STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 1, Random Variables and Applications

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

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

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)?

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).

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)?

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.

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?

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

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

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?

<u>Solution</u>:

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

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

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 2

Computer Simulations of Discrete Random Variables; Discrete Methods

1. Function **rnd** in Statistics Toolbox; special functions **rand** and **randn**.

2. Using a Standard Uniform $U(0,1)$ random number generator, write Matlab codes that simulate the following common discrete probability distributions:

a. Bernoulli Distribution $Bern(p)$, with parameter $p \in (0,1)$:

$$X \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

b. Binomial Distribution $B(n,p)$, with parameters $n \in \mathbb{N}, p \in (0,1)$:

$$X \begin{pmatrix} k \\ C_n^k p^k q^{n-k} \end{pmatrix}_{k=\overline{0,n}}$$

c. Geometric Distribution $Geo(p)$, with parameter $p \in (0,1)$:

$$X \begin{pmatrix} k \\ pq^k \end{pmatrix}_{k \in \mathbb{N}}$$

d. Negative Binomial Distribution $NB(n,p)$ with parameters $n \in \mathbb{N}, p \in (0,1)$:

$$X \begin{pmatrix} k \\ C_{n+k-1}^k p^n q^k \end{pmatrix}_{k \in \mathbb{N}}$$

e. Poisson Distribution $\mathcal{P}(\lambda)$ with parameter $\lambda > 0$:

$$X \begin{pmatrix} k \\ \dfrac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k \in \mathbb{N}}$$

Optional

f. Discrete Uniform Distribution $U(m)$ with parameter $m \in \mathbb{N}$:

$$X \begin{pmatrix} k \\ \dfrac{1}{m} \end{pmatrix}_{k=\overline{1,m}}$$

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 3

Computer Simulations of Random Variables and Monte Carlo Studies; Inverse Transform Method, Rejection Method, Special Methods

1.
a) Use the DITM to generate a $Geo(p), p \in (0,1)$, variable.

b) Then use that to generate a $NB(n,p), n \in \mathbb{N}, p \in (0,1)$, variable.

2.
a) Use the ITM to generate an $Exp(\lambda), \lambda > 0$, variable.

b) Then use that to generate a $Gam(\alpha, \lambda), \alpha \in \mathbb{N}, \lambda > 0$, variable (a Gamma $Gam(\alpha, \lambda)$ variable is the sum of α independent $Exp(1/\lambda)$ variables).

3. Use a special method to generate a $Poiss(\lambda), \lambda > 0$, variable.

4 Use the rejection method to approximate π (see Example 7.2, Lecture 4).

5. Application: Forecasting for new software release
An IT company is testing a new software to be released. Every day, software engineers find a random number of errors and correct them. On each day t, the number of errors found, X_t, has a Poisson(λ_t) distribution, where the parameter λ_t is the lowest number of errors found during the previous k days,

$$\lambda_t = \min\{X_{t-1}, X_{t-2}, \ldots, X_{t-k}\}.$$

If some errors are still undetected after $tmax$ days (i.e. if not all errors are found in $tmax - 1$ days), the software is withdrawn and goes back to development. Generate a Monte Carlo study to estimate
a) the time it will take to find all errors;
b) the total number of errors found in this new release;
c) the probability that the software will be sent back to development.
(Try $k = 4, [X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}] = [10, 5, 7, 6], tmax = 10.$)

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 4, Markov Chains, Applications and Simulations

1. (Operating mode) A computer system can operate in two different modes. Every hour, it remains in the same mode or switches to a different mode according to the transition probability matrix

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix}.$$

a) If the system is in Mode I at 5:30 pm, what is the probability that it will be in Mode I at 8:30 pm on the same day?
b) In the long run, in which mode is the system more likely to operate?

2. (Genetics) An offspring of a black dog is black with probability 0.6 and brown with probability 0.4. An offspring of a brown dog is black with probability 0.2 and brown with probability 0.8. Rex is a brown dog. What is the probability that his grandchild is black?

3. (Traffic lights) Every day, student A takes the same road from his home to the university. There are 4 street lights along his way, and he noticed the following pattern: if he sees a green light at an intersection, then 60% of the time the next light is also green (otherwise, red), and if he sees a red light, then 70% of the time the next light is also red (otherwise, green).
a) If the first light is green, what is the probability that the third light is red?
b) Student B has *many* street lights between his home and the university, but he notices the same pattern. If the first street light on his road is green, what is the probability that the last light is red?

4. (Shared device) A computer is shared by 2 users who send tasks to it remotely and work independently. At any minute, any connected user may disconnect with probability 0.5, and any disconnected user may connect with a new task with probability 0.2. Let $X(t)$ be the number of concurrent users at time t.
a) Find the transition probability matrix.
b) Suppose there are 2 users connected at 10:00 a.m. What is the probability that there will be 1 user connected at 10.02?
c) How many connections can be expected by noon?

