SOR 201 Examples 4

1. (i) Consider a sequence of n independent Bernoulli trials, each with probability of success p. Let I_i denote the indicator random variable associated with a success in the ith trial ($i = 1, \ldots, n$); also let $X = I_1 + I_2 + \cdots + I_n$.

   (a) Define I_i and hence find $\mathrm{E}(I_i), \mathrm{Var}(I_i)$. Why are $I_1, I_2, \ldots, I_n$ independent random variables ?

   (b) What does the random variable X represent ? What is the probability distribution of X ? Calculate the mean and variance of X using the results in (a).

   (ii) A random sample of size n is drawn, sampling without replacement, from a population of N_1 type 1 items and N_2 type 2 items, where $N_1 + N_2 = N$. Let the random variable X denote the number of type 1 items in the sample. Let

$$I_i = \begin{cases} 1 & \text{if the } i\text{th draw gives a type 1 item} \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Name the probability distribution of X and give its probability function. What is the connection between X and the indicator random variables $I_1, I_2, \ldots, I_n$?

   (b) Explain why

$$\mathrm{P}(I_i = 1) = N_1/N \text{ and } \mathrm{P}(I_i = 1, I_j = 1) = N_1(N_1 - 1)/\{N(N-1)\}, i \neq j.$$

   Then determine $\mathrm{E}(I_i)$, $\mathrm{Var}(I_i)$ and $\mathrm{Cov}(I_i, I_j)$, $\quad i \neq j$, and hence find the mean and variance of X.

   (iii) Suppose n cards marked $1, 2 \ldots, n$ are laid out at random in a row. Let A_i be the event that card i appears in the ith position – described as a match in the ith position. Let S_n denote the total number of matches. For the indicator random variables $\{I_i\}$ corresponding to the events $\{A_i\}$, show that

$$\mathrm{E}(I_i) = \frac{1}{n}; \quad \mathrm{Var}(I_i) = \frac{1}{n} - \frac{1}{n^2}; \quad \mathrm{Cov}(I_i, I_j) = \frac{1}{n(n-1)} - \frac{1}{n^2}, i \neq j.$$

   Hence show that $\mathrm{E}(S_n) = 1$ and $\mathrm{Var}(S_n) = 1$.

2. (i) Find the probability generating function (PGF) of the random variable $X \sim \mathrm{Bin}(n, p)$ and hence its mean and variance. If $Y \sim \mathrm{Bin}\ (m, p)$ and X, Y are independent, find the probability distribution of $X + Y$.

   (ii) The count random variable X has PGF

$$G_X(s) = \frac{1 - s^{M+1}}{(M+1)(1-s)}$$

   where M is a positive integer. Find the probability function of X.

   (iii) A player can score 0,1 or 2 points in a game with respective probabilities $\frac{1}{10}, \frac{6}{10}, \frac{3}{10}$. A sequence of n independent games is played, where n is the value obtained by throwing a fair die. Find the PGF of the total sum of scores obtained by the player and the expected total sum.

/continued overleaf

3. (i) Let the random variable X have the geometric distribution with parameter p

$$\text{i.e.} \quad P(X = x) = pq^{x-1}, \qquad x = 1, 2, \ldots.$$

Show that the PGF of X is

$$G_X(s) = ps/(1 - qs), \qquad |qs| < 1$$

and hence find the mean and variance of X.

(ii) Consider a sequence of independent Bernoulli trials, each with probability of success p. Let the random variable Z denote the number of trials required for r successes to occur.

(a) Explain why

$$Z = X_1 + X_2 + \cdots + X_r$$

where $X_1, \ldots, X_r$ are independent random variables, each with the geometric distribution defined in (i).

