

# Probability for Statistics

## Problem Sheet 2

1. Consider a probability space  $(\Omega, \mathcal{F}, \Pr)$  in which

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad \mathcal{F} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}.$$

Determine whether each of the two functions  $X_1, X_2 : \Omega \rightarrow \mathbf{R}$  defined below is a random variable with respect to  $\mathcal{F}$ .

$$X_1(s) = s, \quad X_2(s) = \begin{cases} 0 & s \text{ even} \\ 1 & s \text{ odd} \end{cases} \quad \forall s \in \Omega$$

2. (a) Let  $X : \Omega \rightarrow \mathbf{R}$  be a random variable, and let  $\mathcal{B}$  be the Borel sigma algebra on  $\mathbf{R}$ . Show that  $\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}\}$  is a sigma algebra on  $\Omega$ .
- (b) Consider an experiment in which a fair coin is flipped twice, so that the sample space is  $\Omega = \{HH, HT, TH, TT\}$ . Let  $X : \Omega \rightarrow \mathbf{R}$  take the value 1 if precisely one flip comes up heads, and 0 otherwise. Determine the sigma algebra  $\mathcal{F}_X$ .
- (c) For  $\Omega$  as in the previous part, give an example of a function  $Y : \Omega \rightarrow \mathbf{R}$  and a function  $g$  (with suitable domain) such that  $X = g(Y)$  and  $\mathcal{F}_X \subset \mathcal{F}_Y$ .
3. Suppose  $P$  and  $Q$  are two probability functions defined on the same sample space  $\Omega$  and sigma algebra  $\mathcal{F}$ .
- (a) Show that if  $P(A) = Q(A)$  for all  $A \in \mathcal{F}$  such that  $P(A) \leq \frac{1}{2}$ , then in fact  $P(A) = Q(A)$  for all  $A \in \mathcal{F}$ .
- (b) Show by means of an explicit example that if instead we only have  $P(A) = Q(A)$  for all  $A \in \mathcal{F}$  such that  $P(A) < \frac{1}{2}$ , then  $P$  and  $Q$  need not agree on all of  $\mathcal{F}$ .
4. Let  $(\Omega, \mathcal{F}, \Pr)$  be a probability space and let  $X$  and  $Y$  be random variables with respect to  $\mathcal{F}$ . If  $A \in \mathcal{F}$ , define  $Z : \Omega \rightarrow \mathbf{R}$  by

$$Z(\omega) = \begin{cases} X(\omega) & \omega \in A \\ Y(\omega) & \omega \notin A. \end{cases}$$

- (a) Show that  $Z$  is a random variable with respect to  $\mathcal{F}$ .
- (b) Show that if instead  $A \subseteq \Omega$  is not an event, i.e.  $A \notin \mathcal{F}$ ,  $Z$  need not be a random variable.
5. On the probability space  $(\Omega, \mathcal{F}, \Pr)$ , let  $Z$  be a random variable such that  $\Pr(Z > 0) > 0$ . Explain carefully why there exists  $\delta > 0$  such that  $\Pr(Z \geq \delta) > 0$ .
6. In this question, you will derive the mean and variance of the hypergeometric distribution.
- (a) If  $X \sim \text{BINOMIAL}(n, p)$ , we can write  $X = \sum_{i=1}^n Z_i$ , where  $Z_i \sim \text{BERNOULLI}(p)$  are independent. Use this representation to show that  $E(X) = np$  and  $\text{Var}(X) = np(1 - p)$ .

Suppose now that  $X$  is hypergeometric, representing the distribution of the number of red balls in a sample of size  $n$  drawn without replacement from an urn containing  $r$  red and  $w$  white balls,  $N = r + w$ . In this case,

$$\Pr(X = x) = \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}.$$

As in the binomial case, we can represent  $X$  as a sum of Bernoulli variables:  $X = \sum_{i=1}^n Z_i$ , where  $Z_i$  takes the value 1 if the  $i$ th ball is red and 0 otherwise.

- (b) What is the distribution of the  $Z_i$ ? Are they independent?
- (c) Show that  $E(X) = n \frac{r}{N}$ .
- (d) (Harder) Show that  $\text{Var}(X) = n \frac{r}{N} \frac{w}{N} \frac{N-n}{N-1}$ .

### For discussion

7. For real numbers  $s > 1$ , define the Riemann zeta function as

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

Let  $s > 1$  be fixed, and let the random variable  $X$  have probability mass function

$$f_X(x) = \Pr(X = x) = \frac{1}{x^s} \frac{1}{\zeta(s)}, \quad x \geq 1.$$

Let  $D_k$  be the event that  $X$  is divisible by  $k$ , for  $k \geq 2$ .

- (a) What is  $\Pr(D_k)$ ?
- (b) Show that the events  $\{D_p : p \text{ is prime}\}$  are independent.
- (c) Prove Euler's formula for the zeta function in terms of the prime numbers:

$$\zeta(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

*Hint: You may assume that whenever a collection  $\{A_i : i \in I\}$  of events is independent, so is the collection  $\{A_i^c : i \in I\}$ . Recall also that for a countable collection of independent events,*

$$\Pr\left(\bigcap_{i=1}^{\infty} A_i\right) = \prod_{i=1}^{\infty} \Pr(A_i).$$

8. In this question, we look what happens to the geometric distribution when we pass from discrete to continuous time. Let  $T$  have the waiting time geometric distribution with parameter  $p$ , so that

$$\Pr(T \geq j) = (1 - p)^j, \quad j = 0, 1, 2, \dots$$

We think of  $T$ , which takes non-negative integer values, as the number of units of time we need to wait for an event to occur. When  $p$  is very small,  $T$  typically takes very large values, so we seek to rescale time, so that the waiting times are given in more reasonable units. Let  $M$  be a large number, such that  $a = pM$  and  $t = \frac{j}{M}$  are both small relative to  $M$ . What is the distribution of  $U = \frac{T}{M}$ , in terms of the parameter  $a$ ? What important property has been preserved in this limit?