5. (Weather forecast) Recall the example at the lecture, about Rainbow City with sunny/rainy days (state 1 was "sunny" and state 2 was "rainy"), with transition probability matrix

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

a) If the initial forecast is 80% chance of rain, write a Matlab code to generate the forecast for the next 30 days.
b) In Rainbow City, if there are 7 days or more in a row of sunshine, there is the danger of drought, and if it rains for a week or more, there is the threat of flooding. Local authorities need to be prepared for each situation. Use the code from part a) to conduct a Monte Carlo study for estimating the probability of a water shortage and the probability of flooding.

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 4, Markov Chains, Applications and Simulations

1. (Computer mode) A computer system can operate in two different modes. Every hour, it remains in the same mode or switches to a different mode according to the transition probability matrix

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix}.$$

a) If the system is in Mode I at 5:30 pm, what is the probability that it will be in Mode I at 8:30 pm on the same day?
b) In the long run, in which mode is the system more likely to operate?

Solution:

a) This is a stochastic process, with two states, 1, "the system operates in Mode I" and 2, "the system is in Mode II". So, it is *discrete-state*. The time is measured "every hour", so it is also *discrete-time*. In order to predict the future, we only need to know the present, i.e. how the computer changes modes from one hour to the next, hence, it is also *Markov* and, thus, a *Markov chain*. Also, since the probabilities of switching from one mode to another are the same at any time (hour), it is a *homogeneous* Markov chain.
The initial time is 5:30 pm. Now, 8:30 pm is 3 hours after 5:30 pm, so we want to compute

$$p_{11}^{(3)} = P(X_3 = 1 \mid X_0 = 1),$$

for which we need the 3-step transition probability matrix, $P^{(3)} = P^3$. In Matlab,

```
>> P = [0.4 0.6; 0.6, 0.4]
P =
    0.4000    0.6000
    0.6000    0.4000
>> P^3
ans =
    0.4960    0.5040
    0.5040    0.4960
```

$$P^3 = \begin{bmatrix} p_{11}^{(3)} & p_{12}^{(3)} \\ p_{21}^{(3)} & p_{22}^{(3)} \end{bmatrix} = \begin{bmatrix} 0.496 & 0.504 \\ 0.504 & 0.496 \end{bmatrix}.$$

So that probability is $p_{11}^{(3)} = 0.496$.

b) For the "long run", we need the steady-state distribution. Notice that P (and P^3) has all nonzero entries, so the Markov chain is *regular*, which means a steady-state distribution does exist. We find it by solving the system $\pi P = \pi, \sum_{x=1}^{2} \pi_x = 1$,

$$\begin{aligned} [\pi_1 \;\; \pi_2] \begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix} &= [\pi_1 \;\; \pi_2] \\ \pi_1 + \pi_2 &= 1, \end{aligned}$$

i.e,

$$\left\{ \begin{array}{rcrcl} 0.4\pi_1 & + & 0.6\pi_2 & = & \pi_1 \\ 0.6\pi_1 & + & 0.4\pi_2 & = & \pi_2 \\ \pi_1 & + & \pi_2 & = & 1 \end{array} \right. \text{ or, equivalently, } \left\{ \begin{array}{rcrcl} \pi_1 & - & \pi_2 & = & 0 \\ \pi_1 & + & \pi_2 & = & 1 \end{array} \right.$$

with solution $\pi_1 = \pi_2 = 0.5$. So, in the long run, the pdf of the forecast will be

$$\lim_{h \to \infty} P_h = \pi = [\pi_1 \;\; \pi_2] = [0.5 \;\; 0.5].$$

That means that, in the long run, the system is just as likely to operate in Mode I, as it is to operate in Mode II. (Notice that even after only 3 steps, the transition probabilities were already very close to 0.5.)

2. (Genetics) An offspring of a black dog is black with probability 0.6 and brown with probability 0.4. An offspring of a brown dog is black with probability 0.2 and brown with probability 0.8. Rex is a brown dog. What is the probability that his grandchild is black?

Solution:

This is a stochastic process with 2 states. Let "black" be state 1 and "brown" be state 2. The time is measured from one generation to the next ("offspring"), so discretely. It has the Markov property (only information about the previous generation is needed), so it is a Markov chain. Again, the transition probabilities are stationary (the same at any time), thus, so is the Markov chain (stationary or homogeneous).

