Probability for Statistics
Problem Sheet 2

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

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

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

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

Objective: gain familiarity with the definition of a random variable.

X_1 is not a random variable with respect to $\mathcal{F}$. To see this, consider the Borel set $\{1\} \in \mathcal{B}$. $X_1^{-1}(\{1\}) = \{1\} \notin \mathcal{F}$.

The image of X_2 is the set $\{0,1\}$, so for any Borel set $B \in \mathcal{B}$, $X_2^{-1}(B) = X_2^{-1}(B \cap \{0,1\})$.
So then

$$X_2^{-1}(B) = \begin{cases} \emptyset & 0 \notin B, 1 \notin B \\ \{2,4,6\} & 0 \in B, 1 \notin B \\ \{1,3,5\} & 0 \notin B, 1 \in B \\ \Omega & 0 \in B, 1 \in B. \end{cases}$$

Hence the pre-image of every Borel set is in $\mathcal{F}$, so X_2 is a random variable.

2. (a) Let $X : \Omega \to \mathbf{R}$ 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 Ω.
   (b) Consider an experiment in which a fair coin is flipped twice, so that the sample space is $\Omega = \{HH, HT, TH, TT\}$. Let $X : \Omega \to \mathbf{R}$ take the value 1 if precisely one flip comes up heads, and 0 otherwise. Determine the sigma algebra $\mathcal{F}_X$.
   (c) For Ω as in the previous part, give an example of a function $Y : \Omega \to \mathbf{R}$ and a function g (with suitable domain) such that $X = g(Y)$ and $\mathcal{F}_X \subset \mathcal{F}_Y$.

Objective: Introduce the sigma algebra generated by a random variable. Gain familiarity with the sigma algebra as encoding the information available from an experiment.

(a) There are three properties to check.
   i. First, since $X^{-1}(\emptyset) = \emptyset$, clearly $\emptyset \in \mathcal{F}_X$.
   ii. Now suppose $A \in \mathcal{F}_X$, say $A = X^{-1}(B)$, for $B \in \mathcal{B}$. Since $\mathcal{B}$ is a sigma algebra, $\mathbf{R} \backslash B \in \mathcal{B}$, and $X^{-1}(\mathbf{R} \backslash B) = A^c$, so that $A^c \in \mathcal{F}_X$.
   iii. Suppose now that $A_1, A_2, \ldots \in \mathcal{F}_X$, say $A_i = X^{-1}(B_i)$, for $B_i \in \mathcal{B}$. Then $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$, since $\mathcal{B}$ is a sigma algebra, so then $\bigcup_{i=1}^{\infty} A_i = X^{-1}(\bigcup_{i=1}^{\infty} B_i) \in \mathcal{F}_X$.

(b) *The two possible values of the function X are 0 and 1, so the pre-image of a Borel set B depends on which of these elements it contains.*

$$X^{-1}(B) = \begin{cases} \emptyset & 0 \notin B, 1 \notin B \\ \{HH, TT\} & 0 \in B, 1 \notin B \\ \{HT, TH\} & 0 \notin B, 1 \in B \\ \Omega & 0 \in B, 1 \in B. \end{cases}$$

So then $\mathcal{F}_X = \{\emptyset, \{HH, TT\}, \{HT, TH\}, \Omega\}$.

(c) *Let Y count the number of heads in two flips of the coin. Define $X = g(Y) = Y \mod 2$. Then by considering $Y^{-1}(B)$ for the Borel sets $B = \{0\}, \{1\}, \{2\}$, we see that $\mathcal{F}_Y$ is the set*

$$\{\emptyset, \{HH\}, \{TT\}, \{HT, TH\}, \{HH, TT\}, \{HH, HT, TH\}, \{HT, TH, TT\}, \Omega\}.$$

Reflect: $\mathcal{F}_X$ is the sigma algebra generated by X. It is the smallest sigma algebra with respect to which X is a random variable. We think of a sigma algebra $\mathcal{F}$ as encoding the information we can obtain from an experiment. Even though we might not know which $\omega \in \Omega$ occurs, we do know, for each $E \in \mathcal{F}$, whether or not $\omega \in E$. $\mathcal{F}_X$ is the least information we need to be able to extract from the experiment if we are to be able to determine the value of X for any $\omega \in \Omega$.

3. Suppose P and Q are two probability functions defined on the same sample space Ω 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 $\mathcal{F}$.

