Probability for Statistics
Problem Sheet 1

Questions marked (r) are revision of concepts from last year.

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

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

since for any N, since the coin tosses are independent $\Pr(B_N) = 2^{-N}$.

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 co-countable, 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 co-countable. 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 co-countable 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 $P(\cup_{k=1}^\infty A_k) = 0$. At the same time, $P(A_k) = 0$ for all k, so $\sum_{k=1}^\infty P(A_k) = 0 = P(\cup_{k=1}^\infty A_k)$.*

IF EXACTLY ONE IS CO-COUNTABLE: *Let k_0 be the index of the co-countable set. By the argument in part (a), $\cup_{k=1}^\infty A_k$ is co-countable and $P(\cup_{k=1}^\infty A_k) = 1$. 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, $P(\cup_{k=1}^\infty A_k) = 1 = \sum_{k=1}^\infty P(A_k)$.*

In both cases, P is countably additive.

4. Consider the probability space $(\Omega, \mathcal{F}, \Pr)$ with $A, B \in \mathcal{F}$. Using only the Kolmogorov axioms prove

(a) $\Pr(A) \leq 1$,

(b) If $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$,

(c) $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

Objective: revise manipulations of probability from first year. Practise writing clear derivations from a system of axioms.

(a) Because $A \in \mathcal{F}$, we know $A^C \in \mathcal{F}$ and $\Pr(A^C)$ is defined. We also know that $A \cup A^C = \Omega$ and $A \cap A^C = \emptyset$, so by finite additivity, $\Pr(A) + \Pr(A^C) = \Pr(\Omega)$. By the first axiom $\Pr(A^C) \geq 0$ and by the second axiom $\Pr(\Omega) = 1$. Thus, $\Pr(A) \leq 1$.

(b) Since $A \subseteq B$, $B = A \cup (B \cap A^C)$ and $A \cap (B \cap A^C) = \emptyset$. We can easily deduce that $B \cap A^C \in \mathcal{F}$ and thus has a defined probability. Thus, by finite additivity $\Pr(B) = \Pr(A) + \Pr(B \cap A^C) \geq \Pr(A)$ because $\Pr(B \cap A^C) \geq 0$ by the first axiom.

(c) First we notice that $A \cap B^C$, $A \cap B$, and $B \cap A^C$ are all in $\mathcal{F}$ and thus all have defined probabilities. Note

i. Now $A = (A \cap B^C) \cup (A \cap B)$ and $(A \cap B^C) \cap (A \cap B) = \emptyset$ so, again by finite additivity, $\Pr(A) = \Pr(A \cap B^C) + \Pr(A \cap B)$

ii. Likewise, $\Pr(B) = \Pr(B \cap A^C) + \Pr(A \cap B)$

iii. Finally, $A \cup B = (A \cap B^C) \cup (A \cap B) \cup (B \cap A^C)$ and the three sets on the right hand side are disjoint. Using finite additivity again, $\Pr(A \cup B) = \Pr(A \cap B^C) + \Pr(A \cap B) + \Pr(B \cap A^C)$. Substituting in $\Pr(A)$ and $\Pr(B)$ using the equations derived in (i) and (ii) gives the desired result.

5. (r) Suppose X is a set containing n elements. If A and B are randomly chosen subsets of X, what is the probability that $A \subseteq B$?

Objective: develop understanding of probability by considering an unfamiliar setting.

There are (at least) two ways of computing this probability. First we argue by conditioning on the size of the set A. Suppose $|A| = k$, for $0 \leq k \leq n$. There are $\binom{n}{k}$ subsets of size k and 2^n subsets of X altogether, so

$$\Pr(|A| = k) = \frac{\binom{n}{k}}{2^n}.$$

When A has k elements, the probability that all elements are also found in B is 2^{-k}. Now we compute

$$\Pr(A \subseteq B) = \sum_{k=0}^{n} \Pr(A \subseteq B \mid |A| = k) \Pr(|A| = k) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} 2^{-k}.$$

Use the binomial theorem to evaluate this sum as $\left(\frac{3}{4}\right)^n$.

Reflect: the form of the answer suggests that a more direct approach should be possible. Note that any $x \in X$ is in precisely half of all subsets of X. So each element of X has probability $\frac{1}{2}$ of being in a randomly chosen subset. For two randomly chosen subsets A and B, $A \subseteq B$ precisely when the events $\{x \in A \cap x \notin B\}$ do not occur, for each $x \in X$. By independence, each such event has probability $\frac{1}{4}$, so the desired probability is again $\left(\frac{3}{4}\right)^n$.

6. (r) Consider two coins, of which one is a normal fair coin and the other is biased so that the probability of obtaining a Head is $p > 1/2$.

