)\log n$. This means that $\liminf(\log X_n/\log n) \ge -1$ almost surely. The second claim follows.

6.27. Let us abbreviate $Y = \max_{1\le j\le n}|S_j - E[S_j]|$. Then Kolmogorov's maximal inequality says that

$$P(Y \ge s) \le \frac{\mathrm{Var}(S_n)}{s^2} = \frac{n\sigma^2}{s^2},$$

where we wrote σ^2 for $\mathrm{Var}(X_1)$. Now write

$$E[Y] = \int_0^\infty P(Y \geq s)\,ds.$$

If we plug in the upper bound $n\sigma^2/s^2$ in the integral, the integral blows up and we do not have a good upper bound on $E[Y]$. However, Kolmogorov's bound is not efficient when s is small. So instead, we will replace it by the trivial bound saying a probability is always less than one, to get

$$\begin{aligned} E[Y] &= \int_0^\infty P(Y \geq s)\,ds = \int_0^a P(Y \geq s)\,ds + \int_a^\infty P(Y \geq s)\,ds \\ &\leq \int_0^a ds + \int_a^\infty \frac{n\sigma^2}{s^2}\,ds = a + \frac{n\sigma^2}{a}. \end{aligned}$$

The above holds for all $a > 0$. We can now either minimize the right-hand side over a, or just note that the place where we should switch from the trivial bound to Kolmogorov's bound is when $n\sigma^2/s^2 = 1$, i.e. when $s = \sigma\sqrt{n}$. Plugging this value for a gives the claimed upper bound of $2\sigma\sqrt{n}$.

6.28. Part (1) is simple: the first step of the path has $2d$ options. After that, since the path is self-avoiding, it cannot backtrack its steps and so each following step has at most $2d-1$ options. Hence, we have at most $2d(2d-1)^{n-1}$ such paths.

For part (2), note that the number N_n of open paths is equal to $\sum_{\mathbf{x}\in\mathcal{A}_n} \mathbb{1}\{\mathbf{x} \text{ is open}\}$, where $\mathcal{A}_n$ is the set of self-avoiding paths of length n (starting at 0). But then

$$E[N_n] = E\Big[\sum_{\mathbf{x}\in\mathcal{A}_n} \mathbb{1}\{\mathbf{x} \text{ is open}\}\Big] = \sum_{\mathbf{x}\in\mathcal{A}_n} P\{\mathbf{x} \text{ is open}\}.$$

Now, if a path $\mathbf{x}$ is fixed and is self-avoiding, then the probability it is open is simply p^n. Thus,

$$E[N_n] \leq \sum_{\mathbf{x}\in\mathcal{A}_n} p^n = |\mathcal{A}_n| p^n \leq 2d(2d-1)^{n-1} p^n,$$

where $|\mathcal{A}_n|$ is the cardinality of $\mathcal{A}_n$. Fatou's lemma says that

$$E[\underline{\lim}\, N_n] \leq \underline{\lim}\, E[N_n] = 0.$$

Since $N_n \geq 0$, we deduce that $\underline{\lim}\, N_n = 0$ almost surely. But N_n goes down to the number of infinite self-avoiding paths. Hence, we see that there is zero probability of having an infinite self-avoiding path out of the origin.

We have thus shown that if $p < 1/(2d-1)$ then percolation does not happen and so $p \leq p_c$. Taking $p \to 1/(2d-1)$ proves that $p_c \geq 1/(2d-1)$.

6.29. Abbreviate $S_n = X_1 + \cdots + X_n$. Write

$$\sum_{i=1}^{n} \frac{X_i}{i} = \frac{S_n}{n} + \sum_{i=1}^{n-1} \frac{S_i}{i} \cdot \frac{1}{i+1}.$$

(For this, note that $X_i = S_i - S_{i-1}$, with the convention that $S_0 = 0$. Separate the sum of $S_i/i - S_{i-1}/i$ into the difference of two sums, change variables in the second sum, then recombine.)

The law of large numbers says $S_n/n \to \mu$ almost surely and we know that $\sum_{i=1}^{n} 1/(i+1) \to \infty$. Thus $\sum_{i=2}^{n} \frac{S_i}{i} \cdot \frac{1}{i+1}$ is asymptotically the same as $\sum_{i=2}^{n} \frac{\mu}{i+1}$, meaning that their ratio goes to 1. But by comparison with integrals, this latter sum is itself asymptotically equivalent to $\mu \int_1^n \frac{1}{x}\, dx = \mu \log n$. (CHECK all these calculus claims in this paragraph.)

