STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 1, Random Variables and Applications

1. (Memoryless Property) The Exponential $Exp(\lambda), \lambda > 0$ and Shifted Geometric $Geo(p), p \in (0,1)$ variables "lose memory"; in predicting the future, the past gets "forgotten", only the present matters, i.e. if $X \in Exp(\lambda)$ or $X \in SGeo(p)$,

$$P(X > x + y \mid X > y) = P(X > x),\ \forall x, y \geq 0, \forall x, y \in \mathbb{N}, \text{respectively}.$$

<u>**Solution:**</u>

Exponential

$X \in Exp(\lambda)$, pdf $f(x) = \lambda e^{-\lambda x},\ x \geq 0$, cdf $F(x) = P(X \leq x) = 1 - e^{-\lambda x}, x \geq 0$.

$$\begin{aligned}
P(X > x + y | X > y) &= \frac{P\Big((X > x + y) \cap (X > y)\Big)}{P(X > y)} = \frac{P(X > x + y)}{P(X > y)} \\
&= \frac{1 - F(x + y)}{1 - F(y)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} \\
&= e^{-\lambda x} = 1 - F(x) = P(X > x).
\end{aligned}$$

Shifted Geometric

$X \in SGeo(p)$, pdf $\begin{pmatrix} x \\ pq^{x-1} \end{pmatrix}_{x=1,\dots}$, cdf $F(x) = P(X \leq x) = 1 - q^x, x \geq 0$. The computations go exactly the same way as above.

2. Messages arrive at an electronic message center at random times, with an average of 9 messages per hour. What is the probability of

a) receiving *exactly* 5 messages during the next hour (event A)?

b) receiving *at least* 5 messages during the next hour (event B)?

<u>**Solution:**</u>

Let X denote the number of messages arriving in the next hour. Arriving messages qualify as "rare" events with the arrival rate $\lambda = 9/\text{hr}$. Therefore, X has a Poisson distribution with parameter $\lambda = 9$. Then

a)

$$P(A) = P(X = 5) \overset{\text{Matlab}}{=} \mathit{poisspdf}(5, 9) = 0.0607.$$

b)

$$P(B) = P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - \mathit{poisscdf}(4, 9) = 0.945.$$

3. After a computer virus entered the system, a computer manager checks the condition of all important files. He knows that each file has probability 0.2 to be damaged by the virus, independently of other files. Find the probability that
a) at least 5 of the first 20 files checked, are damaged (event A);
b) the manager has to check at least 6 files in order to find 3 that are undamaged (event B).

Solution:

a) Let X be the number of damaged files, out of the first 20 files checked. This is the number of "successes" (damaged files) out of 20 "trials" (files checked), thus, it has Binomial distribution with $n = 20, p = 0.2$. Thus,

$$P(A) = P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) = 1 - \mathit{binocdf}(4, 20, 0.2) = 0.3704.$$

b) Now consider "success": a file is undamaged, so $p = 0.8$. We can then rephrase our event as
B: check at least 6 in order to find 3 undamaged,
i.e. at least 6 trials to have 3 successes,
i.e. the 3^{rd} success in at least 6 trials,
i.e. the 3^{rd} success after at least 3 failures.

Let X denote the number of files found to be damaged (failures), before the 3^{rd} undamaged (success) one is found. Then X has Negative Binomial distribution with $n = 3$ and $p = 0.8$. So, we want

$$P(B) = P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - \mathit{nbincdf}(2, 3, 0.8) = 0.0579.$$

4. An exciting computer game is released. Sixty percent of players complete all the levels. Thirty percent of them will then buy an advanced version of the game. Among 15 users who completed all levels,
a) what is the probability that at least two people will buy it (event A)?
b) how many of the people who completed all levels are expected to buy the new version?

Solution:

Let X denote the number of people who (after completing all levels) will buy the advanced version of the game (successes), among the mentioned 15 users (trials). Looks like a Binomial distribution

with $n = 15$. For the other parameter, p (probability of success), let us think what information the problem gives. Denote the events:
B : a person buys advanced version of the game (after completing all levels);
C : a person completes all levels.
Then we are given $P(C) = 0.6$ and $P(B|C) = 0.3$ and we need $p = P(B)$. Recall the formula $P(B \cap C) = P(B|C)P(C)$). But $B \cap C = B$. Thus,

$$p \;=\; P(B) \;= P(B \cap C) \;=\; P(B|C)P(C) \;=\; 0.6 \cdot 0.3 \;=\; 0.18,$$

so $X \in Bino(15, 0.18)$. Then
a)

$$P(A) = P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - \mathit{binocdf}(1, 15, 0.18) = 0.7813.$$

b)

$$E(X) = np = 15 \cdot 0.18 = 2.7,$$

between 2 and 3, more probable 3.

5. Consider a satellite whose work is based on block A, independently backed up by a block B. The satellite performs its task until both blocks A and B fail. The lifetimes of A and B are Exponentially distributed with mean lifetime of 10 years. What is the probability that the satellite will work for more than 10 years (event E)?

