

Assignment 2

Due date: Sept. 24, 11:59 pm.

Instructor: Daniel McDonald and Benjamin Bloem-Reddy

Instructions

Academic integrity policy: I encourage you to discuss verbally with other students about the assignment. However, you **must** write your answers by yourself. If you discuss the assignment with anyone (a classmate or anyone else), you must say so at the top of your solutions. Use of AI tools to produce full or partial answers to these questions or to related questions is strictly prohibited. This includes tools like Co-pilot, auto-complete, etc. Also, refrain from looking at answer keys from other schools, previous years, or the Math Stack Exchange. For more information, see: the syllabus or contact the instructor.

1. Solutions may be handwritten or typeset with L^AT_EX (see the course website for a template). It must be legible. Difficult-to-read solutions will be marked down or ignored.
2. *Justify your answers formally. (Err on the side of too much detail so that the teaching team knows exactly what you're doing and why.)*
3. Submit your work by 11:59 pm (Vancouver time) on the due date to Canvas (.pdf).
4. For questions that require computer simulations, you may use any language from the list of (R, Julia, Python), with level of support provided by the teaching team decreasing further in the list. Please include in your PDF submission the relevant plots generated by your simulation code, and append your code (in PDF) to your submission. Please retain the code you used, as you may be asked for it if required.

Questions

1. [Wasserman \(2004\)](#) 4.3.

Solution: We have that $X_1, \dots, X_n$ are iid $\text{Bern}(p)$, so $\mathbb{E}[X_i] = p$ and $\text{Var}(X_i) = p(1-p)$ for all i . The sample mean is $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, so

$$\mathbb{E}[\bar{X}_n] = p \quad \text{and} \quad \text{Var}(\bar{X}_n) = \frac{p(1-p)}{n}.$$

Using Chebyshev's inequality,

$$\mathbb{P}(|\bar{X}_n - p| > \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}.$$

Using Hoeffding's inequality,

$$\mathbb{P}(|\bar{X}_n - p| > \epsilon) \leq 2 \exp(-2n\epsilon^2).$$

When n is large, the exponential term in Hoeffding's inequality decays much faster than the term in Chebyshev's inequality, so the bound from Hoeffding's inequality is smaller.

2. [Wasserman \(2004\)](#) 5.16.

Solution: Let $X_n \sim N(0, 1)$ and $Y_n = -X_n$ for all n . Then $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ in distribution, where $X \sim N(0, 1)$ and $Y \sim N(0, 1)$. However, $X_n + Y_n = 0$ for all n , while $X + Y \sim N(0, 2)$. So, $X_n + Y_n \not\rightsquigarrow X + Y$.

3. [van der Vaart \(1998\)](#) 2.6. (Just consider the case of random variables, rather than random vectors.)

Solution: Since $X_n \rightsquigarrow X$, we have that the CDF $F_{X_n}(x) \rightarrow F_X(x)$ as $n \rightarrow \infty$. Similarly, $F_{Y_n}(y) \rightarrow F_Y(y)$ as $n \rightarrow \infty$.

Because X_n and Y_n are independent for each n , their joint CDF is the product of their individual CDFs:

$$F_{(X_n, Y_n)}(x, y) = F_{X_n}(x)F_{Y_n}(y).$$

Taking the limit as $n \rightarrow \infty$, we get:

$$\lim_{n \rightarrow \infty} F_{(X_n, Y_n)}(x, y) = \lim_{n \rightarrow \infty} (F_{X_n}(x)F_{Y_n}(y)) = \left(\lim_{n \rightarrow \infty} F_{X_n}(x)\right)\left(\lim_{n \rightarrow \infty} F_{Y_n}(y)\right) = F_X(x)F_Y(y).$$

The right-hand side is the joint CDF of two independent random variables X and Y . Therefore,

$$F_{(X, Y)}(x, y) = F_X(x)F_Y(y).$$

This shows that $(X_n, Y_n) \rightsquigarrow (X, Y)$ where X and Y are independent.

4. (a) [van der Vaart \(1998\)](#) 2.10.