The transition probability matrix is

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}.$$

Rex is a brown dog (state 2), so the initial situation is

$$P_0 = [0 \;\; 1].$$

His grandchild is two generations away, so we need the first component of P_2,

$$P_2 \;\; = \;\; P_0 \cdot P^2 \;\; = \;\; [0 \;\; 1] \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} \end{bmatrix} \;\; = \;\; [p_{21}^{(2)} \;\; p_{22}^{(2)}],$$

i.e. $p_{21}^{(2)}$. We could compute the entire matrix P^2, or just that one entry. The entry $p_{21}^{(2)}$ in P^2 is obtained from the second row and first column of P:

$$p_{21}^{(2)} \;\; = \;\; [p_{21} \, p_{22}] \begin{bmatrix} p_{11} \\ p_{21} \end{bmatrix} \;\; = \;\; p_{21} \cdot p_{11} + p_{22} \cdot p_{21} \;\; = \;\; (0.2)(0.6) + (0.8)(0.2) \;\; = \;\; 0.28.$$

3. (Traffic lights) Every day, student A takes the same road from his home to the university. There are 4 street lights along his way, and he noticed the following pattern: if he sees a green light at an intersection, then 60% of the time the next light is also green (otherwise, red), and if he sees a red light, then 70% of the time the next light is also red (otherwise, green).

a) If the first light is green, what is the probability that the third light is red?

b) Student B has *many* street lights between his home and the university, but he notices the same pattern. If the first street light on his road is green, what is the probability that the last light is red?

Solution:

This is a Markov chain with 2 states, "green light" state 1 and "red light" state 2. It is also homogeneous.

The transition probability matrix is

$$P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}.$$

a) The initial situation (first light) is green, so

$$P_0 = [1 \;\; 0].$$

Now, we want the second component of P_2 (the third light is two steps after the first light),

$$P_2 \;=\; P_0 \cdot P^2 \;=\; [1 \;\; 0] \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} \end{bmatrix} \;=\; [p_{11}^{(2)} \;\; p_{12}^{(2)}]],$$

i.e. $p_{12}^{(2)}$. Again, we can compute that directly (without finding the entire matrix P^2), multiplying the first row and second column of P:

$$p_{12}^{(2)} \;=\; [p_{11} \; p_{12}] \begin{bmatrix} p_{12} \\ p_{22} \end{bmatrix} \;=\; p_{11} \cdot p_{12} + p_{12} \cdot p_{22} \;=\; (0.6)(0.4) + (0.4)(0.7) \;=\; 0.52.$$

b) For "many streets" away, we use the steady-state distribution. Since P has all nonzero entries, this Markov chain is regular, so a steady-state distribution exists. We set up the system $\pi P = \pi, \sum_{x=1}^{2} \pi_x = 1$, i.e,

$$\begin{cases} 0.6\pi_1 & + & 0.3\pi_2 & = & \pi_1 \\ 0.4\pi_1 & + & 0.7\pi_2 & = & \pi_2 \\ \pi_1 & + & \pi_2 & = & 1 \end{cases},$$

with solution $\pi_1 = 3/7, \pi_2 = 4/7 \approx 0.5714$. So, after "many streets", the probability of a red light (i.e. that the process is in state 2) is

$$\pi_2 \;=\; 4/7 \;\approx\; 0.5714.$$

4. (Shared device) A computer is shared by 2 users who send tasks to it remotely and work independently. At any minute, any connected user may disconnect with probability 0.5, and any disconnected user may connect with a new task with probability 0.2. Let $X(t)$ be the number of concurrent users at time t.

a) Find the transition probability matrix.

b) Suppose there are 2 users connected at 10:00 a.m. What is the probability that there will be 1 user connected at 10.02?

c) How many connections can be expected by noon?

Solution:

The number of concurrent users at time t, $X(t)$, can take the values $\{0, 1, 2\}$, the time changes by the minute (a discrete set), the probabilities of connecting/disconnecting depend only on the previous value of concurrent users and they are the same at any time (minute), so this is a homogeneous Markov chain with three states: $0, 1$ and 2.

a) Let us find each row of the transition probability matrix P.

For the first row, we want the transition probabilities form state 0 to each of the states $0, 1, 2$,

$$\text{Prow1} = [p_{00}\ \ p_{01}\ \ p_{02}].$$

If $X_0 = 0$, i.e. there are no users at time $t = 0$, then X_1, the number of new connections within the next minute is the number of successes in $n = 2$ trials, with probability of success ("to connect") $p = 0.2$, i.e. has $Bino(2, 0.2)$ distribution. We find it in Matlab with

```
>> Prow1 = binopdf(0:2,2,0.2)
Prow1 =
    0.6400    0.3200    0.0400
```

For the second row, suppose $X_0 = 1$, i.e one user is connected, the other is not. We compute each transition probability.

