

Solution Set 10

Problem 1: Ky-Fan metric

Let X and Y be random variables defined on common probability space $(\Omega, \mathbb{F}, \mathbb{P})$. Define

$$d(X,Y) = \mathbb{E}\left(\log_2\left(1 + \frac{|X-Y|}{1+|X-Y|}\right)\right).$$

a) First, we would like to confirm that $d(X,Y)$ is a distance metric. Show that $d(X,Y)$ satisfies the triangle inequality. That is, $d(X,Z) \le d(X,Y) + d(Y,Z)$ for any X, Y, and Z.

Hint: the function $f(x) = \log_2(1+x)$ is sub-additive, e.g. $f(x+y) \le f(x) + f(y)$.

Next, we would like to check if convergence with respect to $d(X,Y)$ is equivalent to convergence in probability (a distance metric with this property is sometimes called a Ky-Fan metric).

b) Let $(X_n, n \ge 1)$ be sequence of random variables and X be another random variable, all defined on the same probability space $(\Omega, \mathbb{F}, \mathbb{P})$. Show that if $X_n \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} X$ then $\lim_{n\to\infty} d(X_n, X) = 0$.

c) Is the converse true? That is, if $\lim_{n\to\infty} d(X_n, X) = 0$ then $X_n \underset{n\to\infty}{\overset{\mathbb{P}}{\to}} X$. If yes, prove the statement. If no, provide a counter example.

Solution

a) For all $x, y, z \in \mathbb{R}$ we have

$$\begin{aligned}
\log_2\left(1 + \frac{|x-z|}{1+|x-z|}\right) &= \log_2\left(1 + \frac{|x-y+y-z|}{1+|x-y+y-z|}\right) \\
&\le \log_2\left(1 + \frac{|x-y|+|y-z|}{1+|x-y|+|y-z|}\right) \\
&\le \log_2\left(1 + \frac{|x-y|}{1+|x-y|} + \frac{|y-z|}{1+|y-z|}\right) \\
&\le \log_2\left(1 + \frac{|x-y|}{1+|x-y|}\right) + \log_2\left(1 + \frac{|y-z|}{1+|y-z|}\right)
\end{aligned}$$

where the first inequality follows from the fact that $\log_2(1+x)$ is an increasing function in x and the last inequality follows from the hint. Now, since the inequality holds for $X(\omega), Y(\omega), Z(\omega)$ for every $\omega \in \Omega$, we can take the expectation of both sides to get the desired result.

b) Fix $\epsilon > 0$ and note that convergence in probability implies that

$$\lim_{n\to\infty} \mathbb{P}(\{|X_n - X| \geq \epsilon\}) = 0.$$

For simplicity, define $g(x,y) = \log_2\left(1 + \frac{|x-y|}{1+|x-y|}\right)$. We can write

$$\begin{aligned} d(X_n, X) &= \mathbb{E}\left(g(X_n,X)1_{|X_n-X|\geq\epsilon|}\right) + \mathbb{E}\left(g(X_n,X)1_{|X_n-X|<\epsilon}\right) \\ &\leq \mathbb{E}\left(1_{|X_n-X|\geq\epsilon|}\right) + \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right) \\ &= \mathbb{P}\left(\{|X_n - X| \geq \epsilon\}\right) + \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right) \end{aligned}$$

Therefore

$$\lim_{n\to\infty} d(X_n, X) \leq \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right).$$

Since this is true for any ϵ, we can further take a limit as ϵ goes to zero to get the desired result.

c) Yes, the converse is also true. Fix $\epsilon > 0$ and define $\nu = \log_2\left(1+\frac{\epsilon}{1+\epsilon}\right)$. Then

$$\begin{aligned} \mathbb{P}\left(\{|X_n - X| \geq \epsilon\}\right) &= \nu \cdot \frac{1}{\nu}\mathbb{E}\left(1_{|X_n-X|\geq\epsilon|}\right) \\ &\leq \frac{1}{\nu}\mathbb{E}\left(g(X_n,X)1_{|X_n-X|\geq\epsilon}\right) \\ &\leq \frac{1}{\nu} d(X_n, X). \end{aligned}$$

Since for a fixed ϵ, ν is just a constant, we have that

$$\lim_{n\to\infty} \mathbb{P}\left(\{|X_n - X| \geq \epsilon\}\right) = \frac{1}{\nu}\lim_{n\to\infty} d(X_n, X) = 0.$$

Problem 2: Total variation distance

a) Let $\mathbb{P}(\{X = 1\}) = \mathbb{P}(\{X = -1\}) = \frac{1}{2}$ and $Y \sim \mathcal{N}(0,1)$. What is $d_{TV}(X,Y)$?

b) Let $(X_n, n \geq 1)$ be a sequence of random variables and X be another random variable on $(\Omega, \mathbb{F}, \mathbb{P})$. Show that if $\lim_{n\to\infty} d_{TV}(X_n, X) = 0$ then $X_n \underset{n\to\infty}{\overset{d}{\to}} X$.

c) Is the converse true? That is, if $X_n \underset{n\to\infty}{\overset{d}{\to}} X$ then $\lim_{n\to\infty} d_{TV}(X_n, X) = 0$. If yes, prove the statement. If no, provide a counter example.