   (a) *Suppose $A \in \mathcal{F}$. If $P(A) \leq \frac{1}{2}$ then certainly $P(A) = Q(A)$ by hypothesis. If not, then $P(A^c) = 1 - P(A) < \frac{1}{2}$ so $P(A^c) = Q(A^c)$, so $1 - P(A) = 1 - Q(A)$, giving $P(A) = Q(A)$.*

   (b) *Take $\Omega = \{0, 1\}$ and $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$. Define*

   $$P(\{0\}) = P(\{1\}) = \frac{1}{2}$$

   and

   $$Q(\{0\}) = \frac{1}{3} \qquad Q(\{1\}) = \frac{2}{3},$$

   with

   $$P(\emptyset) = Q(\emptyset) = 0, \qquad P(\Omega) = Q(\Omega) = 1.$$

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 $Z : \Omega \to \mathbf{R}$ 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.

(a) Let $B \in \mathcal{B}$, then,

$$\begin{aligned} Z^1(B) = \{\omega \in \Omega : Z(\omega) \in B\} &= \{\omega \in A : X(\omega) \in B\} \cup \{\omega \in A^c : Y(\omega) \in B\} \\ &= (X^{-1}(B) \cap A) \cup (Y^{-1}(B) \cap A^c) \end{aligned}$$

Since $A, A^c \in \mathcal{F}$ and $X^{-1}(B), Y^{-1}(B) \in \mathcal{F}$ by definition of random variable, it follows that $Z^{-1}(B) \in \mathcal{F}$.

(b) Consider $\Omega = \{0, 1\}$ and the trivial sigma algebra $\mathcal{F} = \{\emptyset, \Omega\}$. Let $X, Y : \Omega \to \mathbf{R}$ be the constant random variables $X(\omega) = 0, Y(\omega) = 1, \forall \omega \in \Omega$. Clearly, these are both random variables with respect to $\mathcal{F}$. Let $A = \{0\} \in \Omega$ and $A \notin \mathcal{F}$. Then, $Z(0) = X(0) = 0$ and $Z(1) = Y(1) = 1$. In this case, Z is not a random variable with respect to $\mathcal{F}$ since $Z^{-1}(\{0\}) = \{0\} \notin \mathcal{F}$.

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 $\Pr(Z \geq \delta) > 0$.

For $n \geq 1$, define $A_n = \{Z \geq \frac{1}{n}\}$. Then $A_1 \subseteq A_2 \subseteq \ldots$ is an increasing sequence of events, and

$$\{Z > 0\} = \bigcup_{n=1}^{\infty} A_n.$$

Applying the continuity property of $\Pr$, this then gives ,

$$\Pr(\{Z > 0\}) = \Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \Pr(A_n).$$

If it were the case that $\Pr(A_n) = 0$ for all $n \geq 1$, then the right hand limit would be 0, which is a contradiction. Hence there exists $n \geq 1$ such that $\Pr(A_n) > 0$, so taking $\delta = \frac{1}{n}$ suffices.

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 $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.

$\mathrm{E}(Z_i) = p$*, so the result for* $\mathrm{E}(X)$ *follows immediately by linearity of expectation.*

Reflect: How else could you do this calculation? Generating functions, or evaluating a combinatorial identity, would work just as well.

$\mathrm{E}(Z_i^2) = p$*, so* $\mathrm{Var}(Z_i) = p - p^2 = p(1-p)$*. The result for* $\mathrm{Var}(X)$ *then follows from the independence of the Z_i.*

Reflect: As we see from the next part, independence really is a necessary assumption here.

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 ith ball is red and 0 otherwise.

(b) What is the distribution of the Z_i ? Are they independent?

Let an ordered configuration be a sequence of n removed balls which can be chosen by sampling without replacement. By looking at how many balls we have left to choose from after taking each of the n balls, we see that there are $N(N-1)\dots(N-n+1)$ possible ordered configurations. Rewriting, we see that there are $N!/(N-n)!$ ordered configurations of removed balls, and each of these is equally likely. There is also a one-to-one correspondence between ordered configurations with a red ball in position i and those with a red ball in position j: explicitly, there are $r(N-1)!/(N-1-(n-1))! = r(N-1)!/(N-n)!$ of each of these.

Hence $\Pr(Z_i = 1) = \frac{r(N-1)!/(N-n)!}{N!/(N-n)!} = \frac{r}{N} = \Pr(Z_j = 1)$. *So* $Z_i \sim \textsc{Bernoulli}(\frac{r}{N})$.

