Homework 7

Naman Mishra (22223)

1 October, 2024

Problem 1. Let μ and ν be Borel probability measures on $\mathbb{R}$. Suppose there exists a probability measure θ on $\mathbb{R}^2$ having marginals $\theta \circ \Pi_1^{-1} = \mu$ and $\theta \circ \Pi_2^{-1} = \nu$ such that $\theta\{(x,x) \mid x \in \mathbb{R}\} > 0$. Then show that μ and ν cannot be singular.

Solution. Assume there exists a measurable set A such that $\mu(A) = \nu(A^c) = 1$. Let $U = A \times \mathbb{R}$ and $V = \mathbb{R} \times A^c$. Then $\theta(U) = (\theta \circ \Pi_1^{-1})(A) = 1$ and $\theta(V) = (\theta \circ \Pi_2^{-1})(A^c) = 1$. Thus $\theta(U^c \cup V^c) \leq \theta(U^c) + \theta(V^c) = 0$, so $\theta(U \cap V) = 1$ (de Morgan). But $U \cap V = A \times A^c$ lies entirely outside the diagonal, since $(x, x) \in A \times A^c \implies A \cap A^c \neq \varnothing$. Thus $\theta\{(x,x) \mid x \in \mathbb{R}\} \leq \theta((U \cap V)^c) = 0$, a contradiction.

(More directly, the diagonal lies inside $U^c \cup V^c$, but my brain finds it easier to see that it lies outside $U \cap V$.) ∎

Problem 2. Place r_m balls in m bins at random and count the number of empty bins Z_m. Fix $\delta > 0$. If $r_m > (1+\delta) m \log m$, show that $\mathbf{P}(Z_m > 0) \to 0$ while if $r_m < (1-\delta) m \log m$, show that $\mathbf{P}(Z_m > 0) \to 1$.

Solution. This is simply by the coupon collector problem.

Let $X_n \sim \mathrm{Unif}([m])$ be iid, and note that

$$Z_m \sim m - \#\{X_1, \ldots, X_{r_m}\}.$$

Write

$$Z_m = \sum_{k=1}^{m} Z_{m,k}, \quad \text{where } Z_{m,k} = \prod_{i=1}^{r_m} \mathbf{1}_{X_i \neq k}$$

indicates that bin k is empty. Now by independence of X_is,

$$\mathbf{E}[Z_{m,k}] = \mathbf{P}\{X_1 \neq k\}^{r_m} = \left(1 - \frac{1}{m}\right)^{r_m},$$
$$\mathbf{E}[Z_{m,k}Z_{m,\ell}] = \mathbf{P}\{X_1 \neq k, \ell\}^{r_m} = \left(1 - \frac{2}{m}\right)^{r_m}$$

for each $k \neq \ell$. Thus

$$\begin{aligned}\mathbf{E}[Z_m] &= \sum_{k=1}^{m} \mathbf{E}[Z_{m,k}] = m\left(1 - \frac{1}{m}\right)^{r_m} \\ \mathbf{E}[Z_m^2] &= \mathbf{E}\left[\sum_{k=1}^{m} Z_{m,k} + \sum_{k\neq\ell} Z_{m,k}Z_{m,\ell}\right] \\ &= m\left(1 - \frac{1}{m}\right)^{r_m} + m(m-1)\left(1 - \frac{2}{m}\right)^{r_m}.\end{aligned}$$

For $r_m > (1+\delta)m\log m$, we have

$$\mathbf{E}[Z_m] \leq m\left(1 - \frac{1}{m}\right)^{(1+\delta)m\log m} \leq m(e^{-1/m})^{(1+\delta)m\log m} = \frac{1}{m^\delta} \to 0.$$

By Markov's inequality,

$$\mathbf{P}(Z_m > 0) = \mathbf{P}(Z_m \geq 1) \leq \mathbf{E}[Z_m] \to 0.$$

For $r_m < (1-\delta)m\log m$, we have

$$e^{(-\frac{1}{m}-\frac{1}{m^2})r_m} \leq \left(1 - \frac{1}{m}\right)^{r_m} \leq e^{-\frac{1}{m}r_m},$$
$$\left(1 - \frac{2}{m}\right)^{r_m} \leq e^{-\frac{2}{m}r_m}.$$