Solution:
Both lifetimes T_A and T_B have Exponential distribution with parameter $\lambda = 1/10$ years^{-1} (because the mean is $E(T_A) = E(T_B) = 1/\lambda = \mu = 10$ years). We want

$$\begin{aligned}
P(E) &= P\Big((T_A > 10) \cup (T_B > 10)\Big) \\
&= 1 - P\Big(\overline{(T_A > 10) \cup (T_B > 10)}\Big) \\
&= 1 - P\Big(\overline{(T_A > 10)} \cap \overline{(T_B > 10)}\Big) \\
&= 1 - P\Big((T_A \leq 10) \cap (T_B \leq 10)\Big) \\
&\overset{\text{ind}}{=} 1 - P(T_A \leq 10)P(T_B \leq 10) \\
&= 1 - F_{T_A}(10)F_{T_B}(10) \\
&= 1 - (\mathit{expcdf}(10, 10))^2 \;=\; 0.6004.
\end{aligned}$$

6. Compilation of a computer program consists of 3 blocks that are processed sequentially, one after the other. Each block takes Exponential time with the mean of 5 minutes, independently of other blocks. Compute the probability that the entire program is compiled in less than 12 minutes (event A). Use the Gamma-Poisson formula to compute this probability two ways.

Solution:

Let T denote the total compilation time. Then T is the sum of three independent Exponential variables (the times for each block) with parameter $\lambda = \frac{1}{5}$, therefore it has a Gamma distribution with parameters $\alpha = 3$ and $1/\lambda = 5$. So,

Direct way:

$$P(A) = P(T < 12) = F_T(12) = gamcdf(12, 3, 5) = 0.4303.$$

With the Gamma-Poisson formula: $P(T \le t) = P(X \ge \alpha)$, where X has Poisson distribution with parameter $\lambda t = \frac{1}{5} \cdot 12 = 2.4$. Since T is a continuous random variable, $P(T < 12) = P(T \le 12)$. Then

$$\begin{aligned} P(A) &= P(X \ge 3) = 1 - P(X < 3) = 1 - P(X \le 2) \\ &= 1 - F_X(2) = 1 - poisscdf(2, 2.4) = 0.4303. \end{aligned}$$

Caution!! with strict (or not strict) inequalities for the Poisson variable!

7. On the average, a computer experiences breakdowns every 5 months. The time until the first breakdown and the times between any two consecutive breakdowns are independent Exponential random variables. After the third breakdown, a computer requires a special maintenance.

a) Find the probability that a special maintenance is required within the next 9 months (event A);

b) Given that a special maintenance was *not* required during the first 12 months (event B), what is the probability that it will not be required within the next 4 months (event C)?

Solution:

Since computer breakdowns are rare events (they cannot occur simultaneously), the time between two consecutive breakdowns has Exponential distribution. Since breakdowns occur every 5 months, their frequency is $\lambda = 1/5 \text{ months}^{-1}$, so the distribution is $Exp(1/5)$. Now, since breakdowns are independent of each other, the time T until the third breakdown is the sum of 3 $Exp(1/5)$ variables and, hence, has a $Gamma(3, 5)$ distribution.

a) Then we have

$$P(A) = P(T \leq 9) = F_T(9) = gamcdf(9, 3, 5) = 0.2694.$$

b)

$$P(C|B) = P(T > 16 \mid T > 12) = \frac{P(T > 16)}{P(T > 12)} = \frac{1 - F_T(16)}{1 - F_T(12)} = \frac{1 - gamcdf(16, 3, 5)}{1 - gamcdf(12, 3, 5)} = 0.6668.$$

8. A test for a certain viral infection is 95% reliable for infected patients (i.e. it gives a correct positive result) and 99% reliable for the healthy ones (i.e. gives a correct negative result). It is known that 4% of the population is infected with that virus.
a) How reliable is the test in general (i.e. what is the probability that it shows a correct result)?
b) If a patient got a positive result, how likely is it that she truly is infected?

Solution:

Denote the events
C: the test gives a correct result,
PR: the test gives a positive result,
V: a person has the virus (is infected).
What is given:

$$P(C|V) = P(PR|V) = 0.95,\ P(C|\overline{V}) = P(\overline{PR}|\overline{V}) = 0.99 \text{ and } P(V) = 0.04.$$

a) What we want is $P(C)$ (without any condition).
Notice that $\{V, \overline{V}\}$ form a partition of the sample space. By the Total Probability Rule, we have

$$\begin{aligned} P(C) &= P(C|V)P(V) + P(C|\overline{V})P(\overline{V}) \\ &= 0.95 \times 0.04 + 0.99 \times 0.96 = 0.9884. \end{aligned}$$

b) Here, we want $P(V|PR)$, which is given by

$$P(V|PR) = \frac{P(V \cap PR)}{P(PR)}.$$

The numerator is

$$P(V \cap PR) = P(V)P(PR|V) = 0.04 \times 0.95 = 0.038.$$

For the denominator, we use the TPR again, with the same partition $\{V, \overline{V}\}$:

$$\begin{aligned} P(PR) &= P(PR|V)P(V) + P(PR|\overline{V})P(\overline{V}) \\ &= P(PR|V)P(V) + \left[1 - P(\overline{PR}|\overline{V})\right]P(\overline{V}) \\ &= 0.95 \times 0.04 + 0.01 \times 0.96 \;\; = \;\; 0.0476. \end{aligned}$$

Thus, the probability that the patient is indeed infected, is

$$P(V|PR) = \frac{0.038}{0.0476} = 0.7983.$$