Math 302, Assignment 2

(1) In a small town, there are three bakeries. Each of the bakeries bakes twelve cakes per day. Bakery 1 has two different types of cake, bakery 2 three different types, and bakery 3 four different types. Every bakery bakes equal amounts of cakes of each type.
You randomly walk into one of the bakeries, and then randomly buy two cakes.
  (a) What is the probability that you will buy two cakes of the same type?
  (b) Suppose you have bought two different types of cake. Given this, what is the probability that you went to bakery 2?

Solution: (a) Define the events $F_i = \{\text{choose bakery } i\}$, and $E = \{\text{buy different cakes}\}$. Then $\mathbb{P}(F_i) = \frac{1}{3}$, and we compute the conditional probabilities

$$\mathbb{P}(E|F_1) = \frac{\binom{6}{1}^2}{\binom{12}{2}} = \frac{6}{11}$$
$$\mathbb{P}(E|F_2) = \frac{3 \cdot \binom{4}{1}^2}{\binom{12}{2}} = \frac{8}{11}$$
$$\mathbb{P}(E|F_3) = \frac{6 \cdot \binom{3}{1}^2}{\binom{12}{2}} = \frac{9}{11}$$

By the law of total probability,

$$\mathbb{P}(E) = \frac{1}{3}\left(\frac{6}{11} + \frac{8}{11} + \frac{9}{11}\right) = \frac{23}{33}.$$

Therefore $\mathbb{P}(\text{buy same type}) = \frac{10}{33}$.
(b) By Bayes,

$$\mathbb{P}(F_2|E) = \mathbb{P}(E|F_2)\frac{\mathbb{P}(F_2)}{\mathbb{P}(E)} = \frac{8}{23}$$

(2) An assembly line produces a large number of products, of which 1% are faulty in average. A quality control test correctly identifies 98% of the faulty products, and 95% of the flawless products. For every product that is identified as faulty, the test is run a second time, independently.
  (a) Suppose that a product was identified as faulty in both tests. What is the probability that it is, indeed, faulty?
  (b) What if the quality control test is only performed once?

Solution: Let

$$\begin{aligned}F &= \text{the event that a product is faulty,}\\ E_1 &= \text{the event that the first test result is "faulty"}\\ E_2 &= \text{the event that the second test result is "faulty"}\end{aligned}$$

Then, $\mathbb{P}(F) = 1\%$ and, since the second test is independent of the first,

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2|F) &= \mathbb{P}(E_1|F)\mathbb{P}(E_2|F) = 0.98^2\\ \mathbb{P}(E_1 \cap E_2|F^c) &= \mathbb{P}(E_1|F^c)\mathbb{P}(E_2|F^c) = 0.05^2\end{aligned}$$

By Bayes' theorem,

$$\begin{aligned}\mathbb{P}(F|E_1 \cap E_2) &= \frac{\mathbb{P}(E_1 \cap E_2|F)\mathbb{P}(F)}{\mathbb{P}(E_1 \cap E_2|F)\mathbb{P}(F) + \mathbb{P}(E_1 \cap E_2|F^c)\mathbb{P}(F^c)}\\ &= \frac{0.98^2 \cdot 0.01}{0.98^2 \cdot 0.01 + 0.05^2 \cdot 0.99} \approx 80\%.\end{aligned}$$

If that test had only be run once, we would get

$$\mathbb{P}(F|E_1) = \frac{0.98 \cdot 0.01}{0.98 \cdot 0.01 + 0.05 \cdot 0.99} \approx 17\%,$$

so even a very reliable test cannot identify a rare fault with satisfactory accuracy in a single try.

(3) Let M be an integer chosen uniformly from $\{1, \dots, 100\}$. Decide whether the following events are independent:
  (a) $E = \{M \text{ is even}\}$ and $F = \{M \text{ is divisible by } 5\}$
  (b) $E = \{M \text{ is prime}\}$ and $F = \{\text{at least one of the digits of } M \text{ is a } 2\}$
  (c) Can you replace the number 100 by a different number, in such a way that your answer to (a) changes? (E.g., if your answer was "dependent", try to change the number 100 in such a way your answer becomes "independent").
  (d) Show that if M is selected uniformly from $\{1, 2, \dots, N\}$ for N very large, and p, q are fixed prime numbers, then the events $\{p|M\}$ and $\{q|M\}$ are roughly independent, i.e.

$$\mathbb{P}(pq|M) \approx \mathbb{P}(p|M)\mathbb{P}(q|M),$$

where $a|b$ means a divides b, and $\approx$ means the two sides of the above equation differ by something tending to 0 as $N \to \infty$. Does this still hold if p, q are allowed to be any integers?