This gives

$$\begin{aligned}
\frac{\mathbf{E}[Z_m]^2}{\mathbf{E}[Z_m^2]} &\geq \frac{m^2 e^{2(-\frac{1}{m}-\frac{1}{m^2})r_m}}{me^{-\frac{1}{m}r_m} + m(m-1)e^{-\frac{2}{m}r_m}} \\
&= \frac{me^{-\frac{2}{m^2}r_m}}{e^{\frac{1}{m}r_m} + (m-1)} \\
&\geq \frac{me^{-\frac{2}{m^2}(1-\delta)m\log m}}{e^{\frac{1}{m}(1-\delta)m\log m} + m - 1} \\
&= \frac{m \cdot m^{-2(1-\delta)\frac{1}{m}}}{m^{1-\delta} + m - 1} \\
&= \frac{\left(m^{-\frac{1}{m}}\right)^{2(1-\delta)}}{m^{-\delta} + 1 - m^{-1}} \to 1
\end{aligned}$$

since $\lim_{x\to\infty} x^{\frac{1}{x}} = 1$.

By the Paley-Zygmund inequality,

$$\mathbf{P}(Z_m > 0) \geq \frac{\mathbf{E}[Z_m]^2}{\mathbf{E}[Z_m^2]} \to 1. \qquad \blacksquare$$

Recap of product measures

If $X_1, X_2, \ldots : (\Omega, \mathcal{F}, \mathbf{P}) \to \mathbb{R}$ are measurable functions, so is $X = (X_1, X_2, \ldots)$. The push-forward $\mu = \mathbf{P} \circ X^{-1}$ is then a probability measure on $\mathbb{R}^{\mathbb{N}}$. Obviously, $\mu_k = \mathbf{P} \circ X_k^{-1}$ is a probability measure on $\mathbb{R}$. But $X_k = \Pi_k \circ X$, so $\mu_k = \mu \circ \Pi_k^{-1}$.

> Why? Change of variables gives that if a random variable X has distribution μ, then $Y = g(X)$ has distribution $\mu \circ g^{-1}$, because
>
> $$\mathbf{P}(Y^{-1}(A)) = \mathbf{P}((g \circ X)^{-1}(A)) = \mathbf{P}(X^{-1}(g^{-1}(A))) = \mu(g^{-1}(A)).$$

In other words, $\mu_k(A) = \mu(\mathbb{R}^{k-1} \times A \times \mathbb{R}^{\mathbb{N}})$. If X_k's are known to be independent, then

$$\mathbf{P}\{X_1 \in A_1, \ldots, X_n \in A_n\} = \prod_{k=1}^{n} \mu_k(A_k)$$

for any $A_k \in \mathcal{B}(\mathbb{R})$. Since cylinder sets are a π-system generating $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$, this uniquely determines μ.

Definition 0.1 (product measure). Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i \in I$, be probability spaces indexed by an arbitrary set I. Let $\Omega = \prod_{i\in I} \Omega_i$ and let $\mathcal{F}$ (usually denoted $\bigotimes_{i\in I} \mathcal{F}_i$) be the σ-algebra generated by all finite-dimensional cylinders (equivalently, the smallest σ-algebra on Ω for which all the projections $\Pi_i \colon \Omega \to \Omega_i$ are measurable). If μ is a probability measure on $(\Omega, \mathcal{F})$ such that for any cylinder set $A = \Pi_{i_1}^{-1}(A_{i_1}) \cap \cdots \cap \Pi_{i_n}^{-1}(A_{i_n})$ for some $A_{i_r} \in \mathcal{F}_{i_r}$,

$$\mu(A) = \prod_{r=1}^{k} \mu_{i_r}(A_{i_r}),$$

then we say that μ is the product of μ_i, $i \in I$, and write $\mu = \bigotimes\limits_{i\in I} \mu_i$.

