36-752, Spring 2018
Homework 2 Solution

Due Thu March 1, by 5:00pm in Jisu's mailbox.

Points: 100 pts total for the assignment.

1. Let P and Q two probability measures on some measurable space $(\Omega, \mathcal{F})$ and let μ be any σ-finite measure on that space such that both P and Q are absolutely continuous with respect to μ (for example, you may take $\mu = P + Q$). Let $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ be the corresponding Radon-Nykodim derivatives.
The total variation distance between P and Q is defined as

$$d_{\text{TV}}(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|.$$

This is a very strong notion of distance between probability distributions: if $d_{\text{TV}}(P, Q) < \epsilon$, for some small $\epsilon > 0$, then *any* probabilistic statement involving P will differ by at most ϵ from the corresponding statement involving Q.

   (a) Show that d_{TV} is a metric over the set of all probability measure on $(\Omega, \mathcal{F})$. In particular, $d_{\text{TV}}(P, Q) = 0$ if and only if $P = Q$.

   (b) Show that $d_{\text{TV}}(P, Q) = 1$ if and only if P and Q are mutually singular.

   (c) Prove that the following equivalent representation of the total variation distance:

$$d_{\text{TV}}(P, Q) = \frac{1}{2} \int_{\Omega} |p - q| d\mu.$$

   Thus, the total variation distance is half the L_1 distance between densities.
   Hint: in the definition of total variation distance you may want to take $A = \{q \geq p\}$. Show that the supremum is achieved by this set...

   (d) **Total variation distance and hypothesis testing.** Let X be a random variable taking values in some measurable space $(S, \mathcal{A})$. Suppose we are interested in testing the null hypothesis that the distribution of X (a probability measure on $(S, \mathcal{A})$!) is P versus the alternative hypothesis that it is Q. We do so by devising a *test* ϕ, which is a measurable function from S into $\{0, 1\}$ such that $\phi(x) = 1$ (resp. $\phi(x) = 0$) signifies that the null hypothesis is rejected (resp. not rejected) if X takes on the value x. To measure the performance of a given test function ϕ we evaluate its risk, defined as the sum of type I and type II errors:

$$R_{P,Q}(\phi) = \int_S \phi dP + \int_S (1 - \phi) dQ.$$

   Show that

$$\inf_{\phi} R_{P,Q}(\phi) = 1 - d_{\text{TV}}(P, Q),$$

where the infimum si over all test functions.
The above result formalizes the intuition that the closer P and Q are, the harder it is to tell them apart using any test function. In particular, $R_{P,Q}(\phi) = 0$ – i.e., it is possible to perfectly discriminate between P and Q – if and only if the two probability measures are mutually singular.
Hint: use the Neymann-Pearson approach and take ϕ to be the indicator function of the set $\{q \geq p\}$.

Points: 10 pts = 2 + 3 + 3 + 2.

Solution.
Let $A_0 := \{\omega \in \Omega : q(\omega) \geq p(\omega)\}$.We first show that

$$d_{TV}(P,Q) = Q(A_0) - P(A_0) = P(A_0^\complement) - Q(A_0^\complement).$$

Note that $q - p \geq 0$ on A_0 and $q - p < 0$ on $A_0^\complement$. Hence for any $A \in \mathcal{F}$,

$$\begin{aligned}
Q(A) - P(A) &= \int_A (q-p)d\mu \\
&= \int_{A\cap A_0} (q-p)d\mu + \int_{A\cap A_0^\complement} (q-p)d\mu \\
&\leq \int_{A\cap A_0} (q-p)d\mu \\
&\leq \int_{A_0} (q-p)d\mu \\
&= Q(A_0) - P(A_0).
\end{aligned}$$

Similarly,

$$\begin{aligned}
P(A) - Q(A) &= \int_A (p-q)d\mu \\
&= \int_{A\cap A_0} (p-q)d\mu + \int_{A\cap A_0^\complement} (p-q)d\mu \\
&\leq \int_{A\cap A_0^\complement} (p-q)d\mu \\
&\leq \int_{A_0^\complement} (p-q)d\mu \\
&= P(A_0^\complement) - Q(A_0^\complement).
\end{aligned}$$

And $P(A_0)+P(A_0^\complement) = Q(A_0)+Q(A_0^\complement) = 1$ gives $Q(A_0)-P(A_0) = P(A_0^\complement)-Q(A_0^\complement)$. And hence for all $A \in \mathcal{F}$,

$$|P(A) - Q(A)| \leq Q(A_0) - P(A_0) = P(A_0^\complement) - Q(A_0^\complement).$$

Hence taking sup over $A \in \mathcal{F}$ gives

$$d_{TV}(P,Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)| \le Q(A_0) - P(A_0) = P(A_0^{\complement}) - Q(A_0^{\complement}).$$

And since $A_0 \in \mathcal{F}$, $Q(A_0) - P(A_0) = P(A_0^{\complement}) - Q(A_0^{\complement}) \le d_{TV}(P,Q)$ is trivial, and hence

$$d_{TV}(P,Q) = Q(A_0) - P(A_0) = P(A_0^{\complement}) - Q(A_0^{\complement}).$$

(a)
Let P, Q, R be probability measures on $(\Omega, \mathcal{F})$. First, $d_{TV}(P,Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)| \ge 0$. Second,

$$\begin{aligned} d_{TV}(P,Q) = 0 &\iff \sup_{A\in\mathcal{F}} |P(A) - Q(A)| = 0 \\ &\iff \text{for all } A \in \mathcal{F},\ P(A) = Q(A) \\ &\iff P = Q. \end{aligned}$$

Third, $d_{TV}(P,Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)| = \sup_{A\in\mathcal{F}} |Q(A) - P(A)| = d_{TV}(Q,P)$. Forth,

$$\begin{aligned} d_{TV}(P,R) &= \sup_{A\in\mathcal{F}} |P(A) - R(A)| \\ &\le \sup_{A\in\mathcal{F}} (|P(A) - Q(A)| + |Q(A) - R(A)|) \\ &\le \sup_{A\in\mathcal{F}} |P(A) - Q(A)| + \sup_{A\in\mathcal{F}} |Q(A) - R(A)| \\ &= d_{TV}(P,Q) + d_{TV}(Q,R). \end{aligned}$$

