Probability and Statistics

Exercise sheet 5

Exercise 5.1 Let X be a real-valued random variable defined on a probability space $(\Omega, \mathcal{A}, P)$. For a fixed integer $k \in \{1, 2, ...\}$ show that $E(X^k)$ exists if and only if $E[(X - E(X))^k]$ exists. In other words, you need to show that

$$E(|X|^k) < \infty \;\Leftrightarrow\; E[|X - E(X)|^k] < \infty.$$

(The case $k = 1$ is trivially true).

Solution 5.1 Let $k \geq 2$ be an integer.

• Suppose that $E(X^k)$ exists, i.e. $E[|X|^k] < \infty$. Using Lyapunov's inequality we have (we take $\alpha = 1$ and $\beta = k$) that

$$E[|X|] \leq (E[|X|^k])^{\frac{1}{k}} < \infty$$

and hence $E[X] = \mu$ exists.

Now, by Minkowski's inequality it follows that

$$E[|X - \mu|^k]^{\frac{1}{k}} \leq (E[|X|^k])^{\frac{1}{k}} + |\mu| < \infty$$

which implies that $E[(X - \mu)^k]$ exists.

• Now, suppose that $E[(X - \mu)^k]$ exists. Then, $E[|X - \mu|^k] < \infty$. By Minkowski's inequality, we have

$$\begin{aligned} E[|X|^k]^{\frac{1}{k}} &= (E[|X - \mu + \mu|^k])^{\frac{1}{k}} \\ &\leq (E[|X - \mu|^k])^{\frac{1}{k}} + \mu \\ &< \infty. \end{aligned}$$

This means that $E(X^k)$ exists.

Exercise 5.2 (Proving Jensen's inequality).

Let φ be a convex function defined on an interval (a, b) with $-\infty \leq a < b \leq +\infty$. Consider some random variable X such that $P(X \in (a, b)) = 1$. Assume that $E(X)$ and $E(\varphi(X))$ exist, that is, $E(|X|) < \infty$ and $E(|\varphi(X)|) < \infty$. Here, we recall that φ is convex on (a, b) if $\forall x, y \in (a, b)$ and $\lambda \in [0, 1]$

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y)$$

and that it is strictly convex if $\forall x, y \in (a, b)$ such that $x \neq y$ and $\lambda \in (0, 1)$

$$\varphi(\lambda x + (1 - \lambda)y) < \lambda\varphi(x) + (1 - \lambda)\varphi(y).$$

(a) Show that for any $c \in (a, b)$, we can find a linear function l such that

$$\varphi(x) \geq l(x) \quad \forall x \in (a, b)$$

and

$$\varphi(c) = l(c).$$

Hint: You may assume that φ admits left and right derivatives, i.e. the limits

$$\varphi_+(x) := \lim_{\epsilon \to 0^+} \frac{\varphi(x+\epsilon) - \varphi(x)}{\epsilon}$$

and

$$\varphi_-(x) := \lim_{\epsilon \to 0^-} \frac{\varphi(x+\epsilon) - \varphi(x)}{\epsilon}$$

both exist for any $x \in (a,b)$. Explain why $\varphi_+(x) \geq \varphi_-(x)$ for $x \in (a,b)$. Then, show that for any $s \in [\varphi_-(c), \varphi_+(c)]$ you can construct a line with slope s with the desired properties.

(b) Show that under the given assumptions on X we have

$$E[\varphi(X)] \geq \varphi(E(X)).$$

Hint: Use your result from (a). Given the random variable X, what is a reasonable value to choose for c?

(c) Suppose now that φ is strictly convex. Show that we have equality if and only if $P(X = E(X)) = 1$, that is, X is a degenerate random variable.

Solution 5.2

(a) Let $x \in (a,b)$ and $b - x > t \geq \epsilon > 0$. Then

$$\begin{aligned} \varphi(x+\epsilon) &= \varphi\left(\frac{\epsilon}{t}(x+t) + \left(1 - \frac{\epsilon}{t}\right)x\right) \\ &\leq \frac{\epsilon}{t}\varphi(x+t) + \left(1 - \frac{\epsilon}{t}\right)\varphi(x) \\ \Leftrightarrow \frac{\varphi(x+\epsilon) - \varphi(x)}{\epsilon} &\leq \frac{\varphi(x+t) - \varphi(x)}{t} \end{aligned}$$

which means that as $\epsilon \searrow 0$ the function

$$\epsilon \mapsto \frac{\varphi(x+\epsilon) - \varphi(x)}{\epsilon}$$

is non-increasing. (i.e. flipping it around, if ϵ increases this function is non-decreasing).