$$\begin{aligned}
p_{10} &= P\big(\text{(the connected user disconnects)} \cap \text{(the disconnected user does not connect)}\big) \\
&\overset{ind}{=} 0.5 \cdot 0.8 \;=\; 0.4, \\
p_{11} &= P\Big[\big(\text{(the connected user does not disconnect)} \cap \text{(the disconnected user does not connect)}\big) \\
&\quad \cup \big(\text{(the connected user does disconnect)} \cap \text{(the disconnected user does connect)}\big)\Big] \\
&= 0.5 \cdot 0.8 + 0.5 \cdot 0.2 \;=\; 0.5, \\
p_{12} &= P\big(\text{(the connected user does not disconnect)} \cap \text{(the disconnected use does connect)}\big) \\
&= 0.5 \cdot 0.2 \;=\; 0.1.
\end{aligned}$$

So, row 2 of P is

```
Prow2  = [0.4    0.5    0.1]
Prow2 =
    0.4000    0.5000    0.1000
```

Finally, if $X_0 = 2$, i.e. both users are connected, then no new users can connect and the number of *disconnections*, Y, is $Bino(2, 0.5)$ distributed. That means that the number of *concurrent users* will be $2 - Y$, so the corresponding probabilities are

```
>> Prow3 = binopdf(2:-1:0,2,0.5)
Prow3 = 0.2500    0.5000    0.2500
```

Then the transition probability matrix is

```
>> P = [Prow1; Prow2; Prow3]
P =

    0.6400    0.3200    0.0400
    0.4000    0.5000    0.1000
    0.2500    0.5000    0.2500
```

b) Initial situation, at 10:00 a.m., there are 2 users connected, so

$$P_0 = [0 \;\; 0 \;\; 1].$$

At 10:02, after two steps, we want

$$P_2 = P_0 \cdot P^2,$$

in Matlab,

```
>> P_2 = [0 0 1] * P^2
P_2 =
    0.4225    0.4550    0.1225
```

So, the probability that at 10:02 there is one user connected is the second component of the vector above,

$$p_{21}^{(2)} = 0.455.$$

c) Noon is *many* minutes after 10:00, so we use the steady-state distribution, which we know exists, since P has all non-zero entries. We find it in Matlab. We write the system in the form

$$Ax = b,$$

with A the coefficients matrix, so a 4×3 (singular) matrix and b the right-hand side constant vector (of dimension 4×1). To set up A, notice that for the equations $\pi P = \pi$, the coefficients on the left are those of the *transpose* of P, P^T (in Matlab the transpose is prime, '). Then we have to subtract the π_1, π_2, π_3 from the right-hand side, i.e. the identity matrix I_3 and finally add a row of 1's, the coefficients from the last equation $\sum_{x=0}^{2} \pi_x = 1$. So the matrix A is

```
>> A = [P' - eye(3); 1 1 1]
A =
   -0.3600    0.4000    0.2500
    0.3200   -0.5000    0.5000
    0.0400    0.1000   -0.7500
    1.0000    1.0000    1.0000
```

and the vector b

```
>> b = [0; 0; 0; 1]
b =
     0
     0
     0
     1
```

Then we solve the system by

```
>> x = A\b
x =
    0.5102
    0.4082
    0.0816
```

So, the steady-state distribution is $\pi = [\pi_0 \;\; \pi_1 \;\; \pi_2] = [0.5102 \;\;\; 0.4082 \;\;\; 0.0816]$.
Then the expected nr. of connections by noon is the expected value of the steady-state distribution, i.e. the random variable with pdf

$$\begin{pmatrix} 0 & 1 & 2 \\ 0.5102 & 0.4082 & 0.0816 \end{pmatrix},$$

so,

$$0 \cdot 0.5102 + 1 \cdot 0.4082 + 2 \cdot 0.0816 = 0.5714,$$

between 0 and 1, (slightly) more probable 1.

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 5, Counting Processes

1. On the average, 6 airplanes per minute land at a certain international airport. Assume the number of landings is modeled by a Binomial counting process.
a) What frame length should be used to guarantee that the probability of a landing does not exceed 0.1?
b) Using the chosen frames, compute the probability of no landings during the next half a minute;
c) Using the chosen frames, compute the probability of more than 170 landed airplanes during the next 30 minutes.

2. Messages arrive at a communications center according to a Binomial counting process with 30 frames per minute. The average arrival rate is 40 messages per hour. How many messages can be expected to arrive between 10 a.m. and 10:30 a.m.? What is the standard deviation of that number of messages?

3. An internet service provider offers special discounts to every third connecting customer. Its customers connect to the internet according to a Poisson process with the rate of 5 customers per minute. Compute
a) the probability that no offer is made during the first 2 minutes;
b) the probability that no customers connect for 20 seconds;
c) expectation and standard deviation of the time of first offer.

