1. In a school that has n students, each student could be affected by a rare disease with probability p. A test for the disease shows positive with probability 0.99 when applied to an ill person, and with probability 0.02 when applied to a healthy person.

a) What is the probability that a person has the disease given that the test shows positive? What is the probability that a person has the disease given the test is negative?

Answer. Let H = "a person is healthy", H^c = "a person is ill"; A = "test positive", A^c = "test negative.

We are given that $\mathbf{P}(H^c) = p, \mathbf{P}(H) = 1-p, \mathbf{P}(A|H) = 0.02, \mathbf{P}(A|H^c) = 0.99$. Hence $\mathbf{P}(A^c|H) = 0.98, \mathbf{P}(A^c|H^c) = 0.01$.

By Bayes formula (which includes total probability in the denominator), a tree could be used as well,

$$\begin{aligned}\mathbf{P}(H^c|A) &= \frac{\mathbf{P}(A|H^c)\,\mathbf{P}(H^c)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|H^c)\,\mathbf{P}(H^c)}{\mathbf{P}(A|H^c)\,\mathbf{P}(H^c) + \mathbf{P}(A|H)\,\mathbf{P}(H)} \\ &= \frac{.99p}{.99p + .02(1-p)} = \frac{99p}{97p+2} =: p_1\end{aligned}$$

Similarly,

$$\mathbf{P}(H^c|A^c) = \frac{0.01p}{0.01p + 0.98(1-p)} = \frac{p}{98-97p} =: p_2$$

b) To each student that test is applied. Let X be the number of ill students in that school, Y be the number of students whose test is positive.

Find $\mathbf{E}(X|Y=k), \mathbf{E}(X|Y)$. Hint. $X = U + V$, where U is number of ill students whose test positive, and V is number of ill students whose test negative.

Answer. Let U be number of ill students whose test positive, and V be number of ill students whose test negative. Then $X = U + V$, and given $Y = k$, U is binomial(k, p_1), and V is binomial($n-k$, p_2). Hence

$$\begin{aligned}\mathbf{E}(X|Y=k) &= \mathbf{E}(U|Y=k) + \mathbf{E}(V|Y=k) = kp_1 + (n-k)\,p_2, \\ \mathbf{E}(X|Y) &= Yp_1 + (n-Y)\,p_2.\end{aligned}$$

Comment. $\mathbf{E}(X|Y)$ is the best mean square estimate of X based on Y. For instance, if actually $Y = 10$, then this estimate is the number $10p_1 + (n-10)\,p_2$. The parts a), b) are our hw7 faulty robots problem: "robots"= "students", "faulty" = "ill", "passed the test" = "negative test".

c) K. is a student at that school. After being tested positive, the second test applied to K. was negative. What is the probability that K. is ill?

Answer. Let $B =$ "2nd test negative". Then, using conditional independence, we get

$$\begin{aligned}
\mathbf{P}\left(H^c|AB\right) &= \frac{\mathbf{P}\left(AB|H^c\right)\mathbf{P}\left(H^c\right)}{\mathbf{P}\left(AB|H^c\right)\mathbf{P}\left(H^c\right)+\mathbf{P}\left(AB|H\right)\mathbf{P}\left(H\right)} = \frac{0.01\cdot 0.99p}{0.01\cdot 0.99p+0.98\cdot 0.02\left(1-p\right)} \\
&= \frac{99p}{99p+196\left(1-p\right)}.
\end{aligned}$$

`Alternatively`, since the first test was positive, by Bayess for conditional probability $\mathbf{P}\left(\cdot|A\right)$ ("actual" probability),

$$\begin{aligned}
\mathbf{P}\left(H^c|AB\right) &= \frac{\mathbf{P}\left(B|AH^c\right)\mathbf{P}\left(H^c|A\right)}{\mathbf{P}\left(B|AH^c\right)\mathbf{P}\left(H^c|A\right)+\mathbf{P}\left(B|AH\right)\mathbf{P}\left(H|A\right)} \\
&= \frac{\mathbf{P}\left(B|H^c\right)\mathbf{P}\left(H^c|A\right)}{\mathbf{P}\left(B|H^c\right)\mathbf{P}\left(H^c|A\right)+\mathbf{P}\left(B|H\right)\mathbf{P}\left(H|A\right)} = \frac{0.01p_1}{0.01p_1+0.98\left(1-p_1\right)} \\
&= \frac{p_1}{p_1+98\left(1-p_1\right)} = \frac{p_1}{98-97p_1}.
\end{aligned}$$

2. N distinct balls are placed into n distinct boxes at random with all n^N ways equally likely, $n > 7$.

a) Let X be the number of empty boxes. Calculate $\mathbf{E}(X)$ and $\mathrm{Var}(X)$.