Hence d_{TV} is a metric over the set of all probability measure on $(\Omega, \mathcal{F})$.
(b)
Let $A_0 = \{\omega \in \Omega :\ q(\omega) \ge p(\omega)\}$ be from above. When $d_{TV}(P,Q) = 1$, then

$$d_{TV}(P,Q) = Q(A_0) - P(A_0) = P(A_0^{\complement}) - Q(A_0^{\complement}) = 1.$$

Then

$$\begin{aligned} P(A_0) &= Q(A_0) - 1 \le 0, \\ Q(A_0^{\complement}) &= P(A_0^{\complement}) - 1 \le 0, \end{aligned}$$

and hence $P(A_0) = Q(A_0^{\complement}) = 0$. Since $A_0 \cap A_0^{\complement} = \emptyset$, P and Q are mutually singular. When P and Q are mutually singular, there exists $B \in \mathcal{F}$ with $P(B) = 0$ and $Q(B^{\complement}) = 0$. Then $P(B^{\complement}) = Q(B) = 1$, and hence

$$Q(B) - P(B) = P(B^{\complement}) - Q(B^{\complement}) = 1.$$

Then

$$d_{TV}(P,Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)| \geq |P(B) - Q(B)| = 1.$$

And since $d_{TV}(P,Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)| \leq 1$,

$$d_{TV}(P,Q) = \sup_{A\in\mathcal{F}} |P(A) - Q(A)| = 1.$$

(c)

Let $A_0 = \{\omega \in \Omega : \, q(\omega) \geq p(\omega)\}$ be from above. Then

$$\begin{aligned} d_{TV}(P,Q) = Q(A_0) - P(A_0) &= \int_{\{q\geq p\}} (q-p)d\mu \\ &= \int_{\{q\geq p\}} |p-q| d\mu, \end{aligned}$$

where last line is from that $q - p = |p - q|$ on $\{q \geq p\}$. And similarly,

$$\begin{aligned} d_{TV}(P,Q) = P(A_0^\complement) - Q(A_0^\complement) &= \int_{\{p>q\}} (p-q)d\mu \\ &= \int_{\{p>q\}} |p-q| d\mu, \end{aligned}$$

where last line is from that $p - q = |p - q|$ on $\{p > q\}$. And hence

$$\begin{aligned} d_{TV}(P,Q) &= \frac{1}{2}\int_{\{q\geq p\}} |p-q| d\mu + \frac{1}{2}\int_{\{p>q\}} |p-q| d\mu \\ &= \frac{1}{2}\int_{\Omega} |p-q| d\mu. \end{aligned}$$

Also, note that

$$\begin{aligned} d_{TV}(P,Q) &= \int_{\{q\geq p\}} (q-p)d\mu \\ &= \int_{\{q\geq p\}} (q - \min\{p,q\})d\mu + \int_{\{p>q\}} (q - \min\{p,q\})d\mu \\ &= \int_{\Omega} q d\mu - \int_{\Omega} \min\{p,q\} d\mu \\ &= 1 - \int_{\Omega} \min\{p,q\} d\mu. \end{aligned}$$

And hence

$$d_{\mathrm{TV}}(P,Q) = \frac{1}{2}\int_{\Omega} |p-q| d\mu$$

(d)

Since $\phi : S \to \{0,1\}$, $\phi(\omega) = 1_{\{\omega:\phi(\omega)=1\}}(\omega)$. And hence

$$\begin{aligned}
R_{P,Q}(\phi) &= \int_S \phi dP + \int_S (1-\phi) dQ \\
&= 1 - \left(\int_S \phi dQ - \int_S \phi dP \right) \\
&= 1 - \left(\int_S 1_{\{\phi=1\}} dQ - \int_S 1_{\{\phi=1\}} dP \right) \\
&= 1 - (Q(\{\phi = 1\}) - P(\{\phi = 1\})) \\
&\geq 1 - d_{TV}(P,Q),
\end{aligned}$$

and the inequality holds if and only if $Q(\{\phi = 1\}) - P(\{\phi = 1\}) = d_{TV}(P,Q)$. Also, when $\phi(\omega) = 1_{\{\omega:q(\omega)\geq p(\omega\}}(\omega)$, then $\{\phi = 1\} = \{q \geq p\}$ implies $Q(\{\phi = 1\}) - P(\{\phi = 1\}) = d_{TV}(P,Q)$, and hence $R_{P,Q}(\phi) = 1 - d_{TV}(P,Q)$ holds. And hence

$$\inf_\phi R_{P,Q}(\phi) = 1 - d_{TV}(P,Q).$$

2. Another way of quantifying how close two probability measures on some measurable space $(\Omega, \mathcal{F})$ are is to compute their Kullback-Liebler (KL) divergence, defined as

$$K(P,Q) = \begin{cases} \int \log \frac{dP}{dQ} dP & \text{if } P << Q \\ \infty & \text{otherwise.} \end{cases}$$

If P and Q are both absolutely continuous with respect to a σ-finite measure μ, then, assuming $P << Q$, $K(P,Q) = \int \log \left(\frac{p(\omega)}{q(\omega)} \right) p(\omega) d\mu(\omega)$ where p and q are the μ-densities of P and Q, respectively. In general $K(P,Q)$ is not a metric over the space of probability measures on $(\Omega, \mathcal{F})$: K is not symmetric!

(a) Use Jensen inequality to show that $K(P,Q) \geq 0$ with equality if and only if $P = Q$.

(b) Take $(\Omega, \mathcal{F})$ to be $(\mathbb{R}^k, \mathcal{B}^k)$ and let $\mathcal{P} = \{P_\theta, \theta \in \mathbb{R}^k\}$ where P_θ is the multivariate k-dimensional normal distribution with covariance matrix Σ and mean θ (thus, Σ is the common covariance matrix of all the P_θ's). Compute $K(P_{\theta_1}, P_{\theta_2})$ for all P_{θ_1} and P_{θ_2} in $\mathcal{P}$. Conclude that, over $\mathcal{P}$, K behaves like a metric (which one?).