4. On the average, Mr. X drinks and drives once in 4 years. He knows that

- every time he drinks and drives, he is caught by the police;
- according to the law of his state, the third time he is caught drinking and driving, he loses his driver's license;
- a Poisson counting process models such "rare events" as drinking and driving.

What is the probability that Mr. X will keep his driver's license for at least 10 years?

5. Simulation and illustration of Binomial and Poisson counting processes.
a) Given sample path size N_B and probability of arrival p, simulate a Binomial counting process $X(t)$.
Application: For a frame size of 1 second, simulate the number of airplane landings from Problem 1., for 1 minute.
b) Given frequency λ and a time frame $[0, T_{max}]$, simulate a Poisson counting process $X(t)$.
Application: Simulate the number of internet connections from Problem 3., for a period of half an hour.

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 5, Counting Processes

1. On the average, 6 airplanes per minute land at a certain international airport. Assume the number of landings is modeled by a Binomial counting process.
a) What frame length should be used to guarantee that the probability of a landing does not exceed 0.1?
b) Using the chosen frames, compute the probability of no landings during the next half a minute;
c) Using the chosen frames, compute the probability of more than 170 landed airplanes during the next 30 minutes.

Solution:
a) We have $\lambda = 6/\,\mathrm{min}$. So, if we want $p \le 0.1$, then

$$\Delta \;=\; \frac{p}{\lambda} \le \frac{0.1}{6}\,\mathrm{min} \;=\; 1\,\mathrm{sec.}$$

b) Let $\Delta = 1$ sec. In $t = 1/2$ min. $= 30$ sec., there are $n = \dfrac{t}{\Delta} = \dfrac{30}{1} = 30$ frames. The number of landings $X(30)$ during 30 frames has Binomial distribution with parameters $n = 30$ and $p = 0.1$. We want

$$P(X(30) = 0) \;=\; \mathit{binopdf}(0, 30, 0.1) \;=\; 0.0424.$$

c) Similarly, in 30 minutes $= 1800$ sec., there are $n = \dfrac{1800}{1} = 1800$ frames. Thus, the number of landings $X(1800)$ during the next half hour has Binomial distribution with parameters $n = 1800$ and $p = 0.1$. We want to compute

$$P(X(1800) > 170) \;=\; 1 - P(X(1800) \le 170) \;=\; 1 - \mathit{binocdf}(170, 1800, 0.1) \;=\; 0.7709.$$

2. Messages arrive at a communications center according to a Binomial counting process with 30 frames per minute. The average arrival rate is 40 messages per hour. How many messages can be expected to arrive between 10 a.m. and 10:30 a.m.? What is the standard deviation of that number of messages?

Solution:

The arrival rate is $\lambda = 40/\text{hr} = \frac{2}{3}/$ min. and the frame length is $\Delta = \frac{1}{30}$ min. Then, the probability of a new message arriving during any given frame is

$$p = \lambda\Delta = \frac{2}{3} \cdot \frac{1}{30} = \frac{1}{45}.$$

Between 10 and 10:30 a.m., there are $t = 30$ minutes, so there are $n = \frac{t}{\Delta} = \frac{30}{1/30} = 900$ frames, hence, the number of new messages $X(900)$ arriving during this time has Binomial distribution with $n = 900$ and $p = 1/45$, so

$$\begin{aligned} E(X) &= np = 20 \text{ messages}, \\ \sigma(X) &= \sqrt{np(1-p)} = 4.4222 \text{ messages}. \end{aligned}$$

3. An internet service provider offers special discounts to every third connecting customer. Its customers connect to the internet according to a Poisson process with the rate of 5 customers per minute. Compute

a) the probability that no offer is made during the first 2 minutes;

b) the probability that no customers connect for 20 seconds;

c) expectation and standard deviation of the time of first offer.

Solution:

a) We have $\lambda = 5/\text{min}$. In $t = 2$ minutes, the number of connections X has Poisson distribution with parameter $\lambda t = 5 \cdot 2 = 10$. No offer is made if there are *fewer* than three connections during that time. So, we want

$$P(\text{no offer}) = P(X < 3) = P(X \le 2) = \mathit{poisscdf}(2, 10) = 0.0028.$$

b) The time between connections (interarrival time) T has Exponential(λ) distribution. No customers connect for 20 sec. $= 1/3$ minutes, if the interarrival time exceeds that. So, we want

$$P(T > 1/3) = 1 - P(T \le 1/3) = 1 - \mathit{expcdf}(1/3, 1/5) = 0.1889.$$

Or we can express λ in seconds, $\lambda = 5/60 = 1/12/\text{sec}$. Then we compute

$$P(T > 20) = 1 - P(T \le 20) = 1 - \mathit{expcdf}(20, 12) = 0.1889.$$

Alternatively, we want 0 connections (arrivals) in $t = 20$ seconds $= 1/3$ minutes. Just like in part a), the number of connections X has Poisson distribution with parameter $\lambda t = 5 \cdot 1/3 = 5/3$. Then we compute