The variables Z_i and Z_j for $i \neq j$ are clearly not independent since

$$\Pr(Z_i = 1, Z_j = 1) = \frac{r}{N}\frac{(r-1)}{N-1} \neq \Pr(Z_i = 1)\Pr(Z_j = 1)$$

for $i \neq j$.

(c) Show that $\mathrm{E}(X) = n\frac{r}{N}$.

This follows immediately from the previous answer, using linearity of expectation.

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(Z_i) = n\frac{r}{N}.$$

(d) (Harder) Show that $\mathrm{Var}(X) = n\frac{r}{N}\frac{w}{N}\frac{N-n}{N-1}$.

$$\begin{aligned}
\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{n} Z_i\right) &= \sum_{i=1}^{n} \mathrm{Var}(Z_i) + 2\sum_{i<j} \mathrm{Cov}(Z_i, Z_j) \\
&= n\frac{r}{N}\frac{w}{N} + n(n-1)\left(\frac{r}{N}\frac{(r-1)}{N-1} - \frac{r^2}{N^2}\right) \\
&= n\frac{r}{N}\left[\frac{w}{N} + (n-1)\left(\frac{r-1}{N-1} - \frac{r}{N}\right)\right] \\
&= n\frac{r}{N}\left[\frac{w}{N} + (n-1)\frac{r-N}{N(N-1)}\right] \\
&= n\frac{r}{N}\frac{w}{N}\left[1 - \frac{n-1}{N-1}\right] = n\frac{r}{N}\frac{w}{N}\frac{N-n}{N-1}.
\end{aligned}$$

where the first equality follows since $\mathrm{E}[Z_i] = \frac{r}{N}$ for all i, and for any $i < j$, $\mathrm{E}[Z_iZ_j] = \frac{r}{N}\frac{r-1}{N-1}$ since once we have chosen the first ball, we are left with $r-1$ red balls in $N-1$ total balls, hence $\mathrm{Cov}(Z_i, Z_j) = \frac{r}{N}\frac{(r-1)}{N-1} - \frac{r^2}{N^2}$.

Reflect: Look carefully at the form of this expression. Why is it zero when $N = n$? Why is it equal to the expression for the binomial variance when $n = 1$?

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

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

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

(a) What is $\Pr(D_k)$?
If X is divisible by k then $X = km$, for some positive integer m, so sum over all such numbers.

$$\Pr(D_k) = \frac{1}{\zeta(s)} \sum_{m=1}^{\infty} \left(\frac{1}{mk}\right)^s = \frac{1}{k^s}\frac{1}{\zeta(s)} \sum_{m=1}^{\infty} \frac{1}{m^s} = \frac{1}{k^s}.$$

(b) Show that the events $\{D_p : p \text{ is prime}\}$ are independent.
Let $p_1, \ldots p_r$ be distinct primes. Then the event $D_{p_1} \cap \ldots \cap D_{p_r}$ occurs if and only if X can be written as a product $p_1 \ldots p_r m$ for some positive integer m. As in the previous part, this then gives

$$\Pr(\cap_{k=1}^r D_k) = \frac{1}{\zeta(s)} \sum_{m=1}^{\infty} \left(\frac{1}{p_1 \ldots p_r m}\right)^s = \prod_{k=1}^{r} \frac{1}{p_k^s}.$$

Note that the right hand side is the product of the probabilities of the intersected events and we are done.

(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,

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

The events $\{D_p : p \text{ is prime}\}$ are independent, and so, by the hint, are the events $\{D_p^c : p \text{ is prime}\}$. Note that D_p^c is the event that X is not divisible by p, which has probability $1 - \frac{1}{p^s}$. Again using the hint, we have that

$$\Pr\left(\bigcap_{p \text{ prime}} D_p^c\right) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right).$$

Now X can only fail to be divisible by all primes when X takes the value 1, and $\Pr(X = 1) = \frac{1}{\zeta(s)}$. Equating the two probabilities and taking reciprocals gives Euler's formula.

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

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

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 occcur. 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?

$$\Pr(U \geq u) = \Pr\left(\frac{T}{M} \geq u\right) = \Pr(T \geq uM) = (1-p)^{uM} = \left(1 - \frac{a}{M}\right)^{uM} \approx e^{-au}.$$

So U has the exponential distribution with rate a. This limiting distribution inherits the memoryless property.