Solution worksheet 02

Ilia, Dudnik
`ilia.dudnik@fau.de`

Mader, Benedikt
`benedikt.mader@fau.de`

Ganzmann, Tobias
`tobias.ganzmann@fau.de`

Aakash Ram, Chandran
`aakash.chandran@fau.de`

November 10, 2021

Exercise 1

In unsorted sets with n different items, we perform sequential and random searches for one particular item (of which we know that it is included exactly once in each set).

Sequential search means that we pick one item after the other from the set until the desired item is found.

Random search means that we check an arbitrary item from the set, but leave it in the set, and recheck until the desired item is found.

1. What is the **mean number of checks** (including accidental rechecks in random search) until the item is found in a set **in either case** (with sequential and with random search)? Which **distribution** does the random variable "number of checks" have **in each case**? Simply give the names and characteristic parameters of these distributions.

   In either cases the distributions should be the probability distribution of the number X of trials needed to get one success, supported on the set {1, 2, ..., n}.

   - sequential search

     There is no difference in having success exactly on the first check or on the 10th check. Hence, this is the *uniform distribution* with $a = 1$ and $b = n$, $X \in \{1, 2, ..., n\}$. The probability density function in this case:

     $$f_X(X = x|n) = \frac{1}{n-1}.$$

     The mean of the uniform distribution: $\frac{a+b}{2}$. Hence,

     $$E[X] = \frac{n+1}{2}.$$

   - random search

     This is *geometric distribution* with $p = \frac{1}{n}$ and $X \in \mathbb{R}$. Given that,

     $$f_X(X = x|n) = \frac{1}{n} \cdot (1 - \frac{1}{n})^{x-1}.$$

The mean of the geometric distribution: $\frac{1}{p}$. Hence,

$$E[X] = n.$$

2. For the distribution arising from random search: show **formally** that this distribution enjoys the **memoryless property** (in analogy to the bottom of slide 33 of the exercise video "Basics of Probability Theory")! Interpret this memoryless property for random search **informally** (i.e., in your own words).

   - formally

     $f_X(x)$ for the random search is memoryless, then $P(X <= y + z|X > y) = P(X <= z)$. Hence,

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

   - informally

     The probability of getting a number in 30ty tries, considering the fact that 20 tries have already happened, is the same as the probability pf happenning it in 30ty tries from the very beginning.

Exercise 2

In technical systems and software programs, processes are often started at the same time and run in parallel. Before the next phase of the system/program can be entered, sometimes all of these processes have to be finished or sometimes simply a single one. We define the runtime of process **k** with the random variable X_k.

Let $X_1, X_2, \dots, X_n$ be (mutually) independent and let $Y = \mathbf{min}\{X_1, X_2, \dots, X_n\}$ and $Z = \mathbf{max}\{X_1, X_2, \dots, X_n\}$. What do Y and Z model?

- Y models the start time of next phase in case of completion at least one process
- Z models the start time of next phase in case of completion all processes

(a) Show that the distribution functions F_Y of Y and F_Z of Z are given by $F_Y(t) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(t))$ and $F_Z(t) = \prod_{i=1}^{n} F_{X_i}(t)$, respectively.

   - Z is basically when every from $X_1, X_2, \dots, X_n$ should happen simultaneously. According to Z description: $P(Z \le t) = P(X_1 \le t \bigcap X_2 \le t \bigcap \cdots \bigcap X_n \le t)$. Since $X_1, X_2, \dots, X_n$ are mutually independent, then $P(Z \le t) = P(X_1 \le t) \cdot P(X_2 \le t) \cdot \cdots \cdot P(X_n \le t) \Longrightarrow F_Z(t) = \prod_{i=1}^{n} F_{X_i}(t)$.

- Y is basically when only one them happens and that will be enought for us, which can be formulated as negation of probability that none of the $X_1, X_2, \dots, X_n$ happened simultaneously. Let us find the probability that all processes are still running: $\prod_{i=1}^{n}(1 - F_{X_i}(t))$, the proof of this is the same as one above. Hence, Y probability will be 1 - this probability $\implies F_Y(t) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(t))$

(b) Specialize (and simplify) these distribution functions for **independent and identically exponentially distributed** $X_i (i = 1, \dots, n)$! What distribution do you get for **Y**? What is the value of **E[Y]** in this case (exponentially distributed X_i)? Describe this observation in your own words.