$$P(X = 0) \ = \ poisspdf(0, 5/3) \ = \ 0.1889.$$

c) The time T_3 of the third connection (arrival) (and therefore, the first offer) is the sum of 3 independent $Exp(5)$ times, so it has Gamma distribution with parameters $\alpha = 3$ and $\lambda = 1/5$. Then

$$\begin{aligned} E(T_3) &= \alpha\lambda = 3/5 = 0.6 \text{ min}, \\ \sigma(T_3) &= \sqrt{V(T_3)} = \sqrt{\alpha\lambda^2} = 0.3464 \text{ min}. \end{aligned}$$

4. On the average, Mr. X drinks and drives once in 4 years. He knows that

- every time he drinks and drives, he is caught by the police;
- according to the law of his state, the third time he is caught drinking and driving, he loses his driver's license;
- a Poisson counting process models such "rare events" as drinking and driving.

What is the probability that Mr. X will keep his driver's license for at least 10 years?

Solution:

The arrival rate of drinking and driving is $\lambda = 1/4$/year. Let X be the number of times Mr. X is caught drinking and driving during 10 years. Then X has Poisson distribution with parameter $\lambda t = (1/4)(10) = 2.5$. Keeping his driver's license is equivalent to being caught *less* than three times in 10 years. Then,

$$\begin{aligned} P(\text{Mr. X keeps his driver's license}) &= P(X < 3) = P(X \leq 2) \\ &= poisscdf(2, 2.5) = 0.5438. \end{aligned}$$

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 6, Queuing Systems

1. Performance of a car wash center is modeled by a B1SQP with 2-minute frames. Cars arrive every 10 minutes, on the average, and the average service time is 6 minutes. There are no cars at the center at 10:00 a.m., when the center opens. What is the probability that at 10:04 one car is being washed and another is waiting?

2. A metered parking lot with two parking spaces is modeled by a Bernoulli two-server queuing system with capacity limited by two cars and 30-second frames. Cars arrive at the rate of one car every 4 minutes and each car is parked for 5 minutes, on the average.
a) find the transition probability matrix for the number of parked cars;
b) find the steady-state distribution for the number of parked cars;
c) what fraction of the time are both parking spaces vacant?
d) what fraction of arriving cars will not be able to park?
e) every 2 minutes of parking costs 25 cents; assuming all drivers use all the parking time they pay for, how much money is the parking lot going to raise every 24 hours?

3. Trucks arrive at a weigh station according to a Poisson process with average rate of 1 truck every 10 minutes. Inspection times are Exponential with the average of 3 minutes. When a truck is on the scale, the other arrived trucks stay in line waiting for their turn. Compute
a) the expected number of trucks at the weigh station at any time;
b) the proportion of time when the weigh station is empty;
c) the expected time each truck spends at the station, from arrival to departure;
d) the fraction of time there are fewer than 2 trucks in the weigh station.

4. A toll area on a highway has three toll booths and works as an M/M/3 queuing system. On the average, cars arrive at the rate of one car every 5 seconds, and it takes 12 seconds to pay the toll, not including the waiting time. Compute the fraction of time when there are ten or more cars waiting in the line.

5. Sports fans tune to a local sports radio station according to a Poisson process with the rate of three fans every two minutes and listen to it for an Exponential amount of time with the average of 20 minutes.
a) what queuing system is the most appropriate for this situation?
b) compute the expected number of concurrent listeners at any time;
c) find the fraction of time when 40 or more fans are tuned to this station.

<u>**Simulations**</u>

6. Messages arrive at an electronic mail server according to a Poisson process with the average frequency of 5 messages per minute. The server can process only one message at a time and messages are processed on a "first come – first serve" basis. It takes an Exponential amount of time M_1 to process any text message, plus an Exponential amount of time M_2, independent of M_1, to process attachments (if there are any), with $E(M_1) = 2$ seconds and $E(M_2) = 7$ seconds. Forty percent of messages contain attachments. Use Monte Carlo methods to estimate
a) the expected response time of this server;
b) the expected waiting time of a message before it is processed.

7. A small clinic has several doctors on duty, but only one patient is seen at a time. Patients are scheduled to arrive at equal 15-minute intervals, are then served in the order of their arrivals and each of them needs a Gamma time with the doctor, that has parameters $\alpha = 4$ and $\lambda = 10/3 \text{ min}^{-1}$. Use Monte Carlo simulations to estimate
a) the probability that a patient has to wait before seeing the doctor;
b) the expected waiting time for a patient;

STATISTICAL COMPUTATIONAL METHODS

Seminar Nr. 6, Queuing Systems

1. Performance of a car wash center is modeled by a B1SQP with 2-minute frames. Cars arrive every 10 minutes, on the average, and the average service time is 6 minutes. There are no cars at the center at 10:00 a.m., when the center opens. What is the probability that at 10:04 one car is being washed and another is waiting?