(c) Take $(\Omega, \mathcal{F})$ to be $(\mathbb{R}, \mathcal{B})$ and, for any $\theta > 0$, let P_θ be the distribution Uniform$(0,\theta)$. Compute $K(P_{\theta_1}, P_{\theta_2})$ for all $\theta_1, \theta_2 \in (0, \infty)$.

Points: 10 pts = 5 + 2 + 3.

Solution.

(a)

First, when P is not absolutely continuous with respect to Q, then $P \neq Q$ and $K(P,Q) = \infty > 0$ holds.

Second, when $P \ll Q$, let $\Omega_0 := \left\{\omega : \frac{dP}{dQ}(\omega) > 0\right\}$. Then on Ω_0, $P \ll Q$ and $Q \ll P$ as well, and $\frac{dQ}{dP} = \left(\frac{dP}{dQ}\right)^{-1}$ on Ω_0. And also

$$P(\Omega \backslash \Omega_0) = \int_{\Omega \backslash \Omega_0} dP = \int_{\Omega \backslash \Omega_0} \frac{dP}{dQ} dQ = 0,$$

and hence $P(\Omega_0) = P(\Omega) = 1$. Hence applying Jensen inequality on a convex function $\varphi(x) = -\log x$ gives

$$\begin{aligned}
\int_\Omega \log \frac{dP}{dQ} dP &= \int_{\Omega_0} -\log\left(\frac{dQ}{dP}\right) dP \\
&\geq -\log\left(\int_{\Omega_0} \frac{dQ}{dP} dP\right) \\
&= -\log\left(\int_{\Omega_0} dQ\right) \\
&\geq -\log\left(\int_{\Omega} dQ\right) = 0.
\end{aligned}$$

Now, the second equality holds if and only if $\int_{\Omega \backslash \Omega_0} dQ = 0$, i.e. $Q(\Omega \backslash \Omega_0) = 0$ and $Q(\Omega_0) = 1$. Also, since $\varphi(x) = -\log x$ is strictly convex, the first equality holds if and only if $\frac{dQ}{dP}$ is a point mass under P on Ω_0, i.e. if and only if there exists $a \in \mathbb{R}$ such that $\frac{dQ}{dP} = a$ a.e. under P on Ω_0. Then if two equality holds, then

$$\int_{\Omega_0} \frac{dQ}{dP} dP \begin{cases} = \int_{\Omega_0} a dP = aP(\Omega_0) = a, \\ = \int_{\Omega_0} dQ = Q(\Omega_0) = 1, \end{cases}$$

and hence $a = 1$. Then for all $A \in \Omega$,

$$\begin{aligned}
P(A) = \int_A dP &= \int_{A \cap \Omega_0} dP \\
&= \int_{A \cap \Omega_0} \frac{dP}{dQ} dQ = \int_{A \cap \Omega_0} dQ \\
&= Q(A \cap \Omega_0) = Q(A),
\end{aligned}$$

and hence $P = Q$.

And when $P = Q$, then $\frac{dP}{dQ} = 1$ a.e. $[P]$, and hence $K(P,Q) = 0$.

Hence $K(P,Q) \geq 0$ holds with equality if and only if $P = Q$.
(b)
Let $p_\theta(x) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{1}{2}(x-\theta)^\top \Sigma^{-1}(x-\theta)\right)$ be the pdf of P_θ on $\mathbb{R}^k$. Then

$$\begin{aligned}
\frac{dP_{\theta_1}}{dP_{\theta_2}}(x) &= \frac{p_{\theta_1}(x)}{p_{\theta_2}(x)} \\
&= \exp\left(-\frac{1}{2}(x-\theta_1)^\top \Sigma^{-1}(x-\theta_1) + \frac{1}{2}(x-\theta_2)^\top \Sigma^{-1}(x-\theta_2)\right) \\
&= \exp\left((\theta_1-\theta_2)^\top \Sigma^{-1}\left(x - \frac{\theta_1+\theta_2}{2}\right)\right).
\end{aligned}$$

Let $X \sim P_{\theta_1}$, then $\mathbb{E}_{P_{\theta_1}}[X] = \theta_1$, and hence

$$\begin{aligned}
KL(P_{\theta_1}, P_{\theta_2}) &= \mathbb{E}_{P_{\theta_1}}\left[\log\left(\frac{dP_{\theta_1}}{dP_{\theta_2}}(X)\right)\right] \\
&= \mathbb{E}_{P_{\theta_1}}\left[(\theta_1-\theta_2)^\top \Sigma^{-1}\left(X - \frac{\theta_1+\theta_2}{2}\right)\right] \\
&= \frac{1}{2}(\theta_1-\theta_2)^\top \Sigma^{-1}(\theta_1-\theta_2).
\end{aligned}$$

Hence $KL(P_{\theta_1}, P_{\theta_2}) = KL(P_{\theta_2}, P_{\theta_1})$, i.e. KL is symmetric. In fact, $\sqrt{2KL(\cdot,\cdot)}$ is a Mahalanobis distance.
(c)
When $\theta_1 > \theta_2$, then $P_{\theta_1}((\theta_2,\theta_1)) > 0$ but $P_{\theta_2}((\theta_2,\theta_1)) = 0$, so P_{θ_1} is not absolutely continuous with respect to θ_2, and hence $KL(P_{\theta_1}, P_{\theta_2}) = \infty$. When $\theta_1 \leq \theta_2$, let $p_\theta(x) = \frac{1}{\theta} I_{(0,\theta)}(x)$ be the pdf of P_θ on $\mathbb{R}$. Then

$$\begin{aligned}
\frac{dP_{\theta_1}}{dP_{\theta_2}}(x) = \frac{p_{\theta_1}(x)}{p_{\theta_2}(x)} &= \frac{\frac{1}{\theta_1} I_{(0,\theta_1)}(x)}{\frac{1}{\theta_2} I_{(0,\theta_2)}(x)} \\
&= \frac{\theta_2}{\theta_1} I_{(0,\theta_1)}(x).
\end{aligned}$$

