

**Homework 4**

Due Tuesday, October 13

Please either give the assignment to Loraine at the CDS or send it via email to the graders **before noon**.

1. *Basketball* A data analyst working for the Golden State Warriors determines that Stephen Curry has a 60% of making his next shot if he has made the previous one, but only 30% if he has missed it; the shots before the previous one don't seem to matter. Random vectors or sequences with this structure are called Markov chains.
  - a. What is the probability that Curry makes his 14th shot if he has made his 2nd and 12th shots (we don't know whether he made or missed the rest)?
  - b. The probability that he makes his first shot is 40%. What is the probability that he made it if he did not make his 3rd or his 9th (we don't know whether he made or missed the rest)?
  - c. Regardless of whether he makes that first shot or not, what is the expected number of shots that he makes in a row after that?

2. *Model* Consider the joint pmf

$$p_{N,C,K}(n, c, k) = \begin{cases} \frac{1}{30} \binom{n}{k} c^k (1-c)^{n-k} & \text{for } n \in \{1, \dots, 20\}, k \in \{0, 1, \dots, n\}, c = \frac{1}{4}, \\ \frac{1}{60} \binom{n}{k} c^k (1-c)^{n-k} & \text{for } n \in \{1, \dots, 20\}, k \in \{0, 1, \dots, n\}, c = \frac{4}{5}. \end{cases} \quad (1)$$

- a. Compute  $p_N$  and  $p_C$ , the marginal pmfs of  $N$  and of  $C$ . Are  $N$  and  $C$  independent? Justify your answer.
  - b. Give a real-life example that could correspond to this model, i.e. invent some situation involving quantities that could be reasonably modeled as  $N$ ,  $C$  and  $K$ . You can use coin flips, but you are welcome to be more creative.
  - c. Write the expression for the pmf of  $N$  given  $K$  and  $C$  and give it an interpretation according to your example of part (b) (don't worry about simplifying the expression).
  - d. Use Matlab, Julia or NumPy (or whatever you want) to plot the distribution of  $N$  given  $K = 1$ ,  $K = 10$ ,  $K = 19$  and  $C = 1/4$  on one graph and  $K = 1$ ,  $K = 10$ ,  $K = 19$  and  $C = 4/5$  on another graph. Interpret the results in terms of your example.
3. *Router*. At a router, each packet is routed to one of two different outgoing connections. In order to decide how to allocate bandwidth to the two connections you need to characterize the distribution of the packets over a unit of time, which you fix to be a second. After analyzing the data you decide that the number of packets arriving at the router is Poisson with parameter  $\lambda$  packets/second. A fraction  $0 < p < 1$  of the traffic is routed to connection 1 on average. You decide to model this by assuming that each packet is routed to connection 1 with probability  $0 < p < 1$ , independent of the total number of packets and of how other packets are routed.
    - a. Find the mean and variance of the number of packets routed through connection 1.
    - b. Find the pmf of the number of packets routed through connection 1. What distribution is it? (Hint: Use the identity  $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ .)

4. *Cheap GPS* Mary is running a marathon and she asks her friend Jennifer to monitor her position along the race. Jennifer buys a cheap tracking device, but realizes that the data are very noisy. Assume Mary is running on a straight line and model the device measurements, which are taken every second, as

$$X_i = d_i + Z_i, \quad i = 1, 2, \dots, \quad (2)$$

where  $Z_1, Z_2, \dots$  is an iid sequence with mean zero and standard deviation  $\delta = 1$  and  $d_1, d_2, \dots$  is the deterministic sequence of locations in meters.

- a. Jennifer defines the precision of the location estimate as the smallest  $\Delta$  such that the probability of the error being larger than  $\Delta$  is smaller than 1%. Compute  $\Delta_1$ , the precision of using  $X_i$  as an estimate of  $d_i$ .
- b. Jennifer is not satisfied. She decides to compute

$$Y_i = \sum_{j=i-m}^i X_j, \quad i = m, m+1, \dots, \quad (3)$$

which is called a running mean. Find the precision  $\Delta_2$  of using  $X_i$  as an estimate of  $d_i$  for  $i = m, m+1, \dots$  (i.e. find  $\Delta_2$  such that we are guaranteed that  $|Y_i - d_i| > \Delta_2$  is at most 1%). Assume that Mary does not run backwards and that her maximum speed is 2 meters/second. (Hint: Use the triangle inequality  $|a + b| \leq |a| + |b|$  and the identity  $\sum_{i=1}^k i = k(k+1)/2$ .)

- c. Evaluate  $\Delta_2$  for  $m = 4$ . Why is Jennifer doing this?
  - d. What is the best precision  $\Delta_2$  that Jennifer can achieve using a running mean? Remember that  $m$  has to be an integer.
5. *Iterated expectation for random vectors* .
- a. Show that for any disjoint subvectors indexed by  $\mathcal{I}, \mathcal{J} \subseteq \{1, 2, \dots, n\}$ ,  $\mathcal{I} \cap \mathcal{J} = \emptyset$ ,

$$\mathbb{E}(\mathbb{E}(\mathbf{X}_{\mathcal{I}}|\mathbf{X}_{\mathcal{J}})) = \mathbb{E}(\mathbf{X}_{\mathcal{I}}). \quad (4)$$

How have you defined  $\mathbb{E}(\mathbf{X}_{\mathcal{I}}|\mathbf{X}_{\mathcal{J}})$ ? What kind of object is it?

- b. Compute the mean and variance of  $K$  in Problem 2.