Revision Exercises for MA 413 – Statistics for Data Science

This sheet will remind you of basic concepts from probability and statistics that you will need for the material in the course, and can be considered as "warm-up" exercises.

1, Set Theory Let Ω be the sample space and let $B_1, B_2, \ldots,$ be events in this space. Define $C_n = \cup_{i=n}^{\infty} B_i$ and $D_n = \cap_{i=n}^{\infty} B_i$.

(a) Demonstrate that $C_1 \supset C_2 \supset \ldots$ and that $D_1 \subset D_2 \subset \ldots$.

(b) Show that $\omega \in \cap_{i=1}^{\infty} C_i$ if and only if ω belongs to an infinite number of events B_1, B_2,...

(c) Demonstrate that $\omega \in \cup_{n}^{\infty} D_n$ if and only if ω belongs to all of the events B_1, B_2,..., except possibly finitely many.

and coins, difficult Assume that we are tossing a fair coin, and terminate the tossing when we get exactly two heads in a row. Write down the sample space Ω for the number of tosses required. What is the probability that we need exactly k tosses?

3, Set Theory Assume that the events A and B are independent. Prove that A^c and B^c are independent also, where as usual we define $A^c = \Omega/A$.

4, Set Theory Suppose m events form a partition of the same space Ω, that is the events are disjoint and $\cup_{i=1}^{m} A_i = \Omega$. Assume that $\Pr(B) > 0$. Prove that if $\Pr(A_1|B) < P(A_1)$ then $\Pr(A_i|B) > P(A_i)$ for some $i = 2, \ldots, m$.

5, Dice and coins Assume that we are flipping a coin n times and let θ be the probability of getting a head in a single coin flip. Let X be the number of heads we get in the n independent flips. This means that X has a binomial distribution. Intuition suggests that if n is sufficiently large then X should be close to np. By using the histogram in matlab/r and generating Binomials in a pseduonumber generator average X and explore the distribution. Try especially what happens if $p << 1/n$, $p = 1/n$ or if $p \approx 0.5$.

6, PMFs Take Y such that $\Pr(Y = 2) = \Pr(Y = 3) = 1/10$ and that $\Pr(Y = 5) = 8/10$. a) Plot the CDF of Y. Use F to find $\Pr(1.5 < Y < 4.2)$ and $\Pr(1.5 < Y < 4.2)$.

7, PDFs Define Y to have the PDF given by

$$f_Y(y) = \begin{cases} 1/4 & \text{if} \quad 0 < y < 1 \\ 3/8 & \text{if} \quad 3 < y < 5 \\ 0 & \text{o/w} \end{cases} .$$

(a) Find the CDF of Y.

(b) Let $Z = 1/Y$. Find the pdf of $f_Z(z)$ for Z.

8, PDFs Let X have CDF F. Find the CDF of $Y = \max\{0, X\}$.

9, PDFs Let $X \sim \text{Exp}(\eta)$. Determine $F(x)$ and its inverse.

10, PDFs Let

$$f_{X,Y}(x, y) = \begin{cases} C \cdot (x + y^3) & 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{o/w} \end{cases} . \tag{1}$$

Find $\Pr(X < 1/2|Y \le 1/2)$.

11, PDFs Let $X, Y \sim \text{Uniform}(0, 1)$ and let them be independently generated. Find the PDF of the variables $U = X - Y$ and $V = X/Y$.