Problem 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, $d(X,Z) \leq d(X,Y) + d(Y,Z)$ for any X, Y, and Z.

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

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 $(X_n, n \geq 1)$ 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.

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.

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?

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!

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 $\mathbb{P}(\{S_{100} > S_0/10\})$)

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?)