36-752, Spring 2018
Homework 2

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

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
   $$d_{\mathrm{TV}}(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|.$$
   This is a very strong notion of distance between probability distributions: if $d_{\mathrm{TV}}(P, Q) < \epsilon$, 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, $d_{\mathrm{TV}}(P, Q) = 0$ if and only if $P = Q$.

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

   (c) Prove that the following equivalent representation of the total variation distance:
   $$d_{\mathrm{TV}}(P, Q) = \frac{1}{2} \int_{\Omega} |p - q| d\mu.$$
   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
   $$\inf_{\phi} R_{P,Q}(\phi) = 1 - d_{\mathrm{TV}}(P, Q),$$
   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 $\{q \geq p\}$.

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 $\theta_1, \theta_2 \in (0, \infty)$.

3. **Riemann versus Lebesgue Integral.** Let $f(x) = \frac{(-1)^n}{n}$ if $n - 1 \leq x < n$ 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.*

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.

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 $[\mu]$.*

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

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

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.$[\mu]$.

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.

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 $x \in \mathbb{R}^k$).

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.