Kenneth DeMason, September 10, 2023

The following is a compilation of the problems I attempted from Gordan Zitkovic's M385C - Probability Theory I Fall 2022 course at UT Austin. These problems come from his course notes.

I tend to be very thorough with my solutions. This is especially apparent in the first few homework sets, where I was still gauging how much work was expected of us. Needless to say, you do not need to show every single step as I did. Please do not hesitate to reach out for clarification.

Only problems listed with a subtitle are included in the table of contents. You can search for the remaining ones by using useful keywords.

Contents

HW 1

Problem 1.1. Let $(S, \mathcal{S})$ be a measurable space, and let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{S}$. Show that the following sets are also in $\mathcal{S}$.

1. The set C_1 of all $x \in S$ such that $x \in A_n$ for at least 5 different values of n.
2. The set C_2 of all $x \in S$ such that $x \in A_n$ for exactly 5 different values of n.
3. The set C_3 of all $x \in S$ such that $x \in A_n$ for all but finitely many n ("finitely many" includes none).
4. The set C_4 of all $x \in S$ such that $x \in A_n$ for at most finitely many values of n.

Solution: We express each C_i as the countable union/intersection of measurable sets, which by properties of σ-algebras implies that the C_i are measurable.

1. Let $\mathcal{I}^k_{\geq 5} = \{I \in \mathcal{P}(\{1,...,k\}) \mid |I| \geq 5\}$. Note that $\mathcal{I}^k_{\geq 5}$ is finite for all k (in fact empty for $k = 1,...,4$). Then we can write C_1 as
$$C_1 = \bigcup_{k=5}^{\infty} \bigcup_{I\in\mathcal{I}^k_{\geq 5}} \bigcap_{i\in I} A_i.$$
2. Let $\mathcal{I}^k_5 = \{I \in \mathcal{P}(\{1,...,k\}) \mid |I| = 5\}$. Note that $\mathcal{I}^k_5$ is finite for all k (in fact empty for $k = 1,...,4$). Then we can write C_2 as
$$C_2 = \bigcup_{k=5}^{\infty} \bigcup_{I\in\mathcal{I}^k_5} \left(\bigcap_{i\in I} A_i \cap \bigcap_{j\in\mathbb{N}\setminus I} A_j^c \right).$$
3. C_3 is precisely the set of $x \in S$ such that x is eventually in all the A_n. That is,
$$C_3 = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n.$$
This is also the lim inf of the A_n.
4. If x is in A_n for at most finitely many n, then $x \in A_n^c$ for all but finitely many n. Thus
$$C_4 = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c.$$

Problem 1.2: (Atomic structure of algebras). A **partition** of a set S is a family $\mathcal{P}$ of non-empty subsets of S with the property that each $x \in S$ belongs to exactly one $A \in \mathcal{P}$.

1. How many algebras are there on the set $S = \{1, 2, 3\}$?
2. By constructing a bijection between the two families, show that the number of different algebras on a finite set S is equal to the number of different partitions of S. Note: The elements of the partition corresponding to an algebra are said to be its **atoms**.
3. Does there exist an algebra with 754 elements?

Solution:

1. There are five, as follows
$$\mathcal{S}_1 = \{\varnothing, \{1\}, \{2,3\}, \{1,2,3\}\}, \quad \mathcal{S}_2 = \{\varnothing, \{2\}, \{1,3\}, \{1,2,3\}\}, \quad \mathcal{S}_3 = \{\varnothing, \{3\}, \{1,2\}, \{1,2,3\}\}$$
$$\mathcal{S}_4 = \{\varnothing, \{1,2,3\}\}, \quad \mathcal{S}_5 = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$
2. Let $\mathcal{S}$ be an algebra of S. We form a partition as follows: For $s \in S$ define A_s by
$$A_s := \bigcap_{s\in A\in\mathcal{S}} A.$$
Since $s \in A_s$ by definition each is nonempty. We just need to show if $t \in S$ then
$$A_s \cap A_t = \begin{cases} A_s & \text{if } t \in A_s \\ \varnothing & \text{if } t \notin A_s \end{cases}.$$

Suppose first $t \in A_s$. Then if $s \in A$ for some $A \in \mathcal{S}$ it must also be that $t \in A$. Hence,

$$A_s := \bigcap_{s \in A \in \mathcal{S}} A \subseteq \bigcap_{t \in A \in \mathcal{S}} A = A_t.$$

On the other hand, since $t \in A_s$ it is a valid candidate for the sets used to define A_t. Thus,

$$A_t := \bigcap_{t \in A \in \mathcal{S}} A \subseteq A_s.$$

Together the above two imply $A_s = A_t$. Next suppose $r \in A_s \cap A_t$. Let $A \in \mathcal{S}$ be such that $s \in A$. Then since $r \in A_s$ we necessarily have $r \in A$. If $t \notin A$ then $t \in A^c$, but also because $r \in A_t$ we must have $r \in A^c$, a contradiction. So $s \in A$ implies $t \in A$, and by applying the same logic as above we get $A_s \subseteq A_t$. Reversing roles of s and t proves the other containment, so $A_s = A_t$. In other words, either $A_s \cap A_t = A_s = A_t$ or is empty.

So there is a map $F : \mathscr{S} \to \mathscr{P}$ where $\mathscr{S}$ and $\mathscr{P}$ denote the collections of algebras on S and partitions of S respectively. On the other hand there is a natural map $G : \mathscr{P} \to \mathscr{S}$ defined by $G(\mathcal{P}) = \sigma(\mathcal{P})$ (here the notion of σ-algebra and algebra coincide). We claim that $G \circ F = \mathrm{Id}_{\mathscr{S}}$ while $F \circ G = \mathrm{Id}_{\mathscr{P}}$.

To show the former note that for any algebra $\mathcal{S}$ on S we have by construction $F(\mathcal{S}) \subset \mathcal{S}$. Thus $G(F(\mathcal{S})) \subseteq \mathcal{S}$. Now let $A \in \mathcal{S}$ and write $A = \{s_i\}_{i=1}^n$. Since $A_{s_i} \subseteq A$ for all $i = 1, ..., n$ it follows that $\cup_i A_{s_i} \subseteq A$. By our enumeration of elements of A, and recalling that $s_i \in A_{s_i}$, it is clear that $A \subseteq \cup_i A_{s_i}$. Hence $A = \cup_i A_{s_i} \in G(F(\mathcal{S}))$ since for all i we have $A_{s_i} \in F(\mathcal{S})$. Thus $\mathcal{S} \subseteq G(F(\mathcal{S}))$.

To show the latter, for any partition $\mathcal{P}$ of S write $\mathcal{P} = \{P_i\}_{i=1}^n$. Let $s \in P_i$. Since A_s is defined by intersecting all the sets in $G(\mathcal{P})$ containing s, and $P_i \in G(\mathcal{P})$, it follows that $A_s \subseteq P_i$. Since all the P_j are disjoint, $G(\mathcal{P})$ is given by

$$G(\mathcal{P}) = \left\{ \bigcup_{j \in I} P_j \;\middle|\; I \subseteq \{1, ..., n\} \right\}.$$

(if $I = \varnothing$ we interpret the union as the empty union). In particular if $A \in G(\mathcal{P})$ then $A = \cup_{j \in I} P_j$ for some $I \subset \{1, ..., n\}$. Because $s \in P_i$ and $s \notin P_j$ for $j \neq i$ it follows that $s \in A$ if and only if $i \in I$, or $P_i \subseteq A$. Thus $P_i \subseteq A$ for any $A \in G(\mathcal{P})$ with $s \in A$, and as A_s is the intersection of all such sets we have $P_i \subseteq A_s$.

3. First since $G(F(\mathcal{S})) = \mathcal{S}$ it follows that every algebra on S is generated by a partition. So it suffices to consider $G(\mathcal{P})$ for some partition of S.

   In the above description of $G(\mathcal{P})$ there are 2^n many choices of I. Since every algebra on $\mathcal{S}$ is generated this way, they must all have 2^k many elements for some $k \in \mathbb{N}$. 754 is not of this form, so there is not an algebra with 754 elements.

Problem 1.3: (Notes Problem 1.18). Show that $f : \mathbb{R} \to \mathbb{R}$ is measurable if it is either monotone or convex.

Solution:

1. Let $f : \mathbb{R} \to \mathbb{R}$ be monotone non-decreasing (since $-f$ is measurable whenever f is, it suffices to consider such functions only). Set $l = \lim_{x \to -\infty} f(x)$ and consider (a, ∞) where possibly $a, l = -\infty$. There are two cases:
   - $a \leq l$: In this case we have $f^{-1}((a, \infty)) = \mathbb{R}$ since f being monotone non-decreasing implies for $x \in \mathbb{R}$ that $l \leq f(x) < \infty$.
   - $l < a$: We have to account for jump discontinuities in this case. Let

     $$t = \inf \{x \in \mathbb{R} \mid f(x) \geq a\}.$$

     If $f(t) > a$ then $f^{-1}((a, \infty)) = [t, \infty)$, otherwise $f^{-1}((a, \infty)) = (t, \infty)$.
2. Convex functions on $\mathbb{R}$ are continuous, hence measurable. To see this recall that Lipschitz implies continuous, and since continuity is a local property we need only check that convexity

implies locally Lipschitz. To this end let $z \in \mathbb{R}$, $\delta > 0$, and consider $(x, y) \subset \mathbb{R}$ containing z. Set

$$L := \sup_{w \in (x,y)} \left\{ \left| \frac{f(x) - f(w)}{x - w} \right|, \left| \frac{f(w) - f(y)}{w - y} \right| \right\}.$$

We show that f is Lipschitz on (x, y) with Lipschitz constant L. That is, for all $w_1, w_2 \in (x, y)$ we have

$$|f(w_2) - f(w_1)| \leq L|w_2 - w_1|.$$

Without loss of generality assume $w_1 < w_2$. Then we can write $w_1 = (1-t)w_2 + tx$, and by convexity of f

$$f(w_1) \leq (1-t)f(w_2) + tf(x) = f(w_2) + t[f(x) - f(w_2)].$$

Rearranging this and applying $t = (w_1 - w_2)/(x - w_2)$, noting $w_1 - w_2 < 0$, we get

$$\frac{f(w_1) - f(w_2)}{w_1 - w_2} \geq \frac{f(x) - f(w_2)}{x - w_2} \geq -L.$$

By the same logic we get

$$\frac{f(w_2) - f(w_1)}{w_2 - w_1} \leq \frac{f(y) - f(w_1)}{y - w_1} \leq L.$$

The two together show

$$\left| \frac{f(w_2) - f(w_1)}{w_2 - w_1} \right| \leq L,$$

as desired.

Problem 1.4: (Notes Problem 1.10). One can obtain the product σ-algebra $\mathcal{S}$ on $\{-1, 1\}^{\mathbb{N}}$ as the Borel σ-algebra corresponding to a particular topology which makes $\{-1, 1\}^{\mathbb{N}}$ compact. Here is how. Start by defining a mapping $d : \{-1, 1\}^{\mathbb{N}} \times \{-1, 1\}^{\mathbb{N}} \to [0, \infty)$ by

$$d(\mathbf{s}^1, \mathbf{s}^2) = 2^{-i(\mathbf{s}^1, \mathbf{s}^2)}, \quad \text{where } i(\mathbf{s}^1, \mathbf{s}^2) = \inf\{i \in \mathbb{N} \mid s_i^1 \neq s_i^2\},$$

for $\mathbf{s}^j = (s_1^j, s_2^j, ...)$, $j = 1, 2$.

1. Show that d is a metric on $\{-1, 1\}^{\mathbb{N}}$.
2. Show that $\{-1, 1\}^{\mathbb{N}}$ is compact under d. Hint: Use the diagonal argument.
3. Show that each cylinder of $\{-1, 1\}^{\mathbb{N}}$ is both open and closed under d.
4. Show that each open ball is a cylinder.
5. Show that $\{-1, 1\}^{\mathbb{N}}$ is separable, i.e. it admits a countable dense subset.
6. Conclude that $\mathcal{S}$ coincides with the Borel σ-algebra on $\{-1, 1\}^{\mathbb{N}}$ under the metric d.

Solution:

1.
   - By definition d maps into $[0, \infty)$, so we need only check that $d(\mathbf{s}^1, \mathbf{s}^2) = 0$ implies $\mathbf{s}^1 = \mathbf{s}^2$. If $d(\mathbf{s}^1, \mathbf{s}^2) = 0$ then necessarily $i(\mathbf{s}^1, \mathbf{s}^2) = \infty$, i.e. $s_i^1 = s_i^2$ for all $i \in \mathbb{N}$.
   - We have $i(\mathbf{s}^1, \mathbf{s}^2) = i(\mathbf{s}^2, \mathbf{s}^1)$ since the definition of $i(\cdot, \cdot)$ is symmetric. Hence d is too.
   - Given $\mathbf{s}^j$ for $j = 1, 2, 3$ compare $i(\mathbf{s}^1, \mathbf{s}^3)$ and $i(\mathbf{s}^2, \mathbf{s}^3)$ to $i(\mathbf{s}^1, \mathbf{s}^2)$. Since $\mathbf{s}^1$ and $\mathbf{s}^2$ differ at $i(\mathbf{s}^1, \mathbf{s}^2)$, and they can only take one of two values, it must be that $i(\mathbf{s}^k, \mathbf{s}^3) \leq i(\mathbf{s}^1, \mathbf{s}^2)$ for $k = 1$ or $k = 2$. Hence,

   $$d(\mathbf{s}^1, \mathbf{s}^2) = 2^{-i(\mathbf{s}^1, \mathbf{s}^2)} \leq 2^{-i(\mathbf{s}^k, \mathbf{s}^3)} < 2^{-i(\mathbf{s}^1, \mathbf{s}^3)} + 2^{-i(\mathbf{s}^2, \mathbf{s}^3)} = d(\mathbf{s}^1, \mathbf{s}^3) + d(\mathbf{s}^2, \mathbf{s}^3).$$

2. Let $\{\mathbf{s}^j\}_{j=1}^{\infty} \subset \{-1, 1\}^{\mathbb{N}}$. Construct $\mathbf{s}$ as follows: Consider the sequence of first entries $\{s_1^j\}_{j=1}^{\infty}$. As a bounded sequence there exists a convergent subsequence, which after possibly relabelling we continue to write as $\{s_1^j\}_{j=1}^{\infty}$. Consider now $\{s_2^j\}_{j=1}^{\infty}$ and extract a convergent subsequence, etc. In this way we end up with $\{\mathbf{s}^j\}_{j=1}^{\infty}$ such that for each $i \in \mathbb{N}$ the sequence $\{s_i^j\}_{j=1}^{\infty}$ converges. Let $s_i = \lim_{j \to \infty} s_i^j$ and define $\mathbf{s} = (s_1, s_2, ...)$.

   Since the sequences $\{s_i^j\}_{j=1}^{\infty}$ take only finitely many values, for them to converge they must eventually be constant. Let $j_i = 1 + \sup\{j \in \mathbb{N} \mid s_i^j \neq s_i\}$ (the empty supremum is interpreted as zero). It follows that $i(\mathbf{s},\mathbf{s}^j) > k$ for all $j \geq J_k := \max\{j_1, ..., j_k\}$. In particular for any $k \in \mathbb{N}$ there exists J_k such that for $j \geq J_k$ we have

$$d(\mathbf{s},\mathbf{s}^j) = 2^{-i(\mathbf{s},\mathbf{s}^j)} < 2^{-k}.$$

So, $d(\mathbf{s},\mathbf{s}^j) \to 0$.

3. Recall that the cylinders can be described in the following way: For each $C \in \mathcal{C}$ there exists an $n \in \mathbb{N}$ and $B \subset \{-1,1\}^n$ such that

$$C = \{\mathbf{s} \in \{-1,1\}^{\mathbb{N}} \mid (s_1, ..., s_n) \in B\}$$

Let $\{\mathbf{s}^j\}_{j=1}^{\infty} \subset C$ be such that $\mathbf{s}^j \to \mathbf{s} \in \{-1,1\}^{\mathbb{N}}$. We first show that in fact $\mathbf{s} \in C$, so that C is closed. Since $d(\mathbf{s}^j,\mathbf{s}) \to 0$, in particular there exists a $J \in \mathbb{N}$ such that for $j \geq J$,

$$d(\mathbf{s}^j,\mathbf{s}) = 2^{-i(\mathbf{s}^j,\mathbf{s})} < 2^{-n},$$

in other words $i(\mathbf{s}^j,\mathbf{s}) > n$. This implies for and $j \geq J$ and $i \leq n$ we have $s_i^j = s_i$. But because $\mathbf{s}^j \in C$ for all j we have $(s_1^j, ..., s_n^j) \in B$. Hence $(s_1, ..., s_n) \in B$, and $\mathbf{s} \in C$.

Next we show C is open. Let $\mathbf{s} \in C$ and choose $0 < \epsilon < 2^{-n}$. Then if $\mathbf{s}' \in \{-1,1\}^{\mathbb{N}}$ is such that $d(\mathbf{s},\mathbf{s}') < \epsilon < 2^{-n}$ we have, by the same logic as above that $i(\mathbf{s},\mathbf{s}') > n$. Again by the same logic this means $s_i' = s_i$ for $i = 1, ..., n$ and in particular as $(s_1, ..., s_n) \in B$ we have $(s_1', ..., s_n') \in B$. Thus $\mathbf{s}' \in C$, and C is open.

4. Let $\mathbf{s} \in \{-1,1\}^{\mathbb{N}}$ and $\epsilon > 0$. Then there exists $n \in \mathbb{N}$ such that $2^{-n-1} \leq \epsilon < 2^{-n}$. If now $\mathbf{s}' \in \{-1,1\}^{\mathbb{N}}$ is such that

$$2^{-i(\mathbf{s},\mathbf{s}')} = d(\mathbf{s},\mathbf{s}') < \epsilon < 2^{-n}$$

then $i(\mathbf{s},\mathbf{s}') > n$ implies $s_i = s_i'$ for $i = 1, ..., n$. Hence,

$$B_\epsilon(\mathbf{s}) = \{\mathbf{s}' \in \{-1,1\}^{\mathbb{N}} \mid (s_1', ..., s_n') \in B\} \in \mathcal{C}$$

where $B = \{(s_1, ..., s_n)\}$.

5. Any compact metric space (X, d) is separable. To see this Let $r_n = 1/n$ for $n \in \mathbb{N}$ and consider $\{B_{r_n}(x)\}_{x \in X}$. As an open cover of X, owing to compactness it admits a finite subcover $\{B_{r_n}(x_n^k)\}_{k=1}^{k_n}$. Let

$$D = \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{k_n} \{x_n^k\}.$$

Then D is countable and dense by construction.

6. We have that $\sigma(\mathcal{C}) = \mathcal{S}$. The Borel σ-algebra on $(\{-1,1\}^{\mathbb{N}}, d)$, being generated by the open sets, thus contains $\sigma(\mathcal{C})$ since every cylinder is open.

Next we show $\mathcal{B}(\{-1,1\}^{\mathbb{N}}, d) \subseteq \sigma(\mathcal{C})$ by showing $\tau \subset \sigma(\mathcal{C})$, where τ is the topology on $\{-1,1\}^{\mathbb{N}}$ induced by d. To do this let (X, d) be any separable metric space, D a countable dense subset, U open, and

$$\mathscr{B} := \{B_r(x) \subset U \mid x \in D,\ r \in \mathbb{Q}\}.$$

We show that

$$U = \bigcup_{B \in \mathscr{B}} B.$$

Since each $B \in \mathscr{B}$ is a subset of U, it suffices to show the forward inclusion. So let $x \in U$, then there exists $R > 0$ such that $B_R(x) \subset U$. On the other hand since D is dense there exists an $x' \in D$ such that $d(x, x') < r < R/2$ with $r \in \mathbb{Q}$. Then for any $z \in B_r(x')$ we have

$$d(z, x) \leq d(z, x') + d(x, x') < 2r < R$$

so that

$$x \in B_r(x') \subset B_R(x) \subset U.$$

Hence $B_r(x') \in \mathscr{B}$ and $x \in B_r(x')$ implying the forward inclusion.

In our case, each ball is a cylinder. So given $U \in \tau$ we can express it as the countable union of cylinders. That is, $U \in \sigma(\mathcal{C})$.

HW 2

Problem 2.1: (Stronger separation, Notes Problem 2.7). Let $(S, \mathcal{S}, \mu)$ be a measure space and let $f, g \in L^0(S, \mathcal{S})$ satisfy $\mu(\{x \in S \mid f(x) < g(x)\}) > 0$. Prove or construct a counterexample for the following statement:

"There exist constants $a, b \in \mathbb{R}$ such that $\mu(\{x \in S \mid f(x) \leq a < b \leq g(x)\}) > 0$."

Solution: For notational purposes define

$$E = \{x \in S \mid f(x) < g(x)\}$$
$$E_{a,b} = \{x \in S \mid f(x) \leq a < b \leq g(x)\}$$

so that we know $\mu(E) > 0$ and we either want to find a, b such that $\mu(E_{a,b}) > 0$ or show none exist. We prove the former.

Whenever $x \in E$ we can always find $p, q \in \mathbb{Q}$ such that $f(x) \leq p < q \leq g(x)$; that is $x \in E_{p,q}$ for some $p, q \in \mathbb{Q}$. Hence

$$E = \bigcup_{p,q \in \mathbb{Q}} E_{p,q}.$$

By subadditivity,

$$0 < \mu(E) = \mu\left(\bigcup_{p,q \in \mathbb{Q}} E_{p,q}\right) \leq \sum_{p,q \in \mathbb{Q}} \mu(E_{p,q}).$$

If all the $E_{p,q}$ were μ-null then the sum on the right-hand side would be zero, a contradiction.

Problem 2.2: (A uniform distribution on a circle, Notes Problem 2.11). Let S^1 be the unit circle, and let $f : [0, 1) \to S^1$ be the "winding map"

$$f(x) = (\cos(2\pi x), \sin(2\pi x)), \quad x \in [0, 1).$$

1. Show that the map f is $(B([0, 1)), \mathcal{S}^1)$-measurable, where $\mathcal{S}^1$ denotes the Borel σ-algebra on S^1 (with the topology inherited from $\mathbb{R}^2$).
2. For $\alpha \in (0, 2\pi)$ let R_α denote the (counter-clockwise) rotation of $\mathbb{R}^2$ with center $(0, 0)$ and angle α. Show that $R_\alpha(A) := \{R_\alpha(x) \mid x \in A\}$ is in $\mathcal{S}^1$ if and only if $A \in \mathcal{S}^1$.
3. Let μ^1 be the push-forward of the Lebesgue measure λ by the map f. Show that μ^1 is rotation-invariant, i.e. that $\mu^1(A) = \mu^1(R_\alpha(A))$. Note: The measure μ^1 is called the **uniform measure** (or the **uniform distribution**) on S^1.