Furthermore, by putting $z = \lambda x + (1-\lambda)y$ (note that $z \in [x,y]$ when $x \leq y$), the convexity condition can be shown to be equivalent to

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x} \tag{1}$$

or to

$$\frac{f(y) - f(z)}{y - z} \geq \frac{f(z) - f(x)}{z - x}. \tag{2}$$

Then, by (2), fixing some $a < y < x$, we have that for any $\epsilon \in (0, b-x)$,

$$\frac{\varphi(x+\epsilon) - \varphi(x)}{\epsilon} \geq \frac{\varphi(x) - \varphi(y)}{x - y}.$$

Thus, as a monotone non-increasing (as ϵ decreases) function bounded by below, $\epsilon \mapsto \frac{\varphi(x+\epsilon)-\varphi(x)}{\epsilon}$ admits a limit as $\epsilon \searrow 0$, which is by definition equal to the right derivative of φ at x, that is $\varphi'_+(x)$.

Using similar arguments, we can show that

$$\varphi'_-(x) = \lim_{\epsilon \nearrow 0} \frac{\varphi(x+\epsilon)-\varphi(x)}{\epsilon}$$

exists.

Moreover, by (2),

$$\frac{\varphi(x+\epsilon)-\varphi(x)}{\epsilon} \geq \frac{-\varphi(x-\epsilon)+\varphi(x)}{\epsilon}$$

which implies that as $\epsilon \searrow 0$,

$$\varphi'_+(x) \geq \varphi'_-(x).$$

Now, fix $c \in (a, b)$. For $x > c$, we have

$$\frac{\varphi(x)-\varphi(c)}{x-c} \geq \frac{\varphi(c+\epsilon)-\varphi(c)}{\epsilon}$$

(by (1)) for small $\epsilon > 0$.

Taking $\epsilon \searrow 0$, this implies that

$$\frac{\varphi(x)-\varphi(c)}{x-c} \geq \varphi'_+(c). \tag{3}$$

Similarly, for $x < c$,

$$\begin{aligned} &\frac{\varphi(c)-\varphi(x)}{c-x} \leq \frac{\varphi(c)-\varphi(c-\epsilon)}{\epsilon} \\ \Rightarrow &\frac{\varphi(c)-\varphi(x)}{c-x} \leq \varphi'_-(c). \end{aligned} \tag{4}$$

Then we see from inequalities (3) and (4) that if we take any slope $s \in [\varphi'_-(c), \varphi'_+(c)]$, then

$$\frac{\varphi(x)-\varphi(c)}{x-c} \geq s, \quad x > c$$

and

$$\frac{\varphi(c)-\varphi(x)}{c-x} \leq s, \quad x < c$$

which, combined, yield

$$\varphi(x) - \varphi(c) \geq s(x-c)$$

or

$$\varphi(x) \geq s(x-c) + \varphi(c).$$

Since the inequality holds also for $x = c$, we conclude that for any $s \in [\varphi'_-(c), \varphi'_+(c)]$ the linear function

$$l(x) = s(x-c) + \varphi(c)$$

satisfies the requirements.

(b) Take $c = E(X)$. Then choosing one of the possible l from part (a), we get

$$\begin{aligned} \varphi(x) &\geq l(x) \\ \Rightarrow E[\varphi(X)] &\geq E[l(X)] = l(E(X)) \end{aligned}$$

by linearity of expectation.

