Probability for Statistics
Problem Sheet 1

The first three questions should be accessible once you have watched up to lecture 3. Questions 4 and 5 should be accessible once you have watched up to lecture 6. Question 6 is a skills question, extending ideas that should be familiar from last year. Question 7 is an optional discussion question, for interest.

1. Let Ω be a set.

   (a) Show that the collection $\mathcal{F} = \{\emptyset, \Omega\}$ is a sigma algebra.

   (b) Show that for any subset $E \subseteq \Omega$, $\mathcal{F}_E = \{\emptyset, E, E^c, \Omega\}$ is a sigma algebra.

   (c) Let $\mathcal{F}$ be the collection of all subsets of Ω. Show that $\mathcal{F}$ is a sigma algebra.

   (d) Show that the intersection of two sigma algebras on Ω is a sigma algebra.

   (e) Give an example to show that the union of two sigma algebras on Ω need not be a sigma algebra.

   Objective: to recall definitions of sigma algebras and related concepts. To fill in some claims left as exercises from lectures.

   To show that $\mathcal{F}$ is a sigma algebra we must verify (i) $\emptyset \in \mathcal{F}$; (ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$; and (iii) if $A_1, A_2, \ldots \in \mathcal{F}$ then $\cup_{k=1}^{\infty} A_k \in \mathcal{F}$.

   (a) (i) $\emptyset \in \{\emptyset, \Omega\}$, (ii) $\emptyset^c = \Omega \in \{\emptyset, \Omega\}$ and $\Omega^c = \emptyset \in \{\emptyset, \Omega\}$, and (iii) $\emptyset \cup \Omega = \Omega \in \{\emptyset, \Omega\}$.

   (b) (i) Certainly $\emptyset \in \mathcal{F}_E$, (ii) Clearly $A^c \in \mathcal{F}_E$ for each $A \in \mathcal{F}_E$, (iii) The only non-trivial union to check is $E \cup E^c = \Omega$.

   (c) (i) $\emptyset$ is a subset of any set, so $\emptyset \subseteq \Omega$ and thus $\emptyset \in \mathcal{F}$; (ii) if $A \in \mathcal{F}$, then $A \subseteq \Omega$. But $A \subseteq \Omega$ means that $A^c \subseteq \Omega$, which in turn implies $A^c \in \mathcal{F}$; (iii) If $A_1, A_2, \ldots \in \mathcal{F}$, then each $A_k \subseteq \Omega$ and $\cup_{k=1}^{\infty} A_k \subseteq \Omega$. But this means that $\cup_{k=1}^{\infty} A_k \in \mathcal{F}$. **Reflect: trivial, since the sigma algebra axioms are closure properties for collections of subsets of Ω. The power set clearly contains all subsets of Ω, so instantly satisfies the axioms.**

   (d) (Simpler version of the argument given in lectures.) Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be the two sigma algebras. (i) $\emptyset \in \mathcal{F}_1$ and $\emptyset \in \mathcal{F}_2$, since $\mathcal{F}_1$ and $\mathcal{F}_2$ are both sigma algebras. Thus $\emptyset \in \mathcal{F}_1 \cap \mathcal{F}_2$; (ii) If $A \in \mathcal{F}_1 \cap \mathcal{F}_2$ then $A \in \mathcal{F}_1$. Because $\mathcal{F}_1$ is a sigma algebra, this means $A^c \in \mathcal{F}_1$. By the same reasoning, $A^c \in \mathcal{F}_2$, thus $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$; (iii) if $A_1, A_2, \ldots \in \mathcal{F}_1 \cap \mathcal{F}_2$, then $A_1, A_2, \ldots \in \mathcal{F}_1$. Because $\mathcal{F}_1$ is a sigma algebra, this means $\cup_{k=1}^{\infty} A_k \in \mathcal{F}_1$. By the same reasoning, $\cup_{k=1}^{\infty} A_k \in \mathcal{F}_2$, thus $\cup_{k=1}^{\infty} A_k \in \mathcal{F}_1 \cap \mathcal{F}_2$.
   Reflect: as we saw in lectures, this argument extends to arbitrary intersections of sigma algebras. Indeed, this result forms the basis for our construction of the Borel sigma algebra on R.

   (e) Define $\Omega = \{0, 1, 2\}$, and consider the sigma algebras $\mathcal{F}_0 = \{\emptyset, \{0\}, \{1, 2\}, \{0, 1, 2\}\}$ and $\mathcal{F}_1 = \{\emptyset, \{1\}, \{0, 2\}, \{0, 1, 2\}\}$. Then

   $$\mathcal{F}_0 \cup \mathcal{F}_1 = \{\emptyset, \{0\}, \{1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\},$$

   which is clearly not a sigma algebra because e.g. $\{0\} \cup \{1\} \notin \mathcal{F}_0 \cup \mathcal{F}_1$.

