1. Question: Basic Inequalities (*very elementary*)

Let X be an exponential random variable with parameter $\lambda = 12$, i.e. with density

$$f_X(x) = \begin{cases} 12e^{-12x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

1.1. Use Markov's inequality to find an upper bound for $P(X > 6)$.

1.2. Use Chebyshev's inequality to find an upper bound for $P(X > 6)$.

1.3. Explicitly compute the probability above and compare with the upper bounds you derived.

Solution:

1.1.

$$P(X > 6) \leq \frac{E[X]}{6} = \frac{1}{72}.$$

1.2. From Chebyshev's inequality, we have

$$P(|X - E[X]| > t) \leq \frac{\text{Var}(X)}{t^2}.$$

Then

$$P(X > 6) \leq P\left(\left|X - \frac{1}{12}\right| > 6 - \frac{1}{12}\right) \leq \frac{\text{Var}(X)}{(71/12)^2} = \frac{1}{71^2}.$$

1.3.

$$P(X > 6) = \int_6^\infty 12e^{-12x}dx = 1 - e^{-12x}\Big|_6^\infty = e^{-72} \approx 5.3801862e - 32.$$

We notice that the exact probability is much smaller than the upper bounds we derived in **1.1** and **1.2**, but that is not completely fine! The nature of upper bounds is to find expressions that will *always* be larger that what they are bounding, even in "extreme" cases.

2. Question: Transformations of Several Random Variables (*elementary*)

2.1. Let X and Y be independent random variables with cumulative distribution functions F_X and F_Y, respectively. Show that the cumulative distribution function of $X + Y$ is

$$F_{X+Y}(t) = \int F_Y(t - x)d\mathbb{P}_X(x). \qquad (\star)$$

2.2. The concept of **2.1** is also referred to as *convolution*. Specifically write out the pmf and pdf of $X + Y$ when X and Y are discrete and continuous RVs, respectively.
Hint: You may use that in ($\star$), integration and differentiation are interchangeable by the dominated convergence theorem and mean value theorem.

2.3. Let X be a uniform distribution on $[0, 1]$, i.e. $f_X(x) = \frac{1}{1-0}\mathbb{1}_{x\in[0,1]}$, and Y be a uniform distribution on $[1, 2]$, i.e. $f_Y(x) = \frac{1}{2-1}\mathbb{1}_{x\in[1,2]}$. Find f_Z for $Z := X + Y$.

2.4. Determine the cdf of the random variable $Z := \min\{X, Y\}$ for independent random variables X and Y. What does the pdf look like if X and Y are continuous?

Solution:

2.1. Note that

$$\begin{aligned} F_{X+Y}(t) &= \int_{x+y\leq t} d\mathbb{P}_X(x)d\mathbb{P}_Y(y) \\ &= \int \left(\int_{y\leq t-x} d\mathbb{P}_Y(y) \right) d\mathbb{P}_X(x) \\ &= \int F_Y(t-x)d\mathbb{P}_X(x), \end{aligned}$$

where the second equality follows from Fubini's theorem.

2.2. For continuous X and Y:

$$\begin{aligned} f_{X+Y}(t) &= \frac{\partial}{\partial t} F_{X+Y}(t) \\ &= \frac{\partial}{\partial t} \int F_Y(t-x)d\mathbb{P}_X(x) \overset{\text{Hint}}{=} \int_{x\in\Omega_X} \frac{\partial}{\partial t}\Big(f_X(x)F_Y(t-x)\Big)dx \overset{(*)}{=} \int_{x\in\Omega_X} f_X(x)f_Y(t-x)dx \end{aligned}$$

Where $(*)$ follows from the chain rule:

$$\frac{\partial}{\partial t}F_Y(t-x) = f_Y(t-x)\cdot\frac{\partial}{\partial t}(t-x) = f_Y(t-x)\cdot 1 = f_Y(t-x).$$