Since $l(E(X)) = l(c) = \varphi(c)$, it follows that

$$E[\varphi(X)] \geq \varphi(E(X)).$$

(c) Equality holds if and only if

$$E[\varphi(X) - l(X)] = 0.$$

Now, $\varphi(X) - l(X) \geq 0$ by construction of l. Thus, equality holds if and only if

$$P(\varphi(X) = l(X)) = 1.$$

But we can consider the two events $X = E[X]$ and $X \neq E[X]$ (note that in the first case we know $\varphi(X) = l(X)$ by construction):

$$1 = P(\varphi(X) = l(X)) = P(X = E[X]) + P(\varphi(X) = l(X), X \neq E[X]). \tag{5}$$

However, we can check that $\varphi(x) > l(x)$ for any $x \neq E[X] = c$ if φ is strictly convex. To see this, note that in (1) and (2) the inequalities become strict if φ is strictly convex. Therefore, choosing for example $x > c$, we get that for small $\epsilon > 0$,

$$\frac{\varphi(x) - \varphi(c)}{x - c} > \frac{\varphi(c + \epsilon) - \varphi(c)}{\epsilon} \geq \varphi'_+(c).$$

This implies that

$$\varphi(x) > \varphi(c) + (x - c)\varphi_+(c) \geq \varphi(c) + s(x - c).$$

A similar argument can be done for $x < c$, which proves our statement above, that $\varphi(x) > l(x)$ for $x \neq c$. But then we conclude that $\varphi(X) = l(X), X \neq E[X]$ is impossible (has probability 0), and finally from (5) we get

$$P(X = E[X]) = 1.$$

Exercise 5.3 Suppose you can choose a number $n \geq 1$ and then toss a fair coin n times. You will be given a prize if you get either exactly 7 or exactly 9 heads. What is the "best" choice for the number n?

Solution 5.3 The best n is the one that yields the highest probability to get the prize. It is clear that $n \geq 9$. Note that we want to maximize

$$P(S_n = 7 \text{ or } 9)$$

over the possible n, where $S_n \sim \text{Bin}(n, \frac{1}{2})$.

$$P(S_n \in \{7,9\}) = \frac{1}{2^n}\left(\binom{n}{7} + \binom{n}{9}\right)$$
$$= \begin{cases} 0.072, & n = 9 \\ 0.126, & n = 10 \\ 0.187, & n = 11 \\ 0.247, & n = 12 \\ 0.296, & n = 13 \\ 0.331, & n = 14 \\ 0.349, & n = 15 \\ 0.349, & n = 16 \\ 0.333, & n = 17 \end{cases}$$

One should choose either $n = 15$ or 16 (the probability is exactly the same in both cases).

Exercise 5.4 (A novel way to give a test)

A student takes a 5-answer multiple choice test. His/her grade is determined by the number of questions required to get 5 correct answers. The grading is done as follows:

- Grade A is given if the student only needs 5 questions;
- Grade B is given if the student needs 6 or 7 questions;
- Grade C is given if the student needs 8 or 9 questions;
- Grade F (fail) is given otherwise.

Suppose the student guesses independently at random on each question. What is the most likely grade (i.e. which outcome has the highest probability)?

Solution 5.4 Let N be the number of questions that the student needs to answer to get the 5 correct answers. It is clear that N is Negative Binomial, $\mathrm{NB}(r, p)$. Here, $r = 5$ and $p = \frac{1}{5}$ since the probability that the student gets a question answered correctly is $\frac{1}{5}$ (he/she makes a random guess).

$$\begin{aligned} P(\text{student gets grade A}) &= P(N = 5) \\ &= \binom{4}{4}\left(\frac{4}{5}\right)^0 \frac{1}{5^5} \\ &= \frac{1}{5^5} = 0.00032. \end{aligned}$$

$$\begin{aligned} P(\text{student gets grade B}) &= P(N = 6 \text{ or } 7) \\ &= \binom{5}{4}\left(\frac{4}{5}\right)^1 \frac{1}{5^5} + \binom{6}{4}\left(\frac{4}{5}\right)^2 \frac{1}{5^5} \\ &= \frac{1}{5^5}\left(4 + 15 \times \frac{16}{25}\right) \\ &= 0.004352. \end{aligned}$$

$$\begin{aligned} P(\text{student gets grade C}) &= P(N = 8 \text{ or } 9) \\ &= \frac{1}{5^5}\left(\binom{7}{4}\left(\frac{4}{5}\right)^3 + \binom{8}{4}\left(\frac{4}{5}\right)^4\right) \\ &= \frac{4^3}{5^8}\left(\binom{7}{4} + \binom{8}{4} \times \frac{4}{5}\right) \\ &= 0.0149. \end{aligned}$$

$$\begin{aligned} P(\text{student fails}) &= 1 - P(N < 10) \\ &= 0.980 \end{aligned}$$

Failing the test is the most likely event (not surprisingly, since the student has only 1 chance out of 5 to get each question correctly answered, so it is unlikely that he will get 5 out of the first 9 correct).