Solution:

1. The topology on S^1 is the subspace topology, so $V \subset S^1$ is open if and only if there exists $U \subset \mathbb{R}^2$ open such that $V = U \cap S^1$. The map above, regarded as a map $[0, 1) \to \mathbb{R}^2$ is continuous. Moreover for any $V \subset S^1$ open we have

   $$f^{-1}(V) = f^{-1}(U \cap S^1) = f^{-1}(U),$$

   the last equality holding trivially since f maps into S^1. By continuity we have that $f^{-1}(V)$ is open in $[0, 1)$, and thus lies in $\mathcal{B}([0, 1))$.
2. The map $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ is continuous. For $V \subset S^1$ open we have $V = U \cap S^1$ for $U \subset \mathbb{R}^2$ open and

   $$R_\alpha^{-1}(V) = R_\alpha^{-1}(U) \cap S^1$$

   so that $R_\alpha : \mathbb{R}^2 \to S^1$ is continuous. Finally the inclusion $i_{S^1} : S^1 \to \mathbb{R}^2$ is continuous so that the composition $R_\alpha \circ i_{S^1} : S^1 \to S^1$ is too. Denoting this composition still by R_α we see that $R_\alpha^{-1}(V) \in \mathcal{S}^1$ whenever $V \subset S^1$ is open. So R_α is $(\mathcal{S}^1, \mathcal{S}^1)$-measurable.

   Now let $\alpha \in (0, 2\pi)$ and suppose $R_\alpha(A) \in \mathcal{S}^1$. Then $A = R_\alpha^{-1}(R_\alpha(A)) \in \mathcal{S}^1$. On the other hand suppose $A \in \mathcal{S}^1$. We showed R_α is $(\mathcal{S}^1, \mathcal{S}^1)$-measurable for any $\alpha \in (0, 2\pi)$, so in particular R_β is with $\beta = 2\pi - \alpha$. Finally $R_\alpha(A) = R_\beta^{-1}(A) \in \mathcal{S}^1$.

3. Define $g : [0,2) \to [0,1)$ by

$$g(x) = \begin{cases} x & x < 1 \\ x - 1 & 1 \le x \end{cases}.$$

By translation invariance of the Lebesgue measure we know that if $B \in \mathcal{B}([0,1))$ and $x \in [0,1)$ then $\lambda(g(B+x)) = \lambda(B)$. Note that for $\alpha \in (0, 2\pi)$ and $A \in \mathcal{S}^1$ we have

$$f^{-1}(R_\alpha(A)) = g\left(f^{-1}(A) + \frac{\alpha}{2\pi}\right)$$

and thus by translation invariance,

$$\mu^1(R_\alpha(A)) = \lambda(f^{-1}(R_\alpha(A))) = \lambda(f^{-1}(A)) = \mu^1(A).$$

Problem 2.3: (A change-of-variable formula, Notes Problem 3.19). Let $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \nu)$ be two measure spaces, and let $F : S \to T$ be a measurable function with the property that $\nu = F_{\#}\mu$ (i.e., ν is the push-forward of μ through F). Show that for every $f \in L^0_+(T, \mathcal{T})$ or $f \in L^1(T, \mathcal{T})$, we have

$$\int_T f \; d\nu = \int_S (f \circ F) \; d\mu.$$

Solution: First let $B \in \mathcal{T}$ and consider $f = \chi_B$. Then the above reads exactly as

$$\int_T \chi_B \; d\nu = \nu(B) = \mu(F^{-1}(B)) = \int_S \chi_{F^{-1}(B)} \; d\mu = \int_S \chi_B \circ F \; d\mu.$$

The last equality holds since $\chi_B(F(s)) = 1$ whenever $F(s) \in B$ and $\chi_B(F(s)) = 0$ whenever $F(s) \notin B$. Now notice that the identity in question is linear in f. That is, if f_1, f_2 satisfy the identity and $\alpha \in [0, \infty)$ then

$$\int_T (\alpha f_1 + f_2) \; d\nu = \alpha \int_T f_1 \; d\nu + \int_T f_2 \; d\nu = \alpha \int_S (f_1 \circ F) \; d\mu + \int_S (f_2 \circ F) \; d\mu = \int_S [\alpha f_1 + f_2] \circ F \; d\mu.$$

Thus the identity holds on all non-negative simple functions, and by monotone convergence for all $f \in L^0_+(T, \mathcal{T})$. Precisely, if $f \in L^0_+(T, \mathcal{T})$ then there exists a sequence of non-negative simple functions $\{g_n\}_{n=1}^\infty$ such that $g_n \le g_{n+1}$ and $g_n \to f$. Then by monotone convergence,

$$\int_T f \; d\nu = \lim_{n\to\infty} \int_T g_n \; d\nu = \lim_{n\to\infty} \int_S g_n \circ F \; d\mu = \int_S f \circ F \; d\mu.$$

Finally if $f \in L^1(T, \mathcal{T})$ we split it into its positive and negative parts and apply the previous result to each part.

Problem 2.4: (An integrability criterion, Notes Problem 3.14). Let $(S, \mathcal{S}, \mu)$ be a finite measure space, and let $f \in L^0_+$. Show that

$$\int_S f \; d\mu < \infty \text{ if and only if } \sum_{n\in\mathbb{N}} \mu(\{f \ge n\}) < \infty$$

where, as usual, $\{f \ge n\} = \{x \in S \mid f(x) \ge n\}$. Hint: Approximate f from below and from above by a piecewise constant function.

Solution: Consider

$$\phi = \sum_{n=1}^\infty \chi_{\{f \ge n\}}.$$

Notice that $\phi(x) \le f(x) < \phi(x) + 1$. To see this, if $x \in S$ and $N \in \mathbb{N}$ is such that $N \le f(x) < N+1$ then for all $n \le N$ we have $x \in \{f \ge n\}$ and if $n > N$ then $x \notin \{f \ge n\}$. It follows that $\phi(x) = N \le f(x) < N + 1 = \phi(x) + 1$. Then,

$$\sum_{n=1}^\infty \mu(\{f \ge n\}) \le \int_S f \; d\mu < \sum_{n=1}^\infty \mu(\{f \ge n\}) + \mu(S).$$

Owing to the fact that $\mu(S) < \infty$ we conclude the result.

Problem 2.5: (Asymptotic density, Notes Problem 2.12). We say that the subset A of $\mathbb{N}$ **admits asymptotic density** if the limit

$$D(A) = \lim_{n\to\infty} \frac{\#(A \cap \{1, ..., n\})}{n}$$

exists (remember that # denotes the number of elements of a set). Let $\mathcal{D}$ be the collection of all subsets of $\mathbb{N}$ that admit asymptotic density.

1. Is D a finitely-additive set function on $\mathcal{D}$? How about σ-additive?
2. Is $\mathcal{D}$ a σ-algebra?
3. Is $\mathcal{D}$ an algebra?

Solution: Changing the notation, we denote by $\mathcal{H}^0$ the counting measure and $\mathcal{H}^0_n = \mathcal{H}^0 \llcorner \{1, ..., n\}$. For $A \subseteq \mathbb{N}$ define $s_n(A) = \mathcal{H}^0_n(A)/n$. Thus,

$$D(A) = \lim_{n\to\infty} \frac{\mathcal{H}^0_n(A)}{n} = \lim_{n\to\infty} s_n(A).$$

1. Let's start by trying to prove σ-additivity. Given $\{A_k\}_{k=1}^\infty \subset \mathcal{D}$, since $\mathcal{H}^0_n$ is a measure for every $n \in \mathbb{N}$ we have

$$\mathcal{H}^0_n\left(\bigcup_{k=1}^\infty A_k\right) = \sum_{k=1}^\infty \mathcal{H}^0_n(A_k).$$

Dividing both sides by n yields

$$s_n\left(\bigcup_{k=1}^\infty A_k\right) = \sum_{k=1}^\infty s_n(A_k)$$

and as $n \to \infty$ we get

$$D\left(\bigcup_{k=1}^\infty A_k\right) = \lim_{n\to\infty} \sum_{k=1}^\infty s_n(A_k).$$

The question now becomes: can we exchange the limit and summation? In general this is not possible. Noting that sums are really integrals (with respect to the counting measure), we may try to appeal to a convergence theorem. Neither appear to hold (the sequence $s_n(A)$ need not be monotone for a given $A \in \mathcal{D}$, and each $s_n(A) \leq 1$ but the constant function 1 is not integrable with respect to the counting measure). For example consider $A_k = \{k\}$ so that

$$s_n(A_k) = \begin{cases} 0 & n < k \\ 1/n & n \geq k \end{cases}$$

and thus $D(A_k) = 0$. However, $\bigcup_k A_k = \mathbb{N}$ which has $D(\mathbb{N}) = 1$. Hence D is not σ-additive on $\mathcal{D}$ as

$$1 = D(\mathbb{N}) = D\left(\bigcup_{k=1}^\infty A_k\right) \neq \sum_{k=1}^\infty D(A_k) = 0.$$

Of course if the union is finite then we can immediately exchange the limit and sum. Thus for $\{A_k\}_{k=1}^K \subset \mathcal{D}$ disjoint we have

$$D\left(\bigcup_{k=1}^K A_k\right) = \lim_{n\to\infty} \sum_{k=1}^K s_n(A_k) = \sum_{k=1}^K \lim_{n\to\infty} s_n(A_k) = \sum_{k=1}^K D(A_k).$$

Hence D is a finitely additive set function on $\mathcal{D}$. Since all the $A_k \in \mathcal{D}$, the right hand side is finite. It follows that $\cup_{k=1}^K A_k$ admits asymptotic density, and $\mathcal{D}$ is an algebra.

2. Note for any finite $A \subset \mathbb{N}$ we have $A \in \mathcal{D}$, since if $n \in \mathbb{N}$ is such that all $x \in A$ have $x \leq n$ then $s_n(A) = \mathcal{H}^0(A)/n$, which goes to zero. Now let

$$A_k = \{n \mid a(k) \leq n < b(k)\}, \quad A = \bigcup_{k=1}^\infty A_k$$

where $a(1) = 1$ and $a(k) = 1 + 2 \cdot 3^{k-2}$ for $k \geq 2$, and $b(1) = 2$ and $b(k) = 1 + 4 \cdot 3^{k-2}$ for $k \geq 2$. Another way to see the construction of A is we alternate between including and omitting natural numbers in chunks of length 1, then $2 \cdot 3^{k-1}$. That is,

$$A = \{1, 3, 4, 7, 8, 9, 10, 11, 12, 19, ..., 36, ...\}.$$

Now set $\bar{a}(k) = 4 \cdot 3^{k-2}$ and $\bar{b}(k) = 2 \cdot 3^{k-1}$, with $\bar{a}(1) := 1$. We compute $s_{\bar{a}(k)}(A)$ and $s_{\bar{b}(k)}(A)$ as follows:

$$\begin{aligned} s_{\bar{a}(k)}(A) &= \frac{1 + \sum_{i=1}^{k-1} 2 \cdot 3^{i-1}}{\bar{a}(k)} = \frac{3^{2-k}}{4} \left(1 + 2 \sum_{i=1}^{k-1} 3^{i-1}\right) \\ &= \frac{3^{2-k}}{4} \left(1 + 2\left(\frac{3^{k-1} - 1}{2}\right)\right) = \frac{3}{4} \\ s_{\bar{b}(k)}(A) &= \frac{1 + \sum_{i=1}^{k-1} 2 \cdot 3^{i-1}}{\bar{b}(k)} = \frac{3^{1-k}}{2} \left(1 + 2 \sum_{i=1}^{k-1} 3^{i-1}\right) = \frac{1}{2}. \end{aligned}$$

where $\sum_{i=1}^{0}$ is interpreted as the empty sum. Since $\bar{a}(k)$ and $\bar{b}(k)$ are increasing sequences, it follows that $s_n(A)$ is 3/4 infinitely often and 1/2 infinitely often. In fact $1/2 \leq s_n(A) \leq 3/4$ so that $\limsup_{n\to\infty} s_n(A) = 3/4$ and $\liminf_{n\to\infty} s_n(A) = 1/2$. Thus the limit does not exist, and $A \notin \mathcal{D}$.

3. See end of 1. We only need to check that $\mathcal{D}$ is closed under finite unions of disjoint sets as we can disjointify if necessary.

HW 3

Problem 3.1: (Notes Problem 3.5). Let $(S, \mathcal{S}, \mu)$ be a measure space, and suppose $f \in \mathcal{L}^1_+$ is such that $\int f \, d\mu = c > 0$. Show that for each $\alpha > 0$ the limit

$$\lim_{n\to\infty} \int n \log\left(1 + \left(\frac{f}{n}\right)^\alpha\right) d\mu$$

exists in $[0, \infty]$ and compute its value. *Hint*: Consider the cases $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$ separately. In the case $\alpha > 1$, prove and use the inequality $\log(1 + x^\alpha) \leq \alpha x$, valid for $x \geq 0$.

Solution: We prove the hint first. Consider $F_\alpha(x) = 1 + x^\alpha$ and $G_\alpha(x) = e^{\alpha x}$. We equivalently prove $F_\alpha(x) \leq G_\alpha(x)$ for $x \geq 0$. Observing $F_\alpha(0) = G_\alpha(0)$ it suffices to show $F'_\alpha(x) \leq G'_\alpha(x)$ for $x \geq 0$. Indeed, $F'_\alpha(x) = \alpha x^{\alpha-1}$ and $G'_\alpha(x) = \alpha e^{\alpha x}$, but exponentials grow faster than polynomials [valid when $\alpha \geq 1$].

The idea will be to use a convergence theorem, and so we should compute what

$$\lim_{n\to\infty} n \log\left(1 + \left(\frac{f(x)}{n}\right)^\alpha\right)$$

is for various values of α. To this end let

$$\ell_{n,\alpha}(x) = n \log\left(1 + \left(\frac{x}{n}\right)^\alpha\right).$$

We have that $\lim_{n\to\infty} \ell_{n,\alpha}(0) = 0$ for any $\alpha > 0$. To compute the limit at other values of x let's use the hint in the following weak form: $\log(1 + x) \leq x$. Applying this with $1/x - 1$ yields

$$-\log(x) = \log\left(1 + \left[\frac{1}{x} - 1\right]\right) \leq \frac{1}{x} - 1, \quad \Rightarrow \quad \log(x) \geq 1 - \frac{1}{x} = \frac{x-1}{x}.$$

Applying this now to $1 + (x/n)^\alpha$ yields

$$\log\left(1 + \left(\frac{x}{n}\right)^\alpha\right) \geq \frac{(x/n)^\alpha}{1 + (x/n)^\alpha} = \frac{x^\alpha}{n^\alpha + x^\alpha}.$$

which holds for all $\alpha > 0$. In turn we have the following chain of inequalities:

$$\frac{nx^\alpha}{n^\alpha + x^\alpha} \leq n \log\left(1 + \left(\frac{x}{n}\right)^\alpha\right) \leq \frac{nx^\alpha}{n^\alpha} = \frac{x^\alpha}{n^{\alpha-1}},$$

by another application of the (weak version of the) hint. The squeeze theorem then tells us if $0 < \alpha < 1$ the limit is ∞, if $\alpha = 1$ it is x, and if $\alpha > 1$ the limit is zero. Hence,

$$\ell_\alpha(x) := \lim_{n\to\infty}\left[n\log\left(1+\left(\frac{x}{n}\right)^\alpha\right)\right] = \begin{cases} 0 & x = 0 \\ \begin{cases} \infty & 0<\alpha<1 \\ x & \alpha = 1 \\ 0 & \alpha > 1 \end{cases} & x \neq 0 \end{cases}$$

Applying Fatou's lemma we see that

$$\int_S \ell_\alpha(f(x))\ d\mu(x) = \int_S \lim_{n\to\infty} \ell_{n,\alpha}(f(x))\ d\mu(x) \le \liminf_{n\to\infty} \int_S \ell_{n,\alpha}(f(x))\ d\mu(x).$$

To compute the left hand side we break into three cases.

- $0 < \alpha < 1$: In this case we have

  $$\int_S \ell_\alpha(f(x))\ d\mu(x) = \int_{\{f=0\}} 0\ d\mu(x) + \int_{\{f\neq 0\}} \infty\ d\mu(x) = \infty\cdot\mu(\{f\neq 0\}) = \infty.$$

  We concluded that $\mu(\{f\neq 0\}) > 0$ as $\int_S f(x)\ d\mu(x) > 0$. So by Fatou's lemma,

  $$\infty \le \liminf_{n\to\infty}\int_S \ell_{n,\alpha}(f(x))\ d\mu(x) \le \limsup_{n\to\infty}\int_S \ell_{n,\alpha}(f(x))\ d\mu(x) \le \infty$$

  where the latter two inequalities are trivial. It follows that the limit exists and

  $$\lim_{n\to\infty}\int_S n\log\left(1+\left(\frac{f(x)}{n}\right)^\alpha\right)\ d\mu(x) = \int_S \ell_\alpha(f(x))\ d\mu(x) = \infty.$$

- $\alpha = 1$: In this case

  $$\int_S \ell_\alpha(f(x))\ d\mu(x) = \int_{\{f=0\}} 0\ d\mu(x) + \int_{\{f\neq 0\}} f(x)\ d\mu(x) = \int_S f(x)\ d\mu(x) = c.$$

  Since $\log(1+x) \le x$ we get

  $$\limsup_{n\to\infty}\int_S n\log\left(1+\frac{f(x)}{n}\right)\ d\mu(x) \le \limsup_{n\to\infty}\int_S f(x)\ dx = c.$$

  Thus, by Fatou

  $$\begin{aligned} c = \int_S \ell_\alpha(f(x))\ d\mu(x) &\le \liminf_{n\to\infty}\int_S n\log\left(1+\frac{f(x)}{n}\right)\ d\mu(x) \\ &\le \limsup_{n\to\infty}\int_S n\log\left(1+\frac{f(x)}{n}\right)\ d\mu(x) = c. \end{aligned}$$

  So the limit exists and

  $$\lim_{n\to\infty}\int_S n\log\left(1+\frac{f(x)}{n}\right)\ d\mu(x) = \int_S \ell_1(f(x))\ d\mu(x) = c.$$

- $\alpha > 1$: In this case $\ell_\alpha(x) \equiv 0$ so that

  $$\int_S \ell_\alpha(f(x))\ d\mu(x) = 0.$$

  The sequence is monotone *decreasing* after a certain point, so we cannot apply monotone convergence. However, owing to the hint we have that

  $$\ell_{n,\alpha}(f(x)) = n\log\left(1+\left(\frac{f(x)}{n}\right)^\alpha\right) \le \alpha f(x)$$

  where $\alpha f(x)$ is integrable, and we can apply dominated convergence. Thus the inequality in Fatou's lemma is an equality and

  $$\lim_{n\to\infty}\int_S n\log\left(1+\left(\frac{f(x)}{n}\right)^\alpha\right)\ d\mu(x) = \int_S \ell_\alpha(f(x))\ d\mu(x) = 0.$$

Altogether we have

$$\lim_{n\to\infty}\int_S n\log\left(1+\left(\frac{f(x)}{n}\right)^{\alpha}\right)\,d\mu(x)=\begin{cases}\infty & 0<\alpha<1\\ c & \alpha=1\\ 0 & \alpha>1\end{cases}$$

Problem 3.2: (Convergence in measure, Notes Problem 4.12). A sequence $\{f_n\}_{n=1}^{\infty}$ in $\mathcal{L}^0$ is said to **converge in measure** toward $f\in\mathcal{L}^0$ if

$$\text{for all } \epsilon>0,\quad \mu(\{|f_n-f|\ge\epsilon\})\to 0 \text{ as } n\to\infty.$$

Assume that $\mu(S)<\infty$ (parts marked by (†) are true without this assumption).

1. Show that the mapping
$$d(f,g)=\int\frac{|f-g|}{1+|f-g|}\,d\mu,\quad f,\ g\in\mathcal{L}^0,$$
defines a pseudo metric on $\mathcal{L}^0$ and that convergence in d is equivalent to convergence in measure.
2. Show that $f_n\to f$, a.e., implies that $f_n\to f$ in measure. Give an example which shows that the assumption $\mu(S)<\infty$ is necessary.
3. Give an example of a sequence which converges in measure, but not a.e.
4. (†) For $f\in\mathcal{L}^0$ and a sequence $\{f_n\}_{n=1}^{\infty}\subset\mathcal{L}^0$, show that $f_n\to f$ a.e. if the convergence in measure "happens fast", i.e. if
$$\sum_{n=1}^{\infty}\mu(\{|f_n-f|\ge\epsilon\})<\infty,\quad\text{for all } \epsilon>0.$$
5. (†) Show that each sequence convergent in measure has a subsequence which converges a.e.
6. (†) Show that each sequence convergent in $\mathcal{L}^p$, for some $p\in[1,\infty)$, converges in measure.
7. For $p\in[1,\infty)$, find an example of a sequence which converges in measure, but not in $\mathcal{L}^p$.
8. Let $\{f_n\}_{n=1}^{\infty}\subset\mathcal{L}^0$ with the property that any of its subsequences admits a (further) subsequence which converges a.e. to $f\in\mathcal{L}^0$. Show that $f_n\to f$ in measure.
9. Let $\Phi:\mathbb{R}^2\to\mathbb{R}$ be a continuous function, and let $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ be two sequences in $\mathcal{L}^0$. If $f,\ g\in\mathcal{L}^0$ are such that $f_n\to f$ and $g_n\to g$ in measure, then
$$\Phi(f_n,g_n)\to\Phi(f,g)\ \text{ in measure.}$$
Note: $\Phi=+$ or $\Phi=\times$ are particularly useful.