Solution: (a) $\mathbb{P}(E)\mathbb{P}(F) = (1/2)(1/5) = 1/10 = \mathbb{P}(E \cap F)$, so E and F are independent.
(b) $\mathbb{P}(E) = 25/100$, $\mathbb{P}(F) = 19/100$, and $\mathbb{P}(E \cap F) = 3/100$. The events E and F are not independent since $\mathbb{P}(E \cap F) \neq \mathbb{P}(E)\mathbb{P}(F)$.
(c) Already replacing it by 101 makes the events dependent: $\frac{50}{101} \cdot \frac{20}{101} \neq \frac{10}{101}$. The fact that E and F of (a) were dependent was a pure coincidence, and had nothing to do with number theoretic properties of divisibility by 2 and 5. Whether one could change the 100 in part (b) to make the events

mentioned there independent would be a much harder question! (d) This is a consequence of the fact that there are approximately N/a numbers between 1 and N that are divisible by a. If N itself is divisible by pq, the equality holds exactly. Implicit in the statement is the fact that $\{p|M\} \cap \{q|M\} = \{pq|M\}$, which only holds if p, q are relatively prime.

(4) Let X be a discrete random variable with values in $\mathbb{N} = \{1, 2, \ldots\}$. Prove that X is geometric with parameter $p = \mathbb{P}(X = 1)$, i.e.

$$\mathbb{P}(X = k) = p(1-p)^{k-1} \text{ for } k \in \mathbb{N},$$

if and only if the *memoryless property*

$$\mathbb{P}(X = n + m \,|\, X > n) = \mathbb{P}(X = m)$$

holds.
Hint: Use $\mathbb{P}(X = k) = \mathbb{P}(X = k + 1|X > 1)$ repeatedly.

Solution: We first show that a geometric RV has the memoryless property: We learned that $\mathbb{P}(X = m) = p(1-p)^{m-1}$ and that $\mathbb{P}(X > m) = (1-p)^m$, therefore by the definition of conditional probability we obtain

$$\begin{aligned} \mathbb{P}(X = n + m \,|\, X > n) &= \frac{\mathbb{P}(X = n + m)}{\mathbb{P}(X > n)} = \frac{p(1-p)^{n+m-1}}{(1-p)^n} \\ &= p(1-p)^{m-1} = \mathbb{P}(X = m) \end{aligned}$$

Now we show that the memoryless property implies that $\mathbb{P}(X = k) = p(1-p)^{k-1}$ with $p = \mathbb{P}(X = 1)$. Using the law of total probability and the hint, for $k > 1$ we have

$$\begin{aligned} P(X = k) &= P(X = k|X > 1)P(X > 1) + P(X = k|X = 1)P(X = 1) \\ &= P(X = k|X > 1)(1 - P(X = 1)) + 0 \\ &= P(X = k - 1)(1 - P(X = 1)) = \cdots = P(X = 1)(1 - P(X = 1))^{k-1}. \end{aligned}$$

Therefore, X is Geom(p) with $p = P(X = 1)$.

(5) Let X take values $\{1, 2, 3, 4, 5\}$, and have p.m.f. given by

Table 1. The p.m.f. of X

k	1	2	3	4	5
$\mathbb{P}(X = k)$	1/7	1/14	3/14	2/7	2/7

(a) Calculate $\mathbb{P}(X \leq 3)$
(b) Calculate $\mathbb{P}(X < 3)$
(c) Calculate $\mathbb{P}(X < 4.12|X > 1.6)$

Solution:
a) $\mathbb{P}(X \leq 3) = 1/7 + 1/14 + 3/14 = 3/7$.

b) $\mathbb{P}(X < 3) = 1/7 + 1/14 = 3/14$.
c)

$$\begin{aligned}\mathbb{P}(X < 4.12 \,|\, X > 1.6) &= \frac{\mathbb{P}(1.6 < X < 4.12)}{\mathbb{P}(X > 1.6)} = \frac{\mathbb{P}(2 \leq X \leq 4)}{\mathbb{P}(X \geq 2)} \\ &= \frac{1/14 + 3/14 + 2/7}{1/14 + 3/14 + 2/7 + 2/7} = \frac{2}{3}.\end{aligned}$$

(6) Consider the following lottery: There are a total of 10 tickets, of which 5 are "win" and 5 are "lose". You draw tickets until you draw the first "win". Drawing one ticket costs \$2, 2 tickets \$4, 3 tickets \$8, and so on. A winning ticket pays out \$8.