Given $F_{X_i}(t) = 1 - e^{-\lambda t}$, where $\lambda \in \mathbb{R}_{>0}$:

- $F_Z(t) = (1 - e^{-\lambda t})^n$
- $F_Y(t) = 1 - (1 - (1 - e^{-\lambda t}))^n = 1 - e^{-\lambda n t}$. $E[Y] = \int_0^\infty t\lambda n \cdot e^{-\lambda n t})\, dt = \frac{1}{n\lambda}$.
  It is obvious that the more processes the less the average time of Y. Hence, the $E[Y]$ is inversely proportional to n.

Exercise 3

A systems runs by a generator that provides power for the system. If the generator fails, the system has a battery, which can supply it with power **for exactly five more days**. Let the time to failure X for the generator be exponentially distributed with **expectation 1700** days, while Y denotes the time to the complete failure of the system.

(a) What is the **probability** that the generator fails **within the first 1700 day**? What is the quantile with respect to this probability value **for random variable Y**?

The expected value of an exponentially distributed random variable X: $\frac{1}{\lambda} \implies \lambda = \frac{1}{1700}$.
$F_X(X \le 1700) = 1 - e^{-\frac{1700}{1700}} \approx 0.6321$.
For Y quantile of 0.6321 will also be at the mean of distribution, when x = 1705.
$F_Y(t) = 1 - e^{-\lambda(y-5)} = 1 - e^{-\lambda(1705-5)} = 0.6321$

(b) Compute the **coefficient of variation for both X and Y**!

- $C_X = \frac{\sigma_X}{E[X]} = \frac{\sqrt{Var[X]}}{E[X]} = 1$
- $C_Y = \frac{\sigma_Y}{E[Y]} = \frac{\sqrt{Var[Y]}}{E[Y]} = \frac{1700}{1705}$

(c) What is the (coefficient of) **correlation between X and Y**?

$$
\begin{aligned}
p_{X,Y} &= \frac{C_{X,Y}}{\sigma_X \sigma_Y} \\
&= \lambda^2(E[XY] - E[X]E[Y]) = 1
\end{aligned}
$$

Honestly, we do not know how to calculate $E[XY]$. This is because, in our point of view, $E[XY] = \int_0^\infty \int_{x+5}^{x+5} x \cdot y \cdot \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(y-5)}\, dydx = 0$, but in this case $p_{X,Y} \neq 1$. Since $Y = X + 5$, it is obvious that $p_{X,Y} = 1$.

Exercise 4

Assume a single-server queue with one service unit (one packet can be processed at each time). For this single-server queue, the interarrival times of packets are exponentially distributed (with rate λ) and the service time for serving one packet is exponentially distributed (with rate μ). This is also called M/M/1 queue. The mean number of customers in this system (including the service unit) is computed as:

$$N = \frac{U}{1 - U}, \text{ where } U = \frac{\lambda}{\mu}.$$

Besides, if we assume that the service time is generally distributed (arbitrary / any distribution), then the queue is called M/G/1 queues. We assume that the service times have the same mean as above ($\frac{1}{\mu}$). Then the mean waiting time (i.e., mean delay in the queue excluding the service unit) is given by:

$$W = \frac{U(c_s^2 + 1)}{2\mu(1 - U)}$$

where c_s is the coefficient of variation of the service time distribution.

(a) Under which condition (on the arrival rate λ) do the mean delay (D) in the system (including the service unit) and the mean number (N) of customers in the system have the same numeric values for an M/M/1 queue?

Let S - the service time, then $D = W + E[S] = \frac{U(c_s^2+1)}{2\mu(1-U)} + \frac{1}{\mu}$.
Since, $c_s = 1$: $D = \frac{U(1+1)}{2\mu(1-U)} + \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)} + \frac{1}{\mu}$.

$N = \frac{U}{1-U} = \frac{\lambda}{\mu-\lambda}$. Pooling everything together: $\frac{\lambda}{\mu(\mu-\lambda)} + \frac{1}{\mu} = \frac{\lambda}{\mu-\lambda} \implies \lambda = 1$.

(b) For identical mean service times, by which factor do the mean waiting times $W^{M/M/1}$ and $W^{M/D/1}$ differ for an M/M/1 and an M/D/1 queue? (D stands for deterministic / constant service time).

$W^{M/M/1} = \frac{U(1+1)}{2\mu(1-U)} = \frac{U}{\mu(1-U)}$.
Since $c_{const} = 0$, $W^{M/D/1} = \frac{U(0+1)}{2\mu(1-U)} = \frac{U}{2\mu(1-U)}$.
Hence, $W^{M/M/1} = 2W^{M/D/1}$.