Solution a) Letting $A = \{-1, 1\}$ we see that $d_{TV}(X,Y) = 1$, which is the maximum value possible. In other words, X and Y have essentially complementary supports (and this would be true for any pair of discrete and continuous random variables).

b) For any t for which $F_X(t)$ is continuous, we have

$$|F_{X_n}(t) - F_X(t)| = |\mathbb{P}(\{X_n \leq t\}) - \mathbb{P}(\{X \leq t\})| \leq \sup_{A\in\mathcal{B}(\mathbb{R})} |(\mathbb{P}(X_n \in A) - \mathbb{P}(X \in A))| = d_{TV}(X_n, X)$$

Thus, we see that $|F_{X_n}(t) - F_X(t)| \to 0$, and thus X_n converges in distribution to X.

c) No, the converse is not true. Consider Z_n iid with $\mathbb{P}(\{Z_1 = 1\}) = \mathbb{P}(\{Z_1 = -1\}) = \frac{1}{2}$ and define $S_n = Z_1 + Z_2 + \cdots + Z_n$. The random variable $X_n = \frac{S_n}{\sqrt{n}}$ converges to $X \sim \mathcal{N}(0,1)$ by CLT. However, X_n is a discrete random variable for any finite n. By an argument analogous to part a) we see that $d_{TV}(X_n, X) = 1$ for all n.

Problem 3: Convergence in L^p

a) Given a sequence of random variables $(X_n, n \geq 1)$, a random variable X, and $r \geq 1$, we say that X_n converges to X in rth mean (written $X_n \underset{n\to\infty}{\overset{L^r}{\to}} X$) if $\mathbb{E}(|X_n^r|) < \infty$ for all n and

$$\mathbb{E}(|X_n - X|^r) \underset{n\to\infty}{\to} 0.$$

Show that if $r > s \geq 1$ then,

$$X_n \underset{n\to\infty}{\overset{L^r}{\to}} X \Rightarrow X_n \underset{n\to\infty}{\overset{L^s}{\to}} X.$$

b) Suppose that $X_n \underset{n\to\infty}{\overset{L^1}{\to}} X$. Show that $\mathbb{E}(X_n) \underset{n\to\infty}{\to} \mathbb{E}(X)$. Is the converse true?

Solution

a) Assume that $r > s \geq 1$ and that

$$\mathbb{E}(|X_n - X|^r) \underset{n\to\infty}{\to} 0.$$

We can use Jensen's inequality to show

$$\mathbb{E}(|X_n - X|^s)^{\frac{r}{s}} \leq \mathbb{E}(|X_n - X|^r)$$

by applying the convex function $f(x) = x^{r/s}$. Then

$$\mathbb{E}(|X_n - X|^s)^{\frac{r}{s}} \leq \mathbb{E}(|X_n - X|^r) \underset{n\to\infty}{\to} 0 \Rightarrow \mathbb{E}(|X_n - X|^s)^{\frac{r}{s}} \underset{n\to\infty}{\to} 0 \Rightarrow \mathbb{E}(|X_n - X|^s) \underset{n\to\infty}{\to} 0.$$

And therefore $X_n \underset{n\to\infty}{\overset{L^s}{\to}} X$.

b) We have that

$$|\mathbb{E}(X_n) - \mathbb{E}(X)| = |\mathbb{E}(X_n - X)| \leq \mathbb{E}(|X_n - X|) \underset{n\to\infty}{\to} 0.$$

Therefore $\mathbb{E}(X_n) \underset{n\to\infty}{\to} \mathbb{E}(X)$.

The converse is not true. Consider the sequence $(X_n, n \geq 1)$ of i.i.d. Bernoulli(p) random variables, with $0 < p < 1$. Then $\mathbb{E}(X_n) \underset{n\to\infty}{\to} \mathbb{E}(X)$. However

$$\mathbb{E}(|X_n - X|) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}.$$

This does not converge to zero as n goes to infinity.