6.30. By definition $\mathcal{B}_n f(x) = E[f(S_n)]$, where S_n is a Binomial(n, x). To study monotonicity of this polynomial, we need to couple the Binomial random variables for different x's. To do this, note that if $U_1, \ldots, U_n$ are independent Uniform$(0,1)$ random variables, then $S_n(x) = \sum_{k=1}^{n} \mathbb{1}\{U_k \leq x\}$ is a Binomial(n, x). Hence,

$$\mathcal{B}_n f(x) = E\Big[f\Big(\sum_{k=1}^{n} \mathbb{1}\{U_k \leq x\}\Big)\Big].$$

But now we see that increasing x increases the indicators, which increases the sum, and hence the value of f does not go down (since f is nondecreasing). Since expectations preserve inequalities, we see that $\mathcal{B}_n f$ also does not go down, meaning it is a nondecreasing function.

6.37. The fact in question is surprising because one is then tempted to say that for $a < b$, $F(b) - F(a) = \int_a^b F'(x)\, dx = 0$ and so F is constant, which we know it cannot be!

For part (1), observe that the map

$$\sum_{j=1}^{\infty} \zeta_j/3^j \mapsto \sum_{j=1}^{\infty} \zeta_j/2^{j+1}$$

puts C in bijection with $[0,1]$ (through binary expansions). Hence, C is uncountable.

Let C_n be the set of numbers of the form $\sum_{j\geq 1} \zeta_j/3^j$ with $\zeta_j \in \{0,2\}$ for all $j \leq n$. Observe that if $x \in C_n$ then both $x/3$ and $(x+2)/3$ are in C_{n+1}. Conversely, if $x \in C_{n+1}$, then $x = \sum_{j=1}^{n+1} \zeta_j/3^j$ and hence either $\zeta_1 = 0$ and $3x \in C_n$ or $\zeta_1 = 2$ and then $3x - 2 \in C_n$. This says that $C_{n+1} = (C_n/3) \cup (2/3 + C_n/3)$. Now, observe that $C_n/3$ and $2/3 + C_n/3$ are disjoint, since numbers in the former are $\leq 1/3$ and in the latter are $\geq 2/3$. Using shift-invariance and scaling property of the Lebesgue measure we see that $m(C_{n+1}) = 2m(C_n)/3$. Since $C_1 = [0,1/3] \cup [2/3,1]$, we have that $m(C_1) = 2/3$ and thus $m(C_n) = (2/3)^n \to 0$. But C_n are nonincreasing and $C = \cap_n C_n$. Hence, $m(C) = 0$.

For part (2), fix $y \in [0,1)$ and note that there exists a unique sequence $\zeta_j \in \{0,1,2\}$ such that $y = \sum_{j=1}^{\infty} \zeta_j/3^j$, with $\zeta_j \in \{0,1\}$ infinitely often. (Allowing expansions that eventually stabilize at $\zeta_j = 2$ ruins uniqueness, as $\sum_{j=k}^{\infty} 2/3^j = 1/3^{j-1}$.) Abbreviate $Y = \sum_{j\geq 1} X_j/3^j$. Then

$$P(Y = y) \leq P(X_j = \zeta_j : j \leq n) = 1/2^n \to 0.$$

Similarly, $Y = 1$ means $X_j = 2$ for all $j \geq 1$ and has probability 0. Hence, Y has a continuous CDF.

On the other hand, if $y \notin C$, then $y = \sum_{j=1}^{\infty} \zeta_j/3_j$ with $\zeta_j = 1$ for some $j \geq 1$. Let j_0 be the smallest such index. Let $a = \sum_{j=1}^{j_0-1} \zeta_j/3^j$ ($a = 0$ if $j_0 = 1$) and $b = a + 2/3^{j_0}$. Then $y \in (a,b)$ (CHECK). Furthermore,

$$P(a < Y < b) = P(X_{j_0} = 1) = 0.$$

This shows that the CDF of Y is constant in a neighborhood of y and thus its derivative at y exists and equals 0. Since C has measure 0, its complement has full measure and we have shown that $F' = 0$ almost surely.

Here again we have that $P(Y \in C) = 1$ (by construction of Y) while $m(C) = 0$, and that $P(Y = y) = 0$ for all y. In other words, Y is another example satisfying Theorem 6.20. In fact, one can show that the CDF of the Y in Theorem 6.20 is also such that $F' = 0$ almost surely. This is a characterization of a singular random variable (i.e. a random variable that does not have a discrete nor an absolutely continuous part).