Theorem 0.2. *Let $\mu_n \in \mathcal{P}(\mathbb{R})$. Then $\mu = \bigotimes_{n\in\mathbb{N}} \mu_n$ exists and is unique.*

Proof. Let X_n be independent random variables with distribution μ_n. Then $X = (X_n)_n$ is a random variable with distribution μ. ■

0.1 Fubini's theorem

Theorem 0.3 (Fubini-Tonelli theorem). *Let $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$, $i = 1, 2$, be probability spaces and let $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, and $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$. Let $Y \colon \Omega \to \mathbb{R}$ be a random variable that is either positive or integrable with respect to $\mathbf{P}$. Then $Y(\omega_1, \cdot) \colon \Omega_2 \to \mathbb{R}$ is a random variable on Ω_2 for each $\omega_1 \in \Omega_1$, and is either positive or integrable (with respect to $\mathbf{P}_2$) for almost every ω_1 $[\mathbf{P}_1]$. Further, the function $\omega_1 \mapsto \int_{\Omega_2} Y(\omega_1, \omega_2) \,\mathrm{d}\,\mathbf{P}_2(\omega_2)$ is a random variable on Ω_1, and is either positive or integrable. Finally,*

$$\int_{\Omega_1} \left[\int_{\Omega_2} Y(\omega_1, \omega_2) \,\mathrm{d}\,\mathbf{P}_2(\omega_2) \right] \mathrm{d}\,\mathbf{P}_1(\omega_1) = \int_{\Omega} Y \,\mathrm{d}\,\mathbf{P}\,.$$

1 The Radon-Nikodym theorem and conditional probability

Consider a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let $X \colon \Omega \to \mathbb{R}$ be a non-negative random variable with expectation 1. Define $\mathbf{Q}(A) := \mathbf{E}[X \mathbf{1}_A]$ for $A \in \mathcal{F}$.

Then $\mathbf{Q}$ is a probability measure on $(\Omega, \mathcal{F})$.

- $\mathbf{Q}(\varnothing) = \mathbf{E}[X\mathbf{1}_\varnothing] = 0$.
- $\mathbf{Q}(A \sqcup B) = \mathbf{E}[X\mathbf{1}_{A\sqcup B}] = \mathbf{E}[X\mathbf{1}_A + X\mathbf{1}_B] = \mathbf{Q}(A) + \mathbf{Q}(B)$.
- If $A_n \uparrow A$, then $\mathbf{1}_{A_n} \uparrow \mathbf{1}_A$, so by MCT $\mathbf{Q}(A_n) \uparrow \mathbf{Q}(A)$.

Definition 1.1. Two measures μ and ν on the same $(\Omega, \mathcal{F})$ are said to be *mutually singular* and write $\mu \perp \nu$, if these is a set $A \in \mathcal{F}$ such that $\mu(A) = \nu(A^c) = 0$. We say that μ is *absolutely continuous* to ν and write $\mu \ll \nu$ if $\mu(A) = 0$ whenever $\nu(A) = 0$.

Theorem 1.2 (Radon-Nikodym theorem)**.** *Suppose μ and ν are two finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, then $\mathrm{d}\nu = f\,\mathrm{d}\mu$ for some $f \in L^1(\mu)$.*

Lemma 1.3. *Let μ_1 and μ_2 be two probability measures on a common measurable space $(\Omega, \mathcal{F})$, and let $\mu_p = p\mu_1 + (1-p)\mu_2$ for $p \in [0,1]$. Let X be any random variable on $(\Omega, \mathcal{F})$. Then $\int X\,\mathrm{d}\mu_p = p\int X\,\mathrm{d}\mu_1 + (1-p)\int X\,\mathrm{d}\mu_2$.*

Proof. From the construction of expectation,

$$\int X\,\mathrm{d}\,\mathbf{P} = \lim_{n\to\infty} \sum_{k=1}^{n2^n-1} \frac{k}{2^n}\,\mathbf{P}\left\{\frac{k}{2^n} \le X < \frac{k+1}{2^n}\right\},$$

we have

$$\begin{aligned}\int X\,\mathrm{d}\mu_p &= \lim_{n\to\infty} \sum_{k=1}^{n2^n-1} \frac{k}{2^n}\mu_p\left\{\frac{k}{2^n} \le X < \frac{k+1}{2^n}\right\} \\ &= \lim_{n\to\infty} \sum_{k=1}^{n2^n-1} \frac{k}{2^n}(p\mu_1 + (1-p)\mu_2)\left\{\frac{k}{2^n} \le X < \frac{k+1}{2^n}\right\} \\ &= p\int X\,\mathrm{d}\mu_1 + (1-p)\int X\,\mathrm{d}\mu_2.\end{aligned}$$

■

Problem 3. Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and let $\theta = \frac{1}{2}\mu + \frac{1}{2}\nu$.