Solution:

1. We first show d is a pseudo-metric.
   - $d:\mathcal{L}^0\times\mathcal{L}^0\to[0,\infty)$ and $d(f,f)=0$: Evidently, since the integrand is non-negative, we have $d(f,g)\ge 0$. To show it is finite notice that
   $$\frac{|f-g|}{1+|f-g|}\le 1$$
   so that $d(f,g)\le\mu(S)<\infty$. That $d(f,f)=0$ is clear, since the integrand is zero.
   - $d(f,g)=d(g,f)$: Symmetry is obvious since
   $$d(f,g)=\int_S\frac{|f-g|}{1+|f-g|}\,d\mu=\int_S\frac{|g-f|}{1+|g-f|}\,d\mu=d(g,f).$$
   - $d(f,h)\le d(f,g)+d(g,h)$: To prove the triangle inequality first note that the map $t\mapsto t/(1+t)$ is monotone increasing for non-negative t. To see this write
   $$\frac{t}{1+t}=1-\frac{1}{1+t}$$

and note that $t \mapsto 1/(1+t)$ is monotone decreasing. Hence,

$$\begin{aligned}\frac{|f-h|}{1+|f-h|} &\le \frac{|f-g|+|g-h|}{1+|f-g|+|g-h|} \\ &= \frac{|f-g|}{1+|f-g|+|g-h|} + \frac{|g-h|}{1+|f-g|+|g-h|} \\ &\le \frac{|f-g|}{1+|f-g|} + \frac{|g-h|}{1+|g-h|}\end{aligned}$$

where the last inequality holds since we are decreasing the demoninator. Integrating both sides with respect to μ yields

$$\begin{aligned}d(f,h) &= \int_S \frac{|f-h|}{1+|f-h|}\, d\mu \\ &\le \int_S \frac{|f-g|}{1+|f-g|}\, d\mu + \int_S \frac{|g-h|}{1+|g-h|}\, d\mu = d(f,g)+d(g,h).\end{aligned}$$

We now show that the topology generated by d coincides with the topology of convergence in measure. Note that $f_n \xrightarrow{m} f$ if and only if for every $\epsilon > 0$ we have

$$0 = \lim_{n\to\infty} \mu\left(\{|f_n - f| \ge \epsilon\}\right) = \lim_{n\to\infty} \mu\left(\left\{\frac{1}{1+\epsilon} \ge \frac{1}{1+|f_n-f|}\right\}\right).$$

Define

$$E_{n,\epsilon} := \left\{\frac{1}{1+\epsilon} \ge \frac{1}{1+|f_n-f|}\right\}.$$

Now,

$$\begin{aligned}d(f_n,f) = \int_S \frac{|f_n-f|}{1+|f_n-f|}\, d\mu &= \int_S \left[1 - \frac{1}{1+|f_n-f|}\right]\, d\mu \\ &= \mu(S) - \int_S \frac{1}{1+|f_n-f|}\, d\mu.\end{aligned}$$

Splitting the integral gives

$$\begin{aligned}\int_S \frac{1}{1+|f_n-f|}\, d\mu &= \int_{E_{n,\epsilon}} \frac{1}{1+|f_n-f|}\, d\mu + \int_{E_{n,\epsilon}^c} \frac{1}{1+|f_n-f|}\, d\mu \\ &\ge \int_{E_{n,\epsilon}} \frac{1}{1+|f_n-f|}\, d\mu + \int_{E_{n,\epsilon}^c} \frac{1}{1+\epsilon}\, d\mu \\ &= \int_{E_{n,\epsilon}} \frac{1}{1+|f_n-f|}\, d\mu + \frac{\mu(S)}{1+\epsilon}.\end{aligned}$$

If $f_n \xrightarrow{m} f$ then $\mu(E_{n,\epsilon}) \to 0$ for any $\epsilon > 0$. Hence,

$$\lim_{n\to\infty} \int_S \frac{1}{1+|f_n-f|}\, d\mu \ge \frac{\mu(S)}{1+\epsilon}.$$

Applying this in the above yields

$$\lim_{n\to\infty} d(f_n,f) = \mu(S) - \lim_{n\to\infty} \int_S \frac{1}{1+|f_n-f|}\, d\mu \le \left(\frac{\epsilon}{1+\epsilon}\right)\mu(S)$$

and hence $f_n \to f$ in d. We must show now that if $\{f_n\} \in \mathcal{L}^0$ is such that $d(f_n, f) \to 0$ then $f_n \xrightarrow{m} f$. We've seen that the map $t \mapsto t/(1+t)$ is increasing. Thus for $\epsilon > 0$,

$$\begin{aligned} d(f_n, f) &= \int_S \frac{|f_n - f|}{1 + |f_n - f|}\, d\mu \\ &= \int_{\{|f_n - f| \geq \epsilon\}} \frac{|f_n - f|}{1 + |f_n - f|}\, d\mu + \int_{\{|f_n - f| < \epsilon\}} \frac{|f_n - f|}{1 + |f_n - f|}\, d\mu \\ &\geq \int_{\{|f_n - f| \geq \epsilon\}} \frac{\epsilon}{1+\epsilon}\, d\mu + C_{n,\epsilon} = \left(\frac{\epsilon}{1+\epsilon}\right) \mu(\{|f_n - f| \geq \epsilon\}) + C_{n,\epsilon} \end{aligned}$$

where $C_{n,\epsilon} \geq 0$. But since $d(f_n, f) \to 0$, and each of the above terms on the right-hand side are non-negative, they must both go to zero. In particular,

$$\lim_{n\to\infty} \mu(\{|f_n - f| \geq \epsilon\}) = 0$$

and thus $f_n \xrightarrow{m} f$.

2. Fix $\epsilon > 0$ and let

$$E_n = \{|f_n - f| \geq \epsilon\}, \qquad F_n = \bigcup_{k=n}^{\infty} E_k.$$

In this way we have constructed a decreasing sequence $\{F_n\}_{n=1}^{\infty}$, where also $\mu(F_1) \leq \mu(S) < \infty$. By dominated convergence of sets,

$$\mu(F_n) \to \mu\left(\bigcap_{n=1}^{\infty} F_n\right) = \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right).$$

Note that if $D = \{x \in S \mid f_n(x) \text{ does not converge to } f(x)\}$ then $\mu(D) = 0$ and

$$\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k \subseteq D$$

since if x is in the former set, then $|f_n(x) - f(x)| \geq \epsilon$ infinitely often and there is no hope of convergence. Thus, by monotonicity

$$\lim_{n\to\infty} \mu(F_n) = 0.$$

Fix now $\eta > 0$. Then there exists $N \in \mathbb{N}$ such that for $n \geq N$, $\mu(F_n) < \eta$. By monotonicity, since $E_k \subset F_N$ for all $k \geq N$,

$$\mu(\{|f_k - f| \geq \epsilon\}) = \mu(E_k) \leq \mu(F_N) < \eta$$

for all $k \geq N$. That is, $\mu(\{|f_k - f| \geq \epsilon\}) \to 0$.

For a counterexample when $\mu(S) = \infty$, consider the Lebesgue measure on $\mathbb{R}$ and let $f_n = \chi_{[n-1,n]}$. Then $f_n \to 0$ pointwise everywhere, yet for any $\epsilon \in (0, 1]$ we have

$$E_n = \{|f_n - f| \geq \epsilon\} = [n-1, n],$$

which has unit Lebesgue measure.

3. We consider the so-called typewriter sequence. The simple idea is for something to fail to converge pointwise, you want it to oscillate between values. For example, the alternating sequence between $\chi_{[0,1/2)}$ and $\chi_{[1/2,1)}$. Ideally we want this to converge in measure to the zero function, but (as in the previous counterexample) the measures of the intervals need do shrink. The typewriter sequence is defined by

$$f_n(x) = \chi_{E_{n,k}}(x), \quad \text{where } 2^k \leq n < 2^{k+1} \text{ and } E_{n,k} = \left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right).$$

How does this sequence work? Fix a $k \in \mathbb{N}$ and consider $f_{2^k}, f_{2^k+1}, \ldots, f_{2^{k+1}-1}$. Noting that

$$\bigcup_{n=2^k}^{2^{k+1}-1} E_{n,k} = [0, 1)$$

and the $E_{n,k}$ are disjoint (for $n = 2^k, ..., 2^{k+1} - 1$), we see that for any $x \in [0, 1)$ there are infinitely many n, k such that $x \in E_{n,k}$. It follows that f_n does not converge (anywhere) to the zero function. Yet, for $0 < \epsilon \leq 1$ we have

$$\{|f_n| \geq \epsilon\} = E_{n,k}$$

and

$$|\{|f_n| \geq \epsilon\}| = |E_{n_k}| = \frac{1}{2^k} < \frac{2}{n}.$$

Consequently, $f_n \xrightarrow{m} 0$. We do not need to check for $\epsilon > 1$ as the super-level sets are empty.

4. Defining E_n^ϵ, D as before as in 2 (here we need to keep track of the parameter ϵ). Borel-Cantelli says that if

$$\sum_{n=1}^{\infty} \mu(E_n^\epsilon) < \infty,$$

which is exactly the assumption we have, then

$$\mu\left(\limsup_{n\to\infty} E_n^\epsilon\right) = 0.$$

Note that $f_n(x)$ does not converge to $f(x)$ if there exists an $\epsilon > 0$ such that for all $n \in \mathbb{N}$ there exists an $k \geq n$ such that $|f_k(x) - f(x)| \geq \epsilon$. Hence,

$$D \subset \bigcup_{\epsilon>0} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|f_k - f| \geq \epsilon\} = \bigcup_{\epsilon>0} \limsup_{n\to\infty} E_n^\epsilon.$$

The right-hand side is not a countable union, but we can always replace it with

$$D \subset \bigcup_{q\in\mathbb{Q}} \limsup_{n\to\infty} E_n^q$$

as we can always find a rational number less than the extracted ϵ. Finally,

$$\mu(D) \leq \sum_{q\in\mathbb{Q}} \mu\left(\limsup_{n\to\infty} E_n^q\right) = 0$$

and thus $f_n \to f$ a.e.

5. Let $\epsilon > 0$ and for $k \in \mathbb{N}$ choose n_k such that

$$\mu(\{|f_{n_k} - f| \geq \epsilon\}) < \frac{1}{2^k}.$$

Then,

$$\sum_{k=1}^{\infty} \mu(\{|f_{n_k} - f| \geq \epsilon\}) < \sum_{k=1}^{\infty} \frac{1}{2^k} = 1 < \infty$$

and thus $f_{n_k} \to f$ a.e. by 4.

6. Applying Hölder's inequality with q a conjugate exponent to p yields

$$d(f, g) = \int_S (|f - g|)\left(\frac{1}{1 + |f - g|}\right) d\mu \leq \|f - g\|_{\mathcal{L}^p} \left\|\frac{1}{1 + |f - g|}\right\|_{\mathcal{L}^q}.$$

If $\mu(S) < \infty$ the right-hand side is bounded by $\|f - g\|_{\mathcal{L}^p} \mu(S)^{1/q}$ since $1/(1 + |f - g|) \leq 1$. Then $f_n \to f$ in $\mathcal{L}^p$ implies

$$d(f_n, f) \leq \mu(S)^{1/q} \|f_n - f\|_{\mathcal{L}^p} \to 0.$$

There's another neat way to do this using Chebyshev's inequality:

$$\mu(\{|f_n - f| \geq \epsilon\}) = \mu(\{|f_n - f|^p \geq \epsilon^p\}) \leq \frac{1}{\epsilon^p} \int_S |f_n - f|^p \, d\mu = \frac{1}{\epsilon^p} \|f_n - f\|_{\mathcal{L}^p}^p.$$

7. It suffices by 2 to find a sequence $f_n \to f$ a.e. but f_n does not converge to f in $\mathcal{L}^p$. Consider $f_n(x) = n^{1/p}\chi_{[0,1/n]}$, and let our measure space be $[0,1]$ with the Lebesgue measure. Evidently $f_n \to 0$ a.e. Computing the $\mathcal{L}^p$ norms we see,
$$\|f_n\|_{\mathcal{L}^p} = \left(\int_{[0,1]} n\chi_{[0,1/n]}\right)^{1/p} = \left(n \cdot \frac{1}{n}\right)^{1/p} = 1.$$
Hence f_n does not converge to 0 in $\mathcal{L}^p$.
8. Recall that every sequence $\{x_n\}_{n=1}^\infty$ of real numbers, such that all of its subsequences admit a convergent subsequence to $x \in \mathbb{R}$, in fact converges to x itself. Let $\epsilon > 0$ and set
$$x_n = \mu(\{|f_n - f| \geq \epsilon\}).$$
Recall by 2 that if $f_n \to f$ a.e. then $f_n \xrightarrow{m} f$. Let $\{f_{n_k}\}_{k=1}^\infty$ be a subsequence of $\{f_n\}_{n=1}^\infty$. Then there exists a further subsequence $\{f_{n_{k_l}}\}_{l=1}^\infty$ such that $f_{n_{k_l}} \to f$ a.e. Hence $f_{n_{k_l}} \xrightarrow{m} f$, implying $x_{n_{k_l}} \to 0$. Thus for every subsequence $\{x_{n_k}\}_{k=1}^\infty$ of $\{x_n\}_{n=1}^\infty$ there exists a convergent further subsequence to 0. By the recollection at the beginning of the problem, it follows that $x_n \to 0$, i.e. $f_n \xrightarrow{m} 0$.
9. We show first that $h_n = (f_n, g_n)$ converges in measure to (f, g). To see this note if $\epsilon > 0$ and
$$x \in \{|(f_n, g_n) - (f, g)| \geq \epsilon\}$$
then
$$(f_n(x) - f(x))^2 + (g_n(x) - g(x))^2 \geq \epsilon^2.$$
By the pigeonhole principle
$$x \in \{|f_n - f| \geq \epsilon^2/2\} \cup \{|g_n - g| \geq \epsilon^2/2\}.$$
Hence we've shown
$$\{|(f_n, g_n) - (f, g)| \geq \epsilon\} \subset \{|f_n - f| \geq \epsilon^2/2\} \cup \{|g_n - g| \geq \epsilon^2/2\}$$
and thus
$$\mu(\{|(f_n, g_n) - (f, g)| \geq \epsilon\}) \leq \mu(\{|f_n - f| \geq \epsilon^2/2\}) + \mu(\{|g_n - g| \geq \epsilon^2/2\}) \to 0.$$
We now show if $h_n \xrightarrow{m} h$ and Φ is continuous then $\Phi(h_n) \xrightarrow{m} \Phi(h)$. To do this we'll apply 8. Since $h_n \xrightarrow{m} h$, for subsequence $h_{n_k} \xrightarrow{m} h$ we can find a further subsequence $h_{n_{k_l}} \to h$ a.e. by 5. But continuous functions preserve almost everywhere convergence, so $\Phi(h_{n_{k_l}}) \to \Phi(h)$ a.e. Hence for any subsequence $\Phi(h_{n_k})$ we have found a further subsequence $\Phi(h_{n_{k_l}})$ which converges to $\Phi(h)$ a.e. Applying 8, we see that $\Phi(h_n) \xrightarrow{m} \Phi(h)$.

Midterm

Problem M.1. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions on the measure space $(S, \mathcal{S}, \mu)$ where μ is a finite measure. Consider the following two statements:

1. Every subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ of $\{f_n\}_{n\in\mathbb{N}}$ has a further subsequence $\{f_{n_{k_l}}\}_{l\in\mathbb{N}}$ which converges a.e. to 0.
2. $f_n \to 0$, in measure.

Show that $1 \Rightarrow 2$. For extra credit, show that $2 \Rightarrow 1$.

Note: You can use, without proof, the fact that the a.e.-convergence implies convergence in measure to the same limit (when μ is finite).

Solution: We prove instead the more general equivalence where 0 is replaced by a generic measurable function f.

i. (1 $\Rightarrow$ 2) Recall that every sequence $\{x_n\}_{n=1}^{\infty}$ of real numbers, such that all of its subsequences admit a convergent subsequence to $x \in \mathbb{R}$, in fact converges to x itself. Let $\epsilon > 0$ and set

$$x_n = \mu(\{|f_n - f| \geq \epsilon\}).$$

Note that if $f_n \to f$ a.e. then $f_n \xrightarrow{m} f$. Let $\{f_{n_k}\}_{k=1}^{\infty}$ be a subsequence of $\{f_n\}_{n=1}^{\infty}$. Then by hypothesis there exists a further subsequence $\{f_{n_{k_l}}\}_{l=1}^{\infty}$ such that $f_{n_{k_l}} \to f$ a.e. Hence $f_{n_{k_l}} \xrightarrow{m} f$, implying $x_{n_{k_l}} \to 0$. Thus for every subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ of $\{x_n\}_{n=1}^{\infty}$ there exists a convergent further subsequence to 0. By the recollection at the beginning of the problem, it follows that $x_n \to 0$, i.e. $f_n \xrightarrow{m} f$.

ii. (2 $\Rightarrow$ 1) We show first that convergence in measure implies the existence of a subsequence which converges a.e. Define

$$E_n^{\epsilon} := \{|f_n - f| \geq \epsilon\}, \quad D = \{x \in S \mid f_n(x) \text{ does not converge to } f(x)\}.$$

Note that $f_n(x)$ does not converge to $f(x)$ if there exists an $\epsilon > 0$ such that for all $n \in \mathbb{N}$ there exists a $k \geq n$ such that $|f_k(x) - f(x)| \geq \epsilon$. Hence,

$$D \subset \bigcup_{\epsilon>0} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|f_k - f| \geq \epsilon\} = \bigcup_{\epsilon>0} \limsup_{n\to\infty} E_n^{\epsilon}.$$

The right-hand side is not a countable union, but we can always replace it with

$$D \subset \bigcup_{q\in\mathbb{Q}} \limsup_{n\to\infty} E_n^{q}$$

as we can always find a rational number less than the extracted ϵ. Finally,

$$\mu(D) \leq \sum_{q\in\mathbb{Q}} \mu\left(\limsup_{n\to\infty} E_n^{q}\right)$$

If $\mu(D) = 0$ then $f_n \to f$ a.e. Borel-Cantelli says that if

$$\sum_{n=1}^{\infty} \mu(E_n^{\epsilon}) < \infty,$$

then

$$\mu\left(\limsup_{n\to\infty} E_n^{\epsilon}\right) = 0.$$

So, it suffices to find a subsequence f_{n_k} such that for every $\epsilon > 0$

$$\sum_{k=1}^{\infty} \mu(E_{n_k}^{\epsilon}) < \infty.$$

Since $f_n \xrightarrow{m} f$ we know that $\mu(E_n^{\epsilon}) \to 0$ for all $\epsilon > 0$. Choose $n_{k+1} \geq n_k$ such that $\mu(E_{n_k}^{1/k}) < 1/2^k$. Let $\epsilon > 0$ and let $k_\epsilon \in \mathbb{N}$ be the smallest natural number such that $1/k_\epsilon < \epsilon$. Observe that for any $k \geq k_\epsilon$, $1/k < \epsilon$ and so

$$\mu(E_n^{\epsilon}) = \mu(\{|f_n - f| \geq \epsilon\}) \leq \mu\left(\left\{|f_n - f| \geq \frac{1}{k}\right\}\right) = \mu(E_n^{1/k})$$

for any $n \in \mathbb{N}$. Hence,

$$\sum_{k=1}^{\infty} \mu(E_{n_k}^{\epsilon}) \leq \sum_{k=1}^{k_\epsilon - 1} \mu(E_{n_k}^{1/k}) + \sum_{k=k_\epsilon}^{\infty} \mu(E_{n_k}^{1/k}) = \sum_{k=1}^{k_\epsilon - 1} \mu(E_{n_k}^{1/k}) + \sum_{k=k_\epsilon}^{\infty} \frac{1}{2^k} < \infty.$$

By the above, $f_{n_k} \to f$ a.e. Now since $f_n \xrightarrow{m} f$, any subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ will also converge in measure to f. Applying the above, we can find a further subsequence $\{f_{n_{k_l}}\}_{l=1}^{\infty}$ which converges a.e.

Problem M.2. Let $f \in \mathcal{L}^0_+(S, \mathcal{S}, \mu)$, where μ is a finite measure, be such that

$$\limsup_{t\to\infty} [t^{p_0}\mu(\{f > t\})] < \infty, \quad \text{where } p_0 > 1.$$

Show that $f \in \mathcal{L}^p(S, \mathcal{S}, \mu)$ for all $p \in [1, p_0)$.

Solution: Let $\epsilon > 0$ and set

$$L := \limsup_{t\to\infty}[t^{p_0}\mu(\{f > t\})] = \lim_{T\to\infty} \sup_{t\geq T} [t^{p_0}\mu(\{f > t\})].$$

Then there exists a $T_0 \geq 0$ such that for all $T \geq T_0$

$$\left| \sup_{t\geq T} [t^{p_0}\mu(\{f > t\})] - L \right| < \epsilon.$$

To estimate $\|f\|_{\mathcal{L}^p(S)}$ for $p \in [1, p_0)$ we need to introduce $t^{p_0}\mu(\{f > t\})$. This looks awfully familiar to the layer-cake formula,

$$\|f\|^p_{\mathcal{L}^p(S)} = \int_S f^p \, d\mu = p \int_0^\infty t^{p-1}\mu(\{f > t\}) \, dt,$$

except the exponent on t is incorrect. We can fix this by multiplying and dividing by an appropriate factor. In the process we also split the integral, as we only have a bound on $t^{p_0}\mu(\{f > t\})$ when $t \geq T_0$. Thus,

$$\begin{aligned}
\|f\|^p_{\mathcal{L}^p(S)} &= p \int_0^{T_0} t^{p-1}\mu(\{f > t\}) \, dt + p \int_{T_0}^\infty t^{p-1}\mu(\{f > t\}) \, dt \\
&= p \int_0^{T_0} t^{p-1}\mu(\{f > t\}) \, dt + p \int_{T_0}^\infty t^{p-1}\frac{t^{p_0+1-p}}{t^{p_0+1-p}}\mu(\{f > t\}) \, dt \\
&= p \int_0^{T_0} t^{p-1}\mu(\{f > t\}) \, dt + p \int_{T_0}^\infty \frac{1}{t^{p_0+1-p}} \cdot [t^{p_0}\mu(\{f > t\})] \, dt.
\end{aligned}$$

Finally we have $\sup_{t\geq T_0} t^{p_0}\mu(\{f > t\}) < \epsilon + L$, particularly $t^{p_0}\mu(\{f > t\}) < \epsilon + L$ for $t \geq T_0$. Additionally $\mu(\{f \geq t\}) \leq \mu(S) < \infty$ by monotonicity and the finiteness assumption. Hence,

$$\|f\|^p_{\mathcal{L}^p(S)} \leq p\mu(S) \int_0^{T_0} t^{p-1} \, dt + p[\epsilon + L] \int_{T_0}^\infty \frac{dt}{t^{p_0-p+1}} = \mu(S)T_0^p + \frac{p[\epsilon + L]}{(p_0 - p)T_0^{p_0-p}} < \infty.$$

Remark: The map

$$T \mapsto \mu(S)T^p + \frac{p[\epsilon + L]}{(p_0 - p)T^{p_0-p}}$$

for $T \geq 0$ is minimized when

$$T = T_m := \left(\frac{\epsilon + L}{\mu(S)}\right)^{1/p_0}.$$

If $T_0 \leq T_m$, then repeating the same work above with T_m instead of T_0 (valid since the bound on the supremum holds for $T \geq T_0$) yields

$$\begin{aligned}
\|f\|^p_{\mathcal{L}^p(S)} &\leq \mu(S)\left(\frac{\epsilon + L}{\mu(S)}\right)^{p/p_0} + \frac{p[\epsilon + L]}{p_0 - p}\left(\frac{\epsilon + L}{\mu(S)}\right)^{(p-p_0)/p_0} \\
&= (\epsilon + L)^{p/p_0}\mu(S)^{1-p/p_0} + \left(\frac{p}{p_0 - p}\right)(\epsilon + L)^{p/p_0}\mu(S)^{1-p/p_0} \\
&= \left(\frac{p_0}{p_0 - p}\right)(\epsilon + L)^{p/p_0}\mu(S)^{1-p/p_0}.
\end{aligned}$$

Addendum: We technically haven't proven (or talked about) the layer-cake formula in class, so let us prove it here. Let ν on $\mathcal{B}([0, \infty))$ be defined by

$$\nu(E) = \int_E p \cdot z^{p-1} \, dz.$$

It is clear that $\nu \ll \mathcal{L}^1 \llcorner [0,\infty)$ and that $d\nu/d(\mathcal{L}^1 \llcorner [0,\infty))(t) = pt^{p-1}$. Note that $\nu(\{t\}) = 0$ for any $t \in [0,\infty)$ so that

$$\nu([0,t)) = \nu([0,t]) = \int_0^t p \cdot z^{p-1}\ dz = z^p\Big|_{z=0}^{z=t} = t^p.$$

Since $\mu \otimes \nu$ is the product of two σ-finite measures and $g(x,t) := \chi_{f>t}(x) = \chi_{[0,f(x))}(t)$ is non-negative, Tonelli's theorem guarantees that

$$\begin{aligned}\int_0^\infty \mu(\{f>t\})\ d\nu(t) &= \int_0^\infty \left[\int_S \chi_{\{f>t\}}(x)\ d\mu(x)\right] d\nu(t)\\ &= \int_S \left[\int_0^\infty \chi_{[0,f(x))}(t)\ d\nu(t)\right] d\mu(x)\\ &= \int_S \nu([0,f(x)))\ d\mu(x) = \int_S f(x)^p\ d\mu(x).\end{aligned}$$

On the other hand, applying the Radon-Nikodym derivative yields

$$\int_S f(x)^p\ d\mu(x) = \int_0^\infty \mu(\{f>t\}) \frac{d\nu}{d(\mathcal{L}^1 \llcorner [0,\infty))}(t)\ dt = p\int_0^\infty t^{p-1} \cdot \mu(\{f>t\})\ dt.$$

HW 4

Problem 4.1: (The "layered" representation, Notes Problem 5.10). Let ν be a measure on $\mathcal{B}([0,\infty))$ such that $N(u) = \nu([0,u)) < \infty$ for all $u \in \mathbb{R}$. Let $(S, \mathcal{S}, \mu)$ be a σ-finite measure space. For $f \in L^0_+(S)$, show that

1. $\int N \circ f\ d\mu = \int_{[0,\infty)} \mu(\{f>u\})\ d\nu(u)$.
2. For $p > 0$ we have $\int f^p\ d\mu = p\int_{[0,\infty)} u^{p-1}\mu(\{f>u\})\ du$ where du represents integration with respect to the Lebesgue measure.

