Probability and Statistics

Exercise sheet 6

Exercise 6.1 (Binomial disguised)

Suppose that each day the price of a stock moves up 12.5 cents with probability $\frac{1}{3}$ and down 12.5 cents with probability $\frac{2}{3}$.

If the movements of the stock from one day to another are independent, what is the probability that after 10 days the stock has its original price?

Solution 6.1 Let X_i be the increase (or decrease) of the stock price at day i. Then,

$$X_i = \begin{cases} +12.5 \text{ cents with probability } \frac{1}{3}, \\ -12.5 \text{ cents with probability } \frac{2}{3}. \end{cases}$$

Consider the transformed random variables

$$Y_i = \frac{X_i}{25} + \frac{1}{2}$$

for $i = 1, ..., 10$.

Then, $Y_i \in \{0, 1\}$ with $P(Y_i = 0) = \frac{2}{3}$ and $P(Y_i = 1) = \frac{1}{3}$. Hence, $Y_1, ..., Y_{10}$ are i.i.d Bernoulli$\left(\frac{1}{3}\right)$.

The event we are interested in is

$$\{\text{the stock has its original price after 10 days}\} = \left\{\sum_{i=1}^{10} X_i = 0\right\} = \left\{\sum_{i=1}^{10} Y_i = 5\right\}.$$

Since $\sum_{i=1}^{10} Y_i \sim \text{Bin}\left(10, \frac{1}{3}\right)$, it follows that

$$\begin{aligned} P(\text{The stock has its original price after 10 days}) &= \binom{10}{5} \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5 \\ &= \frac{10!}{(5!)^2} \frac{2^5}{3^{10}} \\ &\approx 0.137. \end{aligned}$$

Exercise 6.2 (The coffee is on me if...)

Barbara, Christa and Cecilia are going to have coffee at the local coffee shop. They will each toss a fair coin, and if one of them comes out as the "odd woman out" (the one with the different outcome), then she pays for all three. They keep tossing until an odd woman is found.

What is the probability that a decision will be reached with 2 rounds of tosses?

Optional: Can you generalize that to n people, coins with probability p of obtaining heads and the question being if a decision is reached within k rounds of tosses?

Solution 6.2 For the i^{th} round of tosses of each of the fair coins, define

$$X_i := \begin{cases} 0 & \text{if that round results in HHH or TTT,} \\ 1 & \text{otherwise.} \end{cases}$$

Since all tosses are independent, it follows that $X_1, X_2, \ldots$ are i.i.d $\sim$ Bernoulli(p_0) with

$$\begin{aligned} p_0 &= 1 - P(\text{HHH or TTT}) \\ &= 1 - 2 \times \left(\frac{1}{2}\right)^3 = \frac{3}{4}. \end{aligned}$$

Thus, if X is the number of tosses necessary to reach a decision, then $X \sim \text{Geo}(\frac{3}{4})$.
Hence,

$$P(X = 2) = \left(1 - \frac{3}{4}\right)^{2-1} \frac{3}{4} = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}.$$

For the general case, suppose now that the coins have probability p of landing on heads and that n people are involved. We define once again, for the i^{th} round of tosses, the random variable

$$X_i := \begin{cases} 0 & \text{if exactly one person gets a result different from all the others,} \\ 1 & \text{otherwise.} \end{cases}$$

It is easy to see that $X_i = 1$ if and only if the outcome of the n tosses is contained in

$$\bigcup_{1 \le j \le n} \{HH \ldots H \underbrace{T}_{j^{\text{th}}} H \ldots H\} \cup \{TT \ldots T \underbrace{H}_{j^{\text{th}}} T \ldots T\}$$

and hence

$$P(X_i = 1) = np^{n-1}(1-p) + n(1-p)^{n-1}p =: \tilde{p}$$

(when $p = \frac{1}{2}$, $n = 3$ we get exactly $\frac{3}{4}$).

Thus, again letting X be the number of rounds necessary for reaching a decision, we get that

$$P(X = k) = (1 - \tilde{p})^{k-1}\tilde{p}.$$

For example, letting $n = 5$ and $p = \frac{2}{3}$, we have $\tilde{p} \approx 0.370$ and

$$\begin{aligned} P(X = 2) &\approx 0.233, \\ P(X = 3) &\approx 0.147, \\ P(X = 4) &\approx 0.092. \end{aligned}$$

Exercise 6.3 (Domination of the minority)

In a small town of Alaska, there are 60 Republicans and 40 Democrats. 10 are selected at random for a council. What is the probability that there will be more Democrats than Republicans?

Solution 6.3 Let X be the number of Democrats in the council and Y be the number of Republicans in the council. The event we consider is

$$\{X > Y\} = \bigcup_{k=6}^{10} \{X = k\}.$$

Note that $X \sim \text{Hypergeo}(10, 40, 100)$. Thus,

$$P(X > Y) = \sum_{k=6}^{10} \frac{\binom{40}{k}\binom{60}{10-k}}{\binom{100}{10}} \approx 0.154.$$

Exercise 6.4 A Poisson process with rate λ per time unit is a random process that has to satisfy the following properties:

1. The number of arrivals in every fixed interval of time of length t has the Poisson distribution with rate λt.
2. The number of arrivals in every collection of disjoint time intervals are independent.

Suppose that the arrival time of customers to some store A is a Poisson process with rate 1 per hour.

(a) What is the probability that the first customer arrives after time t?

(b) What is the probability that "strictly more than 2 customers come during the first hour"?

(c) Fix a time $t > 0$. What is the probability that "a customer comes exactly at time t"? Does this mean that nobody comes at any time? How should we interpret what seems to be a contradiction?