Solution:

We have

$$\begin{aligned} \Delta &= 2 \text{ min.}, \\ \lambda_A &= 1/\mu_A = 1/10 \text{ min}^{-1}, \\ \lambda_S &= 1/\mu_S = 1/6 \text{ min}^{-1}, \text{ so} \\ p_A &= \lambda_A \Delta = 1/5 \text{ and} \\ p_S &= \lambda_S \Delta = 1/3. \end{aligned}$$

There are no cars at the beginning (10:00 a.m), so the value of $X_0 = 0$, i.e. the initial distribution of X_0 is

$$P_0 = [1\ 0\ 0 \ldots].$$

There are 2 frames between 10:00 am and 10:04 am. We want the conditional probability

$$P(X_2 = 2 \mid X_0 = 0).$$

Since the number of cars at the wash center can change by at most 1 during each frame, this probability equals

$$\begin{aligned} p_{02}^{(2)} &= [\text{1st row of } P][\text{3rd column of } P] = [1 - p_A\ \ p_A\ \ 0] \begin{bmatrix} 0 \\ p_A(1 - p_S) \\ (1 - p_A)(1 - p_S) + p_A p_S \end{bmatrix} \\ &= p_A^2 (1 - p_S) = \frac{1}{25} \cdot \frac{2}{3} = \frac{2}{75} = 0.0267. \end{aligned}$$

2. A metered parking lot with two parking spaces is modeled by a Bernoulli two-server queuing system with capacity limited by two cars and 30-second frames. Cars arrive at the rate of one car every 4 minutes and each car is parked for 5 minutes, on the average.
a) find the transition probability matrix for the number of parked cars;
b) find the steady-state distribution for the number of parked cars;
c) what fraction of the time are both parking spaces vacant?
d) what fraction of arriving cars will not be able to park?
e) every 2 minutes of parking costs 25 cents; assuming all drivers use all the parking time they pay for, how much money is the parking lot going to raise every 24 hours?

Solution:
This is a B2SQS with

$$C = 2, k = 2, \Delta = 1/2 \text{ min}, \lambda_A = 1/4 \text{ min}^{-1} \text{ and } \lambda_S = 1/5 \text{ min}^{-1}.$$

a) $p_A = \lambda_A \Delta = 1/8$ and $p_S = \lambda_S \Delta = 1/10$. There are 3 states $\{0, 1, 2\}$ and the transition probabilities are

$$\begin{aligned}
p_{00} &= 1 - p_A = 7/8, \\
p_{01} &= p_A = 1/8, \\
p_{02} &= 0; \\
p_{10} &= p_S(1 - p_A) = 7/80, \\
p_{11} &= (1 - p_A)(1 - p_S) + p_A p_S = 4/5, \\
p_{12} &= p_A(1 - p_S) = 9/80; \\
p_{20} &= p_S^2(1 - p_A) = 7/800, \\
p_{21} &= 2p_S(1 - p_A)(1 - p_S) + p_S^2 p_A = 127/800, \\
p_{22} &= (1 - p_A)(1 - p_S)^2 + 2p_A p_S(1 - p_S) + p_A(1 - p_S)^2 = 333/400.
\end{aligned}$$

So, the transition probability matrix is

$$P = \begin{bmatrix} 7/8 & 1/8 & 0 \\ 7/80 & 4/5 & 9/80 \\ 7/800 & 127/800 & 333/400 \end{bmatrix}.$$

b) The steady-state distribution is

$$\pi = [\pi_0 \;\; \pi_1 \;\; \pi_2] = [0.3089 \quad 0.4135 \quad 0.2777].$$

c) That would be

$$P(X=0) \;=\; \pi_0 \;=\; 0.3089,$$

or 30.89% of the time.

d) A car cannot park if both spaces are taken, so

$$P(X=2) \;=\; \pi_2 \;=\; 0.2777,$$

or 27.77% of cars.

e) The expected number of parked cars is

$$E(X) \;=\; \sum_0^2 x\pi_x \;=\; 0 \cdot 0.3089 + 1 \cdot 0.4135 + 2 \cdot 0.2777 \;=\; 0.9689.$$

Then the total revenue in 24 hours is

$$E(X) \cdot 24 \cdot 60 \cdot 0.25/2 \;=\; 174.4020 \text{ dollars.}$$

3. Trucks arrive at a weigh station according to a Poisson process with average rate of 1 truck every 10 minutes. Inspection times are Exponential with the average of 3 minutes. When a truck is on the scale, the other arrived trucks stay in line waiting for their turn. Compute
a) the expected number of trucks at the weigh station at any time;
b) the proportion of time when the weigh station is empty;
c) the expected time each truck spends at the station, from arrival to departure;
d) the fraction of time there are fewer than 2 trucks in the weigh station.