(b) Do the same for $o_{\mathbb{P}}(1)O_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$.

(c) Explain carefully why $o_{\mathbb{P}}(1) + O_{\mathbb{P}}(1) \neq o_{\mathbb{P}}(1)$: give a counterexample, or show where your argument for (a) fails.

Solution: Let X_n and Y_n be random vectors such that $X_n = o_{\mathbb{P}}(1)$ and $Y_n = O_{\mathbb{P}}(1)$, and let $\|\cdot\|$ denote a norm on the space of random vectors.

Since $X_n = o_{\mathbb{P}}(1)$, for every $\epsilon > 0$ and any $\delta > 0$, $\mathbb{P}(\|X_n\| > \epsilon) < \delta$ for all sufficiently large n .

Since $Y_n = O_{\mathbb{P}}(1)$, for every $\delta > 0$, there exists a constant $M(\delta) > 0$ such that $\mathbb{P}(\|Y_n\| > M(\delta)) < \delta$ for all sufficiently large n .

(a) First, we want to show that $X_n + Y_n = O_{\mathbb{P}}(1)$.

By the union bound (Boole's inequality), we have for any $a, b > 0$:

$$\mathbb{P}(\|X_n + Y_n\| > a + b) \leq \mathbb{P}(\|X_n\| > a) + \mathbb{P}(\|Y_n\| > b).$$

Because $X_n = o_{\mathbb{P}}(1)$, we can choose a large enough such that $\mathbb{P}(\|X_n\| > a) < \delta/2$ for sufficiently large n . Similarly, since $Y_n = O_{\mathbb{P}}(1)$, we can choose $b = M(\delta/2)$ such that $\mathbb{P}(\|Y_n\| > b) < \delta/2$ for sufficiently large n . Therefore, we have:

$$\mathbb{P}(\|X_n + Y_n\| > a + b) < \delta/2 + \delta/2 = \delta$$

for sufficiently large n . This shows that $X_n + Y_n = O_{\mathbb{P}}(1)$.

(b) Let $Z_n = X_n Y_n$. Fix $\epsilon > 0$ and $\delta > 0$; we will show $\mathbb{P}(\|X_n Y_n\| > \epsilon) < \delta$ for all n sufficiently large.

Since $Y_n = O_{\mathbb{P}}(1)$, we can choose $M(\delta) < \infty$ and N_1 such that

$$\mathbb{P}(\|Y_n\| > M(\delta)) < \frac{\delta}{2} \quad \text{for all } n \geq N_1.$$

Since $X_n = o_p(1)$, apply the definition with the specific threshold $\epsilon/M > 0$: there exists N_2 such that

$$\mathbb{P}\left(\|X_n\| > \frac{\epsilon}{M}\right) < \frac{\delta}{2} \quad \text{for all } n \geq N_2.$$

We claim that for every n ,

$$\{\|X_n Y_n\| > \epsilon\} \subseteq \left\{\|X_n\| > \frac{\epsilon}{M}\right\} \cup \{\|Y_n\| > M\}.$$

To see this, suppose ω lies in the complement of the right-hand side, i.e. $\|X_n(\omega)\| \leq \epsilon/M$ **and** $\|Y_n(\omega)\| \leq M$. Then

$$\|X_n(\omega)Y_n(\omega)\| = \|X_n(\omega)\| \|Y_n(\omega)\| \leq \frac{\epsilon}{M} \cdot M = \epsilon,$$

so $\omega \notin \{\|X_n Y_n\| > \epsilon\}$. Taking complements gives the claimed inclusion.

For $n \geq N := \max(N_1, N_2)$, monotonicity and subadditivity of probability give

$$P(\|X_n Y_n\| > \epsilon) \leq P\left(\|X_n\| > \frac{\epsilon}{M}\right) + P(\|Y_n\| > M) < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Since $\delta > 0$ was arbitrary, this shows

$$P(\|X_n Y_n\| > \epsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and since $\epsilon > 0$ was also arbitrary, $X_n Y_n = o_p(1)$.