Solution:

1. First note that $\nu([0,\infty)) \leq 1$, and so is a finite measure. Next consider $g : S \times [0,\infty)$ by $g(x,u) = \chi_{\{f>u\}}(x) = \chi_{[0,f(x))}(u)$. Since $\mu \otimes \nu$ is the product of two σ-finite measures and $g \geq 0$, Tonelli's theorem guarantees that

$$\begin{aligned}\int_0^\infty \mu(\{f>u\})\ d\nu(u) &= \int_0^\infty \left[\int_S \chi_{\{f>u\}}(x)\ d\mu(x)\right] d\nu(u)\\ &= \int_S \left[\int_0^\infty \chi_{[0,f(x))}(u)\ d\nu(u)\right] d\mu(x)\\ &= \int_S \nu([0,f(x)))\ d\mu(x) = \int_S N(f(x))\ d\mu(x).\end{aligned}$$

2. Let ν on $\mathcal{B}([0,\infty))$ be defined by

$$\nu(E) = \int_E p \cdot z^{p-1}\ dz.$$

It is clear that $\nu \ll \mathcal{L}^1 \llcorner [0,\infty)$ and that $d\nu/d(\mathcal{L}^1 \llcorner [0,\infty))(u) = pu^{p-1}$. Note that $\nu(\{u\}) = 0$ for any $u \in [0,\infty)$ so that

$$\nu([0,u)) = \nu([0,u]) = \int_0^u p \cdot z^{p-1}\ dz = z^p\Big|_{z=0}^{z=u} = u^p.$$

It follows that $N(u) = u^p < \infty$ for all $u \in [0,\infty)$ and so we may apply part 1. Doing so yields

$$\begin{aligned}\int_S f^p\ d\mu = \int_S N \circ f\ d\mu &= \int_0^\infty \mu(\{f>u\})\ d\nu(u)\\ &= \int_0^\infty \mu(\{f>u\}) \frac{d\nu}{d(\mathcal{L}^1 \llcorner [0,\infty))}(u)\ d\mathcal{L}^1(u) = \int_0^\infty pu^{p-1} \cdot \mu(\{f>u\})\ du.\end{aligned}$$

Problem 4.2: (A Dirichlet integral, Notes Problem 5.11).

1. Show that $\int_0^\infty |\sin(x)/x|\, dx = \infty$. Hint: Find a function below $|\sin(x)/x|$ which is easier to integrate.
2. For $a > 0$ let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by
$$f(x,y) = \begin{cases} e^{-xy}\sin(x) & 0 \le x \le a,\ 0 \le y, \\ 0 & \text{otherwise.} \end{cases}$$
Show that $f \in L^1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mathcal{L}^2)$, where $\mathcal{L}^2$ denotes the Lebesgue measure on $\mathbb{R}^2$.
3. Establish the equality
$$\int_0^a \frac{\sin x}{x}\, dx = \frac{\pi}{2} - \cos(a)\int_0^\infty \frac{e^{-ay}}{1+y^2}\, dy - \sin(a)\int_0^\infty \frac{ye^{-ay}}{1+y^2}\, dy.$$
4. Conclude that for $a > 0$,
$$\left| \int_0^a \frac{\sin(x)}{x}\, dx - \frac{\pi}{2} \right| \le \frac{2}{a},$$
so that $\lim_{a\to\infty} \int_0^a \sin(x)/x\, dx = \pi/2$.

Hint: You can use the fact (no need to prove it) that the Lebesgue integral (with respect to the Lebesgue measure) of a continuous function on a compact interval coincides with its Riemann integral.

Solution:

1. Define $f : [0,\infty) \to [0,1]$ by
$$f(x) = \begin{cases} 2/\pi x - 2k & x \in [k\pi, (k+1/2)\pi) \\ -2/\pi x + 2(k+1) & x \in [(k+1/2)\pi, (k+1)\pi). \end{cases}$$
The graph is a sawtooth which lies below $|\sin(x)|$ since all the maxima and minima occur at the same locations and $|\sin(x)|$ is concave. Then,
$$\begin{aligned} \int_0^\infty \left|\frac{\sin(x)}{x}\right| dx &= \sum_{k=0}^\infty \left[\int_{k\pi}^{(k+1/2)\pi} \frac{|\sin(x)|}{x}\, dx + \int_{(k+1/2)\pi}^{(k+1)\pi} \frac{|\sin(x)|}{x}\, dx \right] \\ &\ge \sum_{k=0}^\infty \int_{k\pi}^{(k+1/2)\pi} \left(\frac{2}{\pi} - \frac{2k}{x}\right) dx + \sum_{k=0}^\infty \int_{(k+1/2)\pi}^{(k+1)\pi} \left(-\frac{2}{\pi} + \frac{2(k+1)}{x}\right) dx \\ &= -2\sum_{k=0}^\infty k \int_{k\pi}^{(k+1/2)\pi} \frac{1}{x}\, dx + 2\sum_{k=0}^\infty (k+1) \int_{(k+1/2)\pi}^{(k+1)\pi} \frac{1}{x}\, dx \\ &\ge \sum_{k=0}^\infty \int_{k\pi}^{(k+1/2)\pi} \frac{-2k}{k\pi}\, dx + \sum_{k=0}^\infty \int_{(k+1/2)\pi}^{(k+1)\pi} \frac{2(k+1)}{(k+1)\pi}\, dx \end{aligned}$$

$$\begin{aligned} \int_0^\infty \left|\frac{\sin(x)}{x}\right| dx &= \sum_{k=0}^\infty \left[\int_{k\pi}^{(k+1/2)\pi} \frac{|\sin(x)|}{x}\, dx + \int_{(k+1/2)\pi}^{(k+1)\pi} \frac{|\sin(x)|}{x}\, dx \right] \\ &\ge \sum_{k=0}^\infty \int_{k\pi}^{(k+1/2)\pi} \frac{|\sin(x)|}{x}\, dx \ge \sum_{k=0}^\infty \int_{k\pi}^{(k+1/2)\pi} \left(\frac{2}{\pi} - \frac{2k}{x}\right) dx \end{aligned}$$
where we have used the fact that $x \mapsto 1/x$ is decreasing for $x > 0$.
2. We want to show that
$$\int_{\mathbb{R}^2} |f(x,y)|\, d\mathcal{L}^2(x,y) < \infty.$$
Since $\mathcal{L}^2 = \mathcal{L}^1 \otimes \mathcal{L}^1$ and $|f|$ is non-negative, we can use Tonelli's theorem to conclude that
$$\int_{\mathbb{R}^2} |f(x,y)|\, d\mathcal{L}^2(x,y) = \int_0^a \left[\int_0^\infty e^{-xy}|\sin(x)|\, dy \right] dx = \int_0^\infty \left[\int_0^a e^{-xy}|\sin(x)|\, dx \right] dy.$$

The inner integral is easier to compute:

$$\int_0^a \left[\int_0^\infty e^{-xy}|\sin(x)|\ dy\right] dx = \int_0^a |\sin(x)|\frac{e^{-xy}}{-x}\bigg|_{y=0}^{y=\infty} dx = \int_0^a \frac{|\sin(x)|}{x}\ dx.$$

Then since $|\sin(x)| \le |x|$ we have

$$\int_0^a \frac{|\sin(x)|}{x}\ dx \le \int_0^a 1\ dx = a < \infty.$$

3. What is

$$\int_0^a e^{-xy}\sin(x)\ dx?$$

Integration by parts twice yields

$$\begin{aligned}\int_0^a e^{-xy}\sin(x)\ dx &= -\cos(x)e^{-xy}\big|_{x=0}^{x=a} - \int_0^a ye^{-xy}\cos(x)\ dx \\ &= -\cos(a)e^{-ay} + 1 - y\left(e^{-xy}\sin(x)\big|_{x=0}^{x=a} + \int_0^a ye^{-xy}\sin(x)\ dx\right) \\ &= -\cos(a)e^{-ay} + 1 - ye^{-ay}\sin(a) - y^2\int_0^a e^{-xy}\sin(x)\ dx\end{aligned}$$

Thus,

$$\int_0^a e^{-xy}\sin(x)\ dx = \frac{1}{1+y^2} - \cos(a)\left(\frac{e^{-ay}}{1+y^2}\right) - \sin(a)\left(\frac{ye^{-ay}}{1+y^2}\right).$$

Integrating over $y \ge 0$ finally gives

$$\int_0^\infty \left[\int_0^a e^{-xy}\sin(x)\ dx\right] dy = \frac{\pi}{2} - \cos(a)\int_0^\infty \frac{e^{-ay}}{1+y^2}\ dy - \sin(a)\int_0^\infty \frac{ye^{-ay}}{1+y^2}\ dy.$$

If we could interchange the order of integration (Fubini's theorem) we'd be done, as

$$\begin{aligned}\int_0^\infty \left[\int_0^a e^{-xy}\sin(x)\ dx\right] dy &= \int_0^a \left[\int_0^\infty e^{-xy}\sin(x)\ dy\right] dx \\ &= \int_0^a \frac{\sin(x)e^{-xy}}{-x}\bigg|_{y=0}^{y=\infty} dx = \int_0^a \frac{\sin(x)}{x}\ dx.\end{aligned}$$

So it suffices to show the conditions of Fubini's theorem hold. To do this we just need that $f \in L^0(\mathbb{R}^2)$ and $f^- \in L^1(\mathbb{R}^2)$. As a piecewise continuous function f is measurable, so we just check the latter. Observe that f is negative precisely when $\sin(x)$ is. Hence

$$f^-(x,y) = -e^{-xy}\sin(x)\chi_{\{\sin(\cdot)<0\}\cap[0,a]}(x)\chi_{\mathbb{R}^{\ge 0}}(y).$$

By Tonelli's theorem, we can always interchange the order of integration for non-negative functions. Hence,

$$\begin{aligned}\int_{\mathbb{R}^2} f^-(x,y)\ d\mathcal{L}^2(x,y) &= \int_{\{\sin(\cdot)<0\}\cap[0,a]} \left[\int_0^\infty -e^{-xy}\sin(x)\ dy\right] dx \\ &= \int_{\{\sin(\cdot)<0\}\cap[0,a]} \frac{-\sin(x)}{x}\ dx \le |\{\sin(\cdot)<0\}\cap[0,a]| \le a.\end{aligned}$$

4. We have from 3 that

$$\left|\int_0^a \frac{\sin(x)}{x}\ dx - \frac{\pi}{2}\right| \le \int_0^\infty \frac{(1+y)e^{-ay}}{1+y^2}\ dy$$

since $|\cos(a)|, |\sin(a)| \le 1$. Because $t \mapsto (1+t)/(1+t^2)$ is continuous and has limit 0 as $t \to \infty$, it follows that it is bounded on $[0,\infty)$, say by C. Hence,

$$\left|\int_0^a \frac{\sin(x)}{x}\ dx - \frac{\pi}{2}\right| \le C\int_0^\infty e^{-ay}\ dy = \frac{C}{a}.$$

Problem 4.3: (Push-forward and Radon-Nikodym). Let μ and ν be σ-finite measures on $(S, \mathcal{S})$ with $\nu \ll \mu$. For a measurable space $(R, \mathcal{R})$ and a measurable map $T : S \to R$, let

$$\mu_T = T_{\#}\mu \quad \text{and} \quad \nu_T = T_{\#}\nu$$

denote the push-forwards of μ and ν by T. Show that $\nu_T \ll \mu_T$ and find an expression for (a version of) $d\nu_T/d\mu_T$ in terms of T and $d\nu/d\mu$.

Solution: We first show $\nu_T \ll \mu_T$. Suppose $\mu_T(A) = 0$. Then by definition of the push-forward, $0 = \mu_T(A) = \mu(T^{-1}(A))$. Now since $\nu \ll \mu$ and $T^{-1}(A)$ is μ-null, we have by absolute continuity of ν with respect to μ that $\nu(T^{-1}(A)) = 0$ too. But this is precisely $\nu_T(A)$.

Technically speaking $d\nu_T/d\mu_T$ is an equivalence class of functions, but here I will treat it just as a function. It is defined by the integral equality

$$\int_R f \; d\nu_T = \int_R f \frac{d\nu_T}{d\mu_T} \; d\mu_T$$

for any $f \in \mathcal{L}^0_+(R)$. So we just need to find a function which obeys this property. By the change of variables for push-forward measures, if $f, g \in \mathcal{L}^0_+(R)$ then

$$\int_R f \; d\nu_T = \int_S f \circ T \; d\nu, \qquad \int_R g \; d\mu_T = \int_S g \circ T \; d\mu.$$

Hence,

$$\int_R f \; d\nu_T = \int_S f \circ T \; d\nu = \int_S f \circ T \cdot \frac{d\nu}{d\mu} \; d\mu$$

HW 5

Problem 5.1: (Notes Problem 6.16). An absolutely continuous random variable X is said to have the **standard normal distribution** (denoted by $X \sim N(0, 1)$), if it admits a density of the form

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

1. Show that
$$\int_{\mathbb{R}} f(x) \; dx = 1.$$
Hint: Consider the double integral
$$\int_{\mathbb{R}^2} f(x) f(y) \; dxdy$$
and pass to polar coordinates.
2. For $X \sim N(0, 1)$, show that $\mathbb{E}[|X|^n] < \infty$ for all $n \in \mathbb{N}$. Then compute the n-th moment $\mathbb{E}[X^n]$, for $n \in \mathbb{N}$.
3. A random variable with the same distribution as X^2, where $X \sim N(0, 1)$, is said to have the **χ^2-distribution**. Find an explicit expression for the density of the χ^2-distribution.
4. Let Y have the χ^2-distribution. Show that there exists a constant $\lambda_0 > 0$ such that $\mathbb{E}[\exp(\lambda Y)] < \infty$ for $\lambda < \lambda_0$ and $\mathbb{E}[\exp(\lambda Y)] = \infty$ for $\lambda \geq \lambda_0$.

   Note: For a random variable $Y \in \mathcal{L}^0_+$, the quantity $\mathbb{E}[\exp(\lambda Y)]$ is called the **exponential moment of order λ**.
5. Let $\alpha_0 > 0$ be a fixed, but arbitrary constant. Find an example of a random variable $X \geq 0$ with the property that $\mathbb{E}[X^\alpha] < \infty$ for $\alpha \leq \alpha_0$ and $\mathbb{E}[X^\alpha] = \infty$ for $\alpha > \alpha_0$.

   Note: This is not the same situation as in 4. - this time the critical case α_0 is included in a different alternative.

Solution:

1. Because $f \geq 0$, Tonelli's theorem guarantees that $F(x,y) := f(x)f(y)$ satisfies
$$\int_{\mathbb{R}^2} F(x,y)\ d(x,y) = \int_{\mathbb{R}} f(y) \left[\int_{\mathbb{R}} f(x)\ dx\right] dy = \left[\int_{\mathbb{R}} f(x)\ dx\right]^2.$$
So it suffices to show the former is equal to 1 (again, because $f \geq 0$). Converting to polar coordinates, we have
$$\begin{aligned}\int_{\mathbb{R}^2} F(x,y)\ d(x,y) &= \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\ d(x,y) = \int_0^{2\pi}\left[\int_0^\infty re^{-r^2/2}dr\right] d\theta \\ &= -\frac{1}{2\pi}\int_0^{2\pi} e^{-r^2/2}\Big|_{r=0}^{r=\infty}\ d\theta = \frac{1}{2\pi}\int_0^{2\pi} 1\ d\theta = 1.\end{aligned}$$

2. The expectation is
$$\begin{aligned}\mathbb{E}[|X|^n] &= \int_{\mathbb{R}} |x|^n\ d\mu_X(x) = \int_{\mathbb{R}} |x|^n f(x)\ dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} |x|^n e^{-x^2/2}\ dx = \sqrt{\frac{2}{\pi}}\int_0^\infty x^n e^{-x^2/2}\ dx.\end{aligned}$$
Now set for $t \in [1,\infty)$
$$I(t) = \int_0^\infty x^t e^{-x^2/2}\ dx = \int_0^\infty x^{t-1}\cdot xe^{-x^2/2}\ dx.$$
Integration by parts yields for $t > 1$
$$I(t) = -x^{t-1}e^{-x^2/2}\Big|_{x=0}^{x=\infty} + (t-1)\int_0^\infty x^{t-2}\cdot e^{-x^2/2}\ dx = (t-1)I(t-2).$$
In particular, for $n \geq 2$
$$I(n) = \begin{cases} I(0)\prod_{i=1}^k (2i-1) & n = 2k \\ I(1)\prod_{i=1}^k (2i) & n = 2k+1. \end{cases}$$
Observing that
$$\prod_{i=1}^k (2i-1)\prod_{i=1}^k (2i) = (2k)!, \qquad \prod_{i=1}^k (2i) = 2^k k!,$$
the above reduces to
$$I(n) = \begin{cases} I(0)\cdot \dfrac{(2k)!}{2^k k!} & n = 2k \\ I(1)\cdot 2^k k! & n = 2k+1 \end{cases}$$
which is actually valid for all $n \geq 0$. We now just need to compute $I(0)$ and $I(1)$. These are simple:
$$\begin{aligned} I(0) &= \int_0^\infty e^{-x^2/2}\ dx = \sqrt{\frac{\pi}{2}}\int_{-\infty}^\infty f(x)\ dx = \sqrt{\frac{\pi}{2}}; \\ I(1) &= \int_0^\infty xe^{-x^2/2}\ dx = -e^{-x^2/2}\Big|_{x=0}^{x=\infty} = 1.\end{aligned}$$
In total:
$$\mathbb{E}[|X|^n] = \sqrt{\frac{2}{\pi}} I(n) = \begin{cases} \dfrac{(2k)!}{2^k k!} & n = 2k \\ 2^k k!\sqrt{\dfrac{2}{\pi}} & n = 2k+1. \end{cases}$$

Thus $\mathbb{E}[|X|^n] < \infty$ for all $n \in \mathbb{N}$. In principle to just estimate the expectation, one could probably just split the integral in a desirable way. The above analysis lets us compute the n-th moments immediately:

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f(x)\, dx = \begin{cases} \dfrac{(2k)!}{2^k k!} & n = 2k \\ 0 & n = 2k+1 \end{cases}$$

owing to the symmetry.

3. Recall that

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

is absolutely continuous (as a function) and thus satisfies the fundamental theorem of calculus:

$$F(b) - F(a) = \int_a^b f(t)\, dt, \qquad F'(x) = f(x).$$

In particular for an absolutely continuous random variable with cdf $F(x)$ and pdf $f(x)$ it follows that $F'(x) = f(x)$. Computing the cdf of X^2 gives

$$F_{X^2}(x) = \mathbb{P}[X^2 \leq x],$$

which is evidently zero if $x < 0$. For $x \geq 0$ we can continue and write

$$\begin{aligned} F_{X^2}(x) &= \mathbb{P}[X^2 \leq x] = \mathbb{P}[-\sqrt{x} \leq X \leq \sqrt{x}] \\ &= \mathbb{P}[X \leq x] - \mathbb{P}[X < -\sqrt{x}] = \mathbb{P}[X \leq \sqrt{x}] - (1 - \mathbb{P}[X \geq -\sqrt{x}]). \end{aligned}$$

To proceed let us write down $\mathbb{P}[X \geq -\sqrt{x}]$ explicitly to rewrite it (one can do this heuristically using a graph of the pdf).

$$\mathbb{P}[X \geq -\sqrt{x}] = \mathbb{P}[X^{-1}([-\sqrt{x}, \infty))] = \mu_X([-\sqrt{x}, \infty)) = \int_{-\sqrt{x}}^{\infty} f(t)\, dt.$$

Owing to the symmetry $f(t) = f(-t)$ we get

$$\int_{-\sqrt{x}}^{\infty} f(t)\, dt = \int_{-\infty}^{\sqrt{x}} f(t)\, dt = \mathbb{P}[X \leq \sqrt{x}].$$

In total:

$$F_{X^2}(x) = 2\mathbb{P}[X \leq \sqrt{x}] - 1 = 2F_X(\sqrt{x}) - 1$$

for $x \geq 0$, and is zero for $x < 0$. Since $F_X'(x) = f(x)$ we have

$$f_{X^2}(x) = F_{X^2}'(x) = 2\frac{d}{dx} F_X(\sqrt{x}) = 2f(\sqrt{x}) \cdot \frac{x^{-1/2}}{2} = \frac{e^{-x/2}}{\sqrt{2\pi x}}.$$

4. Writing out the expectation, we have

$$\mathbb{E}[e^{\lambda Y}] = \int_{-\infty}^{\infty} e^{\lambda y}\, d\mu_Y(y) = \int_{-\infty}^{\infty} e^{\lambda y} f_Y(y)\, dy = \int_0^{\infty} \frac{e^{(\lambda - 1/2)y}}{\sqrt{2\pi y}}\, dy$$

Making the change of variables $u^2 = y$ we get

$$\mathbb{E}[e^{\lambda Y}] = 2\int_0^{\infty} \frac{e^{(\lambda - 1/2)u^2}}{\sqrt{2\pi}}\, du$$

which is clearly infinite when $\lambda \geq 1/2$ and finite whenever $\lambda < 1/2$. Thus $\lambda_0 = 1/2$.

