MEASURE THEORETICAL PROBABILITY I HOMEWORK 4

ANDREAS PETROU-ZENIOU; CHENG PENG; TOMMENIX YU; CONROY LAU
STAT 38100
DUE FEB 8

Discussed with classmates.

Exercise 0.

Proof.

A cdf has at most countable discontinuity points:

Since a cdf function F is non-decreasing and is right continuous, we know that for each discontinuity point x_0 of F

$$\sup_{x \le x_0} F(x) < F(x_0)$$

and hence there exists some $q \in \mathbb{Q}$ such that $q \in \left(\sup_{x \le x_0} F(x), F(x_0) \right)$.

Now, since F is non-decreasing, thus for any two distinct discontinuity points $z < y$, we have $\sup_{x \le y} F(x) \ge F(z)$ and hence

$$\left(\sup_{x \le y} F(x), F(x_0) \right) \cap \left(\sup_{x \le z} F(x), F(x_0) \right) = \emptyset$$

which means that we cannot pick the same rational number corresponding to each discontinuity point, in the same manner as we pick q.

Hence, we've constructed a $1-1$ map from $A \to \mathbb{Q}$ where

$$A : \{x | F \text{ is discontinuous at } x\}$$

which in turn tells us

$$|A| \le |\mathbb{Q}|$$

thus A is at most countable.

The set of continuity points is dense in $\mathbb{R}$:

The set of continuity points, by definition, is $\mathbb{R} \backslash A$. But since A is only countable, for any $x \in \mathbb{R}$ and $\forall \varepsilon > 0$, $B_\varepsilon(x)$ contains more points than A.

This means that for any $x \in \mathbb{R}, \varepsilon > 0$ there are some point $y \in \mathbb{R} \backslash A$ such that $d(x, y) < \varepsilon$. In other words, $\mathbb{R} \backslash A$ is dense in $\mathbb{R}$.

$\square$

Exercise 1. *Prob 1.*

Proof.

$Y = aX + b$ for $a > 0 \quad \Rightarrow \mathrm{Corr}(X, Y) = 1$:

By computation (for any a):

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = \mathbb{E}[aX^2 + bX] - \mathbb{E}[X] \cdot \mathbb{E}[aX + b] \\ &= a\mathbb{E}[X^2] + b\mathbb{E}[X] - \mathbb{E}[X]\,(a\mathbb{E}[X] + b) \\ &= a\mathbb{E}[X^2] - a\mathbb{E}[X]^2 + b\mathbb{E}[X] - b\mathbb{E}[X] = a\,\mathrm{Var}(X)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(X)\,\mathrm{Var}(Y) &= \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) \cdot \left(\mathbb{E}[(aX + b)^2] - \mathbb{E}[aX + b]^2\right) \\ &= \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) \cdot \left(\mathbb{E}[a^2X^2 + 2abX + b^2] - (a\mathbb{E}[X] + b)^2\right) \\ &= \left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) \cdot \left(a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - b^2\right) \\ &= a^2\left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right)^2 = (a\,\mathrm{Var}(X))^2.\end{aligned}$$

Thus, if $a > 0$, then

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{a\,\mathrm{Var}(X)}{a\,\mathrm{Var}(X)} = 1$$

$Y = aX + b$ for $a < 0 \quad \Rightarrow \mathrm{Corr}(X, Y) = 1$:

For a similar computation we have

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{a\,\mathrm{Var}(X)}{\sqrt{a^2\,\mathrm{Var}(X)^2}} = \frac{a\,\mathrm{Var}(X)}{-a\,\mathrm{Var}(X)} = -1$$

$Y = aX + b$ for $a > 0 \quad \Leftarrow \mathrm{Corr}(X, Y) = 1$:

First we show that we can WLOG assume $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. We check this by checking that the only attributes we will use later ($\mathrm{Cov}(X, Y)$ and $\mathrm{Var}(X)$) is invariant under a shift by constant, i.e. it doesn't matter if we take $X = X - \mathbb{E}[X]$.

Let $\mathbb{E}[X] =: -a$ we have

$$\begin{aligned}\mathrm{Cov}(X + a, Y) &= \mathbb{E}[(X + a)Y] - \mathbb{E}[X + a]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] + a\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] - a\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathrm{Cov}(X, Y)\end{aligned}$$

and

$$\mathrm{Var}(X + a) = \mathbb{E}[X + a - \mathbb{E}[X + a]]^2 = \mathbb{E}[X - \mathbb{E}[X]]^2 = \mathrm{Var}(X)$$

which then validates our assertion that we can take $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.

