ST2132 Tutorial 1 2210

1. Let X, Y Z be RVs, and a, b, c be constants. Show that

   (a) $\mathrm{var}(aX + bY + c) = a^2\,\mathrm{var}(X) + b^2\,\mathrm{var}(Y) + 2ab\,\mathrm{cov}(X, Y)$.

   (b) $\mathrm{cov}(aX + bY + c, Z) = a\,\mathrm{cov}(X, Z) + b\,\mathrm{cov}(Y, Z)$

2. Let $x_1, \ldots, x_n$ be realisations from tossing a fair coin, where n is large, say around 10,000.

   (a) Roughly, what are values of the mean and variance of the x's?

   (b) A statistician organises the data into two columns, so that row 1 has x_1, x_2, row 2 has x_3, x_4, etc. Let y_i be the sum of row i. We may think of the y's as realisations of which random variable? Roughly, what are values of the mean and variance of the y's?

   (c) Another statistician creates two columns of data, with row i having x_i, x_i. Let z_i be the sum of row i. We may think of the z's as realisations of which random variable? Roughly, what are values of the mean and variance of the z's?

3. An urn contains 10 identical balls: three are marked **1** and seven are marked **0**. Let X_1 and X_2 be the results of two random draws without replacement. Imagine repeating the experiment many times and recording the outcomes a matrix with 2 columns and many rows. Thus each row contains a realisation of (X_1, X_2).

   (a) Roughly, what is the fraction of **1** in column 2? How about the fraction of rows showing (**1**,**1**)? Among rows with **1** in column 2, roughly what fraction have **1** in column 1?

   (b) How would you use the matrix in (b) to approximate $\mathrm{E}[X_2|x_1]$ and $\mathrm{var}[X_2|x_1]$, where $x_1 = \mathbf{1}$?

4. Let H and T be the number of heads and tails respectively in n tosses of a coin with probability of head equal to p.

   (a) Write down $\mathrm{E}(H), \mathrm{var}(H), \mathrm{E}(T), \mathrm{var}(T)$ in terms of n and p.

   (b) Calculate $\mathrm{cov}(H, T)$.

   (c) Write down the expectation, variance and distribution of the random vector

$$\begin{bmatrix} H \\ T \end{bmatrix}$$

5. A random experiment has r possible outcomes, with probabilities $p_1, \ldots, p_r$. For $1 \leq i \leq r$, let X_i be the number of times outcome i occurs in n runs of the experiment. Let

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_r \end{bmatrix}$$

and where $\mathbf{p} = (p_1, \ldots, p_r)$.

(a) Complete the following: $\mathbf{X} \sim$ __________.

(b) What is the distribution of X_i? Write down $\mathrm{E}(X_i)$ and $\mathrm{var}(X_i)$ in terms of n and $\mathbf{p}$.

(c) For $i \neq j$, what is the distribution of $X_i + X_j$? By manipulating $\mathrm{var}(X_i + X_j)$, obtain an expression for $\mathrm{cov}(X_i, X_j)$ in terms of n, p_i, p_j.

(d) Write down $\mathrm{E}(\mathbf{X})$ and $\mathrm{var}(\mathbf{X})$ in terms of n and $\mathbf{p}$.

6. Let $x_1, \ldots, x_n, z$ be constants.

(a) Show directly, or by a result from class, that

$$\sum_{i=1}^{n} (x_i - z)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(z - \bar{x})^2$$

(b) Deduce that

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n\bar{x}^2$$

This means the vectors $(x_1 - \bar{x}, \ldots, x_n - \bar{x})$ and $(\bar{x}, \ldots, \bar{x})$ are __________________. Fill in a term from linear algebra.