And hence $KL(P_{\theta_1}, P_{\theta_2})$ can be computed as

$$\begin{aligned}
KL(P_{\theta_1}, P_{\theta_2}) &= \int \log\left(\frac{dP_{\theta_1}}{dP_{\theta_2}}\right) dP_{\theta_1} \\
&= \int_0^{\theta_1} \log\left(\frac{\theta_2}{\theta_1}\right) \frac{1}{\theta_1} dx \\
&= \log\left(\frac{\theta_2}{\theta_1}\right).
\end{aligned}$$

And hence

$$KL(P_{\theta_1}, P_{\theta_2}) = \begin{cases} \log\left(\frac{\theta_2}{\theta_1}\right) & \text{if } \theta_1 \le \theta_2, \\ \infty & \text{if } \theta_1 > \theta_2. \end{cases}$$

Remark.
In (a), note that $\int \left(\frac{dP}{dQ}\right)^{-1} dP$ need not equal to 1 unless $P \ll Q$ and $Q \ll P$. For example, suppose P and Q has densities $p(\omega) = I_{(0,1)}(\omega)$ and $q(\omega) = \frac{1}{2} I_{(0,2)}(\omega)$ with respect to a Lebesgue measure μ. Then $\frac{dP}{dQ}(\omega) = 2I_{(0,1)}(\omega)$ a.e. with respect to μ. But $\int \left(\frac{dP}{dQ}\right)^{-1} dP = \int \frac{1}{2} I_{(0,1)} d\mu = \frac{1}{2} \neq 1$, while $\int dQ = 1$.

3. **Riemann versus Lebesgue Integral.** Let $f(x) = \frac{(-1)^n}{n}$ if $n-1 \le x < n$ for $n = 1, 2, \ldots$. We saw in class that $\int_0^\infty f(x)dx$ exists as an improper Riemann integral since

$$\lim_{b\to\infty} \int_0^b f(x) = -\log 2.$$

Show however that f is not Lebesgue integrable over $[0, \infty)$. *Hint: it is enough to show that* $\int_0^\infty |f| = \infty$.
In contrast, show that the function $f\colon [0,1] \to \{0,1\}$ such that $f(x) = 1$ if x is rational and 0 otherwise is such that $\int_0^1 f(x)d\lambda(x) = 0$ but it is not Riemann integrable. *Hint: you will need the result in the next exercise.*

Points: 10 pts.

Solution.
Let $f(x) = \frac{(-1)^n}{n}$ if $n-1 \le x < n$ for $n = 1, 2, \ldots$. Note that for all $x \ge 0$, $n - 1 \le x$ implies $|f(x)| = \frac{1}{n} \le \frac{1}{x+1}$, and hence

$$\int_0^\infty |f(x)|dx \le \int_0^\infty \frac{1}{x+1} dx = [\log(x+1)]_0^\infty = \infty.$$

And hence f is (improper) Riemann integrable but not Lebesgue integrable over $[0, \infty)$.
Now, let $f\colon [0,1] \to \{0,1\}$ such that $f(x) = 1$ if x is rational and 0 otherwise. Then

$$\begin{aligned} \int_0^1 f(x)d\lambda(x) &= \int_0^1 I_{\mathbb{Q}\cap[0,1]}(x)d\lambda(x) \\ &= \lambda(\mathbb{Q} \cap [0,1]) \\ &= \sum_{x\in\mathbb{Q}\cap[0,1]} \lambda(\{x\}) \\ &= \sum_{x\in\mathbb{Q}\cap[0,1]} 0 = 0. \end{aligned}$$

To see whether f is Riemann integrable, let $0 = t_0 < \cdots < t_m = 1$ be a partition. Then since $[t_{i-1}, t_i]$ contains both rational and irrational numbers,

$$\sum_{i=1}^{m} \inf_{t \in [t_{i-1}, t_i]} f(t)(t_i - t_{i-1}) = \sum_{i=1}^{m} 0 \times (t_i - t_{i-1}) = 0,$$
$$\sum_{i=1}^{m} \sup_{t \in [t_{i-1}, t_i]} f(t)(t_i - t_{i-1}) = \sum_{i=1}^{m} 1 \times (t_i - t_{i-1}) = \sum_{i=1}^{m} (t_i - t_{i-1}) = 1.$$

Hence upper Riemann sum is always 1 and lower Riemann sum is always 0. Hence they don't converge as partition becomes finer, and hence f is not Riemann integrable.

4. Show that the Lebesgue measure and any counting measure over a countable subset of $\mathbb{R}^k$ (for example, the set of rationals) are mutually singular.

Points: 10 pts.

Solution.
Let $A_0 = \{x_n : n \in \mathbb{N}\} \subset \mathbb{R}^k$ be a countable subset of $\mathbb{R}^k$. Let λ be the Lebesgue measure and μ_{A_0} be the counting measure over A_0 defined as $\mu_{A_0}(A) = |A \cap A_0|$. Then since $\lambda(\{x_n\}) = 0$ for all $n \in \mathbb{N}$,

$$\lambda(A_0) = \sum_{n=1}^{\infty} \lambda(\{x_n\}) = 0.$$

And $\mu_{A_0}(A_0^{\complement}) = |A_0 \cap A_0^{\complement}| = 0$, and hence λ and μ_{A_0} are mutually singular.

5. Let f be an integrable real-valued function over a measure space $(\Omega, \mathcal{F}, P)$. Show that f is finite almost everywhere $[\mu]$. *Hint: you may assume that $f \geq 0$ then you only need to show that $f < \infty$ almost everywhere $[\mu]$.*

Points: 10 pts.

