Probability and Statistics

Exercise sheet 3

Exercise 3.1 In a clinical trial with two treatment groups, the probability of success (the patient being cured) in one treatment group is $p_1 = 0.5$, and the probability of success in the other is $p_2 = 0.6$. There are 5 patients in each group. Assuming the outcomes for all patients are independent, calculate the probability that the first treatment group has at least as many successes as in the second one.

Solution 3.1 The number of successes in the first group follows the Binomial law $\mathcal{B}(5, 0.5)$, and in the second group follows $\mathcal{B}(5, 0.6)$. The probability to have $k \in \{0, 1, 2, 3, 4, 5\}$ successes in the two cases are respectively

$$P[X = k] = \binom{5}{k} 2^{-5} \quad \text{for the first group}$$
$$P[X = k] = \binom{5}{k} 3^k 2^{5-k} 5^{-5} \quad \text{for the second group}$$

	k=0	k=1	k=2	k=3	k=4	k=5
p=0.5	0.0312	0.1562	0.3125	0.3125	0.1562	0.0312
p=0.6	0.0102	0.0768	0.2304	0.3456	0.2592	0.0778

The probability that the first group has at least as many successes as the second group is given by

$$0.03120 \times 0.0102 + 0.1562 \times (0.0102 + 0.0768) + ... + 0.0312 \times 1 = 0.49553028.$$

Exercise 3.2 Suppose a fair die is rolled once and the observed number is $N \in \{1, ..., 6\}$. Then, a fair coin is tossed N times. Let X be the number of heads obtained.

Find the pmf, cdf and expectation of X. Does the expected value make sense intuitively?

Solution 3.2

- For $k \in \{0, 1, 2, 3, 4, 5, 6\}$, the pmf of X is given by

$$\begin{aligned} p(k) &= P(X = k) \\ &= \sum_{n=1}^{6} P(X = k \text{ and } N = n) \\ &= \sum_{n=1}^{6} P(X = k \mid N = n) P(N = n) \\ &= \frac{1}{6} \sum_{n=1}^{6} P(X = k \mid N = n). \end{aligned}$$

Since the coin is fair,

$$P(X = k \mid N = n) = \frac{\binom{n}{k}}{2^n} \quad (k \leq n)$$

because $\binom{n}{k}$ is the number of possibilities to obtain k heads among n tosses and 2^n is the number of all possibilities (H or T) in n tosses. Also,

$$P(X = k \mid N = n) = 0 \quad (k > n).$$

Thus,

$$\begin{aligned} p(k) &= \frac{1}{6}\sum_{n=1}^{6}\binom{n}{k}\frac{1}{2^n}1_{k\in\{0,1,\ldots,n\}} \\ &= \frac{1}{6}\sum_{n=\max(k,1)}^{6}\binom{n}{k}\frac{1}{2^n}. \end{aligned}$$

- By definition, the cdf F of X is given by

$$\forall x \in \mathbb{R}, \quad F(x) = P(X \le x).$$

Thus,

$$F(x) = \begin{cases} 0, & x < 0 \\ \sum_{j=0}^{k} p(j), & k \le x < k+1 \text{ and } 0 \le k \le 5 \\ 1, & x \ge 6. \end{cases}$$

- We calculate:

$$\begin{aligned} E(X) &= \sum_{k=0}^{6} kp(k) \\ &= \sum_{k=1}^{6} kp(k) \\ &= \frac{1}{6}\sum_{k=1}^{6} k \sum_{n=1}^{6}\binom{n}{k}\frac{1}{2^n}1_{k\in\{1,\ldots,n\}} \\ &= \frac{1}{6}\sum_{n=1}^{6}\frac{1}{2^n}\sum_{k=1}^{n} k\binom{n}{k}. \end{aligned}$$

Note

$$\begin{aligned} k\binom{n}{k} &= k\frac{n!}{k!(n-k)!} \\ &= n\frac{(n-1)!}{(k-1)!(n-k)!} \\ &= n\binom{n-1}{k-1} \end{aligned}$$

so that

$$\sum_{k=1}^{n} k\binom{n}{k} = n\sum_{k=0}^{n-1}\binom{n-1}{k} = n2^{n-1}.$$

It follows that

$$
\begin{aligned}
E(X) &= \frac{1}{6}\sum_{n=1}^{6} \frac{1}{2^n} n 2^{n-1} \\
&= \frac{1}{6}\sum_{n=1}^{6} \frac{n}{2} \\
&= \frac{1}{2}E(N) \\
&= \frac{1}{2} \times 3.5 \\
&= 1.75.
\end{aligned}
$$

The result is expected since one expects to see half of the tosses resulting in heads.