Exercise 5.5 (Generating functions). (Optional)

Let X be some integer-valued random variable, that is $X(\omega) \in \{0, 1, 2, ...\}$ $\forall \omega \in \Omega$, the sample space on which X is defined. The generating function of X is defined as

$$G(s) := \sum_{k=0}^{\infty} s^k P(X = k),$$

for those values of s such that the sum on the right-hand side converges.
Note that G is always well-defined for $|s| \leq 1$ since

$$\sum_{k=0}^{\infty} |s|^k P(X = k) \leq \sum_{k=0}^{\infty} P(X = k) = 1.$$

Also, $G(s) = E[s^X]$ is another expression for G.

(a) Consider a power series $f(s) = \sum_{k=0}^{\infty} a_k s^k$, given a real sequence $(a_k)_{k \geq 0}$ and $s \in \mathbb{R}$ for which $f(s)$ is defined. Suppose that there is some $s_0 \neq 0$ such that $f(s_0)$ is defined, that is $\sum_{k=0}^{\infty} a_k s_0^k$ converges.

Show that f is defined and infinitely differentiable for all s such that $|s| < |s_0|$ and

$$f^{(j)}(s) = \sum_{k=j}^{\infty} a_k k(k-1)...(k-j+1) s^{k-j}$$

for $|s| < |s_0|$.

Hint: You may use the fact that if $(f_n)_{n=1}^{\infty}$ is a sequence of differentiable functions $f_n : (a, b) \to \mathbb{R}$ such that $f_n \to f$ pointwise on (a, b) and $f_n' \to g$ uniformly on (a, b) for some functions $f, g : (a, b) \to \mathbb{R}$, then $f' = g$. You may need to apply this inductively to conclude.

(b) Conclude from (a) that $\forall s : |s| < 1$, the generating function G defined above is infinitely differentiable and compute $G^{(j)}(0)$.

(c) Let $X \sim \text{Unif}\{1, 2, ..., n\}$ for some $n \geq 1$. What is the expression of G_X, the generating function of such an X?

(d) Consider now two random variables X and Y that are i.i.d $\sim \text{Unif}\{1, 2, ..., n\}$. Let $S = X + Y$. What is the expression of G_S, the generating function of S?

(e) Use (b) to find the pmf of S. Is it easier to do it directly?

Solution 5.5

(a) Let $|s| < |s_0|$ where $\sum_{k=0}^{\infty} a_k s_0^k < \infty$. Then, $\lim_{k\to\infty} a_k s_0^k = 0$ and we can find $M > 0$ such that $|a_k s_0^k| \le M$ for any $k \in \{0, 1, ...\}$. Now,

$$\begin{aligned}\sum_{k=0}^{\infty} |a_k||s|^k &= \sum_{k=0}^{\infty} |a_k||s_0|^k \left|\frac{s}{s_0}\right|^k \\ &\le M \sum_{k=0}^{\infty} \left|\frac{s}{s_0}\right|^k \\ &= \frac{M}{1 - \frac{|s|}{|s_0|}} < \infty.\end{aligned}$$

Thus, $f(s)$ is well-defined.

Now let

$$f_n(s) = \sum_{k=0}^{n} a_k s^k, \quad n \in \{1, 2, ...\}$$

Then f_n is differentiable on $\mathbb{R}$ since it is a polynomial, and in particular it is differentiable on $(-|s_0|, |s_0|)$ with derivative

$$f_n'(s) = \sum_{k=1}^{n} k a_k s^{k-1}.$$

For $|s| \le \rho < |s_0|$ we have that

$$\begin{aligned}\sum_{k=1}^{n} k|a_k||s|^{k-1} &= \sum_{k=0}^{n} k|a_k| \left(\frac{|s|}{|s_0|}\right)^{k-1} |s_0|^{k-1} \\ &\le M \sum_{k=1}^{n} k \left(\frac{\rho}{|s_0|}\right)^{k-1} \\ &\underset{n\to\infty}{\longrightarrow} \frac{M}{\left(1 - \frac{\rho}{|s_0|}\right)^2}\end{aligned}$$

since $\frac{\rho}{|s_0|} < 1$.

