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 $X \in SGeo(p)$,

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

Solution:

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 $\begin{pmatrix} x \\ pq^{x-1} \end{pmatrix}_{x=1,\ldots}$, cdf $F(x) = P(X \leq x) = 1 - q^x, x = 0, 1, \ldots$ 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)?

Solution:

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)

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

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), hence, it has a 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. 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

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

5. 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:

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

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

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

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

6. Under good weather conditions, 80% of flights arrive on time. During bad weather, only 50% of flights arrive on time. Tomorrow, the chance of good weather is 60%. What is the probability that your flight will arrive on time?

Solution:

Denote the events
A: the flight arrives on time,
G: there's good weather.

What is given:

$$P(A|G) = 0.8,\ P(A|\overline{G}) = 0.5 \text{ and } P(G) = 0.6 \ \big(P(\overline{G}) = 0.4\big).$$

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

$$\begin{aligned} P(A) &= P(A|G)P(G) + P(A|\overline{G})P(\overline{G}) \\ &= 0.8 \cdot 0.6 + 0.5 \cdot 0.4 \;=\; 0.68. \end{aligned}$$