For discrete X and Y, $(\star)$ of course still holds, but since the pmf is not defined as the derivative of the cdf, getting from the formula for $F_{X+Y}(t)$ to the formula for $p_{X+Y}(t)$ is not straightforward. So, we take the direct approach:

$$\begin{aligned} p_{X+Y}(t) &= \sum_{x\in\Omega_X} \mathbb{P}(X=x, Y=t-x) \overset{\text{Independence}}{=} \sum_{x\in\Omega_X} \mathbb{P}(X=x)\mathbb{P}(Y=t-x) \\ &= \sum_{x\in\Omega_X} p_X(x)p_Y(t-x). \end{aligned}$$

Note that this corresponds nicely to the following result of $(\star)$ in the discrete case:

$$F_{X+Y}(t) = \sum_{x\in\Omega_X} p_X(x)F_Y(t-x).$$

2.3. Using the formula from **2.1.**, or rather **2.2**,

$$\begin{aligned} f_Z(t) &= \int_0^1 f_X(x)f_Y(t-x)dx \\ &= \int_0^1 f_Y(t-x)dx \quad = \int_0^1 1_{t-x\in[1,2]}dx = \int_0^1 1_{x\in[t-2,t-1]}dx \end{aligned}$$

The last integral follows from the following: $t-x\in[1,2]$ when $2\geq t-x\geq 1$, which implies $t-1\geq x\geq t-2$. When $t\in[1,2], x$ can only be in $[t-2,t-1]\cap[0,1]$ if $x\in[0,t-1]$.

Therefore, we get

$$f_Z(t) = \begin{cases} t-1 & t\in[1,2] \\ 3-t & t\in[2,3] \end{cases}.$$

2.4.

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z\leq z) \\ &= 1-\mathbb{P}(Z>z) \\ &= 1-\mathbb{P}(\min\{X,Y\}>z) \\ &= 1-\mathbb{P}(X>z, Y>z) \\ \text{because of independence} \quad &= 1-(1-F_X(z))(1-F_Y(z)) \\ &= F_X(z)+F_Y(z)-F_X(z)F_Y(z). \end{aligned}$$

If f_X, f_Y are the densities of continuous X, Y with cdfs $F_X(z)$, $F_Y(z)$, then taking the derivative yields:

$$f_X(z) + f_Y(z) - f_X(z)F_Y(z) - F_X(z)f_Y(z).$$

3. Question: Convergence of Random Variables (*elementary*)

3.1. Consider a sequence of random variables $(X_n : n \in \mathbb{N})$ such that $X_n \xrightarrow{L^p} X$, then $X_n \xrightarrow{P} X$.

3.2. Let $X_1, \ldots, X_n$ be IID with finite mean $\mu = \mathbb{E}(X_1)$ and finite variance $\sigma^2 = \mathbb{V}(X_1)$. Let $\bar{X}_n$ be the sample mean and let S_n^2 be the sample variance.

(i) Show that $\mathbb{E}[\bar{X}] = \mu$ and $\mathbb{E}\left(S_n^2\right) = \sigma^2$. (*You may use that* $E\left(\bar{X}^2\right) = \frac{\sigma^2}{n} + \mu^2$.)

(ii) Show that $S_n^2 \xrightarrow{\text{P}} \sigma^2$. *Hint: Show that* $S_n^2 = c_n n^{-1} \sum_{i=1}^n X_i^2 - d_n \bar{X}_n^2$ *where* $c_n \to 1$ *and* $d_n \to 1$. *Apply the law of large numbers to* $n^{-1} \sum_{i=1}^n X_i^2$ *and to* $\bar{X}_n$. *Then use part (e) of Theorem 11.1.*

3.3. Let $X_1, X_2, \ldots$ be a sequence of random variables such that

$$\mathbb{P}\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2} \quad \text{and} \quad \mathbb{P}(X_n = n) = \frac{1}{n^2}$$

Does X_n converge in probability? Does X_n converge in L^2?

3.4. Construct an example where $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ but $X_n + Y_n$ does not converge in distribution to $X + Y$.