Now, Since $\mathrm{Corr}(X,Y)=1$ we get

$$\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]=\sqrt{\left(\mathbb{E}[X^2]-\mathbb{E}[X]^2\right)\left(\mathbb{E}[Y^2]-\mathbb{E}[Y]^2\right)}$$
$$\Rightarrow \mathbb{E}[XY]=\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

Now we consider

$$\begin{aligned}\mathbb{E}[(tx+Y)^2]&=t^2\mathbb{E}[X^2]+2t\mathbb{E}[XY]+\mathbb{E}[Y^2]\\&=t^2\mathbb{E}[X^2]+2t\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}+\mathbb{E}[Y^2]\\&=\left(t\sqrt{\mathbb{E}[X^2]}+\sqrt{\mathbb{E}[Y^2]}\right)^2.\end{aligned}$$

Thus, there exists $t<0$ (if $t=0$ then conclusion holds trivially) such that

$$\mathbb{E}[(tX+Y)^2]=0$$

which means that $tX+Y=0$, or that $Y=-tX$.

Thus, now shifting back Y with it's original expectation we get

$$Y=aX+\mathbb{E}[Y]=aX+b$$

for $a>0$.

$Y=aX+b$ for $a<0 \quad \Leftarrow \mathrm{Corr}(X,Y)=-1$:

The only difference to the above argument is that

$$\mathbb{E}[XY]=-\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

which means in completing the squares we can take $t>0$, hence $Y=aX+b$ for $a=-t<0$.

$\square$

Exercise 2. *Prob 2.*

Proof.

X_n and X_m are independent for distinct n, m:

To prove the statement, it suffices to prove that $\sigma(X_n)$ and $\sigma(X_m)$ are independent. But note that since the range of X_n and X_m are $I := \{0, 1, \dots, 9\}$, a finite set, so we can partition all Borel sets in $\mathbb{R}$ into 2^{10} equivalent classes such that each contains a different portion of I. In other words, $\forall H \subset I$, define

$$E_H := \left\{ B \in \mathcal{B}(\mathbb{R}) \middle| H \subset B, I \backslash H \cap B = \emptyset \right\}$$

then we've defined an equivalence class on $\mathcal{B}(\mathbb{R})$, with 2^{10} classes. The fact that this is an equivalence class is clear by definition (either $0 \in C$ or not, and so on).

But note that $A \in \sigma(X_n)$ means that $A = X_n^{-1}(E_H)$ for $H \subset I$. For which we note that we can use the theorem (which is essentially $\pi - \lambda$) that if 2 collections of sets are independent, then so is their generated σ-algebra. That is, we define

$$A \in \mathcal{A} := \{\{0\}, \{1\}, \dots, \{9\}\}$$

and since $\sigma(\mathcal{A}) = I$, we know that if

$$S_n := \left\{ A \in \mathcal{A} \middle| X_n^{-1}(E_A) \right\}, \quad S_m := \left\{ A \in \mathcal{A} \middle| X_m^{-1}(E_A) \right\}$$

such that S_n and S_m are independent, then so is $\sigma(S_n) = \sigma(X_n)$ and $\sigma(S_m) = \sigma(X_m)$.

So it suffices us to prove that for any $N_A := X_n^{-1}(A \in \mathcal{A})$ and $M_B := X_m^{-1}(B \in \mathcal{A})$, they are independent. We can compute their probability using the translation invariance of Lebesgue measure.

WLOG we can assume that $m < n$, so $N_{\{k\}}$ is the set of numbers whose n-th digit is $k \in I$. Let

$$N_k := \left[\frac{k}{10^n}, \frac{k+1}{10^n} \right]$$

be an interval. Then we know from basic number theory (everything's finite here) that

$$N_{\{k\}} = \bigcup_{i=1}^{10^{n-1}} \left(\frac{i-1}{10^{n-1}} + N_k \right)$$

where the plus in side is just a translation by $\dfrac{i-1}{10^{n-1}}$. Now by translation invariant and disjoint additive we have

$$\lambda(N_{\{k\}}) = \lambda\left(\bigcup_{i=1}^{10^{n-1}} \left(\frac{i-1}{10^{n-1}} + N_k \right) \right) = 10^{n-1}\lambda(N_k) = 10^{n-1}\frac{1}{10^n} = \frac{1}{10}.$$

Similarly, we get $\lambda(M_{\{k\}}) = \dfrac{1}{10}$. So we compute $\lambda(N_{\{k\}} \cap M_{\{l\}})$, if the value is equal to the product of the measure of the two sets, which is $\dfrac{1}{100}$, then we are done.