5. As a heuristic we consider the following example. Let X be a random variable such that

$$\mu_X = \sum_{n=1}^{\infty} \frac{1}{n^2}\, \delta_n.$$

Then the expectations are

$$\mathbb{E}[X^\alpha] = \sum_{n=1}^{\infty} \frac{n^\alpha}{n^2} = \begin{cases} \zeta(2-\alpha) & \alpha < 1 \\ \infty & \alpha \geq 1. \end{cases}$$

which is almost what we want for $\alpha_0 = 1$. The idea now is to try and modify by an exponential factor to perturb the edge case.

Let Y be a random variable such that

$$\mu_Y = \sum_{n=1}^{\infty} a_n \delta_n, \qquad a_n > 0.$$

With $X = e^Y$ the distribution becomes

$$\mu_X = \sum_{n=1}^{\infty} a_n \delta_{e^n}$$

and so the moments are computed as

$$\mathbb{E}[X^\alpha] = \int_{-\infty}^{\infty} x^\alpha \; d\mu_X(x) = \sum_{n=1}^{\infty} a_n e^{\alpha n}.$$

At this point the rate of decay of the a_n matters, because if $\alpha > 0$ then the exponential will dominate any polynomial decay. Let $\alpha_0 > 0$ be given and set $a_n = c \cdot e^{-\alpha_0 n}/n^2$ where c is a normalization constant, i.e.

$$c^{-1} = \sum_{n=1}^{\infty} \frac{e^{-\alpha_0 n}}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \quad \Rightarrow \quad \sum_{n=1}^{\infty} a_n = 1.$$

Then substituting into the above we have

$$\mathbb{E}[X^\alpha] = c \sum_{n=1}^{\infty} \frac{e^{(\alpha-\alpha_0)n}}{n^2} = \begin{cases} C_\alpha & \alpha \leq \alpha_0 \\ \infty & \alpha > \alpha_0 \end{cases}$$

for some $C_\alpha < \infty$. The contribution $1/n^2$ is necessary to establish convergence when $\alpha = \alpha_0$.

Problem 5.2: (Notes Problem 6.17). A random variable is said to have **exponential distribution** with parameter $\lambda > 0$ – denoted by $X \sim \text{Exp}(\lambda)$ – if its distribution function F_X is given by

$$F_X(x) = 0, \quad \text{and} \quad F_x(x) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0.$$

1. Compute $\mathbb{E}[X^\alpha]$, for $\alpha \in (-1, \infty)$. Combine your result with the result of part 3. of the previous problem to show that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

where Γ is the Gamma function.

2. Remember that the conditional probability $\mathbb{P}[A \mid B]$ of A, given B, for $A, B \in \mathcal{F}$, $\mathbb{P}[B] > 0$ is given by

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Compute $\mathbb{P}[X \geq x_2 \mid X \geq x_1]$, for $x_2 > x_1 > 0$ and compare it to $\mathbb{P}[X \geq (x_2 - x_1)]$.

Note: This can be interpreted as follows: the knowledge that the bulb stayed functional until x_1 does not change the probability that it will not burn in the next $x_2 - x_1$ units of time; bulbs have no memory (at least in probability textbooks).

Conversely, suppose that Y is a random variable with the property that

$$\mathbb{P}[Y \geq y_2 \mid Y \geq y_1] = \mathbb{P}[Y \geq (y_2 - y_1)], \quad \text{for all } y_2 > y_1 > 0.$$

Show that $Y \sim \text{Exp}(\lambda)$ for some $\lambda > 0$.

Hint: You can use the following fact: let $\phi : (0, \infty) \to \mathbb{R}$ be a Borel-measurable function such that $\phi(y) + \phi(z) = \phi(y + z)$ for all $y, z > 0$. Then there exists a constant $u \in \mathbb{R}$ such that $\phi(y) = uy$ for all $y > 0$. Interestingly, ϕ does not have to be linear if the measurability hypothesis is omitted.

Solution:

1. We recall that $f_X(x) = F'_X(x)$ so that
$$\mathbb{E}[X^\alpha] = \int_{-\infty}^{\infty} x^\alpha F'_X(x)\ dx = \int_0^\infty \lambda x^\alpha e^{-\lambda x}\ dx = \frac{1}{\lambda^\alpha}\int_0^\infty x^\alpha e^{-x}\ dx.$$
Moreover, by definition
$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\ dx, \qquad \alpha > 0$$
so
$$\mathbb{E}[X^\alpha] = \frac{\Gamma(\alpha+1)}{\lambda^\alpha}.$$
When $\lambda = 1$ and $\alpha = -1/2$ we have
$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{e^{-x}}{\sqrt{x}}\ dx = \int_0^\infty \frac{e^{-x/2}}{\sqrt{2x}}\ dx = \sqrt{\pi}\int_0^\infty \frac{e^{-x/2}}{\sqrt{2\pi x}}\ dx = \sqrt{\pi}.$$
We recognize the last integral as the integral of the pdf of the χ^2-distribution, so that the integral is just 1.
2. By definition
$$\mathbb{P}[X \geq x_2 \mid X \geq x_1] = \frac{\mathbb{P}[\{X \geq x_2\} \cap \{X \geq x_1\}]}{\mathbb{P}[X \geq x_1]} = \frac{\mathbb{P}[X \geq x_2]}{\mathbb{P}[X \geq x_1]} = \frac{1 - \mathbb{P}[X \leq x_2]}{1 - \mathbb{P}[X \leq x_1]}.$$
To justify the above, we must remark that $X \sim \text{Exp}(\lambda)$ and thus μ_X is atomless. Computing $1 - \mathbb{P}[X \leq x]$ for $x > 0$ yields
$$1 - \mathbb{P}[X \leq x] = 1 - F_X(x) = 1 - [1 - e^{-\lambda x}] = e^{-\lambda x}.$$
Particularly,
$$\mathbb{P}[X \geq x_2 \mid X \geq x_1] = \frac{e^{-\lambda x_2}}{e^{-\lambda x_1}} = e^{-\lambda(x_2 - x_1)} = \mathbb{P}[X \geq (x_2 - x_1)].$$

   On the other hand, let F_Y be the cdf of Y. The same work above holds with the following modification
$$\mathbb{P}[Y \geq y_2 \mid Y \geq y_1] = \frac{1 - \mathbb{P}[Y < y_2]}{1 - \mathbb{P}[Y < y_1]},$$
as we do not know that μ_Y is atomless. Combined with the assumption, we get
$$1 - \mathbb{P}[Y < (y_2 - y_1)] = \frac{1 - \mathbb{P}[Y < y_2]}{1 - \mathbb{P}[Y < y_1]} \quad .$$
Let $p(y) = 1 - \mathbb{P}[Y < y]$ for $y > 0$ – we implicitly assume $p : (0, \infty) \to (0, \infty)$ since if $p(y) = 0$ then the definition for conditional probability was meaningless (we need $\mathbb{P}[Y \geq y_1] > 0$). The above says for any $y, z > 0$ that $p(y)p(z) = p(y + z)$. To see this replace $y_2 - y_1$ by y and y_1 by z, then rearrange. Defining $\phi(y) = \log(p(y))$ we get $\phi(y) + \phi(z) = \phi(y + z)$. Per the hint, it means that there exists a $u \in \mathbb{R}$ such that $\phi(y) = uy$ for all $y > 0$. With $\lambda = -u$ we see that $1 - \mathbb{P}[Y < y] = e^{-\lambda y}$, or that $\mathbb{P}[Y < y] = 1 - e^{-\lambda y}$ for $y > 0$.

   Since $\mathbb{P}[Y < y]$ is monotone in y and $\lim_{y \to 0^+} \mathbb{P}[Y < y] = 0$, it follows that $\mathbb{P}[Y < y]$ for $y \leq 0$. Finally let $y_\infty \in \mathbb{R}$ and let $\{y_n\}_{n=1}^\infty \subset \mathbb{R}$ be such that $y_n \to y_\infty$ and $y_{n+1} \leq y_n$, e.g. $y_n = y_\infty + 1/n$. By dominated convergence of sets, since $\{Y < y_n\}$ form a decreasing sequence of sets with $\mathbb{P}[Y < y_n] < \infty$ (trivially), and continuity of $y \mapsto \mathbb{P}[Y < y]$ we get
$$\mathbb{P}[Y < y_\infty] = \lim_{n \to \infty} \mathbb{P}[Y < y_n] = \mathbb{P}\left[\bigcap_{n=1}^\infty \{Y < y_n\}\right] = \mathbb{P}[Y \leq y_\infty].$$

That is to say, we have shown the cdf of Y satisfies $F_Y(y) = 0$ for $y \leq 0$ and $F_Y(y) = 1 - e^{-\lambda y}$ for $y > 0$, i.e. $Y \sim \text{Exp}(\lambda)$.

Problem 5.3: (Scheffe's lemma, Notes Problem 7.6). Let $\{X_n\}_{n\in\mathbb{N}}$ be absolutely continuous random variables with densities f_{X_n}, such that $f_{X_n} \to f$, λ-a.e., where f is the density of the absolutely continuous random variable X and λ denotes the Lebesgue measure on $\mathbb{R}$. Show that $X_n \xrightarrow{\mathcal{D}} X$.

Hint: Show that

$$\int_{\mathbb{R}} |f_{X_n} - f| \; d\lambda \to 0$$

by writing the integrand in terms of $(f - f_{X_n})^+ \leq f$.

Solution: $X_n \xrightarrow{\mathcal{D}} X$ means that $\mu_{X_n} \xrightarrow{w} \mu_X$. Generally this occurs if for all $\phi \in C_b(\mathbb{R})$ we have

$$\int_{\mathbb{R}} \phi \; d\mu_{X_n} \to \int_{\mathbb{R}} \phi \; d\mu_X.$$

For absolutely continuous measures there is a nicer formulation. The above reads

$$\int_{\mathbb{R}} \phi(x) f_{X_n}(x) \; dx \to \int_{\mathbb{R}} \phi(x) f(x) \; dx,$$

which is implied by the stronger requirement that

$$\int_{\mathbb{R}} \phi(x) |f_{X_n}(x) - f(x)| \; dx \to 0.$$

Since $\phi(x)$ is bounded $\|\phi\|_{L^\infty} < \infty$ and we get

$$\int_{\mathbb{R}} \phi(x)|f_{X_n}(x) - f(x)| \; dx \leq \|\phi\|_{L^\infty(\mathbb{R})} \int_{\mathbb{R}} |f_{X_n}(x) - f(x)| \; dx.$$

So it suffices to show that

$$\int_{\mathbb{R}} |f_{X_n}(x) - f(x)| \; dx \to 0.$$

Now observe for a real-valued function g that $|g(x)| = g^+(x) + g^-(x)$ while $g(x) = g^+(x) - g^-(x)$. In turn $|g(x)| = 2g^+(x) - g(x)$. Applying this to $g = f - f_{X_n}$ yields $|f - f_{X_n}| = 2(f - f_{X_n})^+ - (f - f_{X_n})$. Importantly,

$$\int_{\mathbb{R}^n} |f - f_{X_n}| \; dx = 2\int_{\mathbb{R}} (f - f_{X_n})^+(x) \; dx - \int_{\mathbb{R}} f(x) \; dx + \int_{\mathbb{R}} f_{X_n}(x) \; dx = 2\int_{\mathbb{R}} (f - f_{X_n})^+(x) \; dx,$$

as f and the f_{X_n} are pdfs and integrate to 1. The trivial inequality $(f - f_{X_n})^+ \leq f$ lets us apply dominated convergence to conclude that

$$\lim_{n\to\infty} \int_{\mathbb{R}} (f - f_{X_n})^+(x) \; dx = 0$$

as $f_{X_n} \to f$ a.e. implies $(f - f_{X_n})^+ \to 0$ a.e.

HW 6

Problem 6.1: (The multivariate normal distribution, Notes Problem 8.7). The *characteristic function* $\varphi = \varphi_X : \mathbb{R}^n \to \mathbb{C}$ of a random vector $X = (X_1, ..., X_n)$ is given by

$$\varphi(t_1, t_2, ..., t_n) = \mathbb{E}\left[\exp\left(i \sum_{k=1}^n t_k X_k\right)\right]$$

for $t_1, ..., t_n \in \mathbb{R}$. We will use the shortcut $\boldsymbol{t}$ for $(t_1, ..., t_n)$ and $\boldsymbol{t} \cdot \boldsymbol{X}$ for the random variable $\sum_{k=1}^n t_k X_k$. Take for granted the following fact (the proof of which is similar to the proof of the 1-dimensional case): *Random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same distribution if $\varphi_{\boldsymbol{X}_1}(\boldsymbol{t}) = \varphi_{\boldsymbol{X}_2}(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathbb{R}^n$*, and prove the statements below:

1. Random variables X and Y are independent if and only if $\varphi_{(X,Y)}(t_1, t_2) = \varphi_X(t_1)\varphi_Y(t_2)$ for all $t_1, t_2 \in \mathbb{R}$.

2. Random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same distribution if and only if random variables $\boldsymbol{t} \cdot \boldsymbol{X}_1$ and $\boldsymbol{t} \cdot \boldsymbol{X}_2$ have the same distribution for all $\boldsymbol{t} \in \mathbb{R}^n$. (This fact is known as *Wald's device.*)

An n-dimensional random vector $\boldsymbol{X}$ is said to be *Gaussian* (or, to have the *multivariate normal distribution*) if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and a symmetric positive semi-definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ such that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\left(i\langle \boldsymbol{t}, \boldsymbol{\mu}\rangle - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right),$$

where $\boldsymbol{t}$ is interpreted as a column vector, and $()^T$ is transposition. This is denoted as $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. $\boldsymbol{X}$ is said to be *non-degenerate* if $\boldsymbol{\Sigma}$ is positive definite. Note that, according to this definition, every constant random variable is automatically Gaussian.

3. Show that a random vector $\boldsymbol{X}$ is Gaussian, if and only if the random vector $\boldsymbol{t}\cdot\boldsymbol{X}$ is normally distributed (with some mean and variance) for each $\boldsymbol{t} \in \mathbb{R}^n$. *Note*: Be careful, nothing in the second statement tells you what the mean and variance of $\boldsymbol{t} \cdot \boldsymbol{X}$ are.
4. Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a Gaussian random vector. Show that X_k and X_l, $k \neq l$, are independent if and only if they are uncorrelated.
5. Construct a random vector (X, Y) such that both X and Y are normally distributed, but that $\boldsymbol{X} = (X, Y)$ is not Gaussian.
6. Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be a random vector consisting of n independent random variables with $X_i \sim \mathcal{N}(0, 1)$. Let $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ be a given positive semi-definite symmetric matrix, and $\boldsymbol{\mu} \in \mathbb{R}^n$ a given vector. Show that there exists an affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that the random vector $T(\boldsymbol{X})$ is Gaussian with $T(\boldsymbol{X}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Solution:

1. Suppose first that X and Y are independent. By definition of the characteristic function,

$$\begin{aligned}\varphi_{(X,Y)}(t_1, t_2) = \mathbb{E}[e^{i(t_1X+t_2Y)}] &= \int_{\mathbb{R}^2} e^{i(t_1x+t_2y)}\ d\mu_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}^2} e^{it_1x}e^{it_2y}\ d\mu_{(X,Y)}(x, y).\end{aligned}$$

Owing to independence we have that $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$ so that

$$\int_{\mathbb{R}^2} e^{it_1x}e^{it_2y}\ d\mu_{(X,Y)}(x, y) = \int_{\mathbb{R}^2} e^{it_1x}e^{it_2y}\ d(\mu_X \otimes \mu_Y)(x, y)$$

But, Fubini's theorem (valid even in the complex-valued case) yields

$$\begin{aligned}\int_{\mathbb{R}^2} e^{it_1x}e^{it_2y}\ d(\mu_X \otimes \mu_Y)(x, y) &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{it_1x}e^{it_2y}\ d\mu_X(x)d\mu_Y(y) \\ &= \int_{\mathbb{R}} e^{it_1x}\ d\mu_X(x)\int_{\mathbb{R}} e^{it_2y}\ d\mu_Y(y) = \varphi_X(t_1)\varphi_Y(t_2).\end{aligned}$$

On the other hand suppose that $\varphi_{(X,Y)}(t_1, t_2) = \varphi_X(t_1)\varphi_Y(t_2)$ for all $t_1, t_2 \in \mathbb{R}$. The only line above which used independence is the third middle display. Particularly the last display reads $\varphi_{\mu_X\otimes\mu_Y}(t_1, t_2) = \varphi_X(t_1)\varphi_Y(t_2)$. Owing to the assumption, and the notation that $\varphi_{(X,Y)} = \varphi_{\mu_{(X,Y)}}$ we have

$$\varphi_{\mu_{(X,Y)}}(t_1, t_2) = \varphi_{\mu_X\otimes\mu_Y}(t_1, t_2)$$

for all $t_1, t_2 \in \mathbb{R}$. It follows by the provided fact taken for granted that $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$, i.e. X and Y are independent.

2. Let $\boldsymbol{t} = (t_1, ..., t_n) \in \mathbb{R}^n$ and $\boldsymbol{X} = (X_1, ..., X_n)$ so that $\boldsymbol{t} \cdot \boldsymbol{X} = t_1 X_1 + ... + t_n X_n$. Then for $s \in \mathbb{R}$ we have

$$\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(s) = \mathbb{E}[e^{is\boldsymbol{t}\cdot\boldsymbol{X}}] = \mathbb{E}\left[\exp\left(is\left(\sum_{k=1}^{n} t_k X_n\right)\right)\right]$$
$$= \mathbb{E}\left[\exp\left(i\sum_{k=1}^{n} st_k X_n\right)\right] = \varphi_{\boldsymbol{X}}(s\boldsymbol{t}),$$

where $s\boldsymbol{t} := (st_1, ..., st_n)$. Assuming $\boldsymbol{t} \cdot \boldsymbol{X}_1$ and $\boldsymbol{t} \cdot \boldsymbol{X}_2$ have the same distribution for all $\boldsymbol{t} \in \mathbb{R}^n$, it follows that $\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}_1}(s) = \varphi_{\boldsymbol{t}\cdot\boldsymbol{X}_2}(s)$ for all $s \in \mathbb{R}$. In particular by testing at $s = 1$ we see that $\varphi_{\boldsymbol{X}_1}(\boldsymbol{t}) = \varphi_{\boldsymbol{X}_2}(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathbb{R}^n$. By the fact taken for granted, this means precisely that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same distribution. Now, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same distribution then $\varphi_{\boldsymbol{X}_1}(\boldsymbol{t}) = \varphi_{\boldsymbol{X}_2}(\boldsymbol{t})$ for all $\boldsymbol{t} \in \mathbb{R}^n$. Let $\boldsymbol{t} \in \mathbb{R}^n$ and $s \in \mathbb{R}$. Then applying this to $\bar{\boldsymbol{t}} = s\boldsymbol{t}$ we have that

$$\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}_1}(s) = \varphi_{\boldsymbol{X}_1}(s\boldsymbol{t}) = \varphi_{\boldsymbol{X}_2}(s\boldsymbol{t}) = \varphi_{\boldsymbol{t}\cdot\boldsymbol{X}_2}(s),$$

i.e. for any $\boldsymbol{t} \in \mathbb{R}^n$ the functions $\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}_1}$ and $\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}_2}$ are equal. This once again means that $\boldsymbol{t} \cdot \boldsymbol{X}_1$ and $\boldsymbol{t} \cdot \boldsymbol{X}_2$ have the same distribution.