8.3. Recall that a random variable Z is $\sigma(X)$-measurable, if and only if there exists a Borel-measurable function ϕ such that $Z = \phi(X)$, almost surely. Thus, we need to show that for any bounded Borel-measurable function ϕ we have $E[f(X,Y)\phi(X)] = E[g(X)\phi(X)]$. Let μ_1 and μ_2 be the distributions of X and Y, respectively. Since the two are assumed to be independent, the distribution of (X,Y) is $\mu_1 \times \mu_2$. Write

$$\begin{aligned} E[f(X,Y)\phi(X)] &= \iint f(x,y)\phi(x)\,\mu_1(dx)\mu_2(dy) = \int \Big[\int f(x,y)\,\mu_2(dy)\Big]\phi(x)\,\mu_1(dx) \\ &= \int g(x)\phi(x)\,\mu_1(dx) = E[g(X)\phi(X)]. \end{aligned}$$

8.4. Fix a rational number q and compute

$$\begin{aligned} 0 &\ge E[(X-Y)\mathbb{1}\{X \le q < Y\}] \\ &= E[(X-Y)\mathbb{1}\{q < Y\}] - E[(X-Y)\mathbb{1}\{q < X, q < Y\}] \\ &= E[X\mathbb{1}\{q < Y\}] - E[Y\mathbb{1}\{q < Y\}] - E[(X-Y)\mathbb{1}\{q < X, q < Y\}] \\ &= E[X\mathbb{1}\{q < Y\}] - E[E[X|Y]\mathbb{1}\{q < Y\}] - E[(X-Y)\mathbb{1}\{q < X, q < Y\}] \\ &= E[X\mathbb{1}\{q < Y\}] - E[E[X\mathbb{1}\{q < Y\}|Y]] - E[(X-Y)\mathbb{1}\{q < X, q < Y\}] \\ &= E[X\mathbb{1}\{q < Y\}] - E[X\mathbb{1}\{q < Y\}] - E[(X-Y)\mathbb{1}\{q < X, q < Y\}] \\ &= -E[(X-Y)\mathbb{1}\{q < X, q < Y\}]. \end{aligned}$$

Similarly,

$$0 \ge E[(Y-X)\mathbb{1}\{Y \le q < X\}] = -E[(Y-X)\mathbb{1}\{q < X, q < Y\}] = E[(X-Y)\mathbb{1}\{q < X, q < Y\}].$$

The two inequalities give $E[(X-Y)\mathbb{1}\{q < X, q < Y\}] = 0$ and the equalities then give $E[(Y-X)\mathbb{1}\{Y \le q < X\}] = E[(X-Y)\mathbb{1}\{X \le q < Y\}] = 0$.

Since $X - Y > 0$ on the event $\{X \le q < Y\}$ it must be that $P(X \le q < Y) = 0$. Similarly, $P(Y \le q < X) = 0$. Next, observe that

$$P(X < Y) = P\Big(\bigcup_{q \in \mathbb{Q}} \{X \le q < Y\}\Big) \le \sum_{q \in \mathbb{Q}} P(X \le q < Y) = 0.$$

We have thus shown that $X \ge Y$ almost surely. The reverse inequality follows similarly.

Here is a faster proof, if we know that $X, Y \in L^2$: first use $E[X|Y] = Y$ to write

$$E[XY] = E[E[XY|Y]] = E[E[X|Y]Y] = E[Y^2].$$

Similarly, using $E[Y|X] = X$ we have

$$E[XY] = E[E[XY|X]] = E[XE[Y|X]] = E[X^2].$$

But then this implies $E[X^2] = E[Y^2] = E[XY]$ and consequently

$$E[(X-Y)^2] = E[X^2] + E[Y^2] - 2E[XY] = E[X^2] - E[Y^2] = 0.$$

This in turn implies $X = Y$ almost surely.

8.6. Let X be $\mathcal{F}$-measurable and Z a $\mathcal{G}$-measurable random variable. Expand

$$E[(X-Z)^2 \,|\, \mathcal{G}] = E[X^2 \,|\, \mathcal{G}] - 2E[XZ \,|\, \mathcal{G}] + E[Z^2 \,|\, \mathcal{G}] = E[X^2 \,|\, \mathcal{G}] - 2ZE[X \,|\, \mathcal{G}] + Z^2$$

and

$$\begin{aligned} \mathrm{Var}(X \,|\, \mathcal{G}) &= E\big[(X - E[X \,|\, \mathcal{G}])^2 \,|\, \mathcal{G}\big] \\ &= E[X^2 \,|\, \mathcal{G}] - 2E\big[XE[X \,|\, \mathcal{G}] \,|\, \mathcal{G}\big] + E\big[E[X \,|\, \mathcal{G}]^2 \,|\, \mathcal{G}\big] \\ &= E[X^2 \,|\, \mathcal{G}] - E[X \,|\, \mathcal{G}]^2. \end{aligned}$$

Putting the two together we see that we are asked to prove that

$$E[X \,|\, \mathcal{G}]^2 - ZE[X \,|\, \mathcal{G}] + Z^2 \geq 0$$

almost surely. But this is always true, because the left-hand side is a perfect square. Taking expected values in what we proved gives

$$E[(X-Z)^2] \geq E\big[(X - E[X \,|\, \mathcal{G}])^2\big],$$

which is (8.4).