2. Suppose a fair coin is flipped repeatedly, and that flips are independent. Use the continuity property of the probability function $\Pr$ to show that, with probability 1, the coin will eventually land heads up.

Objective: Understand the continuity property by means of a concrete example.

Let A_n be the event that the coin lands tails on the nth flip. Then by the continuity property applied to the decreasing sequence $B_N = \bigcap_{n=1}^{N} A_n$, we have

$$\Pr(\textit{no heads}) = \Pr\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{N\to\infty} = \Pr\left(\bigcap_{n=1}^{N} A_n\right) = \lim_{N\to\infty} 2^{-N} = 0.$$

Hence the complementary event that there is at least one head has probability 1.

3. Let $\Omega = [0,1]$, the unit interval. Define $\mathcal{F}$ to be the collection of all countable or co-countable subsets of Ω, where a co-countable set is one whose complement is countable.

   (a) Show that $\mathcal{F}$ is a sigma algebra. *[Hint: Is a countable union of countable sets countable?]*

   (b) Define the function $P : \mathcal{F} \to [0,1]$ by

$$P(A) = \begin{cases} 0 & \text{if } A \text{ is countable} \\ 1 & \text{if } A \text{ is co-countable} \end{cases}.$$

   Determine whether or not P is countably additive.

Objective: gain practice working with the defining axioms of a sigma algebra, and the definition of countable additivity for a probability function.

(a) *(i) Clearly $\emptyset \in \mathcal{F}$.*

(ii) If $A \in \mathcal{F}$ then either A is countable or its complement is. In either case, it follows that $A^c \in \mathcal{F}$.

(iii) Suppose $A_1, A_2, \ldots$ is a sequence of sets in $\mathcal{F}$. There are two cases - if all of the sets are countable then, as a countable union of countable sets is countable, their union is also in $\mathcal{F}$. If one of the sets, say A_j is co-countable, then the union is co-countable, since if $x \in (\cup_{i=1}^{\infty} A_i)^c$, then it follows that $x \notin A_j$ so $x \in A_j^c$. So the complement of the union is a subset of a countable set, and hence is countable.

(b) Suppose $\{A_k, k = 1, 2, \ldots\}$ is a countable sequence of pairwise disjoint sets in $\mathcal{F}$. Note that at most one of $\{A_k, k = 1, 2, \ldots\}$ is cocountable, by the following argument. Suppose $A_k \subset \mathcal{F}$ for $k = 1, 2, \ldots$, $A_k \cap A_j = \emptyset$ for $k \neq j$, and that for some k_0, A_{k_0} is cocountable. For $k \neq k_0$, $A_k \cap A_{k_0} = \emptyset$ so $A_k \subseteq A_{k_0}^c$. Because $A_{k_0}^c$ is countable, A_k is also countable. Thus there is at most one cocountable set among $\{A_1, A_2, \ldots\}$. So there are two cases to consider

IF ALL ARE COUNTABLE: *A countable union of countable sets is countable, so $P\left(\cup_{k=1}^{\infty} A_k\right) = 0$. At the same time, $P(A_k) = 0$ for all k, so $\sum_{k=1}^{\infty} P(A_k) = 0 = P\left(\cup_{k=1}^{\infty} A_k\right)$.*

IF EXACTLY ONE IS COCOUNTABLE: *Let k_0 be the index of the cocountable set. By the argument in part (a), $\cup_{k=1}^{\infty} A_k$ is cocountable and $P\left(\cup_{k=1}^{\infty} A_k\right) = 1$. At the same time, $\sum_{k=1}^{\infty} P(A_k) = 1$ because $P(A_{k_0}) = 1$ while $P(A_k) = 0$ for $k \neq k_0$. Thus, $P\left(\cup_{k=1}^{\infty} A_k\right) = 1 = \sum_{k=1}^{\infty} P(A_k)$.*

In both cases, P is countably additive.

4. 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 $\mathcal{F}$.

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

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.

5. (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 $\omega \in \Omega$.

6. (Review and extension of elementary probability.) In this question, you will derive the mean and variance of the hypergeometric distribution.

Objective: Develop familiarity with the idea of representing a random variable as a sum of indicators. Refresh understanding of properties of expectation, and understanding of binomial coefficients. Reflect on the differences between sampling with and without replacement.

(a) (Warm up) If $X \sim \textsc{Binomial}(n, p)$, we can write $X = \sum_{i=1}^n Z_i$, where $Z_i \sim \textsc{Bernoulli}(p)$ 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?
Each of the $N!/(N-n)!$ *ordered configurations of removed balls is equally likely, and there is 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 $\Pr(Z_i = 1) = \frac{r}{N} = \Pr(Z_j = 1)$. *So* $Z_i \sim \text{BERNOULLI}(\frac{r}{N})$.

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

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

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 $n = 1$?

Optional question for group 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.