Solution.
Let $A_n = f^{-1}(\bar{\mathbb{R}} \backslash (-n, n)) = \{\omega \in \Omega : |f(\omega)| \geq n\}$. Then $|f| \geq n$ on A_n, and hence

$$n\mu(A_n) = \int_{A_n} n d\mu \leq \int_{A_n} |f| d\mu \leq \int_{\Omega} |f| d\mu,$$

and hence

$$\mu(A_n) \leq \frac{1}{n} \int_{\Omega} |f| d\mu.$$

Now, $A_n \downarrow A_\infty = \{\omega \in \Omega : |f(\omega)| = \infty\}$, then from the continuity of measure,

$$\begin{aligned}\mu(A_\infty) &= \mu(\lim_n A_n) = \lim_n \mu(A_n) \\ &\leq \lim_n \frac{1}{n} \int_\Omega |f| d\mu = 0.\end{aligned}$$

And hence $\mu(A_\infty) = 0$, and f is finite almost everywhere $[\mu]$.

6. Assume that f and g are simple functions on some measure space $(\Omega, \mathcal{F}, \mu)$. Prove that, for all $a, b \in \mathbb{R}$,

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

Points: 10 pts.

Solution.

Note first that if $f = \sum_{i=1}^m a_i I_{A_i}$ with A_i being disjoint (while a_i's are not necessarily disjoint), then $\int f d\mu = \sum_{i=1}^m a_i \mu(A_i)$. This is since once we rewrite f as $f = \sum_{i=1}^n b_i \sum_{j=1}^{i_k} I_{B_{i,j}} = \sum_{i=1}^n b_i I_{\cup_{j=1}^{i_k} B_{i,j}}$ with b_i's being distinct and $\{(b_i, B_{i,j})\} = \{(a_i, A_i)\}$, then

$$\begin{aligned}\int f d\mu &= \sum_{i=1}^n b_i \mu(\cup_{j=1}^{i_k} B_{i,j}) = \sum_{i=1}^n b_i \sum_{j=1}^{i_k} \mu(B_{i,j}) \\ &= \sum_{i=1}^n \sum_{j=1}^{i_k} b_i \mu(B_{i,j}) = \sum_{i=1}^m a_i \mu(A_i).\end{aligned}$$

Now, since f and g are simple functions, $f = \sum_{i=1}^m a_i I_{A_i}$ and $g = \sum_{j=1}^n b_j I_{B_j}$ for some $a_i, b_j \neq 0$ and $A_i, B_j \in \mathcal{F}$. We let $A_0 = \cup_{j=1}^n B_j - \cup_{i=1}^m A_i$, $B_0 = \cup_{i=1}^m A_i - \cup_{j=1}^n B_j$, and let $a_0 = b_0 = 0$. Then

$$af + bg = \sum_{i=0}^m \sum_{j=0}^n (aa_i + bb_j) 1_{A_i \cap B_j},$$

and $A_i \cap B_j$ are pairwise disjoint, and hence

$$\begin{aligned}
\int (af + bg) d\mu &= \sum_{i=0}^{m} \sum_{j=0}^{n} (aa_i + bb_j)\mu(A_i \cap B_j) \\
&= a \sum_{i=0}^{m} a_i \sum_{j=0}^{n} \mu(A_i \cap B_j) + b \sum_{j=0}^{n} b_j \sum_{i=0}^{m} \mu(A_i \cap B_j) \\
&= a \sum_{i=0}^{m} a_i \mu(A_i) + b \sum_{j=0}^{n} b_j \mu(B_j) \\
&= a \int f d\mu + b \int g d\mu,
\end{aligned}$$

where the second-to-last step is from that $A_i = \cup_{j=0}^{n}(A_i \cap B_j)$ and $B_j = \cup_{i=0}^{m}(A_i \cap B_j)$ and $A_i \cap B_j$ being disjoint.

7. Let $\{f_n\}$ be a sequence of non-negative functions on some measure space $(\Omega, \mathcal{F}, \mu)$. Assume that $\int f_n d\mu \to 0$. Prove or disprove (with a counter-example): $f_n \to 0$ a.e.$[\mu]$.

Points: 10 pts.

Solution.
f_n need not converges to 0 a.e. $[\mu]$. In fact, it is possible that $f_n(\omega) \nrightarrow 0$ for all $\omega \in \Omega$.
Let $\Omega = [0, 1)$, $\mathcal{F}$ =Borel sets of Ω, and μ be the Lebesgue measure. For $n \in \mathbb{N}$, let $m := \lfloor \log_2 n \rfloor$ so that $2^m \leq n < 2^{m+1}$, and define $f_n : [0, 1) \to [0, \infty)$ be

$$f_n(\omega) := I\left(\omega \in \left[\frac{n - 2^m}{2^m}, \frac{n - 2^m + 1}{2^m}\right)\right).$$

Then $\int f_n d\mu = \frac{1}{2^{\lfloor \log_2 n \rfloor}} \leq \frac{2}{n} \to 0$ as $n \to \infty$. However, fix $\omega \in [0, 1)$, then for all $m \in \mathbb{N}$, there exists $0 \leq k < 2^m$ with $\frac{k}{2^m} \leq \omega < \frac{k+1}{2^m}$. Then

$$f_{2^m+k}(\omega) = I\left(\omega \in \left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right) = 1.$$

Since such k exists for each $m \in \mathbb{N}$, $f_n(\omega) = 1$ for infinitely many n, and hence $f_n(\omega) \nrightarrow 0$ for all $\omega \in \Omega$.

8. Let $(\Omega_1, \mathcal{F}_1), \ldots, (\Omega_k, \mathcal{F}_k)$ be k measurable spaces. For $j = 1, \ldots, k$ let $\pi_j \colon \prod_{i=1}^{k} \Omega_i \to \Omega_j$ denote the coordinate projection mapping given by $\pi_i(\omega_1, \ldots, \omega_k) = \omega_j$. Show that the product σ-field $\bigotimes_{i=1}^{k} \mathcal{F}_i$ is the σ-field generated by all the coordinate projections.

Points: 10 pts.

Solution.

Let $\mathcal{F}_0 = \{A_1 \times \cdots \times A_k : A_i \in \mathcal{F}_i\}$, so that $\mathcal{F} = \sigma(\mathcal{F}_0)$. Let $\mathcal{G}_0 = \{\pi_i^{-1}(A_i) : A_i \in \mathcal{F}_i\}$ and let $\mathcal{G} := \sigma(\mathcal{G}_0)$. We would like to show $\mathcal{F} = \mathcal{G}$.

