Exercise session 8 – Prob. & Stat. 2 — Dec 15, 2022

Balls&bins

1. Czech (and Slovak) Birth number BN (rodné číslo) consists of six digits that encode the date of birth (& sex), plus four digits, let us assume these are assigned at random (discuss!). Occasionally, these four digits are used as a password.

Find minimum k such that in a group of k people there is more than 50 % chance that some two people have the same last four digits of their BN.

2. Suppose that n balls are thrown independently and uniformly at random into n bins.

(a) Find the conditional probability that bin 1 has one ball given that exactly one ball fell into the first three bins.

(b) Find the conditional expectation of the number of balls in bin 1 under the condition that bin 2 received no balls.

(c) Write an expression for the probability that bin 1 receives more balls than bin 2. (A formula with a sum is good enough, but you may also try to approximate it – using the Poisson approximation.)

3. Consider the probability that every bin receives exactly one ball when n balls are thrown randomly into n bins.

(a) Give an upper bound on this probability using the Poisson approximation.

(b) Determine the exact probability of this event.

(c) Show that these two probabilities differ by a multiplicative factor that equals the probability that a Poisson random variable with parameter n takes on the value n. Explain why this is implied by a theorem from the lecture.

4. Our analysis of Bucket sort in the last lecture assumed that $n = 2^k$ elements were chosen independently and uniformly at random from the range $[0, 2^\ell)$. Suppose instead that n elements are chosen independently from the range $[0, 2^\ell)$ according to a distribution with the property that any integer $x \in [0, 2^\ell)$ is chosen with probability at most $a/2^\ell$ for some fixed constant $a > 0$. Show that, under these conditions, Bucket sort still runs in a linear expected time.

5. The following problem models a simple distributed system wherein agents contend for resources but "back off" in the face of contention. Balls represent agents, and bins represent resources. The system evolves over rounds. Every round, balls are thrown independently and uniformly at random into n bins. Any ball that lands in a bin by itself is served and removed from consideration. The remaining balls are thrown again in the next round. We begin with n balls in the first round, and we finish when every ball is served.

(a) If there are b balls at the start of a round, what is the expected number of balls at the start of the next round?

(b) Suppose that every round the number of balls served was exactly the expected number of balls to be served. Show that all n balls would be served in $O(\log\log n)$ rounds. (Hint: If x_j is the expected number of balls left after j rounds, show and use that $x_{j+1} \le x_j^2/n$. Estimate x_1 directly.)

6. Consider the following modification of the balls-and-bins process. Let $n = 2^k$ (so $k = \log_2 n$). We k times choose a uniformly random bin, then put one ball into this bin and the next $\frac{n}{k} - 1$ consecutive balls (wrapping back to the beginning, if we go over the last bin). Assume k divides n, for simplicity.

So, again, we distribute n balls, but we do it in $\log_2 n$ rounds, each involving just one random choice.

Argue that the maximum load in this case is only $O(\log\log n/\log\log\log n)$ with probability that approaches 1 as $n \to \infty$.

Conditional expectation

If $\mathbb{E}(X \mid Y = y) = f(y)$, then we define $\mathbb{E}(X \mid Y)$ as $f(Y)$. We have

$$\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X) \qquad \text{(law of iterated expectation).}$$

Recall that if X and Y are continuous, then

$$\mathbb{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx.$$

We also define $var(X|Y)$ as $g(Y)$, if $var(X|Y = y) = g(y)$. We have

$$var(X) = \mathbb{E}(var(X \mid Y)) + var(\mathbb{E}(X \mid Y)) \qquad \text{(law of total variance)}.$$

7. Warm-up: (a) what is $\mathbb{E}(X \mid X)$?
(b) what is $\mathbb{E}(X \mid Y)$, when X and Y are independent?

8. A gambler repeatedly takes part in a game, where with probability $p > 1/2$ he wins the amount he betted and with probability $1 - p$ he loses it. A popular strategy, known as Kelly strategy, is to always bet $2p - 1$ multiple of your current total. What is the expected fortune the gambler obtains after n rounds? (And why is it better then betting everything?)

9. Pat and Nat are dating, and all of their dates are scheduled to start at 9 p.m. Nat always arrives promptly at 9 p.m. Pat is highly disorganized and arrives at a time that is uniformly distributed between 8 p.m. and 10 p.m. Let X be the time in hours between 8 p.m. and the time when Pat arrives. If Pat arrives before 9 p.m., their date will last exactly 3 hours. If Pat arrives after 9 p.m., their date starts when Pat arrives and lasts for a time uniformly distributed between 0 and $3 - X$. Nat gets irritated when Pat is late and will end relationship after the second date where Pat is late by more than 45 minutes. All random variables mentioned are independent.
(a) What is the expected number of hours Nat waits for Pat to arrive?
(b) What is the expected duration of any particular date?
(c) What is the expected number of dates they will have before breaking up?

10. Show that for a discrete or continuous random variable X and any function $g(Y)$ of another random variable Y, we have

$$\mathbb{E}(Xg(Y)|Y) = g(Y)\mathbb{E}(X \mid Y).$$

11. * Let X and Y be independent random variables. Use the law of total variance to show that

$$var(XY) = \mathbb{E}(X)^2 var(Y) + \mathbb{E}(Y)^2\, var(X) + var(X)\, var(Y).$$

12. * We toss n times a biased coin whose probability of heads, denoted by q, is the value of a random variable Q with given mean μ and positive variance σ^2. Let $X_1, \ldots, X_n$ be a Bernoulli random variable that models the outcome of the ith toss (i.e., $X_i = 1$ if the ith toss is a head). We assume that $X_1, \ldots, X_n$ are conditionally independent, given $Q = q$. Let X be the number of heads obtained in the n tosses.
(a) Use the law of iterated expectations to find $\mathbb{E}(X_i)$ and $\mathbb{E}(X)$.
(b) Find $cov(X_i, X_j)$. Are $X_1, \ldots, X_n$ independent?
(c) Use the law of total variance to find $var(X)$. Verify your answer using the covariance result of part (b).