Answer. Let $A_i =$ "ith box is empty", $i = 1, \ldots, n$. Then $X = \sum_{i=1}^n I_{A_i}$, and $\mathbf{E}(X) = \sum_{i=1}^n \mathbf{P}(A_i)$. The sample space Ω is the set of all different placements. Every ball can go into any of N boxes: $\#\Omega = n^N$. The event A_i means that every of N balls goes into any of remaining $n-1$ boxes: $\#A_i = (n-1)^{N}$. Hence

$$\mathbf{P}(A_i) = \frac{(n-1)^N}{n^N} = \left(1 - \frac{1}{n}\right)^N, i = 1, \ldots, n, \mathbf{E}(X) = \sum_{i=1}^n \mathbf{E}(X_i) = n\left(1 - \frac{1}{n}\right)^N.$$

Now

$$\mathrm{var}(X) = \mathbf{E}\left(X^2\right) - (\mathbf{E}(X))^2,$$

and

$$\mathbf{E}\left(X^2\right) = \mathbf{E}\left(\sum_{i=1}^n I_{A_i} + 2\sum_{i<j} I_{A_i A_j}\right) = \mathbf{E}(X) + 2\sum_{i<j} \mathbf{P}\left(A_i \cap A_j\right).$$

$A_i A_j, i < j$, means two distinct boxes (i and j) are empty (any of N balls can go into any of remaining $n-2$ boxes: $\#\left(A_i \cap A_j\right) = (n-2)^N$, and for any $i < j$,

$$\mathbf{P}\left(A_i \cap A_j\right) = \frac{(n-2)^N}{n^N} = \left(1 - \frac{2}{n}\right)^N.$$

So,

$$\begin{aligned}\mathbf{E}\left(X^2\right) &= n\left(1 - \frac{1}{n}\right)^N + 2\sum_{i<j}\left(1 - \frac{2}{n}\right)^N = n\left(1 - \frac{1}{n}\right)^N + 2\binom{n}{2}\left(1 - \frac{2}{n}\right)^N \\ &= n\left(1 - \frac{1}{n}\right)^N + n(n-1)\left(1 - \frac{2}{n}\right)^N\end{aligned}$$

and

$$\mathrm{var}(X) = n\left(1 - \frac{1}{n}\right)^N + n(n-1)\left(1 - \frac{2}{n}\right)^N - n^2\left(1 - \frac{1}{n}\right)^{2N}.$$

Comment. This is elevator problem: "balls" = "people", "boxes" = "floors".

b) Let A be the event that boxes 1 and 2 are both empty, B be the event that boxes 2, 4 are empty, and C be the event that boxes 5, 6, 7 are empty. Find $\mathbf{P}(A \cup B \cup C)$.

Answer. By inclusion-exclusion principle,

$$\begin{aligned}\mathbf{P}(A \cup B \cup C) &= \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) \\ &\quad - \mathbf{P}(B \cap C) + \mathbf{P}(A \cap B \cap C).\end{aligned}$$

As in part a) we find

$$\begin{aligned}\mathbf{P}(A) &= \frac{(n-2)^N}{n^N} = \left(1 - \frac{2}{n}\right)^N = \mathbf{P}(B), \mathbf{P}(C) = \frac{(n-3)^N}{n^N} = \left(1 - \frac{3}{n}\right)^N, \\ \mathbf{P}(A \cap B) &= \frac{(n-3)^N}{n^N} = \left(1 - \frac{3}{n}\right)^N, \mathbf{P}(B \cap C) = \frac{(n-5)^N}{n^N} = \left(1 - \frac{5}{n}\right)^N = \mathbf{P}(A \cap C) \\ \mathbf{P}(A \cap B \cap C) &= \frac{(n-6)^N}{n^N} = \left(1 - \frac{6}{n}\right)^N.\end{aligned}$$

So,

$$\mathbf{P}(A \cup B \cup C) = 2\left(1 - \frac{2}{n}\right)^N - 2\left(1 - \frac{5}{n}\right)^N + \left(1 - \frac{6}{n}\right)^N.$$

3. J has a coin that shows heads with probability p. The number X of heads in n tosses determines how many pages J proofreads in a book. The number of misprints on a page of that book is Poisson random variable with parameter λ and the number of misprints on different pages are independent. Let Y be the number of misprints in X pages of the book.

a) What is $\mathbf{P}(Y = j|X = k), j \geq 0, 0 \leq k \leq n$.