Solution:

A Poisson process of arrivals implies Exponential interarrival times, so the described system is M/M/1 with $\mu_A = 10$ min and $\mu_S = 3$ min. Hence, $\lambda_A = 1/10\,\text{min}^{-1}$, $\lambda_S = 1/3\,\text{min}^{-1}$ and $r = \lambda_A/\lambda_S = 0.3 < 1$.

a) The expected number of trucks at the weigh station is

$$E(X) = \frac{r}{1-r} = 3/7 = 0.4286.$$

b) The proportion of time when the weigh station is empty is

$$P(X=0) = 1-r = 0.7,$$

or 70% of time.

c) This is the expected response time

$$E(R) = \frac{\mu_S}{1-r} = \frac{3}{0.7} = 30/7,$$

or 4.2857 minutes.

d) This is the probability

$$\begin{aligned} P(X<2) &= P(\{X=0\} \cup \{X=1\}) = P(0)+P(1) \\ &= \pi_0 + \pi_1 = (1-r) + r(1-r) \\ &= 1-r^2 = 1-0.09 = 0.91, \end{aligned}$$

or 91% of the time.

4. A toll area on a highway has three toll booths and works as an M / M / 3 queuing system. On the average, cars arrive at the rate of one car every 5 seconds, and it takes 12 seconds to pay the toll, not including the waiting time. Compute the fraction of time when there are ten or more cars waiting in the line.

Solution:

We have

$$\lambda_A = 1/5\,\mathrm{sec}^{-1},\ \lambda_S = 1/12\,\mathrm{sec}^{-1},\ k = 3 \text{ and } r = \lambda_A/\lambda_S = 12/5 = 2.4 < 3.$$

We want to compute

$$P(X_w \geq 10).$$

Since there are 3 toll booths (all busy), this is the same as

$$P(X \geq 13) = \sum_{x=13}^{\infty} \pi_x.$$

Now, for the steady-state distribution, we have

$$\begin{aligned}
\pi_0 &= \frac{1}{1 + r + \dfrac{r^2}{2} + \dfrac{r^3}{6(1 - r/3)}} = 0.0562, \\
\pi_1 &= \ldots, \\
\pi_2 &= \ldots, \\
\pi_x &= \frac{r^3}{3!}\pi_0 \left(\frac{r}{3}\right)^{x-3}, \text{ for any } x \geq 3.
\end{aligned}$$

So,

$$\begin{aligned}
P(X_w \geq 10) &= \frac{r^3\pi_0}{3!} \sum_{x=13}^{\infty} \left(\frac{r}{3}\right)^{x-3} = \frac{r^3\pi_0}{3!} \left(\frac{r}{3}\right)^{10} \left[1 + \left(\frac{r}{3}\right) + \left(\frac{r}{3}\right)^2 + \ldots\right] \\
&= \frac{r^3\pi_0}{6} \cdot \frac{(r/3)^{10}}{1 - r/3} = 0.0695,
\end{aligned}$$

or 6.95% of the time. Note that we used the formula for the sum of a Geometric series with ratio $q < 1$,

$$\sum_{k=a}^{\infty} q^k = q^a \sum_{k=a}^{\infty} q^{k-a} = q^a \sum_{i=0}^{\infty} q^i = \frac{q^a}{1 - q}.$$

5. Sports fans tune to a local sports radio station according to a Poisson process with the rate of three fans every two minutes and listen to it for an Exponential amount of time with the average of 20 minutes.

a) what queuing system is the most appropriate for this situation?

b) compute the expected number of concurrent listeners at any time;

c) find the fraction of time when 40 or more fans are tuned to this station.

Solution:

We have

$$\lambda_A = 3/2 \,\text{min}^{-1}, \; \lambda_S = 1/20 \,\text{min}^{-1}, \; k \approx \infty \text{ and } r = \lambda_A/\lambda_S = 60/2 = 30.$$

a) We have a Poisson process of arrivals (so Exponential interarrival times), Exponential service times and "infinitely" many servers (because any number of people can listen to a radio station simultaneously), therefore, an M / M / ∞ queuing system is the most appropriate to model this situation.

b) The expected number of concurrent listeners is

$$E(X) = r = 30 \text{ listeners.}$$

c) The number of concurrent listeners, X, has Poisson distribution with parameter $r = 30$. So,

$$P(X \geq 40) = 1 - P(X < 40) = 1 - P(X \leq 39) = 1 - poisscdf(39, 30) = 0.0463,$$

or 4.63% of the time.