3. Suppose first that $\boldsymbol{t}\cdot\boldsymbol{X}$ is normally distributed for each $\boldsymbol{t} \in \mathbb{R}^n$. To show that $\boldsymbol{X}$ is Gaussian it suffices to compute its characteristic function. We previously showed that $\varphi_{\boldsymbol{X}}(s\boldsymbol{t}) = \varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(s)$. Taking $s = 1$ we see $\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(1)$. Thus we aim to explicitly compute $\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(s)$ for any $\boldsymbol{t} \in \mathbb{R}^n$. Fixing $\boldsymbol{t} \in \mathbb{R}^n$, since $\boldsymbol{t} \cdot \boldsymbol{X}$ is normally distributed there exist $\mu_{\boldsymbol{t}} \in \mathbb{R}$ and $\sigma_{\boldsymbol{t}} > 0$ such that

$$\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(s) = e^{i\mu_{\boldsymbol{t}}s - 1/2\sigma_{\boldsymbol{t}}^2 s^2}.$$

Moreover, it follows that all the n-th moments of the random variable $\boldsymbol{t} \cdot \boldsymbol{X}$ exist for $n \in \mathbb{N}$. In particular, we may compute the first and second moments as:

$$\mathbb{E}[\boldsymbol{t}\cdot\boldsymbol{X}] = (-i)\frac{d}{ds}\left[e^{i\mu_{\boldsymbol{t}} - 1/2\sigma_{\boldsymbol{t}}^2 s^2}\right]_{s=0} = -i\left[(i\mu_{\boldsymbol{t}} - \sigma_{\boldsymbol{t}}^2 s)e^{i\mu_{\boldsymbol{t}} - 1/2\sigma_{\boldsymbol{t}}^2 s^2}\right]_{s=0} = \mu_{\boldsymbol{t}}$$
$$\mathbb{E}[(\boldsymbol{t}\cdot\boldsymbol{X})^2] = (-i)^2\frac{d^2}{d^2 s}\varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(s)\bigg|_{s=0} = -\frac{d^2}{d^2 s}\left[e^{i\mu_{\boldsymbol{t}}s - 1/2\sigma_{\boldsymbol{t}}^2 s^2}\right]_{s=0}$$
$$= -\frac{d}{ds}\left[\left(i\mu_{\boldsymbol{t}} - \sigma_{\boldsymbol{t}}^2 s\right)e^{i\mu_{\boldsymbol{t}}s - 1/2\sigma_{\boldsymbol{t}}^2 s^2}\right]_{s=0}$$
$$= -\left[-\sigma_{\boldsymbol{t}}^2 e^{i\mu_{\boldsymbol{t}}s - 1/2\sigma_{\boldsymbol{t}}^2 s^2} + \left(i\mu_{\boldsymbol{t}} - \sigma_{\boldsymbol{t}}^2 s\right)^2 e^{i\mu_{\boldsymbol{t}}s - 1/2\sigma_{\boldsymbol{t}}^2 s^2}\right]_{s=0}$$
$$= -\left[-\sigma_{\boldsymbol{t}}^2 + (i\mu_{\boldsymbol{t}})^2\right] = \sigma_{\boldsymbol{t}}^2 + \mu_{\boldsymbol{t}}^2 = \sigma_{\boldsymbol{t}}^2 + \mathbb{E}[\boldsymbol{t}\cdot\boldsymbol{X}]^2.$$

Recalling that $\mathrm{Var}(\boldsymbol{t}\cdot\boldsymbol{X}) = \mathbb{E}[(\boldsymbol{t}\cdot\boldsymbol{X})^2] - \mathbb{E}[\boldsymbol{t}\cdot\boldsymbol{X}]^2$, we see that $\sigma_{\boldsymbol{t}}^2 = \mathrm{Var}(\boldsymbol{t}\cdot\boldsymbol{X})$. In total we have shown for any $\boldsymbol{t} \in \mathbb{R}^n$ that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = \varphi_{\boldsymbol{t}\cdot\boldsymbol{X}}(1) = e^{i\mu_{\boldsymbol{t}} - 1/2\sigma_{\boldsymbol{t}}^2} = e^{i\mathbb{E}[\boldsymbol{t}\cdot\boldsymbol{X}] - 1/2\,\mathrm{Var}(\boldsymbol{t}\cdot\boldsymbol{X})}.$$

For $\boldsymbol{X}$ to be Gaussian there must be $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma}$ an $n\times n$ symmetric, positive semi-definite matrix such that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = e^{i\langle\boldsymbol{t},\boldsymbol{\mu}\rangle - 1/2\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}.$$

So we make the ansatz that there exist $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ such that

$$\langle\boldsymbol{t},\boldsymbol{\mu}\rangle = \mathbb{E}[\boldsymbol{t}\cdot\boldsymbol{X}], \qquad \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} = \mathrm{Var}(\boldsymbol{t}\cdot\boldsymbol{X}).$$

Finding $\boldsymbol{\mu}$ is easy. Noting that $\boldsymbol{\mu}_k = \langle e_k, \boldsymbol{\mu}\rangle$ and that $e_k \cdot X = X_k$ we have

$$\boldsymbol{\mu} = (\mathbb{E}[X_1], ..., \mathbb{E}[X_n]).$$

For $\boldsymbol{\Sigma}$ a generic symmetric $n \times n$ matrix we have that

$$\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} = \sum_{j=1}^{n} t_j \sum_{k=1}^{n} (\boldsymbol{\Sigma})_{jk} t_k = \sum_{j=1}^{n} t_j^2 (\boldsymbol{\Sigma})_{jj} + 2\sum_{j>k}\sum_{k=1}^{n} t_j t_k (\boldsymbol{\Sigma})_{jk}.$$

Thus to determine $\boldsymbol{\Sigma}$ such that $\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} = \operatorname{Var}(\boldsymbol{t}\cdot\boldsymbol{X})$ we must express the latter as above:

$$\begin{aligned}\operatorname{Var}(\boldsymbol{t}\cdot\boldsymbol{X}) &= \mathbb{E}[(\boldsymbol{t}\cdot\boldsymbol{X})^2] - \mathbb{E}[\boldsymbol{t}\cdot\boldsymbol{X}]^2 \\ &= \mathbb{E}[(t_1X_1 + \dots + t_nX_n)^2] - \mathbb{E}[t_1X_1 + \dots + t_nX_n]^2 \\ &= \left(\sum_{j=1}^n t_j^2\mathbb{E}[X_j^2] + 2\sum_{j>k}\sum_{k=1}^n t_jt_k\mathbb{E}[X_jX_k]\right) \\ &\qquad - \left(\sum_{j=1}^n t_j^2\mathbb{E}[X_j]^2 + 2\sum_{j>k}\sum_{k=1}^n t_jt_k\mathbb{E}[X_j]\mathbb{E}[X_k]\right) \\ &= \sum_{j=1}^n t_j^2\left(\mathbb{E}[X_j^2] - \mathbb{E}[X_j]^2\right) + 2\sum_{j>k}\sum_{k=1}^n t_jt_k\left(\mathbb{E}[X_jX_k] - \mathbb{E}[X_j]\mathbb{E}[X_k]\right) \\ &= \sum_{j=1}^n t_j^2 \operatorname{Cov}(X_j,X_j) + 2\sum_{j>k} t_jt_k \operatorname{Cov}(X_j,X_k).\end{aligned}$$

Setting $(\boldsymbol{\Sigma})_{jk} = \operatorname{Cov}(X_j,X_k)$ we have $\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} = \operatorname{Var}(\boldsymbol{t}\cdot\boldsymbol{X}) = \sigma_{\boldsymbol{t}}^2 > 0$. So indeed

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) = e^{i\langle \boldsymbol{t},\boldsymbol{\mu}\rangle - 1/2\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}},$$

for some $\boldsymbol{\mu}\in\mathbb{R}$ and $\boldsymbol{\Sigma}$ an $n\times n$ symmetric positive semi-definite matrix. In particular, these are given by

$$(\boldsymbol{\mu})_k = \mathbb{E}[X_k], \quad (\boldsymbol{\Sigma})_{jk} = \operatorname{Cov}(X_j,X_k).$$

Suppose now that $\boldsymbol{X}$ is Gaussian. Then we can write

$$\varphi_{\boldsymbol{t}\cdot X}(s) = \varphi_{\boldsymbol{X}}(s\boldsymbol{t}) = e^{i\langle s\boldsymbol{t},\boldsymbol{\mu}\rangle - 1/2(s\boldsymbol{t})^T\boldsymbol{\Sigma}(s\boldsymbol{t})} = \varphi_{\boldsymbol{X}}(s\boldsymbol{t}) = e^{i\langle \boldsymbol{t},\boldsymbol{\mu}\rangle s - 1/2\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}s^2}$$

for any $s\in\mathbb{R}$ and $\boldsymbol{t}\in\mathbb{R}^n$, and some $\boldsymbol{\mu}\in\mathbb{R}^n$ and $\boldsymbol{\Sigma}$ an $n\times n$ positive semi-definite matrix. Setting $\mu_{\boldsymbol{t}} = \langle\boldsymbol{t},\boldsymbol{\mu}\rangle$ and $\sigma_{\boldsymbol{t}} = (\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})^{1/2}$, which is well defined since $\boldsymbol{\Sigma}$ is positive semi-definite, we see that $\boldsymbol{t}\cdot\boldsymbol{X} \sim \mathcal{N}(\mu_{\boldsymbol{t}},\sigma_{\boldsymbol{t}})$.

4. Recall that independent random variables are uncorrelated. It suffices then to prove that if $\boldsymbol{X}$ is a random vector then X_k and X_l, for $k\neq l$, are independent if they are uncorrelated. Assuming they are uncorrelated, we have $\operatorname{Cov}(X_j,X_k) = 0$. In particular the random vector $\bar{\boldsymbol{X}} = (X_j,X_k)$ is such that $\bar{\boldsymbol{\Sigma}}$ is diagonal. Hence for $\boldsymbol{t}\in\mathbb{R}^2$ we have

$$\boldsymbol{t}^T\begin{pmatrix}\operatorname{Var}(X_j) & 0\\ 0 & \operatorname{Var}(X_k)\end{pmatrix}\boldsymbol{t} = t_1^2\operatorname{Var}(X_j) + t_2^2\operatorname{Var}(X_k).$$

It follows that

$$\begin{aligned}\varphi_{\bar{\boldsymbol{X}}}(t_1,t_2) &= e^{i(t_1\mathbb{E}[X_j] + t_2\mathbb{E}[X_k]) - 1/2(t_1^2\operatorname{Var}(X_j) + t_2^2\operatorname{Var}(X_k))} \\ &= e^{it_1\mathbb{E}[X_j] - 1/2t_1^2\operatorname{Var}(X_j)}e^{it_2\mathbb{E}[X_k] - 1/2t_2^2\operatorname{Var}(X_k)} = \varphi_{X_j}(t_1)\varphi_{X_k}(t_2).\end{aligned}$$

By part 1, X_j and X_k are independent.

5. We saw in part 3 that $\boldsymbol{X}$ is Gaussian whenever $t_1X + t_2Y$ is normally distributed for any $(t_1,t_2)\in\mathbb{R}$. We thus aim to construct X and Y such that $X+Y$ in particular is not normally distributed.

Let X and X' be independent. Suppose that X is normally distributed with mean zero and

$$\mathbb{P}[X'\geq 0] = \mathbb{P}[X' < 0] = \frac{1}{2}.$$

Let $Y = X\cdot\operatorname{sgn}(X')$. For $B\in\mathcal{B}(\mathbb{R})$ note that

$$[\{X\in B\}\cap\{X'\geq 0\}]\cup[\{X\in -B\}\cap\{X'<0\}] = \{Y\in B\}$$

where $-B = \{-x\in\mathbb{R} \mid x\in B\}$. Since the former two sets are disjoint,

$$\mu_Y(B) = \mathbb{P}[Y\in B] = \mathbb{P}[\{X\in B\}\cap\{X'\geq 0\}] + \mathbb{P}[\{X\in -B\}\cap\{X'<0\}].$$

Owing to independence we have

$$\mathbb{P}[\{X \in B\} \cap \{X' \geq 0\}] = \mathbb{P}[X \in B] \cdot \mathbb{P}[X' \geq 0] = \frac{1}{2}\mathbb{P}[X \in B]$$

since X' is normally distributed. Similarly we have

$$\mathbb{P}[\{X \in -B\} \cap \{X' < 0\}] = \mathbb{P}[X \in B] \cdot \mathbb{P}[X' < 0] = \frac{1}{2}\mathbb{P}[X \in -B] = \frac{1}{2}\mathbb{P}[X \in B],$$

where we use the fact that the pdf of X is symmetric about the origin. Consequently $\mu_Y(B) = \mu_X(B)$ for all $B \in \mathcal{B}(\mathbb{R})$ and Y is normally distributed. However,

$$\begin{aligned}\mathbb{P}[X + Y = 0] &= \mathbb{P}[X(1 + \operatorname{sgn}(X')) = 0] = \mathbb{P}[X = 0] + \mathbb{P}[\{X \neq 0\} \cap \{X' < 0\}] \\ &= \mathbb{P}[X \neq 0] \cdot \mathbb{P}[X' < 0] = \mathbb{P}[X' < 0] = \frac{1}{2}.\end{aligned}$$

Since normal distributions are absolutely continuous, it follows that $X + Y$ is not normally distributed.

6. Affine transformations are characterized as $T(x) = Ax + b$ where $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. Then

$$T(\boldsymbol{X})_k = \sum_{j=1}^n A_{kj}X_j + b_k$$

and the characteristic function of $T(\boldsymbol{X})$ is then given by

$$\begin{aligned}\varphi_{T(\boldsymbol{X})}(\boldsymbol{t}) &= \mathbb{E}\left[\exp\left(i\sum_{k=1}^n t_k T(\boldsymbol{X})_k\right)\right] = \mathbb{E}\left[\exp\left(i\sum_{k=1}^n t_k\left(\sum_{j=1}^n A_{kj}X_j + b_k\right)\right)\right] \\ &= e^{i\langle \boldsymbol{t}, b\rangle}\mathbb{E}\left[\exp\left(i\sum_{j=1}^n\left(\sum_{k=1}^n t_kA_{kj}\right)X_j\right)\right] = e^{i\langle \boldsymbol{t}, b\rangle}\varphi_{\boldsymbol{X}}\left(\sum_{k=1}^n t_kA_{k1}, ..., \sum_{k=1}^n t_kA_{kn}\right).\end{aligned}$$

Since $X_i \sim \mathcal{N}(0,1)$ it follows that

$$\varphi_{\boldsymbol{X}}(\boldsymbol{s}) = e^{-1/2|\boldsymbol{s}|^2}.$$

Hence,

$$\begin{aligned}\varphi_{\boldsymbol{X}}\left(\sum_{k=1}^n t_kA_{k1}, ..., \sum_{k=1}^n t_kA_{kn}\right) &= \exp\left(-\frac{1}{2}\sum_{j=1}^n\left(\sum_{k=1}^n t_kA_{kj}\right)^2\right) \\ &= \exp\left(-\frac{1}{2}\sum_{j=1}^n\left(\sum_{k=1}^n t_k^2A_{kj}^2 + \sum_{m>k}\sum_{k=1}^n t_kt_mA_{kj}A_{mj}\right)\right) \\ &= \exp\left(-\frac{1}{2}\left(\sum_{k=1}^n t_k^2\sum_{j=1}^n A_{kj}^2 + \sum_{m>k}\sum_{k=1}^n t_kt_m\sum_{j=1}^n A_{kj}A_{mj}\right)\right) \\ &= \exp\left(-\frac{1}{2}\boldsymbol{t}^TAA^T\boldsymbol{t}\right).\end{aligned}$$

In total,

$$\varphi_{T(\boldsymbol{X})}(\boldsymbol{t}) = e^{i\langle \boldsymbol{t}, b\rangle - 1/2\boldsymbol{t}^TAA^T\boldsymbol{t}}.$$

If we want $T(\boldsymbol{X}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then we simply need to choose $b = \boldsymbol{\mu}$ and find $A \in \mathbb{R}^{n\times n}$ such that $AA^T = \Sigma$. Equivalently, $A^TA = \Sigma$. Since Σ is positive semi-definite there exists $\Sigma^{1/2}$. As Σ is symmetric so too is $\Sigma^{1/2}$. So, choose $A = \Sigma^{1/2}$.

Problem 6.2: (Notes Problem 8.5). Let μ be a probability measure on $\mathcal{B}(\mathbb{R})$, and let $\varphi = \varphi_\mu$ be its characteristic function.

1. Show that

$$\mu(\{a\}) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-ita}\varphi(t)\ dt.$$

2. Show that μ has no atoms if $\lim_{t\to\infty}|\varphi(t)| = \lim_{t\to-\infty}|\varphi(t)| = 0$.
3. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent random variables with $\mathbb{P}[X_n = \pm 1] = 1/2$ and let $Y = \sum_{k=1}^{\infty} 3^{-k} X_k$. Use the distribution of Y to show the converse of 2 above does not hold.

   Hint: Evaluate $\varphi_Y(t_m)$ along a well-chosen geometric sequence $\{t_m\}_{m=1}^{\infty}$. Use and prove the inequality $\log(\cos(x)) \geq -x^2$ for x small enough, to make sure that $\varphi_Y(t_m) \not\to 0$.

Solution:

1. Mimicking the proof of the inversion formula, we consider the integral

$$I_T(a) = \frac{1}{2T}\int_{-T}^{T} e^{-ita}\varphi(t)\ dt = \frac{1}{2T}\int_{-T}^{T}\int_{-\infty}^{\infty} e^{-ita}e^{itx}\ d\mu(x)dt = \frac{1}{2T}\int_{-T}^{T}\int_{-\infty}^{\infty} e^{it(x-a)}\ d\mu(x)dt.$$

   Note that

$$\frac{1}{2T}\int_{-T}^{T}\int_{-\infty}^{\infty} |e^{it(x-a)}|\ d\mu(x)dt \leq \frac{1}{2T}\int_{-T}^{T}\mu(\mathbb{R})\ dt = 1,$$

   so that Fubini's theorem guarantees

$$\begin{aligned} I_T(a) &= \frac{1}{2T}\int_{-\infty}^{\infty}\int_{-T}^{T} e^{it(x-a)}\ dtd\mu(x) \\ &= \frac{1}{2T}\int_{-\infty}^{\infty}\int_{-T}^{T} (\cos(t(x-a)) + i\sin(t(x-a))\ dtd\mu(x) \\ &= \int_{-\infty}^{\infty}\frac{1}{2T}\int_{-T}^{T}\cos(t(x-a))\ dtd\mu(x). \end{aligned}$$

   We now aim to show that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\cos(t(x-a))\ dt = \begin{cases} 1 & x = a \\ 0 & x \neq a. \end{cases}$$

   Clearly the above holds when $x = a$. Suppose that $x \neq a$, then

$$\int_{-T}^{T}\cos(t(x-a))\ dt = \frac{2\sin(T(x-a))}{x-a} \leq \frac{2}{x-a}.$$

   Consequently

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\cos(t(x-a))\ dt \leq \lim_{T\to\infty}\frac{1}{T(x-a)} = 0.$$

   Hence,

$$\lim_{T\to\infty} I_T(a) = \int_{-\infty}^{\infty}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\cos(t(x-a))\ dtd\mu(x) = \int_{-\infty}^{\infty}\chi_{\{a\}}\ d\mu(x) = \mu(\{a\}).$$

2. Let $\epsilon > 0$. Then there exists a $T > 0$ such that for $|t| \geq T_0$ we have $|\varphi(t)| < \epsilon$. Then for $T > T_0$,

$$\begin{aligned} \frac{1}{2T}\int_{-T}^{T} e^{-ita}\varphi(t)\ dt &= \frac{1}{2T}\left(\int_{-T}^{-T_0} e^{-ita}\varphi(t)\ dt + \int_{-T_0}^{T_0} e^{-ita}\varphi(t)\ dt + \int_{T_0}^{T} e^{-ita}\varphi(t)\ dt\right) \\ &< \frac{1}{2T}\left(\epsilon\int_{-T}^{-T_0} e^{-ita}\ dt + 2T_0 + \epsilon\int_{T_0}^{T} e^{-ita}\ dt\right) \\ &= \frac{1}{2T}\left(2T_0 + \epsilon\left(\frac{e^{iTa} - e^{iT_0a}}{ia}\right) + \epsilon\left(\frac{e^{-iT_0a} - e^{-iTa}}{ia}\right)\right) \\ &= \frac{1}{T}\left(T_0 + \epsilon\left(\frac{\sin(Ta) - \sin(T_0a)}{a}\right)\right) = \epsilon\left(\frac{\sin(Ta)}{Ta}\right) + \frac{C(\epsilon, a)}{T}. \end{aligned}$$

As $\sin(t)/t \to 0$ as $T \to \infty$, it follows that

$$\mu(\{a\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ita}\varphi(t)\,dt = \lim_{T\to\infty} \left(\epsilon \left(\frac{\sin(Ta)}{Ta} \right) + \frac{C(\epsilon,a)}{T} \right) = 0.$$

3. We prove the inequality in the hint first. Let $f : (-\pi/2, \pi/2) \to \mathbb{R}$ be given by $f(x) = \log(\cos(x))$. Then f is analytic and

$$f'(x) = -\tan(x), \quad f''(x) = -\sec(x)^2,$$
$$f'''(x) = -2\sec(x)^2\tan(x), \quad f''''(x) = -4\sec(x)^2\tan(x)^2 - 2\sec(x)^4$$

It follows that

$$f(0) = 0, \quad f'(0) = 0, \quad f''(0) = -1, \quad f'''(0) = 0, \quad f''''(0) = -2.$$

The second order Taylor approximation of $f(x)$ at $x = 0$ is thus $-1/2x^2$, and since the next non-zero derivative is negative, for small x we have $-1/2x^2 \geq f(x)$. Moreover, for any $a < -1/2$ we have $ax^2 < f(x)$ in a neighborhood of zero.

Since all the X_n are independent, we have that

$$\varphi_Y(t) = \prod_{k=1}^{\infty} \varphi_{3^{-k}X_k}(t) = \prod_{k=1}^{\infty} \varphi_{3^{-k}X_k}(t) = \prod_{k=1}^{\infty} \varphi_{X_k}(3^{-k}t).$$

Since $\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2$, it follows that $\mu_{X_n} = 1/2\delta_{-1} + 1/2\delta_1$ for all n, and thus $\varphi_{X_n}(t) = \cos(t)$. Consequently,

$$\varphi_Y(t) = \prod_{k=1}^{\infty} \cos(3^{-k}t).$$

Note now that

$$\varphi_Y(3t) = \prod_{k=1}^{\infty} \cos(3^{-k+1}t) = \cos(t) \prod_{k=2}^{\infty} \cos(3^{-k+1}t) = \cos(t)\varphi_Y(t).$$

We aim to choose t_m so that $\varphi_Y(t_m)$ is independent of m. To this end let $t_m = c \cdot a^m$. Then we need $3t_m = t_{m+1}$, so that $3 \cdot a^m = a^{m+1}$. Hence, $a = 3$. Next we need $\cos(c \cdot a^m) = 1$. Since a^m is always an integer it suffices to choose $c = 0, 2\pi, ...$ We also want $t_m \to \infty$, so choose e.g. $c = 2\pi$. Hence we select $t_m = 2\pi \cdot 3^m$.

To show that $\varphi_Y(t_m) \not\to 0$ it suffices to show $\varphi_Y(2\pi) \neq 0$, as $\varphi_Y(t_m) = \varphi_Y(2\pi)$. Equivalently, we prove $\log(\varphi_Y(2\pi)) > -\infty$. Note that

$$\log(\varphi_Y(t)) = \sum_{k=1}^{\infty} \log(\cos(3^{-k}t))$$

for small t. Per the hint, we want to estimate $\log(\cos(2\pi/3^k))$. The issue is that $2\pi/3^k$ may be too large to apply the hint. Let $M \geq 0$ be such that for $|x| \leq M$ we have $\log(\cos(x)) \geq -x^2$. Choose $K \in \mathbb{N}$ large enough so that $2\pi/3^k \leq M$ for all $k \geq K$. Then,

$$\begin{aligned}
\log(\varphi_Y(2\pi)) &= \sum_{k=1}^{K-1} \log(\cos(2\pi/3^k)) + \sum_{k=K}^{\infty} \log(\cos(2\pi/3^k)) \\
&\geq \sum_{k=1}^{K-1} \log(\cos(2\pi/3^k)) - 4\pi^2 \sum_{k=K}^{\infty} \frac{1}{9^k} > -\infty.
\end{aligned}$$

Problem 6.3: (Notes Problem 8.9).

1. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of normally-distributed random variables converging in distribution towards a random variable X. Show that
   a) $\mathbb{E}[X_n] \to \mathbb{E}[X]$ and $\mathrm{Var}(X_n) \to \mathrm{Var}(X)$, and that
   b) X must be normally distributed itself.

Hint: You can use, without proof, the following fact from analysis: if $\{\mu_n\}_{n=1}^{\infty} \subseteq \mathbb{R}$ has the property that $e^{it\mu_n}$ converges in $\mathbb{R}$ for each t, then the sequence μ_n must be convergent in $\mathbb{R}$.

2. Let X_n be a sequence of random variables in L^p such that $X_n \to X$, a.s., and $\mathbb{E}[|X_n|^p] \to \mathbb{E}[|X|^p]$, for some $X \in L^p$. Show that $X_n \xrightarrow{L^p} X$. *Hint*: Use the inequality $2^p(|X_n|^p + |X|^p) - |X_n - X|^p \geq 0$.
3. Show that for sequences of normal random variables, the almost-sure convergence implies the convergence in L^p for each p.

Solution:

1. Recall that since the X_n are normally distributed there exist $\mu_n \in \mathbb{R}$ and $\sigma_n > 0$ such that
$$\varphi_{X_n}(t) = e^{it\mu_n - 1/2\sigma_n^2 t^2}.$$
   a) Since $x \mapsto e^{itx}$ is continuous and bounded it follows that $\mu_{X_n} \xrightarrow{w} \mu_X$ implying
$$\varphi_{X_n}(t) = \mathbb{E}[e^{itX_n}] \to \mathbb{E}[e^{itX}] = \varphi_X(t).$$
   The goal is to find $\mu \in \mathbb{R}$ and $\sigma \geq 0$ such that
$$\varphi_X(t) = e^{it\mu - 1/2\sigma^2 t^2}.$$
   Analyzing the modulus, we have that
$$|\varphi_{X_n}(t)| = e^{-1/2\sigma_n^2 t^2}$$
   converges to $|\varphi_X(t)|$. In particular, this holds at $t = 1$ so that
$$\sigma_n \to \sqrt{-2\log(|\varphi_X(1)|)} := \sigma \geq 0$$
   Now we establish convergence of the μ_n. Since
$$\begin{aligned}\varphi_X(t) &= \lim_{n\to\infty} \left[e^{it\mu_n - 1/2\sigma_n^2 t^2}\right] \\ &= \left[\lim_{n\to\infty} e^{it\mu_n}\right] \cdot \left[\lim_{n\to\infty} e^{-1/2\sigma_n^2 t^2}\right] = e^{-1/2\sigma^2 t^2}\left[\lim_{n\to\infty} e^{it\mu_n}\right],\end{aligned}$$
   it follows that the last limit exists and is such that
$$\lim_{n\to\infty} e^{it\mu_n} = e^{1/2\sigma^2 t^2}\varphi_X(t).$$
   Via the hint we see that $\{\mu_n\}_{n=1}^{\infty}$ converges to some $\mu \in \mathbb{R}$.

   Showing that $\mu = \mathbb{E}[X]$ and $\sigma = \mathrm{Var}(X)$ is deferred to the next part.

   b) To show that $X \sim \mathcal{N}(\mu, \sigma)$ just note that by continuity of the exponential
$$\varphi_X(t) = \lim_{n\to\infty} \varphi_{X_n}(t) = \lim_{n\to\infty} e^{it\mu_n - 1/2\sigma_n^2 t^2} = e^{it\mu - 1/2\sigma^2 t^2}.$$

2. We say that $X_n \xrightarrow{L^p} X$ if $\mathbb{E}[|X_n - X|^p] \to 0$. Since $X_n \to X$ a.s. we have that $|X_n - X|^p \to 0$ a.s. Moreover since $\mathbb{E}[|X_n|^p] \to \mathbb{E}[|X|^p]$ we have for $\epsilon > 0$ that there exists $N \in \mathbb{N}$ such that for $n \geq N$,
$$\mathbb{E}[|X|^p] - \epsilon < \mathbb{E}[|X_n|^p] < \mathbb{E}[|X|^p] + \epsilon.$$
   In particular, for all $n \in \mathbb{N}$ we have
$$\mathbb{E}[|X_n|^p] \leq \max\{\mathbb{E}[|X_1|^p], ...\mathbb{E}[|X_{N-1}|^p], \mathbb{E}[|X|^p] + \epsilon\} := M_\epsilon.$$
   Finally, via the upper bound we have
$$\mathbb{E}[2^p(|X_n|^p + |X|^p)] \leq 2^p(\mathbb{E}[|X_n|^p] + \mathbb{E}[|X|^p]) < 2^p(\mathbb{E}[|X|^p] + M_\epsilon) < \infty.$$
   Hence the inequality in the hint, $|X_n - X|^p \leq 2^p(|X_n|^p + |X|)$, allows for the use of dominated convergence to conclude.

3. I'm not entirely sure. My best guess is to consider new random variables $Y_n = \mu_n + \sigma_n \chi$ and $Y = \mu + \sigma\chi$ where $\chi \sim \mathcal{N}(0,1)$, which has the same distribution as the X_n and X by problem 5.1 part 6. Via problem 5.3 part 2, it suffices to prove that $\mathbb{E}[|Y_n|^p] \to \mathbb{E}[|Y|^p]$. This seems reasonable, as the almost-sure convergence of X_n to X implies $\mu_n \to \mu$ and $\sigma_n \to \sigma$, so for n large enough $||Y_n|^p - |Y|^p| < \epsilon$. Then,
$$|Y_n|^p \le |Y|^p + |Y_n - Y|^p < |Y|^p + \epsilon$$
for large n, and so the sequence is dominated by $|Y|^p + \epsilon$ (integrable since the measure is finite). Finally $\mathbb{E}[|X_n|^p] = \mathbb{E}[|Y_n|^p] \to \mathbb{E}[|Y|^p] = \mathbb{E}[|X|^p]$ by dominated convergence.

Final

Problem F.1: (Doob's lemma). Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} = \sigma(X)$ be the σ-algebra generated by X. Show that for any random variable Y, measurable with respect to $\mathcal{G}$, there exists a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $Y = h(X)$.

Solution: Suppose first that $Y = \chi_{X^{-1}(B)}$ for some $B \in \mathcal{B}(\mathbb{R})$. Then
$$h(X(\omega)) = Y(\omega) = \chi_{X^{-1}(B)}(\omega) = \begin{cases} 1 & \omega \in X(B) \\ 0 & \omega \notin X(B) \end{cases},$$
i.e. $h(x)$ is an indicator function and must be $h = \chi_B$. By linearity we have that if Y is a simple function then
$$Y = \sum_{i=1}^n a_i \chi_{X^{-1}(B_i)} = \sum_{i=1}^n a_i \chi_{B_i}(X) = \left(\sum_{i=1}^n a_i \chi_{B_i}\right)(X)$$
so that the conclusion holds for simple functions too. Now if $Y \ge 0$ is $\mathcal{G}$-measurable then there exists a sequence $Y_1 \le Y_2 \le ... \le Y$ of simple $\mathcal{G}$-measurable functions $Y_i \ge 0$ such that
$$Y(\omega) = \lim_{i\to\infty} Y_i(\omega)$$
for all $\omega \in \Omega$. Let h_i be the representative for each Y_i, that is $Y_i = h_i(X)$, and define
$$h(x) = \sup_{i\in\mathbb{N}} h_i(x).$$
On the one hand, $Y_i \le Y$ for all $i \in \mathbb{N}$ so that $h_i(X(\omega)) \le Y(\omega)$. Taking the sup over $i \in \mathbb{N}$ gives $h(X(\omega)) \le Y(\omega)$. On the other hand,
$$Y(\omega) = \lim_{i\to\infty} Y_i(\omega) = \limsup_{i\to\infty} h_i(X(\omega)) = \lim_{n\to\infty} \sup_{i\ge n} h_i(X(\omega)) \le h(X(\omega)).$$
Consequently $Y(\omega) = h(X(\omega))$. This implies in particular that h takes finite values. As h is the supremum of $\mathcal{B}(\mathbb{R})$ measurable functions, it too is $\mathcal{B}(\mathbb{R})$ measurable. Finally, for generic Y, write $Y = Y^+ - Y^-$ and find h^+ and h^- such that $Y^\pm = h^\pm(X)$. Then owing to linearity $Y = h(X)$ for $h = h^+ - h^-$.

Problem F.2: (Barndorff-Nielsen's extension of the Borel-Cantelli lemma). Let $\{A_n\}_{n=1}^\infty$ be a sequence of events.

1. Show that
$$(\limsup_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} A_n^c) \subseteq \limsup_{n\to\infty}(A_n \cap A_{n+1}^c).$$
2. If $\liminf_{n\to\infty} \mathbb{P}[A_n] = 0$ and $\sum_n \mathbb{P}[A_n \cap A_{n+1}^c] < \infty$, show that $\mathbb{P}[\limsup_{n\to\infty} A_n] = 0$.

Solution:

1. Recall that
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k.$$
We need to show that for every $n \in \mathbb{N}$ there exists $k_n \ge n$ such that $\omega \in A_{k_n} \cap A_{k_n+1}^c$. By definition, $\omega \in (\limsup_{n\to\infty} A_n) \cap (\limsup_{n\to\infty} A_n^c)$ if for every $n \in \mathbb{N}$ there exists $k_n^1, k_n^2 \ge n$ such that $\omega \in A_{k_n^1} \cap A_{k_n^2}^c$. Without loss of generality assume $k_n^2 > k_n^1$; this can be done, for example, by replacing k_n^2 by $k_{n'}^2$ with $n' > k_n^1$ if necessary. Now define

$f(k_n^1, k_n^2) = \sum_{i=k_n^1}^{k_n^2} \chi_{A_i^c}(\omega)$. Note that there exists $k_n^1 \le i < k_n^2$ for which $\omega \in A_i \cap A_{i+1}^c$ if and only if $f(k_n^1, k_n^2) > 0$; the forward implication is trivial while for the reverse one we may find an $i > k_n^1$ for which $\omega \in A_i^c$, then choose the smallest such i. But $f(k_n^1, k_n^2) > 0$ since $\omega \in \chi_{A^c_{k_n^2}}$.

2. Borel-Cantelli guarantees that $\sum_n \mathbb{P}[A_n \cap A_{n+1}^c] < \infty$ implies $\mathbb{P}[\limsup_{n\to\infty}(A_n \cap A_{n+1}^c)] = 0$. Let $A = \limsup_{n\to\infty} A_n$ and $B = \limsup_{n\to\infty} A_n^c$. By monotonicity and part 1, it follows that $\mathbb{P}[A \cap B] = 0$. Noting that

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - 2\mathbb{P}[A \cap B] = \mathbb{P}[A] + \mathbb{P}[B],$$

it suffices to show that $\mathbb{P}[B] = \mathbb{P}[A \cup B]$. Since $\liminf_{n\to\infty} \mathbb{P}[A_n] = 0$, it follows

$$\limsup_{n\to\infty} \mathbb{P}[A_n^c] = \limsup_{n\to\infty}(1 - \mathbb{P}[A_n]) = 1 - \liminf_{n\to\infty} \mathbb{P}[A_n] = 1.$$

Note that the sets $\cup_{k=n}^{\infty} A_k^c$ are decreasing, so that by dominated convergence (valid in a finite measure space) we have

$$\begin{aligned}\mathbb{P}[B] = \mathbb{P}[\limsup_{n\to\infty} A_n^c] = \mathbb{P}\left[\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k^c\right] &= \lim_{n\to\infty} \mathbb{P}\left[\bigcup_{k=n}^{\infty} A_k^c\right] \\ &= \limsup_{n\to\infty} \mathbb{P}\left[\bigcup_{k=n}^{\infty} A_k^c\right] \ge \limsup_{n\to\infty} \mathbb{P}[A_n^c] = 1.\end{aligned}$$

Accordingly, $\mathbb{P}[B] = 1$. Now for any $\omega \in \Omega$ and $n \in \mathbb{N}$ either $\omega \in A_n$ or $\omega \in A_n^c$. Hence ω is in either infinitely many of the A_n or infinitely many of the A_n^c, i.e. $\omega \in A \cup B$. So $\mathbb{P}[A \cup B] = \mathbb{P}[\Omega] = 1$ and we conclude.

Problem F.3: (A criterion for membership in $L \log L$). Let X be a non-negative random variable, and let F be its cumulative distribution function (cdf). Show that

$$\mathbb{E}[X \log^+(X)] < \infty \quad \text{if and only if} \quad \int_1^{\infty}\int_1^{\infty} (1 - F(uv))\; dudv < \infty,$$

where $\log^+(x) = \max(\log(x), 0)$.

Solution: Observe that, making the change of variables $t = uv$, we get

$$\int_1^{\infty}\int_1^{\infty} (1 - F(uv))\; dudv = \int_1^{\infty} \frac{1}{v} \int_v^{\infty} (1 - F(t))\; dtdv.$$

Let us evaluate the inner integral. By an application of Fubini,

$$\begin{aligned}\int_v^{\infty} (1 - F(t))\; dt = \int_v^{\infty} \mathbb{P}[X > t]\; dt &= \int_v^{\infty}\int_{\Omega} \chi_{\{X>t\}}(\omega)\; d\mathbb{P}dt \\ &= \int_{\Omega}\int_v^{\infty} \chi_{\{X>t\}}(\omega)\; dtd\mathbb{P} = \mathbb{E}\left[\int_v^{\infty} \chi_{\{X>t\}}\; dt\right]\end{aligned}$$

Now,

$$\begin{aligned}\int_v^{\infty} \chi_{\{X>t\}}(\omega)\; dt &= \int_{\mathbb{R}} \chi_{(v,\infty)}(t)\chi_{(-\infty,X(\omega))}(t)\; dt = \int_{\mathbb{R}} \chi_{(v,X(\omega))}(t) \\ &= \begin{cases} 0 & v \ge X(\omega) \\ X(\omega) - v & v < X(\omega) \end{cases} = \begin{cases} 0 & v - X(\omega) \ge 0 \\ X(\omega) - v & X(\omega) - v > 0 \end{cases} \\ &= (X(\omega) - v)^+.\end{aligned}$$

In total,

$$\int_v^{\infty} (1 - F(t))\; dt = \mathbb{E}[(X - v)^+]$$

and so returning to the original integral,

$$\begin{aligned}\int_1^\infty \int_1^\infty (1-F(uv))\ dudv &= \int_1^\infty \frac{1}{v}\mathbb{E}[(X-v)^+]\ dv = \mathbb{E}\left[\int_1^\infty \frac{1}{v}(X-v)^+\ dv\right] \\ &= \int_{-\infty}^\infty \int_1^\infty \frac{1}{v}(x-v)^+\ dvd\mu_X(x) = \int_1^\infty \int_1^x \frac{1}{v}(x-v)\ dvd\mu_X(x) \\ &= \int_1^\infty \int_1^x \left(\frac{x}{v}-1\right)\ dvd\mu_X(x) = \int_1^\infty (x\log(x)-(x-1))\ d\mu_X(x)\end{aligned}$$

On the other hand,

$$\mathbb{E}[X\log^+(X)] = \int_1^\infty x\log(x)\ d\mu_X(x)$$

Evidently $x\log(x) \geq x\log(x) - x + 1$ for $x \geq 1$ so that

$$\mathbb{E}[X\log^+(X)] \geq \int_1^\infty \int_1^\infty (1-F(uv))\ dudv,$$

proving the forward implication. For the reverse assume that

$$\int_1^\infty (x\log(x)-(x-1))\ d\mu_X(x) = \int_1^\infty \int_1^\infty (1-F(uv))\ dudv < \infty.$$

Fix now $x_0 \geq 1$. It suffices to find $C > 0$ such that $C(x\log(x)-x+1) \geq x\log(x)$ on $[x_0,\infty)$ as

$$\infty > C\int_1^\infty (x\log(x)-x+1)\ d\mu_X(x) \geq C\int_1^{x_0} (x\log(x)-x+1) + \int_{x_0}^\infty x\log(x)\ d\mu_X(x).$$

Owing to the fact that both $x\log(x)$ and $x\log(x)-x+1$ are non-negative on $[1,\infty)$ this implies, in particular, that

$$\int_{x_0}^\infty x\log(x)\ d\mu_X(x) < \infty$$

which is enough to conclude

$$\int_1^\infty x\log(x)\ d\mu_X(x) = \int_1^{x_0} x\log(x)\ d\mu_X(x) + \int_{x_0}^\infty x\log(x)\ d\mu_X(x) < \infty.$$

Observe that $x\log(x)$ is monotone increasing on $[1,\infty)$, hence $C(x\log(x)-x+1)$ is too for any $C > 0$. There is one intersection point at $x = 1$, so for another to exist at say $x_0 > 1$ we just need to check that $C(x\log(x)-x+1)$ grows faster than $x\log(x)$ for judiciously chosen $C > 0$. But,

$$\frac{d}{dx}(x\log(x)) = \log(x)+1, \qquad \frac{d}{dx}(C(x\log(x)-x+1)) = C\log(x).$$

Thus the latter eventually grows faster for any $C > 1$, and there is some $x_0 > 1$ for which $C(x\log(x)-x+1) \geq x\log(x)$ on $[x_0,\infty)$.

Problem F.4: (The "Chi-squared" and "Student's t" distributions). Let $\{X_k\}_{k=1}^\infty$ be an iid sequence of standard normal random variables.

1. Given $d \in \mathbb{N}$, the distribution of the random variable $X_1^2 + ... + X_d^2$ is called the **chi-squared distribution with d degrees of freedom**, denoted by $\chi^2(d)$. Compute its pdf.

   Hint: Use the convolutional identity $g_\alpha * g_\beta = g_{\alpha+\beta}$, where $g_\alpha(x) = x^{\alpha-1}\exp(-x/2)/(2^\alpha\Gamma(\alpha))\chi_{\{x>0\}}$, and Γ is the Gamma function.

2. For $n \in \mathbb{N}$, let X be the random (row) vector $X = (X_1, ..., X_n)$ and let M be a $n \times n$ symmetric matrix such that $M^2 = M$. What is the distribution of XMX^T?

   Hint: Use the properties of the multivariate normal from Problem 5.1 in HW5.

3. For $n \geq 2$, what is the joint distribution of $Q^2 := \sum_{i=1}^n (X_i - \bar{X})^2$ and $\bar{X} := 1/n(X_1 + ... + X_n)$?

   Hint: Same hint as in 2 above.

4. Show that there exists a constant C', which depends only on n, such that the pdf of the random variable
$$T = \frac{\sqrt{n}\bar{X}}{\sqrt{Q^2/(n-1)}}$$
is given by
$$f_T(t) = C'\left(1 + \frac{t^2}{d}\right)^{-(d+1)/2} \qquad \text{where } d = n-1.$$
Note: The distribution of T is called the **Student's t distribution with d degrees of freedom** and is denoted by $t(d)$. The value of the constant C' turns out to be $\Gamma((d+1)/2)/(\sqrt{\pi d}\,\Gamma(d/2))$. The only reason we use both $d = n-1$ and n is to be consistent with the standard terminology.

Note: Look up the "Student's t-test" if you are curious about the significance of this problem in statistics.

Solution

1. Recall if X and Y are independent random variables with pdfs f_X and f_Y respectively then
$$f_{X+Y} = f_X * f_Y.$$
So, we just need to compute f_{X^2} where X is a standard normal random variable. Then
$$f_{X_1^2+...+X_d^2} = f_{X^2} * ... * f_{X^2}$$
where the convolution is done d-many times. This computation was already made in HW 5 problem 1, part 3:
$$f_{X^2}(x) = \begin{cases} 0 & x \leq 0 \\ \dfrac{e^{-x/2}}{\sqrt{2\pi x}} & x > 0. \end{cases}$$
which is precisely $g_\alpha(x)$ for $\alpha = 1/2$ (note that $\Gamma(1/2)$ was computed in a previous homework problem to be $1/\sqrt{\pi}$). Hence, per the hint,
$$f_{X_1^2+...+X_d^2}(x) = g_{d/2}(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^{d/2-1}e^{-x/2}}{2^{d/2}\Gamma(d/2)} & x \geq 0. \end{cases}$$

2. Since M is symmetric it is diagonalizable: There exists an orthogonal $n \times n$ matrix P and an $n \times n$ diagonal matrix D such that $M = P^T DP$. Moreover, $M^2 = M$ so that
$$P^T DP = P^T DPP^T DP = P^T D^2 P$$
implying that $D^2 = D$ as P is invertible. It follows that $D_{ii} = 1, 0$. Next, we showed in HW 6 problem 1.6 that if $T(X) = XA$ then
$$\varphi_{T(X)}(t) = e^{-1/2t^T A^T At}.$$
Applying this with $A = P^T$, we see since $PP^T = \text{Id}$ that
$$\varphi_{T(X)}(t) = e^{-1/2t^T PP^T t} = e^{-1/2|t|^2} = \varphi_\chi(t)$$
where $\chi \sim \mathcal{N}(0, \text{Id})$. It follows that $Y = XP^T$ is still a standard multivariate normal. Consequently each $Y_i \sim \mathcal{N}(0,1)$ and are independent. Finally,
$$XMX^T = XP^T DPX^T = YDY^T = \sum_{j \in J} Y_j^2$$
where $J = \{j \in \{1, ..., n\} \mid D_{jj} = 1\}$. Hence $XMX^T \sim \chi^2(d)$ where $d = |J| = \text{Rank}(M)$.