(1) Show that $\mu \ll \theta$ and $\nu \ll \theta$.

(2) If $\mu \perp \nu$, describe the Radon-Nikodym derivative of μ with respect to θ.

Solution.

(1) If $\theta(A) = 0$, then $\mu(A) = \nu(A) = 0$. Thus $\mu \ll \theta$ and $\nu \ll \theta$.

(2) Let $S \in \mathcal{B}(\mathbb{R})$ be such that $\mu(S) = \nu(S^c) = 1$. Choose random variable $X = 2 \times \mathbf{1}_S$. Then

$$\begin{aligned}\int X \mathbf{1}_A \, \mathrm{d}\theta &= \int \mathbf{1}_A \mathbf{1}_S \, \mathrm{d}\mu + \int \mathbf{1}_A \mathbf{1}_S \, \mathrm{d}\nu \\ &= \mu(A \cap S) + \nu(A \cap S) \\ &= \mu(A).\end{aligned}$$

Thus $\mathrm{d}\mu = 2 \times \mathbf{1}_S \, \mathrm{d}\theta$. Similarly $\mathrm{d}\nu = 2 \times \mathbf{1}_{S^c} \, \mathrm{d}\theta$.

■

Problem 4. Decide true or false and justify. Take μ_i, ν_i to be probability measures on $(\Omega_i, \mathcal{F}_i)$.

(1) If $\mu_1 \otimes \mu_2 \ll \nu_1 \otimes \nu_2$, then $\mu_1 \ll \nu_1$ and $\mu_2 \ll \nu_2$.

(2) If $\mu_1 \ll \nu_1$ and $\mu_2 \ll \nu_2$, then $\mu_1 \otimes \mu_2 \ll \nu_1 \otimes \nu_2$.

Solution. Let $\mathfrak{X} = \mathfrak{X}_1 \otimes \mathfrak{X}_2$ for $\mathfrak{X} = \Omega, \mathcal{F}, \mu, \nu$.

(1) True. $\mu_1(A) = \mu(A \times \mathbb{R})$ and $\nu_1(A) = \nu(A \times \mathbb{R})$. Then

$$\nu_1(A) = 0 \iff \nu(A \times \mathbb{R}) = 0 \implies \mu(A \times \mathbb{R}) = 0 \iff \mu_1(A) = 0.$$

(2) True? If $A = A_1 \times A_2$ is a cylinder set, then

$$\nu(A) = 0 \implies \nu_1(A_1)\nu_2(A_2) = 0 \implies \mu_1(A_1)\mu_2(A_2) = 0 \implies \mu(A) = 0.$$

■

Problem 5.

(1) If $\mu_n \ll \nu$ for each n and $\mu_n \xrightarrow{\mathrm{d}} \mu$, then is it necessarily true that $\mu \ll \nu$? If $\mu_n \perp \nu$ for each n and $\mu_n \xrightarrow{\mathrm{d}} \mu$, then is it necessarily true that $\mu \perp \nu$? In either case, justify or give a counterexample.

(2) Suppose X, Y are independent (real-valued) random variables with distribution μ and ν respectively. If μ and ν are absolutely continuous with respect to Lebesgue measure, show that the distribution of $X+Y$ is also absolutely continuous with respect to Lebesgue measure.

Solution.

(1) It may be that $\mu \not\ll \nu$ even when $\mu_n \ll \nu$ for each n. Take $\mu_n = \delta_{\frac{1}{n}}$ and $\nu = \sum_{n=1}^{\infty} \frac{1}{2^n}\mu_n$. Then $\nu \gg \mu_n \xrightarrow{\mathrm{d}} \delta_0 \not\ll \nu$.

The same is a counterexample for singularity, only with $\nu = \delta_0$.

∎