Solution:

3.1. Let $\epsilon > 0$, then from the Markov's inequality applied to random variable $|X_n - X|^p$, we have

$$P\{|X_n - X| > \epsilon\} \leqslant \frac{\mathbb{E}|X_n - X|^p}{\epsilon} \longrightarrow 0.$$

3.2. (i) First,

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum \mathbb{E}[X_i] = \frac{1}{n}n\mu = \mu$$

Furthermore,

$$E\left(s_n^2\right) = \frac{1}{n-1}E\left(\sum X_i^2 + \sum \bar{X}^2 - 2\underbrace{\sum X_i \bar{X}}_{n\bar{X}^2}\right).$$

$$= \frac{n}{n-1}\left(E\left(X_i^2\right) - E\left(\bar{X}^2\right)\right)$$

$$\overset{\text{Hint and } E\left(X_i^2\right)=\sigma^2+\mu^2}{\Rightarrow} E\left(S_n^2\right) = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2\right) = \sigma^2.$$

(ii)

$$S_n^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \underbrace{\frac{n}{n-1}}_{c_n}\frac{1}{n}\sum x_i^2 - \underbrace{\frac{n}{n-1}}_{d_n}\bar{x}^2$$

With $\lim_{n\to\infty} c_n = \lim_{n\to\infty} d_n = 1$ and the LLN, we get:

$$\begin{aligned}\bar{X}_n \xrightarrow{P} \mu &\Rightarrow \bar{X}_n^2 \xrightarrow{P} \mu^2 \\ &\Rightarrow d_n \bar{X}_n^2 \xrightarrow{P} \mu^2 \\ \bar{Y}_n = \frac{1}{n}\sum x_i^2 \xrightarrow{P} E\left(x_i^2\right) &\Rightarrow Y_n \xrightarrow{P} \sigma^2 + \mu^2 \\ &\Rightarrow c_n Y_n \xrightarrow{P} \sigma^2 + \mu^2\end{aligned}$$

and so

$$\begin{aligned}s_n^2 &= c_n Y_n - d_n \bar{X}_n^2 \\ \Rightarrow s_n^2 &\xrightarrow{P} \sigma^2 + \mu^2 - \mu^2 \\ \Rightarrow s_n^2 &\xrightarrow{P} \sigma^2.\end{aligned}$$

3.3.

- X_n does converge in probability, specifically $X_n \xrightarrow{P} 0$, because

$$\mathbb{P}(|X_n| > \varepsilon) \overset{\text{as } \frac{1}{n} \text{ becomes } < \varepsilon \text{ for } n\to\infty}{=} \mathbb{P}(X_n = n) = \frac{1}{n^2} \longrightarrow 0.$$

- Given that convergence in L^p implies convergence in probability, we only need to check whether $X_n \xrightarrow{P} 0$. Since,

$$\begin{aligned}E\left(X_n^2\right) &= \mathbb{P}(X = \frac{1}{n})\frac{1}{n^2} + \mathbb{P}(X = n)n^2 \\ &= \frac{1}{n^2}\left(1 - \frac{1}{n^2}\right) + n^2\frac{1}{n^2} \to 1,\end{aligned}$$

however, this is not the case and, therefore, X_n does not converge in L^2.

3.4. An example would be any $X_1, X_1, \dots$ distributed i.i.d. according to a symmetric distribution $\mathcal{D}$ (such as standard normal $\mathcal{N}(0,1)$) and Y_n defined as

$$Y_n := -X_n.$$

Then, both $X_n \rightsquigarrow Z \sim \mathcal{D}$ and $Y_n \rightsquigarrow Z \sim \mathcal{D}$, but $X_n + Y_n = 0 \neq 2Z$.

4. Question: Miscellaneous Probability Theory *(slighlty more advanced)*

4.1. Let X be a random variable with $\mathbb{E}[X]^2 < \infty$ and let $Y = |X|$. Suppose that X has a Lebesgue density symmetric about 0 . Show that X and Y are uncorrelated (i.e. $\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]EY = 0$), but they are not independent.

4.2. Show that a random variable X is independent of itself if and only if X is constant a.s. Can X and $f(X)$ be independent, where f is a Borel function?