(a) Let X be the number of tickets you draw in the lottery (i.e. the number of tickets until the first win, including the winning ticket). Calculate the p.m.f. of X.
(b) Calculate the expectation $\mathbb{E}\,X$.
(c) Calculate the variance $\sigma^2(X)$.
(d) What are your expected winnings in this game?

Solution:
a) The values that X can take are $\{1, 2, 3, 4, 5, 6\}$, and

$$\mathbb{P}(X = k) = \frac{5}{10}\frac{4}{9}\frac{3}{8}\cdots\frac{5-(k-2)}{10-(k-2)}\frac{5}{N-(k-1)} = \frac{\binom{10-k}{5-1}}{\binom{10}{5}}$$

b) We have

$$\mathbb{E}\,X = \sum_{k=1}^{6} k\frac{\binom{10-k}{5-1}}{\binom{10}{5}} = \frac{11}{6}$$

c) We have

$$\sigma^2(X) = \sum_{k=1}^{6}(k - \tfrac{11}{6})^2\frac{\binom{10-k}{5-1}}{\binom{10}{5}} = \frac{275}{252}$$

d) If the game finishes with k balls, then you gain \$8 and pay $\$2^k$. Therefore, the expected gain is

$$\mathbb{E}(8 - 2^X) = \sum_{k=1}^{6}(8 - 2^k)\frac{\binom{10-k}{5-1}}{\binom{10}{5}} = \frac{185}{63}$$

(7) In a town, there are on average 2.3 children in a family and a randomly chosen child has on average 1.6 siblings. Determine the variance of the number of children in a randomly chosen family.

Solution: Let X be the number of children in a randomly chosen family and let Y be the number of siblings of a randomly chosen child. Let n be the maximal number of children in a family, and assume that there are a_i families with exactly i children for each $i = 0, 1, \ldots, n$. Then the total number of families is $F = \sum_{i=0}^{n} a_i$ and the total number of children is $C = \sum_{i=0}^{n} i a_i$. Thus we have

$$P(X = i) = \frac{a_i}{F} \quad \text{and} \quad \mathbb{P}(Y = i-1) = \frac{i a_i}{C}$$

for all $i = 0, \ldots, n$. The definition of mean and our condition give

$$\mathbb{E}(X) = \sum_{i=0}^{n} i\mathbb{P}(X = i) = \sum_{i=1}^{n} \frac{i a_i}{F} = \frac{C}{F} = 2.3, \tag{0.1}$$

and similarly

$$\mathbb{E}(Y) = \sum_{i=0}^{n} (i-1)\mathbb{P}(Y = i-1) = -1 + \sum_{i=0}^{n} i\mathbb{P}(Y = i-1) = -1 + \sum_{i=0}^{n} \frac{i^2 a_i}{C} = 1.6,$$

so

$$\sum_{i=0}^{n} \frac{i^2 a_i}{C} = 2.6. \tag{0.2}$$

Using (0.1) and (0.2) the second moment of X is

$$\mathbb{E}(X^2) = \sum_{i=0}^{n} i^2 \mathbb{P}(X = i) = \sum_{i=0}^{n} \frac{i^2 a_i}{F} = \frac{C}{F} \sum_{i=0}^{n} \frac{i^2 a_i}{C} = 2.3 \cdot 2.6.$$

Thus

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 2.3 \cdot 2.6 - 2.3^2 = 2.3 \cdot 0.3 = 0.69.$$

(8) Consider a random graph on $n \geq 3$ vertices labeled by $[n] := \{1, \ldots, n\}$ defined as follows: for each i, j draw an edge between i and j w.p. $1/2$ so that the events $A_{i,j} = \{i \text{ and } j \text{ are connected by an edge}\}$ are (mutually) independent for all pairs $\{i, j\}$ with $i \neq j \in [n]$. Let X be the number of triangles in the graph. That is, the number of $\{i, j, k\} \subset [n]$ so that $i \neq j$, $j \neq k$ and $i \neq k$ and each of the three pairs $\{i, j\}, \{i, k\}$ and $\{j, k\}$ has an edge between them.