   (a) Suppose $p = 1$ and a coin is selected at random and flipped n times, with flips mutually independent. Evaluate the conditional probability that the selected coin is the normal one, given that the first n flips are all Heads.

   (b) Now suppose $1/2 < p < 1$ and that again, one of the coins is selected randomly and flipped n times. Let E be the event that the n tosses result in k Heads and $n - k$ Tails, and let F be the event that the coin is fair. Find $\Pr(F|E)$.

   Objective: practice manipulating conditional probabilities and using Bayes' theorem.

   Let E and F be the events that the sequence of n flips results in k Heads, and that the coin is fair respectively. Then

$$\Pr(F|E) = \frac{\Pr(E|F)\Pr(F)}{\Pr(E|F)\Pr(F) + \Pr(E|F^c)\Pr(F^c)}.$$

   (a) Setting $k = n$, $\Pr(E|F) = \left(\frac{1}{2}\right)^n$, $\Pr(E|F^c) = 1$, $\Pr(F) = \Pr(F^c) = \frac{1}{2}$, and hence $\Pr(F|E) = \dfrac{1}{1+2^n}$.

   (b) $\Pr(E|F) = \binom{n}{k}\left(\frac{1}{2}\right)^n$, $\Pr(E|F^c) = \binom{n}{k}p^k(1-p)^{n-k}$, $\Pr(F) = \Pr(F^c) = \frac{1}{2}$, and hence

$$\Pr(F|E) = \frac{1}{1 + 2^n p^k (1-p)^{n-k}}.$$

Optional questions for group discussion

The next two questions are adapted from the work of Kahneman and Tversky, psychologists who studied subjective perceptions of probability. See the Additional Resources section on Blackboard!

7. A city contains two hospitals, one small and one large. The long-term proportion of boys born can be taken to be 50%. On a given day, the proportion of boys born in one of the hospitals is 55%. Which hospital is this more likely to be?

   Objective: to illustrate the relationship between sampling variance and sample size.

   The number of male babies born in a hospital with n births can be approximately modelled as $X \sim \textsc{Binomial}(n, \frac{1}{2})$. This has variance $\frac{n}{4}$. The proportion of boys is then $\frac{X}{n}$, which has variance $\frac{1}{4n}$. So the smaller hospital has the larger variance, and a more extreme deviation, such as 55%, is more plausible there.

8. Tom is an opera buff who enjoys touring art museums when on holiday. Growing up, he enjoyed playing chess with family members and friends. Which of the following two situations is more likely?

   (a) Tom plays trumpet for a major symphony orchestra.

   (b) Tom is a farmer.

Objective: to illustrate the base rate fallacy.

Considering the given information, our preconceived ideas about farmers and professional trumpet players might be expressed as $\Pr(Data|Farmer) < \Pr(Data|Trumpeter)$*, but we are interested in the inverse probabilities,* $\Pr(Farmer|Data)$ *and* $\Pr(Trumpeter|Data)$*. From Bayes' theorem, these probabilities depend not only on the* $\Pr(Data|\cdot)$ *probabilities but also on the prior probabilities of encountering a farmer or a trumpeter in a major symphony orchestra. There are vastly more farmers than symphony trumpet players, so much so that Tom is more likely to be a farmer.*

9. Consider the data in Table 1, comparing two treatments for kidney stones.

Table 1: Success rates for the treatment of kidney stones by open surgery (Treatment A) and percutaneous nephrolithotomy (Treatment B).

	Treatment A	Treatment B
Small stones	81 / 87 (93%)	234 / 270 (87%)
Large stones	192 / 263 (73%)	55 / 80 (69%)
Total	273 / 350 (78%)	289 / 350 (83%)

Which treatment is better? Evaluate the following two responses:

I. 83% of the time, Treatment B was successful, whereas Treatment A was successful only 78% of the time. So Treatment B is better.

II. For patients with small stones, Treatment A was 93% successful whereas Treatment B was 87% successful. Similarly, for large stones, Treatment A was 73% successful, while treatment B was 69% successful. So Treatment A is better.

Overall, larger stones are more difficult to treat. Treatment A was evaluated more on larger stones (263/350 cases), whereas Treatment B was evaluated more on smaller stones (270/350 cases). This means that the overall numbers favouring Treatment B are unreliable, as success of the treatment is confounded with stone size. When we take this additional information into account, treatment A is better.

There is something worrying here, though: how do we know that there isn't some other confounding factor such that, when we control for it, the direction reverses again? Perhaps the real conclusion is that, since the treatment groups are highly imbalanced for the two sizes of stone, the allocation of patients to treatments was apparently not random. In principle, this makes a fair comparison of the treatments impossible from the data alone.