Problem 4: Bernoulli sums

Let $\lambda > 0$ be fixed. For a given $n \geq \lceil \lambda \rceil$, let $X_1^{(n)}, \ldots, X_n^{(n)}$ be i.i.d. Bernoulli (λ/n) random variables and let $S_n = X_1^{(n)} + \ldots + X_n^{(n)}$.

a) Compute $\mathbb{E}(S_n)$ and $\mathrm{Var}(S_n)$ for a fixed value of $n \geq \lceil \lambda \rceil$.

b) Deduce the value of $\mu = \lim_{n\to\infty} \mathbb{E}(S_n)$ and $\sigma^2 = \lim_{n\to\infty} \mathrm{Var}(S_n)$.

c) Compute the limiting distribution of S_n (as $n \to \infty$).

Hint: Use characteristic functions. You might also have a look at tables of characteristic functions of some well known distributions in order to solve this exercise.

For a given $n \geq 1$, let now $Y_1^{(n)}, \ldots, Y_n^{(n)}$ be i.i.d. Bernoulli $(1/n)$ random variables and let

$$T_n = Y_1^{(n)} + \ldots + Y_{\lceil \lambda n \rceil}^{(n)}$$

where $\lambda > 0$ is the same as above.

d) Compute the limiting distribution of T_n (as $n \to \infty$).

e) Is it also the case that either S_n or T_n converge almost surely or in probability towards a limit? Justify your answer!

Solution a) let us compute $\mathbb{E}(S_n) = \sum_{j=1}^n \mathbb{E}(X_j^{(n)}) = n\,\frac{\lambda}{n} = \lambda$ and

$$\mathrm{Var}(S_n) = \sum_{j=1}^n \mathrm{Var}(X_j^{(n)}) = n\,\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) = \lambda - \frac{\lambda^2}{n}$$

b) So $\mu = \lim_{n\to\infty} \mathbb{E}(S_n) = \lambda$ and $\sigma^2 = \lim_{n\to\infty} \mathrm{Var}(S_n) = \lambda$.

c) Let us compute the characteristic function of S_n:

$$\begin{aligned}
\phi_{S_n}(t) &= \mathbb{E}(\exp(itS_n)) = \mathbb{E}(\exp(it\,(X_1^{(n)} + \ldots + X_n^{(n)}))) = \mathbb{E}(\exp(itX_1^{(n)}))\cdots\mathbb{E}(\exp(itX_n^{(n)})) \\
&= \left(\mathbb{E}(\exp(itX_1^{(n)}))\right)^n = \left(e^{it}\,\frac{\lambda}{n} + 1 - \frac{\lambda}{n}\right)^n = \left(1 + \frac{\lambda\,(e^{it}-1)}{n}\right)^n \underset{n\to\infty}{\to} \exp\left(\lambda\,(e^{it}-1)\right)
\end{aligned}$$

This limiting function is the characteristic function of $Z \sim \mathcal{P}(\lambda)$. Indeed, one can check that

$$\phi_Z(t) = \mathbb{E}(\exp(itZ)) = \sum_{k\geq 0} e^{itk}\,\frac{\lambda^k\,e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k\geq 0}\frac{(\lambda\,e^{it})^k}{k!} = \exp(\lambda\,(e^{it}-1))$$

which allows us to conclude that $S_n \underset{n\to\infty}{\overset{d}{\to}} Z$.

d) The computation of the characteristic function is similar here:

$$\mathbb{E}\left(e^{itT_n}\right) = \left(\frac{1}{n}\,e^{it} + \left(1 - \frac{1}{n}\right)\right)^{\lceil \lambda n \rceil} = \left(1 + \frac{1}{n}\,(e^{it}-1)\right)^{\lceil \lambda n \rceil} \underset{n\to\infty}{\to} \exp(\lambda(e^{it}-1))$$

and leads actually exactly to the same result: T_n converges in distribution towards a Poisson random variable Z of parameter λ.

e) No, as each random variable S_n is constructed from a different set of random variables $X_1^{(n)}, \ldots, X_n^{(n)}$, which depends on n. The same holds for the random variables T_n.

Problem 5: The game

Someone proposes to you the following game: start with an initial amount of $S_0 > 0$ francs, of your choice. Then toss a coin: if it falls on heads, you win $S_0/2$ francs; while if it falls on tails, you lose $S_0/2$

francs. Call S_1 your amount after this first coin toss. Then the game goes on, so that your amount after coin toss number $n \geq 1$ is given by

$$S_n = \begin{cases} S_{n-1} + \frac{S_{n-1}}{2} & \text{if coin number } n \text{ falls on heads} \\ \\ S_{n-1} - \frac{S_{n-1}}{2} & \text{if coin number } n \text{ falls on tails} \end{cases}$$

We assume moreover that the coin tosses are independent and fair, i.e., with probability $1/2$ to fall on each side. Nevertheless, you should *not* agree to play such a game: explain why!