The series $\sum_{k=1}^{\infty} k a_k s^{k-1}$ is then uniformly convergent on $[-|\rho|, |\rho|]$ for any $0 < \rho < |s_0|$. Using the hint, we conclude that $f(s) = \sum_{k=0}^{\infty} a_k s^k$ is differentiable on $(-|s_0|, |s_0|)$ and

$$f'(s) = \sum_{k=1}^{\infty} k a_k s^{k-1}$$

for any $s \in (-|s_0|, |s_0|)$.

Applying this argument to $\tilde{a}_k = (k+1)a_{k+1}$ and $\tilde{f}(s) = \sum_{k=0}^{\infty} \tilde{a}_k s^k$ we can show by induction that f is infinitely differentiable on $(-|s_0|, |s_0|)$.

(b) Take $s_0 = 1$. Then $G(s) = \sum_{k=0}^{\infty} s^k P(X = k)$ is well-defined for s_0. Using (a), it follows that G is infinitely differentiable at $s = 0$, with

$$G^{(j)}(s) = \sum_{k=j}^{\infty} k(k-1)...(k-j+1)s^{k-j} P(x = k)$$

for $s \in (-1, 1)$ and $j \in \{0, 1, ...\}$. In particular

$$G^{(j)}(0) = j!P(X = j).$$

(c) For $X \sim \text{Unif}\{1, 2, ..., n\}$,

$$G_X(s) = E[s^X]$$

where s^X is distributed as $\text{Unif}\{s, s^2, ..., s^n\}$, and this is well-defined for any $s \in \mathbb{R}$. Therefore

$$\begin{aligned} G_X(s) &= \frac{1}{n}\sum_{k=1}^{n} s^k, \quad \forall s \in \mathbb{R} \\ &= \frac{s(1-s^n)}{n(1-s)}, \quad \forall s \in \mathbb{R} \setminus \{1\}. \end{aligned}$$

Let X and Y be i.i.d $\sim \text{Unif}\{1, 2, ..., n\}$. Then,

$$\begin{aligned} G_S(s) = E[s^S] &= E[s^{X+Y}] \\ &= E[s^X s^Y] \\ &= E[s^X]E[s^Y] \end{aligned}$$

since $s^X = f(X)$ and $s^Y = g(Y)$ are independent.
It follows that

$$G_S(s) = \left(\frac{1}{n}\sum_{k=1}^{n} s^k\right)^2, \quad \forall s \in \mathbb{R}.$$

(d) It follows from (b) that

$$P(S = j) = \frac{G_S^{(j)}(0)}{j!}, \quad j \in \{0, 1, ...\}.$$

Note that G_S is a polynomial of degree $2n$, hence the support of S is necessarily bounded above by $2n$. It can also be seen directly from the definition of S that

$$S \in \{2, 3, ..., 2n\} = \text{supp}(S).$$

For $j \in \{2, 3, 4, ..., 2n\}$, we can compute $G_s^{(j)}$ using the Leibniz formula for higher derivatives of the product:

$$(fg)^{(k)} = \sum_{j=0}^{k} \binom{k}{j} f^{(j)} g^{(k-j)}.$$

Then,

$$\begin{aligned}
G_S^{(j)}(0) &= \sum_{i=0}^{j} \binom{j}{i} G_X^{(i)}(0) G_Y^{(j-i)}(0) \\
&= \sum_{i=0}^{j} \binom{j}{i} i! P(X=i) \times (j-i)! P(Y=j-i) \\
&= \sum_{i=0}^{j} j! P(X=i) P(Y=j-i) \\
&= j! \sum_{i=0}^{j} P(X=i) P(Y=j-i)
\end{aligned} \tag{6}$$

where $P(X=i) = \frac{1}{n}$ for $i = 1, ..., n$ and similarly $P(Y = j-i) = \frac{1}{n}$ for $j - i = 1, ..., n$, or in other words $i = j-n, j-n+1, .., j-1$.

It follows that in the sum (6), we only get contributions from those terms such that

$$\max(1, j-n) \leq i \leq \min(n, j-1).$$

In conclusion,

$$P(S=j) = \frac{\min(n, j-1) - \max(1, j-n) + 1}{n^2}, \quad j \in \{2, 3, ..., 2n\}.$$

In this case, this computation does not appear to be easier than done directly. Nevertheless, one can see from this expression that S is not uniformly distributed on $\{2, ..., 2n\}$.