($\mathcal{G} \subset \mathcal{F}$):

Since $\pi_i^{-1}(A_i) = \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \cdots \times \Omega_k$, $\mathcal{G}_0 \subset \mathcal{F}_0$, and hence $\mathcal{G}_0 \subset \sigma(\mathcal{F}_0) = \mathcal{F}$. Then since $\mathcal{F}$ is a σ-field containing $\mathcal{G}_0$, $\sigma(\mathcal{G}_0) \subset \mathcal{F}$ as well, i.e. $\mathcal{G} \subset \mathcal{F}$ holds.

($\mathcal{F} \subset \mathcal{G}$):

Note that for all $A_i \in \mathcal{F}_i$, $\prod_{i=1}^k A_i = \bigcap_{i=1}^k \pi_i^{-1}(A_i) \in \mathcal{G}$ since $\mathcal{G}$ is a σ-field. Hence $\mathcal{F}_0 \subset \mathcal{G}$. Then since $\mathcal{G}$ is a σ-field containing $\mathcal{F}_0$, $\sigma(\mathcal{F}_0) \subset \mathcal{G}$ as well, i.e. $\mathcal{F} \subset \mathcal{G}$ holds.

Hence $\mathcal{F} = \mathcal{G}$, i.e. $\bigotimes_{i=1}^k \mathcal{F}_i$ is the σ-field generated by all the coordinate projections.

9. Let λ_k denote the k-dimensional Lebesgue measure and H be a linear subspace in $\mathbb{R}^k$ of dimension no larger than $k-1$. Show that $\lambda_k(H) = 0$. You may proceed as follows:

   (a) show that the Lebesgue measure is translation invariant: for each Borel measurable set A and $x \in \mathbb{R}^k$, $\lambda_k(A) = \lambda_k(x + A)$, where $x + A = \{x + y, y \in A\}$. *Hint: use the good set principle to show that the class of sets A such that $A + x \in \mathcal{B}^k$ for all $x \in \mathbb{R}^k$ coincides with $\mathcal{B}^k$ and then show – using the uniqueness theorem for measures – that any measure μ such that, for any fixed x, $\mu(A) = \lambda_k(A + x)$ for all $A \in \mathcal{B}^k$ coincides with λ_k.*

   (b) Use the fact (which you do not need to prove!) that, for any σ-finite measure μ on $(\Omega, \mathcal{F})$, only countably many disjoint sets in $\mathcal{F}$ can have positive measure to conclude that $\lambda_k(H) = 0$ for any subspace of dimension less than k. (In fact, the same conclusion holds for any affine subspace of dimension less than k, where an affine subspace is a set of the form $x + S = \{x + y, y \in S\}$ for a linear subspace S and a point $x \in \mathbb{R}^k$).

Points: 10 pts = 7 + 3.

Solution.

(a)

We show that for all $x \in \mathbb{R}^k$, $\lambda_k(x + A) = \lambda_k(A)$. Fix any $x \in \mathbb{R}^k$ and let $T_x : \mathbb{R}^k \to \mathbb{R}^k$ be a translation as $T_x(y) = y - x$, then $x + A = T_x^{-1}(A)$. Then since $T_x : \mathbb{R}^k \to \mathbb{R}^k$ is continuous and hence measurable, $A \in \mathcal{B}^k$ implies $x + A \in \mathcal{B}^k$. Now consider the induced measure μ on $(\mathbb{R}^k, \mathcal{B}^k)$ as $\mu(A) = \lambda_k(T_x^{-1}(A))$, i.e. $\mu(A) = \lambda_k(x + A)$. Let $\Pi = \{\prod_{i=1}^k (a_i, b_i] \in \mathcal{B}^k : a_i \leq b_i\}$. Then $\left(\prod_{i=1}^k (a_i, b_i]\right) \cap$

$\left(\prod_{i=1}^k (c_i,\, d_i]\right) = \prod_{i=1}^k (\max\{a_i, c_i\}, \min\{b_i, d_i\}] \in \Pi$, so Π is a π-system. Now, note that $\lambda_k\left(\prod_{i=1}^k (a_i,\, b_i]\right) = \prod_{i=1}^k (b_i - a_i)$ and

$$\mu\left(\prod_{i=1}^k (a_i,\, b_i]\right) = \lambda_k\left(\prod_{i=1}^k (x_i + a_i,\, x_i + b_i]\right) = \prod_{i=1}^k ((x_i+b_i)-(x_i+a_i)) = \prod_{i=1}^k (b_i-a_i),$$

and hence $\lambda_k = \mu$ on Π. Also, note that $\left\{\prod_{i=1}^k (n_i, n_i+1] : n_i \in \mathbb{Z}\right\}$ covers $\mathbb{R}^k$ and $\mu\left(\prod_{i=1}^k (n_i, n_i+1]\right) = \lambda_k\left(\prod_{i=1}^k (n_i, n_i+1]\right) = 1$, and hence both μ and λ_k are σ-finite on Π. Then from uniqueness of the measure, μ and λ_k agree on $\sigma(\Pi) = \mathcal{B}^k$. i.e. $\mu(A) = \lambda_k(x + A) = \lambda_k(A)$ for all $x \in \mathbb{R}^k$ and $A \in \mathcal{B}^k$.
(b)
We show that $\lambda_k(H) = 0$ for any subspace $H \neq \mathbb{R}^k$. Choose $x \in \mathbb{R}^k \backslash H$. Then for $t \in \mathbb{R}$,

$$y \in tx + H \iff y - tx \in H.$$

Hence for $t_1 \neq t_2 \in \mathbb{R}$, $y \in (t_1 x + H) \cap (t_2 x + H)$ implies $y - t_1 x,\, y - t_2 x \in H$, and then

$$x = \frac{(y - t_1 x) - (y - t_2 x)}{t_2 - t_1} \in H,$$

which is a contradiction. Hence $(t_1 x + H) \cap (t_2 x + H) = \emptyset$ if $t_1 \neq t_2$. Then $\{tx + H :\ t \in \mathbb{R}\}$ are uncountably many disjoint sets.
Now, if $\lambda_k(H) > 0$, then from above, $\lambda_k(tx + H) > 0$ for all $t \in \mathbb{R}$. Then $\{tx + H :\ t \in \mathbb{R}\}$ are uncountably many disjoint sets with all positive measures, which is impossible. Hence $\lambda_k(H) = 0$.