8.7. Write $\lambda \mathbb{1}\{X \geq \lambda\} \leq X \mathbb{1}\{X \geq \lambda\} \leq X$ and take conditional expectation.

8.8. We need to show that for every bounded Borel-measurable function ϕ we have

$$E[h(X)\phi(Y)] = E\Big[\frac{\int h(x) f(x,Y)\,dx}{\int f(x,Y)\,dx} \times \phi(Y)\Big]. \tag{1}$$

We can in fact assume ϕ to also be nonnegative, since we really only care about the above being true for indicator functions.

Before we show the above, let us note that dividing by $\int f(x,Y)\,dx$ is OK. Indeed, recall that if (X,Y) has density f, then

$$E[g(Y)] = \iint g(y)f(x,y)\,dx\,dy. \tag{2}$$

Using $g(Y) = \mathbb{1}\{\int f(s,Y)\,ds = 0\}$ we have

$$\begin{aligned} P\Big(\int f(s,Y)\,ds = 0\Big) &= \iint \mathbb{1}\Big\{y : \int f(s,y)\,ds = 0\Big\} f(x,y)\,dx\,dy \\ &= \int \mathbb{1}\Big\{y : \int f(s,y)\,ds = 0\Big\}\Big[\int f(x,y)\,dx\Big]\,dy \\ &= \int 0\,dy = 0. \end{aligned}$$

So $\int f(x,Y)\,dx > 0$ almost surely and we can divide by it.

Now, on to the proof of (1). Use (2) to compute

$$\begin{aligned} E\Big[\frac{\int h(x)f(x,Y)\,dx}{\int f(x,Y)\,dx} \times \phi(Y)\Big] &= \iint \frac{\int h(x)f(x,y)\,dx}{\int f(x,y)\,dx} \cdot \phi(y)f(x',y)\,dx'dy \\ &= \int \Big(\frac{\int h(x)f(x,y)\,dx}{\int f(x,y)\,dx} \cdot \phi(y)\Big[\int f(x',y)\,dx'\Big]\,dy \\ &= \int \Big(\int h(x)f(x,y)\,dx\Big) \cdot \phi(y)\,dy \\ &= \iint h(x)f(x,y)\phi(y)\,dx\,dy \\ &= E[h(X)\phi(Y)]. \end{aligned}$$

In the above, we used Fubini's theorem twice. This is justified because all integrands are nonnegative.

$P(X \le a \,|\, Y = y)$ really stands for the function $\psi(y)$ such that $\psi(Y) = E[\mathbb{1}\{X \le a\} \,|\, Y]$. (We know such a function exists, because the conditional expectation is $\sigma(Y)$-measurable.) And we just showed that

$$E[\mathbb{1}\{X \le a\} \,|\, Y] = \frac{\int \mathbb{1}\{x \le a\}\, f(x,Y)\,dx}{\int f(x,Y)\,dx} = \frac{\int_{-\infty}^{a} f(x,Y)\,dx}{\int f(x,Y)\,dx}\,.$$

Hence,

$$\psi(y) = \frac{\int_{-\infty}^{a} f(x,y)\,dx}{\int f(x,y)\,dx}\,,$$

as claimed.

8.14. First, we verify the martingale property:

$$E[\zeta^{S_{n+1}}\,|\,\mathcal{F}_n] = \zeta^{S_n} E[\zeta^{X_{n+1}}] = \zeta_n^S\Big(\big(\frac{1-p}{p}\big)^1 \times p + \big(\frac{1-p}{p}\big)^{-1} \times (1-p)\Big) = \zeta^{S_n}.$$

Since a martingale's mean is preserved, the fact that it has mean one follows from $E[\zeta^{S_0}] = \zeta^0 = 1$.