Alternatively, one could exploit the following idea to find the result directly. Letting Y be the number of tails obtained, it is clear by the definition of the problem that

$$X + Y = N$$

and by linearity of expectation,

$$E(X) + E(Y) = E(N).$$

A little thought reveals that Y must have the same distribution as X (by symmetry), and therefore that gives right away that

$$E(X) = E(Y) = \frac{E(N)}{2} = 1.75.$$

Exercise 3.3

(a) Suppose that X has pmf

$$P\left(X = \frac{1}{n}\right) = \frac{1}{2^n} \quad (n \geq 1).$$

Find $E(X)$.

(b) Suppose that X has pmf

$$P\left(X = \frac{1}{n}\right) = \frac{1}{2^{n+1}} \quad (n \geq 1)$$

and

$$P\left(X = n\right) = \frac{1}{2^n} \quad (n \geq 2).$$

Find $E(X)$.

Solution 3.3

(a) Using the given pmf,

$$
\begin{aligned}
E(X) &= \sum_{n=1}^{\infty} \frac{1}{n2^n} \\
&= -\log\left(1 - \frac{1}{2}\right) \\
&= \log(2)
\end{aligned}
$$

using the identity

$$\log(1+x) = \sum_{n=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}, \quad x \in (-1,1).$$

(b)

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} \frac{1}{n 2^{n+1}} + \sum_{n \geq 2}^{\infty} \frac{n}{2^n} \\ &= A + B \end{aligned}$$

where

$$A = \frac{1}{2} \log(2)$$

(using the first part), and

$$\begin{aligned} B &= \sum_{n=1}^{\infty} \frac{n}{2^n} - \frac{1}{2} \\ &= \frac{1}{2} \sum_{n=1}^{\infty} \frac{n}{2^{n-1}} - \frac{1}{2} \\ &= \frac{1}{2} \frac{1}{\left(1 - \frac{1}{2}\right)^2} - \frac{1}{2} \\ &= \frac{1}{2}(4-1) = \frac{3}{2} \end{aligned}$$

using

$$\sum_{k=1}^{\infty} k x^{k-1} = \left(\frac{1}{1-x}\right)' = \frac{1}{(1-x)^2} \quad \forall x \in (-1,1).$$

So we conclude

$$E(X) = \frac{\log(2) + 3}{2}.$$

Exercise 3.4

(a) Fix p a positive integer. Give an example of a random variable X taking values in $\{1, 2, 3, ...\}$ such that $E(X^k) < \infty \ \forall k < p$ but $E(X^p) = \infty$.

(b) Let X be some nonnegative random variable and p some positive integer. Show that $E(X^p) \geq E(X)^p$. Can we have equality?

Solution 3.4

(a) Consider a random variable X with pmf

$$P(X = n) = \frac{c_p}{n^{p+1}}, \quad \forall n \geq 1$$

where

$$c_p = \frac{1}{\sum_{n=1}^{\infty} \frac{1}{n^{p+1}}}$$

is well-defined for $p > 0$.

Then,

$$E(X^k) = c_p \sum_{n=1}^{\infty} \frac{n^k}{n^{p+1}} = c_p \sum_{n=1}^{\infty} \frac{1}{n^{p-k+1}}.$$

Since $p + 1 - k > 1$ for $k < p$, $E(X^k) < \infty$.

But $E(X^p) = \infty$, because now $\frac{n^p}{n^{p+1}} = \frac{1}{n}$ and

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

(b) For $p = 1$, we obviously have equality. For $p \geq 2$, the function $\phi(x) = x^p$ is convex on $[0, +\infty)$. It follows from Jensen's inequality that

$$E[\phi(X)] \geq \phi(E(X))$$

which shows what we wanted.