Remark.

In (a), we showed the measurability of $x + A$ by using the measurable function T_x defined as $T_x(y) = y - x$ and using that $x + A = T_x^{-1}(A)$. An alternative is to use good set principle as follows. Let $\mathcal{T} := \left\{A \in \mathcal{B}^k : \text{for all } x \in \mathbb{R}^k,\, x + A \in \mathcal{B}^k\right\}$, and we show that $\mathcal{T} = \mathcal{B}^k$. We first check that $\mathcal{T}$ is σ-algebra. First, $x + \mathbb{R}^k = \mathbb{R}^k \in \mathcal{B}^k$ for all $x \in \mathbb{R}^k$, and hence $\mathbb{R}^k \in \mathcal{T}$. Second, if $A \in \mathcal{T}$, then for all $x \in \mathbb{R}^k$, $x + A \in \mathcal{B}^k$, and $\mathcal{B}^k$ being a σ-field implies $x + A^{\complement} = (x + A)^{\complement} \in \mathcal{B}^k$, and hence $A^{\complement} \in \mathcal{T}$. Third, suppose $\{A_n\}_{n=1}^{\infty} \subset \mathcal{T}$, then for all $x \in \mathbb{R}^k$, $x + A_n \in \mathcal{B}^k$ for all $n \in \mathbb{N}$. Then $\mathcal{B}^k$ being a σ-field implies $x + \bigcup_n A_n = \bigcup_n (x + A_n) \in \mathcal{B}^k$, and hence $\bigcup_n A_n \in \mathcal{T}$. Hence $\mathcal{T}$ is a σ-field. Also, let $\mathcal{O} = \{A \subset \mathbb{R}^k :\ A \text{ open}\}$, then A being open implies that for all $x \in \mathbb{R}^k$, $x + A$ is also open, and hence $x + A \in \mathcal{B}^k$ and hence $\mathcal{O} \subset \mathcal{T}$. Then since $\mathcal{B}^k = \sigma(\mathcal{O})$ and $\mathcal{T}$ is a σ-field, $\mathcal{B}^k \subset \mathcal{T}$ holds. Since $\mathcal{T} \subset \mathcal{B}^k$ from the definition of $\mathcal{T}$, $\mathcal{T} = \mathcal{B}^k$.
Also, checking the *sigma*-finiteness of μ and λ_k is critical in the proof. Suppose we have μ_0 on $(\mathbb{R}^k, \mathcal{B}^k)$ as $\mu_0(A) = \infty$ for all $A \in \mathcal{B}^k$, and we have constructed $\Pi_0 = \{\prod_{i=1}^k (-\infty, a_i] \in \mathcal{B}^k :\ a_i \in \mathbb{R}\}$. Then Π_0 is a π-system and $\mu_0 = \lambda_k$ on Π_0, but μ_0 and λ_k do not agree on $\sigma(\Pi_0) = \mathcal{B}^k$.

10. Use Kolmogorov's extension theorem to demonstrate the existence of a probability distribution over infinite sequences of fair coin tosses. In fact, in this case we can construct such measure explicitly and without relying on Kolmogorov's theorem. Let Ω be the unit interval $(0,1)$ equipped with the σ-field of Borel subsets and the Lebesgue measure P. Let $Y_n(\omega) = 1$ if $[2^n\omega]$ (the integral part of $2^n\omega$) is odd and 0 otherwise. Show that Y_1,Y_2,... are independent with $P(Y_k = 0) = P(Y_k = 1) = 1/2$ for all k. For any $\omega \in \Omega$, the binary sequence $\{Y_n(\omega), n = 1, 2, \ldots\}$ is the corresponding sample path of the process.

Points: 10 pts.

Solution.

Note first that if a random variable Y_n takes either 0 or 1 as its value, $\sigma(Y_n) = \{Y_n^{-1}(I) : I \subset \{0,1\}\}$. Hence showing independence of $Y_1, Y_2, \ldots$ is equivalent to showing that for any $I_{n_i} \subset \{0,1\}$,

$$P\left(\bigcap_{i=1}^{k} Y_{n_i}^{-1}(I_{n_i})\right) = \prod_{i=1}^{k} P(Y_{n_i}^{-1}(I_{n_i})).$$

First, we use Kolmogorov's extension theorem to demonstrate the existence of a probability distribution over infinite sequences of fair coin tosses. Consider $\mathbb{R}^{\mathbb{N}}$. For each $n \in \mathbb{N}$, consider a measurable space $(\mathbb{R}, \mathcal{B})$. Then, for all $v \subset \mathbb{N}$ with $|v| < \infty$, let $(\mathbb{R}^v, \mathcal{B}^v)$ be the corresponding product space and product σ-field. Let P^v be a probability measure on $(\mathbb{R}^v, \mathcal{B}^v)$ defined as for all $B \in \mathcal{B}^v$,

$$P^v(B) = \frac{|B \cap \{0,1\}^v|}{2^{|v|}}.$$

Then P^v is a normalized counting measure with $P^v(\mathbb{R}^v) = 1$, and hence P_v is a probability measure. Also, for any $u \subset v \subset \mathbb{N}$ and for all $B \in \mathcal{B}^u$,

$$\begin{aligned}
\pi_u(P^v)(B) &= P^v(\{x \in \mathbb{R}^v : x_u \in B\}) \\
&= \frac{|\{x \in \mathbb{R}^v : x_u \in B\} \cap \{0,1\}^v|}{2^{|v|}} \\
&= \frac{|\{x \in \mathbb{R}^v : x_u \in B \cap \{0,1\}^u,\, x_{u\setminus v} \in \{0,1\}^{u\setminus v}\}|}{2^{|v|}} \\
&= \frac{|\{x \in \mathbb{R}^v : x_u \in B\} \cap \{0,1\}^u| \times |\{0,1\}|^{|u\setminus v|}}{2^{|u|} \times 2^{|u\setminus v|}} \\
&= P^u(B).
\end{aligned}$$