Fix two positive integers a and b. Let T be the first time $S_n \in \{-a, b\}$. By the optional stopping theorem we have that $E[\zeta^{S_{T\wedge n}}] = 1$. Note that if $m \le T$, then S_m is still between $-a$ and b and therefore $\zeta^{S_m} \le \zeta^{-a} \vee \zeta^b$. Also, the same argument we used in class to show that $P(T < \infty) = 1$ when $p = 1/2$ works for all p (namely, that by the Borel-Cantelli lemma we are bound to eventually get $a+b$ 1-increments and thus exit the interval $[-a, b]$). Bounded convergence now gives us that

$$E[\zeta^{S_T}] = \lim_{n\to\infty} E[\zeta^{S_{T\wedge n}}] = 1.$$

But S_T is either equal to $-a$ or at b. Thus,

$$1 = E[\zeta^{S_T}] = \zeta^{-a}(1 - P(S_T = b)) + \zeta^b P(S_T = b).$$

This gives us that the probability of reaching b before reaching $-a$ is

$$P(S_T = b) = \frac{1-\zeta^{-a}}{\zeta^b - \zeta^{-a}}\,.$$

Let us now prove transience of the random walk. We work out the case $p > 1/2$ (i.e. $\zeta < 1$), the other being similar. In this case, we expect that the random walk drifts to the right. Thus, we compute the probability

$$P(\forall n \ge 1 : S_n \ge 1).$$

This is the probability that the random walk moved to the right of 0 and then never came back to 0.

The random walk has probability p to move from 0 to 1 and then we want to compute the probability that starting at 1 the walk never returns to 0. This latter probability is the same as the probability of starting at 0 the walk never reaches -1. But we know that the probability the walk reaches b before it reaches -1 equals $(1-\zeta^{-1})/(\zeta^b - \zeta^{-1})$. The event that the walk reaches b before it reaches -1 decreases, as $b \nearrow \infty$ to the event that the walk never reaches -1. Therefore, the probability of the latter event equals $(1-\zeta^{-1})/(-\zeta^{-1}) = 1-\zeta = (2p-1)/p$. Consequently, we have

$$P(\forall n \geq 1 : S_n \geq 1) = 2p - 1 > 0.$$

Next, let $p = 1/2$. From 0 the walk can go to 1 or -1 with probability $1/2$. We will prove that from 1 the walk will return to 0 with probability one. The same then holds if the walk starts at -1. Therefore, starting at 0 the walk will return to 0 with probability $1/2 \times 1 + 1/2 \times 1 = 1$, which proves recurrence.

The probability of returning to 0 when started at 1 is the same as that of hitting -1 when started at 0. Now, the probability of hitting b before hitting -1 was shown to equal $1/(b+1)$. As above, this tells us that the probability the walk, started at 0, never reaches -1 equals $\lim_{b\to\infty} 1/(b+1) = 0$. This proves that almost surely the walk will hit -1 eventually, and thus recurrence of the random walk follows.

The number of steps it takes to return to 0, when started at 0, is equal to one plus the number of steps it takes to return to 0 started at 1. (The walk takes one step to go to 1 or to -1 and then the return times from either point to 0 have the same mean.) The mean return time to 0, when started at 1, is the same as the mean hitting time of -1, given we started at 0. To compute this we calculate the mean exit time of the random walk from the interval $[-a, b]$.

Since $S_n^2 - \sigma^2 n$ is a martingale (when $p = 1/2$) we have

$$E[S_{T\wedge n}^2] = \sigma^2 E[T \wedge n].$$

The right-hand side converges to $E[T]$ as $n \to \infty$, by the monotone convergence theorem. On the other hand, since at time $T \wedge n$ the walk is still between -1 and b we see that $|S_{T\wedge n}| \leq b$. Bounded convergence then allows us to take $n \to \infty$ and deduce that

$$\sigma^2 E[T] = E[S_T^2] = 1 \times \frac{b}{1+b} + b^2 \frac{1}{1+b} = b\,.$$

Since $\sigma^2 = 1$ ($X_1^2 = 1$ and $E[X_1] = 0$) we get that the expected time to exit $[-1, b]$ equals b. As $b \to \infty$ this random variable converges to the expected time of hitting -1. Monotone convergence then gives us that this expected time is infinite. The walk is thus null-recurrent.

8.15. This is a direct computation:

$$\frac{E[e^{tS_{n+1}} \,|\, \mathcal{F}_n]}{\prod_{k=1}^{n+1} h_k(t)} = \frac{e^{tS_n} E[e^{tX_{n+1}}]}{\prod_{k=1}^{n+1} h_k(t)} = \frac{e^{tS_n}}{\prod_{k=1}^{n} h_k(t)}\,.$$

The mean 1 property comes again from the fact that $E[e^{tS_0}] = e^0 = 1$.