Answer. Given $X = k$, $Y = \sum_{i=1}^{k} Y_i$, where Y_i is the number of misprints in the ith page; all Y_i are independent Poisson(λ). Hence $Y \sim$Poisson$(k\lambda)$:

$$\mathbf{P}(Y = j|X = k) = e^{-k\lambda}\frac{(k\lambda)^j}{j!} = e^{-k\lambda}\frac{k^j\lambda^j}{j!}, j \geq 0.$$

b) Find $\mathbf{E}(Y|X = k)$, $\mathbf{E}(Y|X)$ and $\mathbf{E}(Y)$. Find Var$(Y|X = k)$, $\mathbf{E}\left(Y^2|X = k\right)$, $\mathbf{E}\left(Y^2|X\right)$ and Var(Y).

Answer. Since given $X = k$, $Y = \sum_{i=1}^{k} Y_i$, where Y_i are independent Poisson(λ), we have $Y \sim$Poisson$(k\lambda)$, and $\mathbf{E}(Y|X = k) = \text{Var}(Y|X = k) = k\lambda$.

`Alternatively,` if we do not know that Y is Poisson,

$$\begin{aligned}
\mathbf{E}(Y|X = k) &= \mathbf{E}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k}\mathbf{E}(Y_i) = k\lambda, \\
\text{Var}(Y|X = k) &= \text{Var}\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k}\text{Var}(Y_i) = k\lambda.
\end{aligned}$$

So,

$$\mathbf{E}\left(Y^2|X = k\right) = Var(Y|X = k) + (k\lambda)^2 = k\lambda + k^2\lambda^2, \mathbf{E}\left(Y^2|X\right) = \lambda X + \lambda^2 X^2,$$

and since X is binomial(n, p),

$$\begin{aligned}
\mathbf{E}(Y|X) &= \lambda X, \mathbf{E}(Y) = \lambda\mathbf{E}(X) = \lambda np, \text{Var}(Y|X) = \lambda X, \\
\text{Var}(Y) &= \mathbf{E}(\text{Var}(Y|X)) + \text{Var}(\mathbf{E}(Y|X)) = \lambda\mathbf{E}(X) + \text{Var}(\lambda X) \\
&= \lambda np + \lambda^2\text{Var}(X) = \lambda np + \lambda^2 npq = \lambda np(1 + \lambda q).
\end{aligned}$$

c) J found 6 misprints in five first pages of that book. What is the probability of at most 1 misprint in first two pages.

Answer. Let V be the number of misprints on first two pages, and U be the number of misprints on the next three pages. The V $\sim$Poisson(2λ) , U $\sim$Poisson(3λ) are independent, and given $U + V = 6$, U is binomial with $n = 6$, $p = \frac{2\lambda}{2\lambda+3\lambda} = \frac{2}{5}$. Hence

$$\mathbf{P}\left(V \leq 1 | U + V = 6\right) = \left(\frac{3}{5}\right)^6 + 6\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)^5.$$

2nd answer (we do not know that the conditional distribution binomial) Let V be the number of misprints on first two pages, and U be the number of misprints on the next three pages. The V $\sim$Poisson(2λ) , U $\sim$Poisson(3λ) are independent, and $V + U$ $\sim$Poisson(5λ). We have to find

$$\mathbf{P}\left(U \leq 1 | U + V = 8\right) = \mathbf{P}\left(U = 0 | U + V = 6\right) + \mathbf{P}\left(U = 1 | U + V = 6\right).$$

We find

$$\begin{aligned}\mathbf{P}\left(U = 0 | U + V = 6\right) &= \frac{\mathbf{P}\left(U = 0, U + V = 6\right)}{\mathbf{P}\left(U + V = 6\right)} = \frac{\mathbf{P}\left(U = 0, V = 6\right)}{\mathbf{P}\left(U + V = 6\right)} \\ &= \frac{\mathbf{P}\left(U = 0\right)\mathbf{P}\left(V = 6\right)}{\mathbf{P}\left(U + V = 6\right)} = \frac{e^{-2\lambda}e^{-3\lambda}\frac{(3\lambda)^6}{6!}}{e^{-5\lambda}\frac{(5\lambda)^6}{6!}} = \left(\frac{3}{5}\right)^6,\end{aligned}$$

and, similarly,

$$\mathbf{P}\left(U = 1 | U + V = 6\right) = \frac{\mathbf{P}\left(U = 1, V = 5\right)}{\mathbf{P}\left(U + V = 6\right)} = 6\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)^5.$$

Comment. For part c), look at the class note of 10/7, Exercises 1,2.