Equality occurs if and only if $X = c$ with probability 1 for some constant $c \geq 0$.

Exercise 3.5 (optional) Let $s \in (1, \infty)$. The Riemann zeta function is defined as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

The goal is to prove that

$$\zeta(s) = \frac{1}{\prod_{i=1}^{\infty} \left(1 - \frac{1}{p_i^s}\right)}$$

with $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ... the prime numbers (in order).

(a) Take $(\mathbb{N}, 2^{\mathbb{N}}, P)$ with

$$P(A) = \frac{1}{\zeta(s)} \sum_{n \in A} \frac{1}{n^s}$$

for any $A \in 2^{\mathbb{N}}$.

Show that $(\mathbb{N}, 2^{\mathbb{N}}, P)$ is a probability space.

(b) For p a prime number, let

$$N_p = \{n \in \mathbb{N} : n \text{ divisible by } p\}.$$

Calculate $P(N_p)$.

(c) Prove that events $\{N_{p_i}\}_{i \geq 1}$ are mutually independent.

(d) Compute $P(\cap_{i \geq 1} N_{p_i}^c)$ and conclude.

Solution 3.5

(a) Clearly $2^{\mathbb{N}}$ is a σ-algebra. P is a probability measure because:

$$P(\{i\}) \geq 0,$$

and for $W_i \subseteq \mathbb{N}$, $i \in \mathbb{N}$, disjoint sets,

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} W_i\right) &= \sum_{n \in \bigcup_{i=1}^{\infty} W_i} f(n) \\ &= \sum_{i=1}^{\infty} \sum_{n \in W_i} f(n) \\ &= \sum_{i=1}^{\infty} P(W_i). \end{aligned}$$

Additionally it is clear from the definition that $\mathbb{P}(\mathbb{N}) = 1$. Then $(\mathbb{N}, 2^{\mathbb{N}}, \mathbb{P})$ is a probability space.

(b) Let p be a prime number, then $N_p = \{np \mid n \in \mathbb{N}\}$. We have

$$\begin{aligned} P(N_p) &= \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{1}{(np)^s} \\ &= \frac{1}{p^s} \left(\frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \frac{1}{n^s} \right) \\ &= \frac{1}{p^s}. \end{aligned}$$

(c) Take $(p_{i_k})_{k=1}^m$ a finite family of prime numbers. Then $\bigcap_{k=1}^m N_{p_{i_k}} = \{n \in \mathbb{N} : n \prod_{k=1}^m p_{i_k}\}$, so

$$\begin{aligned} P\left(\bigcap_{k=1}^{m} N_{p_{i_k}}\right) &= \frac{1}{\zeta(s)} \sum_{n \in N} \frac{1}{n^s \prod_{k=1}^m p_{i_k}^s} \\ &= \frac{1}{\prod_{k=1}^m p_{i_k}^s} \frac{1}{\zeta(s)} \sum_{n \in N} \frac{1}{n^s} \\ &= \prod_{k=1}^{m} \frac{1}{p_{i_k}^s} \\ &= \prod_{k=1}^{m} P(N_{p_{i_k}}), \end{aligned}$$

and the $N_{p_{i_k}}$'s are independent.

(d) Note that $\bigcap_{k=1}^{\infty} N_{p_k}^c = \{1\}$. Then, with question (c),

$$\begin{aligned} \frac{1}{\zeta(s)} &= P\left(\bigcap_{k=1}^{\infty} N_{p_k}^c\right) \\ &\overset{*}{=} \lim_{m \to \infty} P\left(\bigcap_{k=1}^{m} N_{p_k}^c\right) \\ &= \lim_{m \to \infty} \prod_{k=1}^{m} (1 - p_k^{-s}) \\ &= \prod_{k=1}^{\infty} (1 - p_k^{-s}), \end{aligned}$$

Thus,

$$\zeta(s) = \frac{1}{\prod_{k=1}^{\infty}(1 - p_k^{-s})}.$$

* The equality

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{m\to\infty} P\left(\bigcap_{k=1}^{m} A_k\right)$$

and equivalently

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{m\to\infty} P\left(\bigcup_{k=1}^{m} A_k\right)$$

are quite useful and well-known. A simple way to prove them is using the monotone convergence theorem.