8.25. First, recall (or prove!) that $S_n^2 - \sigma^2 n$ is a martingale. So the optional stopping theorem implies that

$$E[S_{T\wedge n}^2] = \sigma^2 E[T \wedge n].$$

Monotone convergence implies that $E[T\wedge n] \to E[T]$ as $n \to \infty$. Then, what is left is to show that $E[S_{T\wedge n}^2] \to E[S_T^2]$.

Note that $S_{T\wedge n}$ converges to S_T almost surely, because $T < \infty$ almost surely. So it is enough to prove that $S_{T\wedge n}$ has a limit in L^2. For then the limit has to be S_T and thus $E[S_{T\wedge n}^2] \to E[S_T^2]$ as desired.

To show that $S_{T\wedge n}$ has a limit in L^2 we need to show it is a Cauchy sequence. For this, write

$$S_{T\wedge n} = \sum_{i=1}^{n} X_i \mathbb{1}\{T \ge i\}$$

and note that $\{T \ge i\}$ is $\mathcal{F}_{i-1}$ measurable and thus

$$E[X_i \mathbb{1}\{T \ge i\} X_j \mathbb{1}\{T \ge j\}] = E[X_i]\, E[\mathbb{1}\{T \ge i\} X_j \mathbb{1}\{T \ge j\}] = 0$$

if $i > j$. Consequently

$$\begin{aligned} E[|S_{T\wedge n} - S_{T\wedge m}|^2] &= E\Big[\Big|\sum_{i=m}^{n} X_i\{T \ge i\}\Big|^2\Big] = \sum_{i=m}^{n} E[X_i^2 \mathbb{1}\{T \ge i\}] \\ &= \sum_{i=m}^{n} E[X_i^2] E[\mathbb{1}\{T \ge i\}] = \sigma^2 \sum_{i=m}^{n} P(T \ge i). \end{aligned}$$

One can make the left-hand side as small as you want by taking m and n large, because the sum from $\sum_{i=1}^{\infty} P(T \geq i) = E[T] < \infty$. The upshot is that $S_{T\wedge n}$ is a Cauchy sequence in L^2 and thus has a limit, which has to match its almost sure limit S_T and then $E[S_{T\wedge n}^2]$ must converge to $E[S_T^2]$.

The above computation gives the impression one could do things more directly. Namely, first note that

$$E[X_i \mathbb{1}\{T \geq i\} X_j \mathbb{1}\{T \geq j\}] = E[X_i]\, E[\mathbb{1}\{T \geq i\} X_j \mathbb{1}\{T \geq j\}] = 0 \qquad (3)$$

if $i > j$. Now write

$$\begin{aligned}
E[S_T^2] &= E\Big[\Big|\sum_{i=1}^{T} X_i\Big|^2\Big] = E\Big[\Big|\sum_{i=1}^{\infty} X_i \mathbb{1}\{T \geq i\}\Big|^2\Big] \\
&= E\Big[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} X_i \mathbb{1}\{T \geq i\} X_j \mathbb{1}\{T \geq j\}\Big] \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} E[X_i \mathbb{1}\{T \geq i\} X_j \mathbb{1}\{T \geq j\}] \\
&= \sum_{i=1}^{\infty} E[X_i^2 \mathbb{1}\{T \geq i\}] \quad \text{(because of (3))} \\
&= \sum_{i=1}^{\infty} E[X_i^2] E[\mathbb{1}\{T \geq i\}] = \sigma^2 E\Big[\sum_{i=1}^{\infty} \mathbb{1}\{T \geq i\}\Big] \\
&= \sigma^2 E[T].
\end{aligned}$$

The problem is with the interchange of the two summations and the expectation, which is justified by truncating (i.e. replacing T with $T \wedge n$) and showing that everything works out fine at the limit. This is essentially what the above solution did.

8.31. Compute

$$E[M_{n+1} \,|\, \mathcal{F}_n] = (n+2)^{-1/2} E[e^{(S_n + X_{n+1})^2/(2n+4)} \,|\, X_1, \ldots, X_n].$$

Given, $X_1, \ldots, X_n$ the random variable S_n is known and X_{n+1} is an independent standard normal. Let us thus compute $E[e^{(s+Z)^2/(2n+4)}]$, where Z is a standard normal. This equals

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(s+z)^2/(2n+4)} e^{-z^2/2}\, dz.$$