(a) Calculate $\mathbb{E}\, X$
(b) Calculate $\sigma^2(X)$

Solution: (a) Let $\xi_{\{i,j,k\}}$ be the indicator of a triangle involving i, j and k (i.e. that all three pairs involving these vertices have an edge between them). That is, it is a random variable that equals 1 on the last event, and otherwise it equals zero. The brackets in the subscript of $\xi_{\{i,j,k\}}$ are used to emphasize that $\xi_{\{1,2,3\}} = \xi_{\{1,3,2\}} = \xi_{\{3,1,2\}} = \xi_{\{3,2,1\}} = \xi_{\{2,1,3\}} = \xi_{\{2,3,1\}}$ (i.e. order doesn't matter here). Then $\mathbb{E}[\xi_{i,j,k}] = \mathbb{P}(\xi_{i,j,k} = 1) \times 1 + \mathbb{P}(\xi_{i,j,k} = 0) \times 0 = p^3$ for all $\{i, j, k\} \in S_3$ where $S_i := \{A \subset [n] : |A| = i\}$. Since the total number of triangles ξ is given by $\sum_{\{i,j,k\} \in S} \xi_{\{i,j,k\}}$, by linearity of expectation we have that $\mathbb{E}[\xi] = \mathbb{E}\sum_{\{i,j,k\}\in S_3}\xi_{\{i,j,k\}} = \sum_{\{i,j,k\}\in S_3}\mathbb{E}\xi_{\{i,j,k\}} = p^3|S_3| = p^3\binom{n}{3}$.

(b) Since $\sigma^2(\xi) = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2$ and since $\mathbb{E}[\xi]$ was already calculated in part (a), we only need to determine $\mathbb{E}[\xi^2]$. For each $A, B \subset [n]$ write ξ_A for the indicator that every pair $a, b \in A$ have an edge between them and write $\xi_{A,B} := \xi_A\xi_B$. Note that $\xi_{A,B} \neq \xi_{A\cup B}$ unless $A = B$. Then

$$\xi^2 = \left(\sum_{A\in S_3}\xi_A\right)\left(\sum_{B\in S_3}\xi_B\right) = \left(\sum_{A,B\in S_3}\xi_{A,B}\right) \tag{0.3}$$

(in the above sum the case $(A,B) = (W,U)$ and the cases $(A,B) = (U,W)$ for $U \neq W \in S_3$ are counted in the sum as two different terms, but as one term when $W = U$).

If $A, B \in S_3$ then $A \cup B \in \cup_{i=3}^6 S_i$. Observe that for $A, B \in S_3$ we have

$$\mathbb{E}[\xi_{A,B}] = \mathbb{P}(\xi_{A,B} = 1)\times 1 + \mathbb{P}(\xi_A = 0)\times 0 = \begin{cases} p^3 & A = B \\ p^5 & |A\cap B| = 2 \\ p^6 & |A\cap B| < 2. \end{cases} \tag{0.4}$$

- The case $A = B$ has $\binom{n}{3}$ different possibilities.
- The case $|A\cap B| = 2$ has $\binom{n}{3}\binom{n-3}{1}\binom{3}{2}$ possibilities ($\binom{n}{3}$ ways to choose A, and then for B we choose one of the $n-3$ elements of $[n]\backslash A$ and two elements from A (there is no double counting because we want to count $(A,B) = (W,U)$ and the cases $(A,B) = (U,W)$ for $U \neq W \in S_3$ as two different cases).
- The case $|A\cap B| < 2$ is the complement of the last two cases and so has $\binom{n}{3}^2 - \binom{n}{3} - \binom{n}{3}\binom{n-3}{1}\binom{3}{2}$ cases.

This together with (0.3) and (0.4) yield that

$$\begin{aligned}\mathbb{E}[\xi^2] &= p^3\times\binom{n}{3} + p^5\binom{n}{3}\binom{n-3}{1}\binom{3}{2} + p^6\left[\binom{n}{3}^2 - \binom{n}{3} - \binom{n}{3}\binom{n-3}{1}\binom{3}{2}\right] \\ &= (p^3 - p^6)\binom{n}{3} + (p^5 - p^6)\binom{n}{3}\binom{n-3}{1}\binom{3}{2} + p^6\binom{n}{3}^2 \\ &= (p^3 - p^6)\binom{n}{3} + (p^5 - p^6)\binom{n}{3}\binom{n-3}{1}\binom{3}{2} + (\mathbb{E}[X])^2.\end{aligned}$$

Hence $\sigma^2(X) = (p^3 - p^6)\binom{n}{3} + (p^5 - p^6)\binom{n}{3}\binom{n-3}{1}\binom{3}{2} = (p^3 - p^6)\binom{n}{3} + (p^5 - p^6)\times\frac{n(n-1)(n-2)(n-3)}{2}$.