(d) Suppose that the arrival time of customers to another store B is a Poisson process with rate μ per hour, and it is independent of the Poisson process of the arrival time of customers of store A. Now, we assume again that A has a general rate of λ per hour, instead of 1.

How can you describe the arrival times of clients to either of the two stores (not caring which store they go to)?

Solution 6.4

(a) Let N_t $(t \geq 0)$ be the number of customers who arrive in $[0, t)$. By definition of a Poisson process, $N_t \sim \text{Poi}(\lambda t)$. Note that

$$\{\text{the first customer arrives after } t\} = \{\text{no customers arrived in } [0, t)\} = \{N_t = 0\}.$$

Thus, it follows that

$$P(\text{the first customer arrives after } t) = P(N_t = 0) = e^{-\lambda t} = e^{-t}$$

since we assume that $\lambda = 1$.

(b) Let N_1 be the number of customers who arrive in $[0, 1)$. Then, $N_1 \sim \text{Poi}(\lambda)$, and

$$\begin{aligned} P(N_1 > 2) &= 1 - P(N_1 = 0) - P(N_1 = 1) - P(N_1 = 2) \\ &= 1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right) \\ &= 1 - e^{-1}\left(1 + 1 + \frac{1}{2}\right) \\ &\approx 0.080. \end{aligned}$$

(c) If we fix time $t \in (0, \infty)$, the number of customers $Y = N_{t+\frac{\epsilon}{2}} - N_{t-\frac{\epsilon}{2}}$ who arrive between $t - \frac{\epsilon}{2}$ and $t + \frac{\epsilon}{2}$ satisfies $Y \sim \text{Poi}(\lambda\epsilon)$.

$$\begin{aligned} P(\text{at least one customer arrives in } \left[t-\frac{\epsilon}{2}, t+\frac{\epsilon}{2}\right] &= P(Y>0) \\ &= 1-P(Y=0) \\ &= 1-e^{-\lambda\epsilon} \\ &\overset{\epsilon\to 0}{\sim} \lambda\epsilon = \epsilon \end{aligned}$$

so that

$$P(\text{customer arrives exactly at } t) = \lim_{\epsilon \searrow 0} P(N_\epsilon > 0) = 0$$

(this step uses the monotone convergence theorem).

This result seems to be in contradiction with the fact that there is a positive probability that at least one client comes in $[0,1)$ (probability equal to $1-e^{-1}>0$), since that event is also the union

$$\{\text{at least one client comes in } [0,1)\} = \bigcup_{t\in[0,1)} \{\text{a client comes at } t\}.$$

This contradiction is alleviated by noting that this is an uncountable union, and so we cannot obtain the probability $1-e^{-1}$ as a sum over a countably infinite number of time points.

(d) Since the sum of two independent Poisson processes with rates λ and μ respectively is a Poisson process with rate $\lambda+\mu$, this process is the one which describes the arrival times of clients to either of the stores.

Exercise 6.5 (optional)

(a) From the set $\{1,2,...,N\}$, with $N \geq 2$ an integer, draw i.i.d random variables $X_1,...,X_n$ with uniform distribution.

What is

$$P\left(\max_{1\leq i\leq n}(X_i) = k\right)$$

for some fixed $k \in \{1,...,N\}$?

Hint: Recall that for any random variable Y,

$$P(Y=y) = F(y) - F(y-) = P(Y\leq y) - P(Y<y).$$

(b) What if now $X_1,...,X_n$ are drawn randomly from $\{1,2,...,N\}$ without replacement?

Solution 6.5 Let $k \in \{1,...,N\}$.

(a) Let

$$M_n = \max_{1\leq i\leq n}(X_i).$$

If F is the cumulative distribution function of M_n, then since $X_1,...,X_n$ are $\overset{\text{iid}}{\sim}$ Unif$\{1,...,N\}$ it follows that

$$\begin{aligned} F(m) &= P(M_n \le m) \\ &= \prod_{i=1}^{n} P(X_i \le m) \\ &= (P(X_1 \le m))^n \\ &= (F_{X_1}(m))^n \end{aligned}$$

where F_{X_1} is the cumulative distribution function of X_1, which is given by

$$F_{X_1}(m) = \begin{cases} 0, & \text{if } m < 1, \\ \frac{1}{N}, & \text{if } 1 \le m < 2, \\ \cdots \\ \frac{j}{N}, & \text{if } j \le m < j+1, \\ \cdots \\ 1, & \text{if } m \ge N. \end{cases}$$

Therefore,

$$\begin{aligned} P(M_n = k) &= P(M_n \le k) - P(M_n < k) \\ &= P(M_n \le k) - P(M_n \le k-1) \\ &= \left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n. \end{aligned}$$

(b) Now, $X_1, ..., X_n$ are drawn randomly from $\{1, 2, ..., N\}$ without replacement. Then, the event

$$\left\{\max_{1\le i\le n} X_i = k\right\} = \{\exists j \in \{1, ..., n\} : X_j = k \text{ and } \{X_1, ..., X_n\} \setminus \{X_j\} \text{ is a subset of } \{1, ..., k-1\}\}.$$

Hence, if $n \le k$

$$P\left(\max_{1\le i\le n} X_i = k\right) = \frac{\binom{k-1}{n-1}}{\binom{N}{n}},$$

and if $n > k$,

$$P\left(\max_{1\le i\le n} X_i = k\right) = 0.$$

Note that the obtained probabilities should add up to one. This gives another way of checking the well-known identity

$$\frac{\sum_{k=n}^{N} \binom{k-1}{n-1}}{\binom{N}{n}} = 1.$$