Since $n > m$ we only need to count how many translation of N_k is there in the intersection. But this is just a finite problem: out of 10^{n-1} choices only 10^{n-2} of them are (by digit expression), so we can simply compute the measure of the intersection is just times translated times thhe measure of N_k:

$$\lambda(N_{\{k\}} \cap M_{\{l\}}) = 10^{n-2}\frac{1}{10^n} = \frac{1}{100} = \lambda(N_{\{k\}}) \cdot \lambda(M_{\{l\}}).$$

Thus, N_A and M_B are independent for any $A, B \in \mathcal{A}$, which means S_n and S_m are independent, which means X_n and X_m are independent.

$\{X_n\}$ is a sequence of independent random variables:

The only thing to do here is to

$\square$

Exercise 3. *Prob 3.*

Proof.

We need to show the following are equivalent:

(1) The tails of X satisfy

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{t^2}{K_1^2}\right)$$

for all $t \geq 0$.

(2) The moments of X satisfy

$$(\mathbb{E}[|X|^p])^{\frac{1}{p}} \leq K_2\sqrt{p}$$

for all $p \geq 1$.

(3) The moment generating function satisfies

$$\mathbb{E}[\exp(\lambda X)] \leq \exp(K_3^2\lambda^2)$$

for all $\lambda \in \mathbb{R}$.

<u>(1) $\Rightarrow$ (2):</u>

First of all, note that $\mathbb{E}[X] = 0$, this means that the integral $\int X d\mathbb{P}$ is well defined, i.e. at least one of X^+ or X^- is integrable. But since they have the same integral (difference is 0), they have to be both finite. Thus X is integrable. This means that X is essentially bounded, or that the essential supremum of X is finite, i.e. $X \leq M$ a.e.. (I do not believe this is true. e.g. normal distribution)

Now we want to show (2), so we try to bound

$$\begin{aligned}\mathbb{E}\left[|X|^p\right] = \int_\Omega |X|^p d\mathbb{P} &= \int_{|X|\geq t} |X|^p d\mathbb{P} + \int_{|X|<t} |X|^p d\mathbb{P} \\ &\leq M\int_{|X|\geq t} d\mathbb{P} + \int_{|X|<t} t^p d\mathbb{P} \leq 2Me^{-\frac{t^2}{K_1^2}} + t^p\end{aligned}$$

and hence we need to give a bound to the following expression

$$\frac{(\mathbb{E}[|X|^p])^{\frac{1}{p}}}{\sqrt{p}} \leq \frac{\left(2Me^{-\frac{t^2}{K_1^2}} + t^p\right)^{\frac{1}{p}}}{\sqrt{p}}$$

uniform in $p \geq 1$. Here note that K_1, M are fixed, and we can adjust t to get what we want. So let's just choose $t = 1$ and we bound the exponential by 1:

$$\frac{\left(2Me^{-\frac{t^2}{K_1^2}} + t^p\right)^{\frac{1}{p}}}{\sqrt{p}} \leq \frac{(2M+1)^{1/p}}{\sqrt{p}}$$

since

$$f(p) := \frac{(2M+1)^{1/p}}{\sqrt{p}}$$

is a continuous function, and $f \to 0$ as $p \to \infty$, we know that there is some N such that when $p \geq N$, $f(p) < \varepsilon < f(1)$. Thus to bound $f(p)$ from above we only need to bound $f(p)$ on $[1, N]$, which is a compact set and hence the maximum is attained. Let K_2 be this maximum then we've constructed a K_2 that makes (2) work.

<u>(2) $\Rightarrow$ (3):</u>

<u>(3) $\Rightarrow$ (1):</u>

$\square$

Exercise 4. *Prob 4.*

Proof. ∎

Exercise 5. *Prob 5.*

Proof.

□

Exercise 6. *Prob 6*

Proof. □