Probability and Statistics

Exercise sheet 7

Exercise 7.1 (Getting the distribution of independent Binomials.)

We recall for this question the following (general) definition:

For X and Y two random variables, not necessarily defined on the same probability space, we say that X and Y have the same distribution (denoted $X \stackrel{d}{=} Y$) if

$$F_X = F_Y \text{ on } \mathbb{R} \tag{1}$$

with F_X and F_Y the cdf's of X and Y, respectively. Note that when X and Y are discrete, (1) is equivalent to $p_X = p_Y$, where p_X and p_Y are the pmf's of X and Y respectively.

Consider now X_1 and X_2 two *independent* random variables such that $X_1 \sim \text{Bin}(n_1, p)$ and $X_2 \sim \text{Bin}(n_2, p)$, with $n_1 \geq 1$, $n_2 \geq 1$ in $\mathbb{N}$ and $p \in (0, 1)$. We want to show that

$$X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p).$$

(a) The hard way:

Assume without loss of generality that $n_1 \leq n_2$. Let $k \in \{0, ..., n_1 + n_2\}$. Show that

$$P(X_1 + X_2 = k) = \left[\sum_{j=\max(0,k-n_2)}^{\min(k,n_1)} \binom{n_1}{j}\binom{n_2}{k-j}\right] p^k (1-p)^{n_1+n_2-k}.$$

Using the fact that the pmf of any random variable with distribution $\text{Hypergeo}(n, D, N)$ has to add up to 1, show that

$$\sum_{j=\max(0,k-n_2)}^{\min(k,n_1)} \binom{n_1}{j}\binom{n_2}{k-j} = \binom{n_1+n_2}{k}$$

and conclude.

(b) A more elegant way:

On $\Omega = \{0, 1\}^{n_1+n_2}$ define the Bernoulli random variables $Y_1, ..., Y_{n_1}, Y_{n_i+1}, ..., Y_{n_1+n_2}$ such that they are all $\stackrel{\text{iid}}{\sim}$ Bernoulli(p).

Here, the Y_i have the natural definition that for each $i \in \{1, ..., n_1 + n_2\}$,

$$Y_i(\omega_1, ..., \omega_{n_1+n_2}) = \omega_i$$

for $(\omega_1, ..., \omega_{n_1+n_2}) \in \{0, 1\}^{n_1+n_2}$, and Ω is equipped with the probability measure P such that

$$P((\omega_1, ..., \omega_{n_1+n_2})) = p^{\omega_1}(1-p)^{1-\omega_1} ... p^{\omega_{n_1+n_2}}(1-p)^{1-\omega_{n_1+n_2}}$$

defined on $\mathcal{A} = 2^\Omega$.

Define the random variables

$$X_1' := Y_1 + ... + Y_{n_1},$$

$$X_2' := Y_{n_1+1} + ... + Y_{n_1+n_2}.$$

Show using a simple argument that

$$X_1 + X_2 \overset{d}{=} X_1' + X_2'$$

(without computing their cdf's or pmf's explicitly).

Conclude now that $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$.

Exercise 7.2

(a) Let

$$f(x) := \frac{1}{x^k} \mathbb{1}_{x \in [1,+\infty)}.$$

For what value of k, if any, is f a density function?

(b) Give an example of a density function f such that $c\sqrt{f}$ cannot be a density function for any $c > 0$.

(c) Let

$$f(x) = c\,|x|\,(1 - x^2)\mathbb{1}_{|x| \leq 1}.$$

1. Find $c > 0$ so that f is a density function.
2. Find the cdf corresponding to this density.
3. Compute $P(X < -\frac{1}{2})$ and $P(|X| \leq \frac{1}{2})$.

Exercise 7.3 (On moment generating functions).

For a random variable X, the moment generating function is defined as

$$\Psi_X(t) := E[e^{tX}]$$

for any $t \in \mathbb{R}$ for which this expectation is finite.

In this question, we will assume that for two random variables X and Y (*not* necessarily defined on the same probability space), we have the equivalence

$$X \overset{d}{=} Y \Leftrightarrow \Psi_X = \Psi_Y \text{ on } (a, b)$$

for some non-empty open interval (a, b) containing 0. (One can prove this equivalence indeed holds, as long as such an interval (a, b) exists where the moment generating functions are defined).

(a) Let $X \sim \text{Bin}(n, p)$. Compute Ψ_X (on its domain of definition). If $X_1 \sim \text{Bin}(n_1, p)$ and $X_2 \sim \text{Bin}(n_2, p)$ are independent, compute $\Psi_{X_1+X_2}$. Can you conclude again that $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$?

(b) Let $X \sim \mathcal{N}(0, 1)$. Compute Ψ_X (on its domain of definition). If $X_1, X_2, ..., X_n$ are $\overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$, compute $\Psi_{X_1+...+X_n}$. What is then the distribution of $X_1 + ... + X_n$?

(c) Let $X \sim \text{Exp}(\lambda)$ for some $\lambda > 0$. Recall that this means that the density of X is given by

$$f(x) = \lambda e^{-\lambda x}\mathbb{1}_{x \in (0,\infty)}.$$

Compute Ψ_X (on its domain of definition).

(d) Let $X \sim G(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$. Compute Ψ_X (on its domain of definition). If $X_1, ..., X_n$ are independent such that $X_i \sim G(\alpha_i, \beta)$ for $\alpha_i > 0, \beta > 0$, compute $\Psi_{X_1+...+X_n}$. What is the distribution of $X_1 + ... + X_n$?

Exercise 7.4 (Optional).
The first goal of this exercise is to show that:

(a) $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ is an absolutely continuous random variable with density f if and only if

$$\forall x \in \mathbb{R} \quad F(x) = \int_{(-\infty,x]} f d\lambda = \int_{-\infty}^{x} f(t)dt.$$

1. Show that the condition is necessary.
2. To show that it is sufficient, define the probability measure

$$\mu_1(B) = \int_B f d\lambda = \int_B f(t)dt, \ \forall B \in \mathcal{B}_\mathbb{R}.$$

Use Carathéodory's extension theorem to show that μ_1 and P_X have to be equal and conclude.

The second goal is to show the following:

(b) A measurable function $f \geq 0$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is the density of some absolutely continuous random variable if and only if

$$\int_\mathbb{R} f(t)dt = 1.$$

1. Show that this condition is necessary.
2. To show that it is sufficient, define

$$F(x) := \int_{-\infty}^{x} f(t)dt, \ x \in \mathbb{R}.$$

Show that:

- F is non-decreasing on $\mathbb{R}$,
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$,
- F is right-continuous on $\mathbb{R}$.

Conclude from this that F has to be the cdf of some random variable X (simply invoke a result from the lecture; no need to give a formal proof).

Finally, show that this X has to be absolutely continuous with density f (for this use (a)).