3. We first compute the distributions of Q^2 and $\bar{X}$ separately. Is there a symmetric matrix M such that $Q^2 = XMX^T$? Computing both sides individually gives

$$\begin{aligned}
Q^2 &= \sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n \left(X_i - \frac{1}{n}\sum_{j=1}^n X_j \right)^2 \\
&= \sum_{i=1}^n \left(X_i^2 - \frac{2}{n}\sum_{j=1}^n X_i X_j + \frac{1}{n^2}\left(\sum_{j=1}^n X_j\right)^2 \right) \\
&= \sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{i=1}^n X_i \sum_{j=1}^n X_j + \frac{1}{n}\left(\sum_{j=1}^n X_j\right)^2 = \sum_{i=1}^n X_i^2 - \frac{1}{n}\left(\sum_{i=1}^n X_i\right)^2 \\
&= \sum_{i=1}^n X_i^2 - \frac{1}{n}\left(\sum_{i=1}^n X_i^2 + 2\sum_{i=1}^n \sum_{j>i} X_i X_j\right) \\
&= \frac{n-1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{i=1}^n \sum_{j>i} X_i X_j;
\end{aligned}$$

$$\begin{aligned}
XMX^T &= (X_1, ..., X_n) \begin{pmatrix} M_{11} & \cdots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{n1} & \cdots & M_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = (X_1, ..., X_n) \begin{pmatrix} \sum_{j=1}^n M_{1j} X_j \\ \vdots \\ \sum_{j=1}^n M_{nk} X_j \end{pmatrix} \\
&= \sum_{i=1}^n \sum_{j=1}^n M_{ij} X_i X_j = \sum_{i=1}^n M_{ii} X_i^2 + 2\sum_{i=1}^n \sum_{j>i} M_{ij} X_i X_j.
\end{aligned}$$

Hence, $M_{ii} = (n-1)/n$ and $M_{ij} = M_{ji} = -1/n$ for $i \neq j$. It follows that $M = \mathrm{Id} - 1/nI$ where I is a matrix of all ones. Note that

$$M^2 = \mathrm{Id} - 2/nI + 1/n^2 I^2 = \mathrm{Id} - 1/nI = M$$

as $I^2 = nI$. Consequently $Q^2 \sim \chi(n-1)^2$ as $\mathrm{Rank}(M) = n-1$ (this follows easily by induction and row reducing).

Since the X_i are iid with $\mu = 0$ and $\sigma = 1$ we have

$$\varphi_{\bar{X}}(t) = (\varphi_{1/n\chi}(t))^n = \left(e^{-1/2(t/n)^2} \right)^n = e^{-1/2t^2/n}$$

where $\chi \sim \mathcal{N}(0,1)$. So $\overline{X} \sim \mathcal{N}(0, 1/\sqrt{n})$.

We now show that Q^2 and $\bar{X}$ are independent. By definition of Q^2, we have that

$$Q^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \tilde{X}\tilde{X}^T$$

where $\tilde{X} = (X_1 - \bar{X}, ..., X_n - \bar{X})$. Since deterministic functions preserve independence it suffices then to show that $\tilde{X}$ and $\bar{X}$ are independent (in the sense that $\tilde{X}_i$ and $\bar{X}$ are independent for all $i = 1, ..., n$). To this end we show that the random vector $(\bar{X}, X_1 - \bar{X}, ..., X_n - \bar{X})$ is a multivariate Gaussian and compute its covariance matrix. First, for any $i = 1, ..., n$ we have that

$$X_i - \bar{X} = X_i - \frac{1}{n}\sum_{j=1}^n X_j = \sum_{j=1}^n c_j X_i$$

where $c_j = -1/n$ for $j \neq i$ and $c_i = (n-1)/n$. Next, similar to the above we have

$$\begin{aligned}\varphi_{X_i-\bar{X}}(t) &= (\varphi_{-1/nX}(t))^{n-1}\varphi_{(n-1)/nX}(t) = \left(e^{-1/2(-t/n)^2}\right)^{n-1} e^{-1/2((n-1)t/n)^2} \\ &= e^{-1/2(n-1)t^2/n^2} e^{-1/2(n-1)^2t^2/n^2} = e^{-1/2((n-1)/n)t^2}\end{aligned}$$

so that $X_i - \bar{X} \sim \mathcal{N}(0, \sqrt{(n-1)/n})$. This implies that $(\bar{X}, X_1 - \bar{X}, ..., X_n - \bar{X}) \sim \mathcal{N}(0, \Sigma)$ where Σ is the covariance matrix.[1] To check for independence, per HW 6 problem 1.4 it suffices to check that $\mathrm{Cov}(\bar{X}, X_i - \bar{X}) = 0$. So,

$$\begin{aligned}\mathrm{Cov}(\bar{X}, X_i - \bar{X}) &= \mathbb{E}[\bar{X}(X_i - \bar{X})] - \mathbb{E}[\bar{X}]\mathbb{E}[X_i - \bar{X}] \\ &= \mathbb{E}[\bar{X}X_i] - \mathbb{E}[\bar{X}^2] = \frac{1}{n}\mathbb{E}\left[X_1X_i + ... + X_nX_i\right] - \frac{1}{n} \\ &= \frac{1}{n}\left(\sum_{j\neq i}\mathbb{E}[X_j]\mathbb{E}[X_i]\right) + \frac{1}{n}\mathbb{E}[X_i^2] - \frac{1}{n} = 0\end{aligned}$$

and $\bar{X}$ is independent from $X_i - \bar{X}$ for $i = 1, ..., n$. The joint distribution then is just the product measure:

$$\mu_{(Q^2,\bar{X})} = \mu_{Q^2} \otimes \mu_{\bar{X}}.$$

4. Suppose U and V are independent absolutely continuous random variables. We claim that

$$f_{UV}(t) = \int_{-\infty}^{\infty} f_U(u) f_V\left(\frac{t}{u}\right)\frac{1}{|u|}\ du.$$

Indeed, from the cdf we have

$$\begin{aligned}F_{UV}(t) = \mathbb{P}[UV \le t] &= \mathbb{P}[UV \le t, U > 0] + \mathbb{P}[UV \le t, U < 0] \\ &= \mathbb{P}[V \le t/U, U > 0] + \mathbb{P}[V \ge t/U, U < 0] \\ &= \int_0^\infty \int_{-\infty}^{t/u} f_U(u)f_V(v)\ dvdu + \int_{-\infty}^0 \int_{t/u}^{\infty} f_U(u)f_V(v)\ dvdu.\end{aligned}$$

Differentiating both sides then yields

$$\begin{aligned}f_{UV}(t) &= \int_0^\infty f_U(u)f_V\left(\frac{t}{u}\right)\frac{1}{u}\ du + \int_{-\infty}^0 f_U(u)f_V\left(\frac{t}{u}\right)\frac{-1}{u}\ du \\ &= \int_{-\infty}^\infty f_U(u)f_V\left(\frac{t}{u}\right)\frac{1}{|u|}\ du.\end{aligned}$$

Next we compute the pdf of $1/\sqrt{Q^2/d}$ (where $d = n-1$): for $x > 0$ we have

$$F_{1/\sqrt{Q^2/d}}(x) = \mathbb{P}\left[\frac{\sqrt{d}}{\sqrt{Q^2}} \le x\right] = \mathbb{P}\left[\frac{d}{x^2} \le Q^2\right] = 1 - \mathbb{P}\left[Q^2 < \frac{d}{x^2}\right] = 1 - F_{Q^2}\left(\frac{d}{x^2}\right)$$

where the last equality holds since $Q^2 \sim \chi^2(d)$ is absolutely continuous. Evidently $F_{1/\sqrt{Q^2/d}}(x) = 0$ for $x \le 0$. Hence,

$$\begin{aligned}f_{1/\sqrt{Q^2/d}}(x) &= \frac{2d}{x^3}f_{Q^2}\left(\frac{d}{x^2}\right) = \frac{2d}{x^3}\left(\frac{(d/x^2)^{d/2-1}e^{-d/(2x^2)}}{2^{d/2}\Gamma(d/2)}\chi_{\{x>0\}}\right) \\ &= \frac{d^{d/2}}{2^{d/2-1}\Gamma(d/2)}\left(\frac{e^{-d/(2x^2)}}{x^{d+1}}\right)\chi_{\{x>0\}}.\end{aligned}$$

[1] Really you should check this for every linear combination of the $\bar{X}$ and $X_i - \bar{X}$, but it follows the same logic as above.

As $\bar{X} \sim \mathcal{N}(0, 1/\sqrt{n})$ it follows that $\sqrt{n}\bar{X} \sim \mathcal{N}(0,1)$. Thus via the computation above:

$$\begin{aligned} f_T(t) = f_{\sqrt{n}\bar{X}/\sqrt{Q^2/d}}(t) &= \int_{-\infty}^{\infty} f_{1/\sqrt{Q^2/d}}(u) f_{\mathcal{X}}\left(\frac{t}{u}\right)\frac{1}{|u|}\ du \\ &= \int_0^\infty \left(\frac{d^{d/2}}{2^{d/2-1}\Gamma(d/2)}\frac{e^{-d/(2u^2)}}{u^{d+1}}\right)\left(\frac{e^{-t^2/(2u^2)}}{\sqrt{2\pi}}\right)\frac{1}{u}\ du \\ &= \frac{d^{d/2}}{2^{d/2-1/2}\sqrt{\pi}\Gamma(d/2)}\int_0^\infty \frac{e^{-(d+t^2)/(2u^2)}}{u^{d+2}}\ du \end{aligned}$$

We aim to turn the integral into a Gamma function (motivated by the expression for C', which contains a $\Gamma((d+1)/2)$ in the numerator). Let us make the change of variables $x = (d+t^2)/(2u^2)$. Then,

$$\begin{aligned} \int_0^\infty \frac{e^{-(d+t^2)/(2u^2)}}{u^{d+2}}\ du &= \frac{1}{d+t^2}\int_0^\infty \frac{e^{-(d+t)^2/(2u^2)}}{u^{d-1}}\ \frac{(d+t^2)du}{u^3} \\ &= \frac{1}{d+t^2}\int_0^\infty \frac{e^{-x}}{[(d+t^2)/(2x)]^{d/2-1/2}}\ dx \\ &= \frac{2^{d/2-1/2}}{(d+t^2)^{d/2+1/2}}\int_0^\infty x^{d/2-1/2}e^{-x}\ dx \\ &= \frac{2^{d/2-1/2}}{d^{d/2+1/2}(1+t^2/d)^{d/2+1/2}}\Gamma\left(\left(\frac{d+1}{2}\right)\right). \end{aligned}$$

Substituting into the above then yields the desired expression:

$$f_T(t) = \frac{\Gamma((d+1)/2)}{\sqrt{\pi d}\,\Gamma(d/2)}\left(1+\frac{t^2}{d}\right)^{-(d+1)/2}.$$

Problem F.5: (A probabilistic proof of Stirling's formula). Let $\{X_n\}_{n=1}^\infty$ be an iid sequence with the Poisson(λ) distribution, i.e., $\mathbb{P}[X_1 = k] = e^{-\lambda}\lambda^k/k!$ for $k \in \mathbb{N}_0$.

1. What is the distribution of $Y_n = X_1 + ... + X_n$, for $n \in \mathbb{N}$?
2. Set $\lambda = 1$ and let $Z_n = Y_n/\sqrt{n} - \sqrt{n}$. Without evaluating it, show that $\mathbb{E}[|Z_n|]$ admits a limit and identify it.

   Hint: Use the fact that, in this case, the function $x \mapsto |x|$ can be used to "test" weak convergence, as if it belonged to $C_b(\mathbb{R})$. Prove this for extra credit.

3. Evaluate $\mathbb{E}[|Z_n|]$ explicitly and derive Stirling's formula

$$\lim_{n\to\infty} \frac{n!}{(n/e)^n\sqrt{2\pi n}} = 1.$$

Solution:

1. We start with the following identity for the convolution of sums of diracs:

$$\begin{aligned} \left(\sum_{n=0}^\infty a_n\delta_{x_n} * \sum_{k=0}^\infty b_k\delta_{y_k}\right)(B) &= \int_{\mathbb{R}}\sum_{n=0}^\infty a_n\delta_{x_n}(B-z)\ d\left(\sum_{k=0}^\infty b_k\delta_{y_k}(z)\right) \\ &= \sum_{k=0}^\infty b_k\int_{\mathbb{R}}\sum_{n=0}^\infty a_n\delta_{x_n}(B-z)\ d\delta_{y_k}(z) \\ &= \sum_{k=0}^\infty\sum_{n=0}^\infty a_nb_k\int_{\mathbb{R}}\delta_{x_n}(B-z)\ d\delta_{y_k}(z) \\ &= \sum_{k=0}^\infty\sum_{n=0}^\infty a_nb_k\delta_{x_n}(B-y_k) = \sum_{k=0}^\infty\sum_{n=0}^\infty a_nb_k\delta_{x_n+y_k}(B). \end{aligned}$$

For a Poisson(λ) distribution, we have that

$$\mu_X = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!}\delta_n.$$

Hence,

$$\mu_{X_1+X_2} = \mu_{X_1} * \mu_{X_2} = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \frac{e^{-2\lambda}\lambda^{n+k}}{n!k!}\delta_{n+k} = \sum_{m=0}^{\infty} c_m\delta_m$$

where

$$c_m = \sum_{j=0}^{m} \frac{e^{-2\lambda}\lambda^m}{j!(m-j)!} = \frac{e^{-2\lambda}\lambda^m}{m!}\sum_{j=0}^{m}\binom{m}{j} = \frac{e^{-2\lambda}(2\lambda)^m}{m!}.$$

It follows that $Y_2 = X_1 + X_2 \sim \text{Poisson}(2\lambda)$. Inductively suppose that $Y_{n-1} \sim \text{Poisson}((n-1)\lambda)$. We will show $Y_n \sim \text{Poisson}(n\lambda)$. Indeed, since $Y_n = Y_{n-1} + X_n$ we have that

$$\begin{aligned}
\mu_{Y_n} = \mu_{Y_{n-1}} * \mu_{X_n} &= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left(\frac{e^{-(n-1)\lambda}((n-1)\lambda)^k}{k!}\right)\left(\frac{e^{-\lambda}\lambda^j}{j!}\right)\delta_{k+j} \\
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{e^{-n\lambda}(n-1)^k\lambda^{k+j}}{k!j!}\delta_{k+j} \\
&= \sum_{m=0}^{\infty} e^{-n\lambda}\lambda^m \sum_{i=0}^{m}\frac{(n-1)^i}{i!(m-i)!}\delta_m \\
&= \sum_{m=0}^{\infty}\frac{e^{-n\lambda}\lambda^m}{m!}\sum_{i=0}^{m}\binom{m}{i}(n-1)^i\delta_m \\
&= \sum_{m=0}^{\infty}\frac{e^{-n\lambda}\lambda^m}{m!}((n-1)+1)^m\delta_m = \sum_{m=0}^{\infty}\frac{e^{-n\lambda}(n\lambda)^m}{m!}\delta_m.
\end{aligned}$$

2. Recall that the central limit theorem says if $\{X_n\}_{n=1}^{\infty}$ is an iid sequence of random variables with $0 < \text{Var}(X_1) < \infty$ then

$$\frac{1}{\sqrt{\sigma^2 n}}\sum_{k=1}^{n}(X_k - \mu) \xrightarrow{\mathcal{D}} \chi$$

where $\chi \sim \mathcal{N}(0,1)$ and $\mu = \mathbb{E}[X_1]$, $\sigma^2 = \text{Var}(X_1)$. Defining $Z_n = Y_n/\sqrt{n} - \sqrt{n}$, we can write this in the above form:

$$\begin{aligned}
Z_n &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k - \sqrt{n} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu + \mu) - \sqrt{n} \\
&= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) + \frac{n\mu}{\sqrt{n}} - \sqrt{n} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) + \sqrt{n}(\mu - 1)
\end{aligned}$$

where $\mu = \sigma = 1$. Note that

$$\begin{aligned}
f_1(x) &:= \sum_{k=0}^{\infty}\frac{kx^k}{k!} = \sum_{k=1}^{\infty}\frac{x^k}{(k-1)!} = \sum_{k=0}^{\infty}\frac{x^{k+1}}{k!} = xe^x; \\
f_2(x) &:= \sum_{k=0}^{\infty}\frac{k^2x^k}{k!} = \sum_{k=1}^{\infty}\frac{kx^k}{(k-1)!} = \sum_{k=0}^{\infty}\frac{(k+1)x^{k+1}}{k!} \\
&= \sum_{k=0}^{\infty}\frac{kx^{k+1}}{k!} + \sum_{k=0}^{\infty}\frac{x^{k+1}}{k!} = (x^2 + x)e^x
\end{aligned}$$

Since

$$\mathbb{E}[X_1] = \int_{-\infty}^{\infty} x \; d\mu_{X_1}(x) = \sum_{k=0}^{\infty} \frac{k}{ek!} = \frac{1}{e} f_1(1) = 1;$$

$$\mathrm{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1] = \int_{-\infty}^{\infty} x^2 \; d\mu_{X_1}(x) - 1 = \sum_{k=0}^{\infty} \frac{k^2}{ek!} - 1 = \frac{1}{e} f_2(1) - 1 = 1,$$

It follows by the central limit theorem that $\mu_{Z_n} \xrightarrow{w} \mu_{\chi}$. Testing weak convergence with $x \mapsto |x|$ (valid per the hint) tells us

$$\mathbb{E}[|Z_n|] = \int_{-\infty}^{\infty} |x| \; d\mu_{Z_n}(x) \to \int_{-\infty}^{\infty} |x| \; d\mu_{\chi}(x)$$

where

$$\begin{aligned}\int_{-\infty}^{\infty} |x| \; d\mu_{\chi}(x) &= \int_{-\infty}^{\infty} \frac{|x| e^{-x^2/2}}{\sqrt{2\pi}} \; dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} x e^{-x^2/2} \; dx \\ &= \sqrt{\frac{2}{\pi}} \left[-e^{-x^2/2} \right]_{x=0}^{x=\infty} = \sqrt{\frac{2}{\pi}}.\end{aligned}$$

3. Computing $\mathbb{E}[|Z_n|]$ explicitly,

$$\mathbb{E}[|Z_n|] = \frac{1}{\sqrt{n}} \mathbb{E}[|Y_n - n|] = \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} |y - n| \; d\mu_{Y_n}(y) = \frac{1}{e^n \sqrt{n}} \sum_{k=0}^{\infty} \frac{n^k |k-n|}{k!}$$

Let's analyze the last sum in more detail.

$$\begin{aligned}\sum_{k=0}^{\infty} \frac{n^k |k-n|}{k!} &= -\sum_{k=0}^{n-1} \frac{n^k (k-n)}{k!} + \sum_{k=n+1}^{\infty} \frac{n^k (k-n)}{k!} \\ &= -2 \sum_{k=0}^{n-1} \frac{n^k (k-n)}{k!} + \sum_{k=0}^{\infty} \frac{n^k (k-n)}{k!}.\end{aligned}$$

The latter sum is easy to evaluate:

$$\sum_{k=0}^{\infty} \frac{n^k (k-n)}{k!} = \sum_{k=0}^{\infty} \frac{k n^k}{k!} - n \sum_{k=0}^{\infty} \frac{n^k}{k!} = n e^n - n(e^n) = 0.$$

The former sum telescopes:

$$\sum_{k=0}^{n-1} \frac{n^k (k-n)}{k!} = \sum_{k=0}^{n-1} \frac{k n^k}{k!} - \sum_{k=0}^{n-1} \frac{n^{k+1}}{k!} = \sum_{k=1}^{n-1} \frac{n^k}{(k-1)!} - \sum_{k=1}^{n} \frac{n^k}{(k-1)!} = -\frac{n^n}{(n-1)!}.$$

Then, the expectation is explicitly computed as

$$\mathbb{E}[|Z_n|] = \frac{1}{e^n \sqrt{n}} \sum_{k=0}^{\infty} \frac{n^k |k-n|}{k!} = \frac{2 n^{n-1/2}}{e^n (n-1)!} = \sqrt{\frac{2}{\pi}} \frac{(n/e)^n \sqrt{2\pi n}}{n!}.$$

By the previous part,

$$\lim_{n \to \infty} \mathbb{E}[|Z_n|] = \sqrt{\frac{2}{\pi}}$$

so that

$$\lim_{n \to \infty} \frac{(n/e)^n \sqrt{2\pi n}}{n!} = 1$$

as desired.

Problem F.6: (Two exercises in conditional expectation).

1. Give an example of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X \in L^1$, and two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$ of $\mathcal{F}$ such that

$$\mathbb{P}\Big[\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H} \big] = \mathbb{E}[X \mid \mathcal{H}] \Big] < 1.$$

2. For $X, Y \in L^2$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, show that the following "self-adjointness" property holds
$$\mathbb{E}\Big[\mathbb{E}[X \mid \mathcal{G}] \cdot Y\Big] = \mathbb{E}\Big[X \cdot \mathbb{E}[Y \mid \mathcal{G}]\Big].$$

Solution:

1. Consider $\Omega = \{a, b, c\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}$ the uniform measure on Ω, and $\mathcal{G} = \sigma(\{a\})$, $\mathcal{H} = \sigma(\{b\})$. For any random variable X on Ω we have that
$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \begin{cases} X(a) & \omega = a \\ (X(b) + X(c))/2 & \omega \in \{b, c\} \end{cases}$$
$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}](\omega) = \begin{cases} X(a)/2 + (X(b) + X(c))/4 & \omega \in \{a, c\} \\ (X(b) + X(c))/2 & \omega = b \end{cases}$$
$$\mathbb{E}[X \mid \mathcal{H}](\omega) = \begin{cases} (X(a) + X(c))/2 & \omega \in \{a, c\} \\ X(b) & \omega = b \end{cases}.$$
So, observe that $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$ if and only if $X(b) = X(c)$. In fact, if $X(b) \neq X(c)$ then $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}](\omega) \neq \mathbb{E}[X \mid \mathcal{H}](\omega)$ for all $\omega \in \Omega$. Thus
$$\mathbb{P}\Big[\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{H}\Big] = 0$$
for such X.
2. Note since $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable, so too is $\mathbb{E}[X \mid \mathcal{G}] \cdot Y$. Hence,
$$\mathbb{E}[X \mid \mathcal{G}] \cdot Y = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot Y \mid \mathcal{G}]$$
Recall that if $ZW \in L^1$ with Z a $\mathcal{G}$-measurable random variable then
$$\mathbb{E}[ZW \mid \mathcal{G}] = Z \cdot \mathbb{E}[W \mid \mathcal{G}].$$
Set $Z = \mathbb{E}[X \mid \mathcal{G}]$ and $W = Y$. Then since $X \in L^2$ we also have $Z \in L^2$, and by Hölder it follows $ZW \in L^1$. Applying the above yields
$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \cdot Y \mid \mathcal{G}] = \mathbb{E}[ZY \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{E}[Y \mid \mathcal{G}].$$
In total
$$\mathbb{E}[X \mid \mathcal{G}] \cdot Y = \mathbb{E}[X \mid \mathcal{G}] \cdot \mathbb{E}[Y \mid \mathcal{G}].$$
The exact same work holds replacing X with Y, but the last quantity is symmetric in X and Y. Hence,
$$\mathbb{E}[X \mid \mathcal{G}] \cdot Y = X \cdot \mathbb{E}[Y \mid \mathcal{G}].$$
Of course, the above only holds almost surely. Taking the expectation of both sides gives true equality:
$$\mathbb{E}\Big[\mathbb{E}[X \mid \mathcal{G}] \cdot Y\Big] = \mathbb{E}\Big[X \cdot \mathbb{E}[Y \mid \mathcal{G}]\Big].$$