Completing the squares in the exponents we get

$$\begin{aligned}\frac{s^2+2sz+z^2-(n+2)z^2}{2n+4} &= -\frac{n+1}{2n+4}\Big(z^2-2\frac{s}{n+1}z-\frac{s^2}{n+1}\Big)\\ &= -\frac{n+1}{2n+4}\Big((z-s/(n+1))^2-\Big(\frac{1}{(n+1)^2}+\frac{1}{n+1}\Big)s^2\Big)\\ &= -\frac{(z-s/(n+1))^2}{2\times\frac{n+2}{n+1}}+\frac{s^2}{2(n+1)}\,.\end{aligned}$$

Plugging back into the above integral we get

$$\frac{1}{\sqrt{2\pi}}e^{s^2/(2n+2)}\int_{-\infty}^{\infty}e^{-(z-s/(n+1))^2/2\sigma^2}\,dz,$$

where $\sigma^2=(n+2)/(n+1)$. Identifying he integrand as part of the pdf of a normal with mean $s/(n+1)$ and variance σ^2, we see that it equals $\sqrt{2\pi\sigma^2}$. Hence, we get that

$$E[e^{(s+Z)^2/(2n+4)}]=\sqrt{\frac{n+2}{n+1}}\,e^{s^2/(2n+1)}.$$

Plugging in $s=S_n$ we get that

$$E[M_{n+1}\,|\,\mathcal{F}_n]=(n+2)^{-1/2}\sqrt{\frac{n+2}{n+1}}e^{S_n^2/(2n+1)}=M_n.$$

This checks the martingale property. The mean one property comes from doing the computation for $n=1$:

$$E[M_1]=\frac{1}{\sqrt{2}}E[e^{X_1^2/4}].$$

The expectation is exactly the integral we just computed with $s=0$ and $n=0$, which we found equals $\sqrt{2}$, making $E[M_1]=1$.

8.34. Recall that $M_T=\sum_n M_n\mathbb{1}\{T=n\}$. Let $A\in\mathcal{F}_T$. Recall that this means that $A \cap \{T = n\} \in \mathcal{F}_n$ for all n. Write

$$\begin{aligned}
E[|M_T|] &= E\Big[\sum_n |M_n| \mathbb{1}\{T = n\}\Big] = \sum_n E[|M_n| \mathbb{1}\{T = n\}] \\
&= \sum_n E\big[|E[Y \,|\, \mathcal{F}_n]| \mathbb{1}\{T = n\}\big] \le \sum_n E\big[E[|Y| \,|\, \mathcal{F}_n] \mathbb{1}\{T = n\}\big] \\
&= \sum_n E\big[E[|Y| \mathbb{1}\{T = n\} \,|\, \mathcal{F}_n]\big] = \sum_n E[|Y| \mathbb{1}\{T = n\}] \\
&= E[|Y| \sum_n \mathbb{1}\{T = n\}] = E[|Y|] < \infty.
\end{aligned}$$

This justifies the multiple applications of Fubini's theorem in the next computation:

$$\begin{aligned}
E[M_T \mathbb{1}_A] &= E\Big[\sum_n M_n \mathbb{1}\{T = n\} \mathbb{1}_A\Big] = \sum_n E[M_n \mathbb{1}\{T = n\} \mathbb{1}_A] \\
&= \sum_n E\big[E[Y \,|\, \mathcal{F}_n] \mathbb{1}\{T = n\} \mathbb{1}_A\big] = \sum_n E\big[E[Y \mathbb{1}\{T = n\} \mathbb{1}_A \,|\, \mathcal{F}_n]\big] \\
&= \sum_n E[Y \mathbb{1}\{T = n\} \mathbb{1}_A] = E[Y \mathbb{1}_A \sum_n \mathbb{1}\{T = n\}] = E[Y \mathbb{1}_A].
\end{aligned}$$

Since $E[M_T \mathbb{1}_A] = E[Y \mathbb{1}_A]$ for all $A \in \mathcal{F}_T$, and we have shown in class that M_T is $\mathcal{F}_T$-measurable, we have that $E[Y \,|\, \mathcal{F}_T] = M_T$.

8.35. One direction comes from the optional stopping theorem. We need to prove the other direction. So assume that $E[M_T] = E[M_1]$ for all bounded stopping times T. We want to prove that M_n is a martingale in the same filtration $\mathcal{F}_n$ for which the stopping times T are defined.