Hints:

First, to ease the notation, define $X_n = +1$ if coin n falls on heads and $X_n = -1$ if coin n falls on tails. That way, the above recursive relation may be rewritten as $S_n = S_{n-1}\,(1 + \frac{X_n}{2})$ for $n \geq 1$.

a) Compute recursively $\mathbb{E}(S_n)$; if it were only for expectation, you could still consider playing such a game, but...

b) Define now $Y_n = \log(S_n/S_0)$, and use the central limit theorem to approximate $\mathbb{P}(\{Y_n > t\})$ for a fixed value of $t \in \mathbb{R}$ and a relatively large value of n. Argue from there why it is definitely not a good idea to play such a game! (computing for example an approximate value of $\mathbb{P}(\{S_{100} > S_0/10\})$)

Solution

a) Let us compute first

$$\mathbb{E}(S_1) = \frac{1}{2}\left(\frac{3S_0}{2} + \frac{S_0}{2}\right) = S_0$$

Assuming now that $\mathbb{E}(S_n) = S_0$ (more precisely, that the expectation stays constant over n coin tosses), let us compute $\mathbb{E}(S_{n+1})$:

$$\begin{aligned}\mathbb{E}(S_{n+1}) &= \mathbb{E}(S_{n+1}|\{X_1 = +1\})\,\mathbb{P}(\{X_1 = +1\}) + \mathbb{E}(S_{n+1}|\{X_1 = -1\})\,\mathbb{P}(\{X_1 = -1\}) \\ &= \frac{1}{2}\left(\mathbb{E}(S_{n+1}|\{S_1 = \frac{3S_0}{2}\}) + \mathbb{E}(S_{n+1}|\{S_1 = \frac{S_0}{2}\})\right) = \frac{1}{2}\left(\frac{3S_0}{2} + \frac{S_0}{2}\right) = S_0\end{aligned}$$

Note: The computation is slightly unorthodox here, but we will see a cleaner way to prove this later in the course.

b) Y_n is the sum of n i.i.d. random variables, as the following computation shows:

$$Y_n = \log\left(\frac{S_n}{S_0}\right) = \log\left(\prod_{j=1}^{n}\left(1 + \frac{X_j}{2}\right)\right) = \sum_{j=1}^{n}\log\left(1 + \frac{X_j}{2}\right)$$

and these random variables are bounded, so by the central limit theorem,

$$\frac{Y_n - n\,\mu}{\sqrt{n}\,\sigma} \underset{n\to\infty}{\overset{d}{\to}} Z \sim \mathcal{N}(0,1)$$

where $\mu = \mathbb{E}(\log(1 + X_1/2)) = \frac{1}{2}\,(\log(3/2) + \log(1/2)) \simeq -0.144$ and

$$\sigma^2 = \mathrm{Var}(\log(1 + X_1/2)) = \frac{1}{2}\,(\log(3/2)^2 + \log(1/2)^2) - \mu^2 \simeq 0.3$$

This is saying that for large n, we have

$$Y_n \simeq -0.144n + \sqrt{0.26n}\,Z \quad \text{in particular:} \quad Y_{100} \simeq -14.4 + 5.4\,Z$$

Therefore

$$\begin{aligned}\mathbb{P}(\{S_{100} > S_0/10\}) &= \mathbb{P}(\{S_{100}/S_0 > 1/10\}) = \mathbb{P}(\{Y_{100} > -\log(10)) \\ &\simeq \mathbb{P}\left(\left\{Z > \frac{-2.3+14.4}{5.4}\right\}\right) = \mathbb{P}(\{Z > 2.24\})\end{aligned}$$

which is roughly 1% (so you can imagine what $\mathbb{P}(\{S_{100} > S_0\})$ looks like ...).

Therefore, the process $(S_n,\, n \geq 1)$, unexpectedly perhaps, "crashes" to zero with high probability as n gets large, even though it seemed a priori a "fair game" with constant expectation. This is an important example among a large class of processes called "martingales"; we will come back to this!