- (c) Let $X_n = 0$ and $Y_n \sim \mathbf{N}(0, 1)$ for all n . Then $X_n = o_{\mathbb{P}}(1)$ and $Y_n = O_{\mathbb{P}}(1)$, but $X_n + Y_n = Y_n \sim \mathbf{N}(0, 1)$, which is not $o_{\mathbb{P}}(1)$.

5. [van der Vaart \(1998\)](#) 2.11. (This problem highlights a subtle defect in the stochastic o -notation, and is something to keep in mind going forward.)

Solution: The issue here is partially one of reading math like reading a sentence, left to right. When we say $o_{\mathbb{P}}(1) = O_{\mathbb{P}}(1)$, we mean that the any sequence of random variables that is $o_{\mathbb{P}}(1)$ is also $O_{\mathbb{P}}(1)$. But it is of course not true that any sequence that is $O_{\mathbb{P}}(1)$ is also $o_{\mathbb{P}}(1)$.

Writing $X_n = o_{\mathbb{P}}(1)$ is a shorthand for saying that the sequence of random variables X_n is among the set of all sequences that are $o_{\mathbb{P}}(1)$. That is, **the notation $o_{\mathbb{P}}(1)$ is a set**, not an object. Building on this, some authors are more careful and write $X_n \in o_{\mathbb{P}}(1)$ rather than $X_n = o_{\mathbb{P}}(1)$ because it is more precise. Using this notation, we can write $o_{\mathbb{P}}(1) \subset O_{\mathbb{P}}(1)$ to indicate that every sequence that is $o_{\mathbb{P}}(1)$ is also $O_{\mathbb{P}}(1)$.

6. Let X be a non-negative random variable. Show that if, for some $p > 0$, $\mathbb{E}[X^{-p}] < \infty$, then $\mathbb{P}(X = 0) = 0$. (Hence, the distribution of X does not have an atom at 0.) Does the Gamma distribution have an atom at zero when its parameters are positive?

Solution: Let $M > 0$, and suppose that $\mathbb{E}[X^{-p}] < \infty$ for some $p > 0$. Then we can write

$$\{X^{-p} \geq M\} = \{X \leq M^{-1/p}\} = \{0 < X \leq M^{-1/p}\} \cup \{X = 0\}.$$

Note that on the event $\{0 < X \leq M^{-1/p}\}$, we have $X^{-p} \geq M$. Therefore,

$$\infty > \mathbb{E}[X^{-p}] \geq \mathbb{E}[M \mathbf{1}\{X^{-p} > M\}] = \mathbb{E}[M \mathbf{1}\{X < M^{-1/p}\}] \geq M \mathbb{P}(X < M^{-1/p}). \quad (1)$$

As $M \rightarrow \infty$, we have $M^{-1/p} \rightarrow 0$, and so $\mathbb{P}(X < M^{-1/p}) \rightarrow \mathbb{P}(X = 0)$ by (downward) continuity of probability.

Now we set up the contradiction. Assume $\mathbb{P}(X = 0) = c > 0$. Then, choose M large enough so that $\mathbb{P}(X < M^{-1/p}) > c/2$. But this implies that $\mathbb{E}[X^{-p}] \geq Mc/2$ by (1). This is a contradiction, because we can take $M \rightarrow \infty$ and make the right-hand side arbitrarily large, while the left-hand side is finite. Therefore, we must have $\mathbb{P}(X = 0) = 0$.

If $X \sim \text{Gamma}(\alpha, \beta)$ with $\alpha, \beta > 0$, then the density of X is

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$

And we have that $\mathbb{E}[X^{-p}] = \int_0^\infty x^{-p} f_X(x) dx$. For this integral to be finite, we need $\alpha - 1 - p > -1$, or equivalently $\alpha > p$. Since $\alpha > 0$, we can always choose p such that $0 < p < \alpha$, and thus $\mathbb{E}[X^{-p}] < \infty$. Therefore, by the previous argument, we conclude that $\mathbb{P}(X = 0) = 0$, and the Gamma distribution does not have an atom at zero when its parameters are positive.

References

- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.
- Larry Wasserman. *All of Statistics*. Springer, New York, 2004.