First, considering the constant stopping time $T = n$ gives $E[M_n] = E[M_1]$ for all n. Now, recall that we want to show that for a fixed n and a fixed $A \in \mathcal{F}_n$ we have

$$E[M_{n+1} \mathbb{1}_A] = E[M_n \mathbb{1}_A].$$

Let

$$T = n \mathbb{1}_{A^c} + (n + 1) \mathbb{1}_A.$$

This is a stopping time because $T = m$ is either $\varnothing \in \mathcal{F}_m$ or $A^c \in \mathcal{F}_n$ (when $m = n$) or $A \in \mathcal{F}_n \subset \mathcal{F}_{n+1}$ (when $m = n + 1$). Thus,

$$E[M_n] = E[M_1] = E[M_T] = E[M_n \mathbb{1}_{A^c}] + E[M_{n+1} \mathbb{1}_A].$$

Rearranging the above gives $E[M_n \mathbb{1}_A] = E[M_{n+1} \mathbb{1}_A]$, as desired.

To characterize a submartingale it is not enough to ask that $E[M_T] \geq E[M_1]$ for every bounded stopping time. To see this, note that if this were true then it would say that any process with $M_1 = 0$ and $M_n \geq 0$ for all n is a submatringale! The right thing to ask for is that $E[M_T] \leq E[M_n]$ for any bounded stopping time T and integer n such that $P(T \leq n) = 1$.

The proof is similar to the above. One direction comes again easily by the optional stopping theorem (the version for a submartingale). For the other direction use the same T as above to deduce that $E[M_{n+1} \mathbb{1}_A] \geq E[M_n \mathbb{1}_A]$ for any $A \in \mathcal{F}_n$, proving M_n is a submartingale. (DO the computation!)

For superartingales the inequality in the above display is simply reversed.

8.39. For (1) write

$$E[e^{tS_n}] = E[e^{tX_1}]^n = \Big(\frac{e^t + e^{-t}}{2}\Big)^n \leq e^{nt^2/2}.$$

To see the last inequality use

$$(2n)! = 2n(2n-1)(2n-2)\cdots 2 \cdot 1 \geq 2n(2n-2)(2n-4)\cdots 2 = 2^n n!$$

to bound

$$\frac{e^t + e^{-t}}{2} = \sum_{k=0}^{\infty} \frac{t^{2n}}{(2n)!} \leq \sum_{k=0}^{\infty} \frac{t^{2n}}{2^n n!} = e^{t^2/2}.$$

For (2) note that

$$S_j \geq nt \Rightarrow tS_j - jt^2/2 \geq tS_j - nt^2/2 \geq nt^2/2 \Rightarrow \frac{e^{tS_j}}{E[e^{tS_j}]} \geq \frac{e^{tS_j}}{e^{jt^2/2}} \geq e^{nt^2/2}.$$

Recall from 8.15 that $M_n = e^{tS_n}/E[e^{tS_n}]$ is a mean 1 martingale. Hence, using the above and Doob's inequality we get

$$P\Big(\max_{j\leq n} S_j \geq nt\Big) \leq P\Big(\max_{j\leq n} M_j \geq e^{nt^2/2}\Big) \leq e^{-nt^2/2} E[M_n] = e^{-nt^2/2}.$$

Alternatively, note that e^{tS_n} is a submartingale and so Doob's inequality still applies. It gives an upper bound that involves the expected value of the plus part of the submartingale, but our submartingale is nonnegative. So we get

$$P\Big(\max_{j\leq n} S_j \geq nt\Big) \leq P\Big(\max_{j\leq n} e^{tS_j} \geq e^{nt^2}\Big) \leq e^{-nt^2} E[e^{tS_n}] \leq e^{-nt^2/2}.$$

Now that (8.40) from the book has been shown to hold, the claim in part (3) follows word for word as in the book (right below (8.40) therein). To get (4) note that if we let $X_n = 2Y_n - 1$, then we are back in the setting of part (1). Therefore, the result in part (3) applies to $S_n = 2T_n - n$, which gives part (4).

8.40. Since $T < \infty$ almost surely we know that $T \wedge n \to T$ almost surely. Consequently, $X_{T\wedge n} \to X_T$ almost surely and hence also in probability. By 4.28(4) uniform integrability and convergence in probability imply convergence in L^1. But then this means that $E[X_{T\wedge n}] \to E[X_T]$. However, the optional stopping theorem says that $E[X_{T\wedge n}] = E[X_1]$ for all n. The claim follows.

8.51. Let (Z_n) be i.i.d. random variables with $P(Z_1 = 1) = P(Z_1 = -1) = 1/2$. Let $X_0 = 0$ and $X_n = Z_1 + \cdots + Z_n$ for $n \geq 1$. We have already seen that this is a mean-zero martingale. Let $T = \inf\{n \geq 0 : X_n = 1\}$. We have seen in Problem 8.14 that $T < \infty$ almost surely and that $E[T] = \infty$. But $X_T = 1$ almost surely and hence $E[X_T] = 1 \neq 0$.