

## Homework 3

### 1. Identifying a pmf from data

Suppose  $\Omega$  is a sample space and  $x : \Omega \rightarrow V$  is a random variable. In this question we will not specify  $\Omega$  and  $x$  directly. Instead we will work with the *induced pmf*  $p^x$  on  $V$ . Here  $V \subset \mathbb{R}$ , given by

$$v = \{v_1, v_2, \dots, v_n\}$$

where

$$v = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 7 \\ 8 \\ 9 \end{bmatrix}$$

The random variable  $x$  has induced pmf  $p^x : V \rightarrow [0, 1]$ . We represent  $p^x$  by a vector  $p^x \in \mathbb{R}^n$ , so that

$$p^x(v_i) = p_i^x$$

Let

$$p^x = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.1 \\ 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}$$

- (a) Find the cdf of  $x$ , and use it to simulate  $x$ . Perform  $m = 10,000$  trials, and collect data points  $y(1), \dots, y(m)$ , where each  $y(i) \in V$ .

Let  $s(j, n)$  denote the frequency of  $v_j$  in the first  $n$  data points, i.e.,

$$s(j, n) = \frac{1}{n} \sum_{i=1}^n I_j(y(i))$$

where  $I_j$  is the indicator function

$$I_j(a) = \begin{cases} 1 & \text{if } a = v_j \\ 0 & \text{otherwise} \end{cases}$$

Plot  $s(j, n)$  against  $n$ , for each  $j$ .

- (b) Plot  $s(4, n)$  against  $n$ , along with the bounds on its 90% confidence intervals given by the Chebyshev inequality.

### 2. The geometric mean and products of IID random variables

- (a) Suppose  $\Omega = \{1, 2, \dots, n\}$  and  $p : \Omega \rightarrow \mathbb{R}$  is the uniform pmf with

$$p(\omega) = \frac{1}{n} \quad \text{for all } \omega \in \Omega$$

Suppose  $x : \Omega \rightarrow \mathbb{R}$  is a random variable. Show that

$$\exp(\mathbb{E} \log x) = \prod_{i=1}^n x(\omega_i)^{1/n}$$

The right-hand side of this expression is the *geometric mean* of the values of  $x$ .

- (b) Suppose  $y_1, y_2, \dots, y_n$  are IID Bernoulli random variables. Each  $y_i$  can take a value of either 6 or  $\frac{1}{3}$ , with

$$\text{Prob}(y_i = 6) = \frac{1}{2} \quad \text{Prob}(y_i = \frac{1}{3}) = \frac{1}{2}$$

Let's define the product random variable  $p_n : \Omega \rightarrow \mathbb{R}$  by

$$p_n = y_1 y_2 \cdots y_n$$

What is the expected value  $E p_n$  as a function of  $n$ ?

- (c) Now define the scalar random variable  $q_n : \Omega \rightarrow \mathbb{R}$  by

$$q_n = \frac{1}{n} \log p_n$$

where  $\log$  is the natural logarithm. Show that  $E q_n = \log \sqrt{2}$ .

- (d) As  $n$  becomes large, describe qualitatively the pmf of  $q_n$ .