Hence $\{P^v : v \subset \mathbb{N}, |v| < \infty\}$ is a consistent set of probability measures. Hence from Kolmogorov's extension theorem, there exists a unique probability measure P on $(\mathbb{R}^{\mathbb{N}}, \otimes_{n\in\mathbb{N}}\mathcal{B})$ such that $\pi_v(P) = P^v$ for all finite $v \subset \mathbb{N}$. Now, define random variables $\{X_n\}_{n\in\mathbb{N}}$ on $\mathbb{R}^N$ as $X_n(\omega) = \omega_n$. Then for any $I_{n_1}, \dots, I_{n_k} \subset \{0,1\}$, let $v_0 := \{n_1, \dots, n_k\}$, then

$$
\begin{aligned}
P\left(\bigcap_{i=1}^k X_{n_i}^{-1}(I_{n_i})\right) &= P\left(\left\{x \in \mathbb{R}^{\mathbb{N}} : x_{n_i} \in I_{n_i}\right\}\right) = P\left(\left\{x \in \mathbb{R}^{\mathbb{N}} : x_{v_0} \in \prod_{i=1}^k I_{n_i}\right\}\right) \\
&= \pi_{v_0}(P)\left(\prod_{i=1}^k I_{n_i}\right) = P^{v_0}\left(\prod_{i=1}^k I_{n_i}\right) \\
&= \frac{\left|\prod_{i=1}^k I_{n_i} \cap \{0,1\}^{v_0}\right|}{2^{|v_0|}} = \frac{\prod_{i=1}^k |I_{n_i}|}{2^k}.
\end{aligned}
$$

And hence

$$
\prod_{i=1}^k P\left(X_{n_i}^{-1}(I_{n_i})\right) = \prod_{i=1}^k \frac{|I_{n_i}|}{2} = \frac{\prod_{i=1}^k |I_{n_i}|}{2^k} = P\left(\bigcap_{i=1}^k X_{n_i}^{-1}(I_{n_i})\right).
$$

Therefore, such $\{X_n : n \in \mathbb{N}\}$ is independent.

Second, we show that $\{Y_n\}_{n\in\mathbb{N}}$, constructed as $Y_n(\omega) = 1$ if $[2^n\omega]$ odd and $Y_n(\omega) = 0$ if $[2^n\omega]$ even, is a sequence of independent variables. Let $N := \max\{n_1, \dots, n_k\}$ and consider the map $Y_{1:N} : \Omega \to \{0,1\}^N$ as $Y_{1:N} = (Y_1, \dots, Y_N)$. For $1 \le n \le N$, define $\pi_n : \{0,1\}^N \to \{0,1\}$ be the n^{th} coordinate map as $\pi_n(y_1, \dots, y_N) = y_n$. Then $Y_n = \pi_n \circ Y_{1:N}$ and hence $Y_n^{-1}(I) = Y_{1:N}^{-1}(\pi_n^{-1}(I))$. Now, define the induced measure Q on $(\{0,1\}^N, 2^{\{0,1\}^N})$ as $Q(J) = P(Y_{1:N}^{-1}(J))$ for all $J \subset \{0,1\}^N$. Then for any $I_{n_1}, \dots, I_{n_k} \subset \{0,1\}$,

$$
\begin{aligned}
P\left(\bigcap_{i=1}^k Y_{n_i}^{-1}(I_{n_i})\right) &= P\left(\bigcap_{i=1}^k Y_{1:N}^{-1}(\pi_{n_i}^{-1}(I_{n_i}))\right) = P\left(Y_{1:N}^{-1}\left(\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i})\right)\right) \\
&= Q\left(\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i})\right),
\end{aligned}
$$

hence showing the independence of P is equivalent to showing that

$$
Q\left(\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i})\right) = \prod_{i=1}^k Q\left(\pi_{n_i}^{-1}(I_{n_i})\right).
$$

Let $\varphi : \{0,1\}^N \to \{0, \dots, 2^N - 1\}$ as $\varphi(y) = \sum_{n=1}^N y_n 2^{N-n}$. Then for all $\omega \in [\frac{\varphi(y)}{2^N}, \frac{\varphi(y)+1}{2^N}) \cap (0,1)$, $Y_n(\omega) = y_n$ holds for all $1 \le n \le N$, and hence $Y_{1:N}(\omega) = y$. Since φ is one-to-one and onto, $Y_{1:N}^{-1}(\{y\}) = [\frac{\varphi(y)}{2^N}, \frac{\varphi(y)+1}{2^N}) \cap (0,1)$, and hence

$$
Q(\{y\}) = P\left([\frac{\varphi(y)}{2^N}, \frac{\varphi(y)+1}{2^N}) \cap (0,1)\right) = \frac{1}{2^N},
$$

i.e. Q is a uniform measure and $Q(J) = \frac{|J|}{2^N}$. Then from $\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i}) = \left\{y \in \{0,1\}^N : y_{n_i} \in I_{n_i}\right\}$, $\left|\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i})\right| = 2^{N-k} \prod_{i=1}^k |I_{n_i}|$, and hence

$$Q\left(\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i})\right) = \frac{2^{N-k} \prod_{i=1}^k |I_{n_i}|}{2^N} = 2^{-k} \prod_{i=1}^k |I_{n_i}|,$$
$$\prod_{i=1}^k Q\left(\pi_{n_i}^{-1}(I_{n_i})\right) = \prod_{i=1}^k \frac{2^{N-1}|I_{n_i}|}{2^N} = 2^{-k} \prod_{i=1}^k |I_{n_i}|,$$

and hence $Q\left(\bigcap_{i=1}^k \pi_{n_i}^{-1}(I_{n_i})\right) = \prod_{i=1}^k Q\left(\pi_{n_i}^{-1}(I_{n_i})\right)$. Therefore, such $\{Y_n : n \in \mathbb{N}\}$ is independent.