(b) Explain why the PGF of Z is given by

$$G_X(s) = \{ps/(1 - qs)\}^r, \qquad |qs| < 1$$

and hence show that

$$P(Z = z) = \binom{z-1}{r-1} p^r q^{z-r}, \qquad z = r, r+1, \ldots.$$

Hint: $1/(1-a)^r = \sum_{i=0}^{\infty} \binom{i+r-1}{i} a^i, \qquad |a| < 1.$

(c) Find the mean and variance of Z.

4. Discrete branching process Consider a population of individuals which can die or reproduce independently of each other with fixed generation time. Suppose the population is of size 1 initially. Let the random variable C denote the number of children of one individual where

$$P(C = k) = \left(\tfrac{1}{2}\right)^{k+1}, \qquad k = 0, 1, 2, \ldots$$

with PGF $G(s)$. Let the random variable X_n be the size of the nth generation with PGF $G_n(s)$.

(a) Find the PGFs $G_0(s), G_1(s), G_2(s)$ and $G_3(s)$.
Hint: Use the result $G_n(s) = G_{n-1}(G(s)), \quad n \geq 1.$

(b) Using the principle of induction, prove that

$$G_n(s) = \frac{n - (n-1)s}{(n+1) - ns}, \qquad n \geq 1.$$

(c) Hence find $P(X_n = 0)$ and $P(X_n = x), \quad x \geq 1$. What is the limit of $P(X_n = 0)$ as $n \to \infty$? Interpret this result.

Additional questions (NOT for handing in)

5. (i) Let I_A, I_B be the indicator random variables for the events A, B respectively, where $\mathrm{P}(A), \mathrm{P}(B) > 0$. Show that

$$\mathrm{Cov}(I_A, I_B) \begin{cases} > 0, \\ = 0, \\ < 0, \end{cases} \quad \text{if } \mathrm{P}(A|B) \begin{cases} > \\ = \\ < \end{cases} \mathrm{P}(A).$$

(ii) Consider n events $A_1, A_2, ...A_n$. Let X be the number of events which occur and define $Y = \begin{cases} 1, & \text{if } X \geq 1 \\ 0, & \text{otherwise.} \end{cases}$

Using indicator r.v.s, prove that $\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathrm{P}(A_i)$.

(iii) For the multinomial distribution with parameters $\{n; p_1, ..., p_k\}$, prove the result $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ by an alternative argument, using the formula

$$\mathrm{Var}(X_i + X_j) = \mathrm{Var}(X_i) + \mathrm{Var}(X_j) + 2\mathrm{Cov}(X_i, X_j).$$

[*Hint*: What is the distribution of $X_i + X_j$?]

6. A bag contains W white balls and B black balls. Balls are taken out one at a time until the first white ball is drawn. Find $\mathrm{E}(X)$, where X is the number of balls withdrawn from the bag.

[*Hint*: Label the black balls $1, 2, ...B$ and let

$$I_i = \begin{cases} 1, & \text{if black ball } i \text{ is withdrawn before any white ball} \\ 0, & \text{otherwise.} \end{cases}$$]

7. Let X be the total score obtained in 3 rolls of a fair die. Show that

$$G_X(s) = \frac{s^3(1-s^6)^3}{6^3(1-s)^3}$$

and derive the value of $\mathrm{P}(X = 14)$. [Use the hint in Qn. 3(ii)(b).]

8. A software representative makes sales to a random number of companies, N, each week, where N is Poisson distributed with parameter $\lambda = \log_e 5$. At company i, the representative sells X_i items, where each X_i has the distribution

$$p_k = \frac{(\frac{4}{5})^k}{k \log_e 5}, \quad k = 1, 2, ...,$$

and $N, X_1, X_2, ...$ are all independent. Find the PGF of T, the total number of items sold in a week, and hence show that T has the modified geometric distribution with parameter $p = \frac{1}{5}$.

[*Hints*: $\log_e(1-x) = -\sum_{k=1}^{\infty} x^k/k \quad \text{for} \quad |x| < 1; \quad e^{-\log_e a} = 1/a$.]