4.3. Let X be a random variable having a cumulative distribution function F with corresponding probability measure $\mathbb{P}$. Show that if $\mathbb{E}[X]$ exists, then

$$\mathbb{E}[X] = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

Solution:

4.1. Let f be the Lebesgue density of X. Then $f(x) = f(-x)$. Since X and $XY = X|X|$ are odd functions of $X, \mathbb{E}[X] = 0$ and $\mathbb{E}[X|X|] = 0$. Hence,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = E(X|X|) - \mathbb{E}[X]\mathbb{E}[|X|] = 0$$

Let t be a positive constant such that $p = P(0 < X < t) > 0$. Then

$$\begin{aligned} P(0 < X < t, Y < t) &= P(0 < X < t, -t < X < t) \\ &= P(0 < X < t) \\ &= p \end{aligned}$$

and

$$\begin{aligned} P(0 < X < t)P(Y < t) &= P(0 < X < t)P(-t < X < t) \\ &= 2P(0 < X < t)P(0 < X < t) \\ &= 2p^2 \end{aligned}$$

i.e., $P(0 < X < t, Y < t) \neq P(0 < X < t)P(Y < t)$. Hence X and Y are not independent.

4.2.

- Suppose that $X = c$ a.s. for a constant $c \in \mathcal{R}$. For any $A \in \mathcal{B}$ and $B \in \mathcal{B}$,

  $$P(X \in A, X \in B) = I_A(c)I_B(c) = P(X \in A)P(X \in B)$$

  Hence X and X are independent. Suppose now that X is independent of itself. Then, for any $t \in \mathcal{R}$,

  $$P(X \leq t) = P(X \leq t, X \leq t) = [P(X \leq t)]^2$$

  This means that $P(X \leq t)$ can only be 0 or 1 . Since $\lim_{t\to\infty} P(X \leq t) = 1$ and $\lim_{t\to-\infty} P(X \leq t) = 0$, there must be a $c \in \mathcal{R}$ such that $P(X \leq c) = 1$ and $P(X < c) = 0$. This shows that $X = c$ a.s.
- Note that if X and $f(X)$ are independent, then so are $f(X)$ and $f(X)$. This is the case because, for any $A, B \in \mathcal{B}$

  $$X,\ f(X) \text{ independent} \Longrightarrow \mathbb{P}(X \in A, f(X) \in B) = \mathbb{P}(X \in A)\mathbb{P}(f(X) \in B)$$

  Now, define the set $A := f^{-1}(C) := \{x : f(x) \in C\}$ for an arbitrary set $C \in \mathcal{B}$. Then

  $$\begin{aligned} \mathbb{P}(f(X) \in C, f(X) \in B) &= \mathbb{P}(X \in A, f(X) \in B) \\ &= \mathbb{P}(X \in A)\mathbb{P}(f(X) \in B) = \mathbb{P}(f(X) \in C)\mathbb{P}(f(X) \in B). \end{aligned}$$

  From the previous result, this occurs if and only if $f(X)$ is constant a.s.

4.3. By Fubini's theorem,

$$\begin{aligned} \int_0^\infty [1 - F(x)]dx &= \int_0^\infty \int_{(x,\infty)} d\mathbb{P}(y)dx \\ &= \int_0^\infty \int_{(0,y)} dxd\mathbb{P}(y) \\ &= \int_0^\infty yd\mathbb{P}(y) \end{aligned}$$

Similarly,

$$\int_{-\infty}^0 F(x)dx = \int_{-\infty}^0 \int_{(-\infty,x]} d\mathbb{P}(y)dx = -\int_{-\infty}^0 yd\mathbb{P}(y)$$

If $\mathbb{E}[X]$ exists, then at least one of $\int_0^\infty yd\mathbb{P}(y)$ and $\int_{-\infty}^0 yd\mathbb{P}(y)$ is finite and

$$\mathbb{E}[X] = \int_{-\infty}^\infty yd\mathbb{P}(y) = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

If you have any questions or feedback, please feel free to contact me via E-mail at hannah.kuempel@stat.uni-muenchen.de!!

Also, thank you to Jun Shao, the authors of the book *All of Statistics: A Concise Course in Statistical Inference*, andMark Hermanwhose exercises this sheet was inspired by.