Note: The random process $(S_n,\, n \geq 1)$ is not unrelated to the following *deterministic* process defined recursively as

$$x_0 \in \mathbb{N}^*, \quad x_{n+1} = \begin{cases} x_n/2 & \text{if } x_n \text{ is even} \\ 3x_n + 1 & \text{if } x_n \text{ is odd} \end{cases}$$

in which an even number gets multiplied by $1/2$ and an odd number gets approximately multiplied by $3/2$ (because it first gets multiplied by 3 and then necessarily divided by 2, as $3x_n + 1$ is even). So if you consider that even and odd numbers appear naturally with probability $1/2$, then the two processes have something in common. But in the deterministic case, one has no proof that the process ultimately reaches the value 1 as n gets large: this is the famous Collatz conjecture, which remains unsolved until now.

Problem 6: The birthday problem

Let $(X_n, n \geq 1)$ be a sequence of i.i.d. random variables, each uniform on $\{1, \ldots, N\}$. Let also

$$T_N = \min\{n \geq 1 \,:\, X_n = X_m \text{ for some } m < n\}$$

(notice that whatever happens, $T_N \in \{2, \ldots, N+1\}$).

a) Show that

$$\mathbb{P}\left(\left\{\frac{T_N}{\sqrt{N}} \leq t\right\}\right) \underset{N\to\infty}{\to} 1 - e^{-t^2/2}, \quad \forall t \geq 0$$

Remarks:
- Approximations are allowed here!

- Please observe that the limit distribution is *not* the Gaussian distribution!

b) *Numerical application:* Use this to obtain a rough estimate of $\mathbb{P}(\{T_{365} \leq 22\})$ and $\mathbb{P}(\{T_{365} \leq 50\})$ (i.e., what is the probability that among 22 / 50 people, at least two share the same birthday?)

Solution

a) Let $k \geq 1$ and note that

$$\{T_N > k\} = \{\text{first } k \text{ draws are all distinct}\}.$$

Hence

$$\mathbb{P}(T_N > k) = \frac{N(N-1)\cdots(N-k+1)}{N^k} = \prod_{j=0}^{k-1}\left(1 - \frac{j}{N}\right).$$

We set $k = k_N = \lfloor t\sqrt{N} \rfloor$ for a fixed $t \geq 0$. We take logarithm on both sides and use the approximation $\log(1-x) = -x + O(x^2)$ as $x \to 0$:

$$\begin{aligned}
\log \mathbb{P}(T_N > k_N) &= \sum_{j=0}^{k_N-1} \log\left(1 - \frac{j}{N}\right) \\
&= \sum_{j=0}^{k_N-1} \left(-\frac{j}{N} + O\Big(\frac{j^2}{N^2}\Big)\right) \\
&= -\frac{1}{N}\sum_{j=0}^{k_N-1} j + O\left(\frac{1}{N^2}\sum_{j=0}^{k_N-1} j^2\right).
\end{aligned}$$

We know the sums are

$$\sum_{j=0}^{k_N-1} j = \frac{k_N(k_N-1)}{2} \quad \text{and} \quad \sum_{j=0}^{k_N-1} j^2 = O(k_N^3).$$

Thus,

$$\log \mathbb{P}(T_N > k_N) = -\frac{k_N(k_N-1)}{2N} + O\Big(\frac{k_N^3}{N^2}\Big).$$

We insert $k_N \sim t\sqrt{N}$ as $N \to \infty$:

$$\begin{aligned}
\frac{k_N(k_N-1)}{2N} &\to \frac{t^2}{2}, \\
\frac{k_N^3}{N^2} &= O(N^{-1/2}) \to 0.
\end{aligned}$$

Hence,

$$\mathbb{P}\left(\frac{T_N}{\sqrt{N}} > t\right) = \mathbb{P}(T_N > k_N) \underset{N\to\infty}{\to} e^{-t^2/2},$$

which implies

$$\mathbb{P}\left(\frac{T_N}{\sqrt{N}} \leq t\right) \underset{N\to\infty}{\to} 1 - e^{-t^2/2}, \qquad t \geq 0.$$

(b) For $N = 365$ we approximate:

$$\mathbb{P}(T_{365} \leq k) \approx 1 - \exp\Big(-\tfrac{1}{2}\big(\tfrac{k}{\sqrt{365}}\big)^2\Big).$$

For $k = 22$:

$$\begin{aligned}
t = \frac{22}{\sqrt{365}} \approx 1.152, &\qquad \frac{t^2}{2} \approx 0.664, \\
\mathbb{P}(T_{365} \leq 22) \approx 1 - e^{-0.663} &\approx 0.485.
\end{aligned}$$

For $k = 50$:

$$\begin{aligned}
t = \frac{50}{\sqrt{365}} \approx 2.617, &\qquad \frac{t^2}{2} \approx 3.424, \\
\mathbb{P}(T_{365} \leq 50) \approx 1 - e^{-3.425} &\approx 0.967.
\end{aligned}$$