

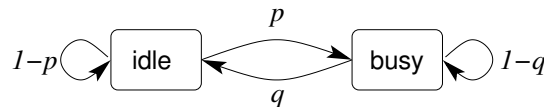
---

**DS n° 2 - Exam (3 hours, course/TD notes not allowed)**


---

**Exercise 1.***Short questions (6 pts)*

1. Consider two events  $A$  and  $B$  with positive measures, and suppose that you are given  $\mathbb{P}(A|B)$ ,  $\mathbb{P}(A|\bar{B})$ ,  $\mathbb{P}(A)$ . Does it set the value  $\mathbb{P}(B)$  in a unique way? If yes, explain how to compute it from  $\mathbb{P}(A|B)$ ,  $\mathbb{P}(A|\bar{B})$ ,  $\mathbb{P}(A)$ . If no, explain why.
2. State one of the asymptotic limit theorems about random variables exposed during the course/TD (specify the type of convergence and the relevant hypotheses).
3. Cite one of the paradoxes studied during the course/TD and give a short description (without proofs).
4. Suppose that you have access to some **Random** function which draws a random real number in  $[0,1]$  uniformly. Provide an algorithm which simulates the exponential law of parameter  $\lambda$ , i.e. whose output is a random variable following this law.
5. A web server receives client requests that follow a Poisson process of intensity  $\lambda$ . Suppose that exactly two requests arrived during the first hour of the day, what is the probability that
  - ▷ both requests arrived during the first 20 minutes?
  - ▷ at least one request arrived during the first 20 minutes?
6. Suppose that time is slotted, thus discrete, and that the use of a communication channel can be modelised by the Markov chain  $(X_n)_{n \in \mathbb{N}}$  below where  $0 < p, q < 1$ . You observe one trajectory  $X_0(\omega), X_1(\omega), X_2(\omega), \dots$  and count the proportion of time when the channel is busy, that is  $f_n(\omega) = \text{card}\{0 \leq k < n | X_k(\omega) = \text{busy}\} / n$ . What can you say about the asymptotic behaviour of  $f_n(\omega)$ ? Does it converge when  $n \rightarrow +\infty$ ? If so, does the limit depend on the trajectory?

**Exercise 2.***Gaussian errors (4 pts)*

You have done  $n$  independent experiments to measure an unknown parameter  $\mu \in \mathbb{R}$ . The measures may be subject to errors, each measure is then modeled by a random variable  $X_i = \mu + \varepsilon_i$ ,  $1 \leq i \leq n$ , where  $\varepsilon_i$  are i.i.d. random variables with a law of mean 0 and variance 1. Given a sample  $(x_1, \dots, x_n)$ ,  $x_i \in \mathbb{R}$ , you choose to use the empirical mean  $\bar{x}_n = (x_1 + \dots + x_n)/n$  to estimate  $\mu$ , and you denote the associated random variable  $\bar{X}_n = (X_1 + \dots + X_n)/n$ .

1. Give a quick justification to the fact that if you make a sufficient number of experiments, you can approach the exact value  $\mu$  as close as you want.

You wish to have the guarantee with high probability ( $\geq 99\%$ ) that the estimation error is small ( $\leq 0,1$ ), that is  $\mathbb{P}(|\bar{X}_n - \mu| \geq 0,1) \leq 0,01$ .

2. Use Chebychev's inequality in order to set a number of experiments providing this guarantee.
3. Same question with the additional assumption that the random variables  $\varepsilon_i$  are normal : can you plan fewer experiments? How many?

**Exercise 3.***A statistical hypothesis test (2 pts)*

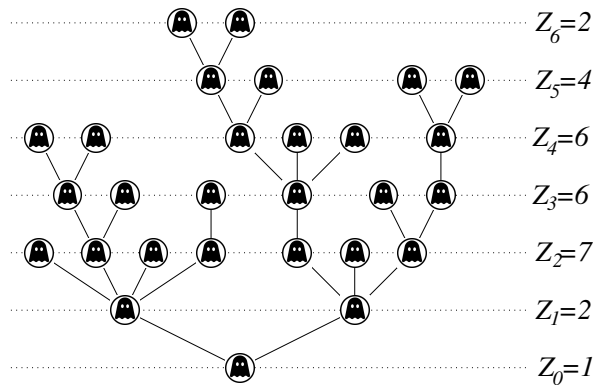
You have just received a sample of  $n$  real numbers  $x_1, \dots, x_n$  generated by  $n$  random variables  $X_1, \dots, X_n$  i.i.d. with uniform law over  $[a, a + 1]$  where  $0 \leq a \leq 1$  is an unknown parameter. You are looking for a test that would allow you to decide with good guarantees whether  $a = 0$  or not. More precisely, you have fixed a small real number  $\alpha$  and you wish that if  $a = 0$ , then your test will fail (that is, state that  $a \neq 0$ ) with probability  $\leq \alpha$ .

Design a test that meets this guarantee and tune its parameters to minimize the other type of error, that is the probability  $\beta$  to state that  $a = 0$  whereas in reality  $a \neq 0$ .

**Exercise 4.***Galton-Watson style virus (4 pts)*

We study the evolution of a computer virus which can replicate day after day (time is discrete here, and the unit is one day). On day 0,  $Z_0$  copies of the virus have infected the system ( $Z_0$  can be a fixed integer or a integer random variable). On day  $n$ , let  $Z_n$  be the number of copies of the virus. The next day, each copy erases itself but also gives birth independently to  $X$  new copies where  $X$  is a integer random variable whose law is known :  $\mathbb{P}(X = k) = p_k$  for  $k \geq 0$ . Its mean  $\mathbb{E}(X)$  is denoted  $\mu$ .

As illustrated below, one can represent this process by a rooted tree where each node is a copy which gives birth to sibling copies (none sometimes if  $p_0 > 0$ , and in the case it is possible that the process ends with no virus left).



1. Let  $(X_i)_{i \in \mathbb{N}^*}$  be a family of i.i.d. real random variables and  $Z$  an integer random variable independent of the  $X_i$ . Prove the weak Wald's lemma :  $\mathbb{E}(\sum_{i=1}^Z X_i) = \mathbb{E}(Z)\mathbb{E}(X_1)$ .

*Suggestion* : first condition by  $\{Z = z\}$  for  $z \in \mathbb{N}^*$  by writing  $\mathbb{E}(\sum_{i=1}^Z X_i) = \sum_z \mathbb{E}(\sum_{i=1}^Z X_i | Z = z) \mathbb{P}(Z = z)$ .

*Reminder* :  $\mathbb{E}(Y|Z = z) = \sum_y y \mathbb{P}(Y = y|Z = z)$ .

2. Suppose that  $Z_0 = 1$ , give a formula for  $\mathbb{E}(Z_n)$  as a function of  $\mu$  and  $n$ .

3. Show that if  $\mu < 1$ , then  $\mathbb{P}(Z_n = 0) \rightarrow 1$  when  $n \rightarrow +\infty$  (i.e. the virus tends to fully disappear).

4. Consider the probability generating function  $G(x) = \sum_{k=0}^{+\infty} p_k x^k$  associated with the number of replicates of a copy, as well as  $G_n(x) = \sum_{k=0}^{+\infty} \mathbb{P}(Z_n = k) x^k$  associated with the total number of replicates on day  $n$ . Show the relation  $G_n(x) = G_{n-1}(G(x))$  for all  $n \geq 1$ .

5. You wish to describe precisely what is happening in the special case of Bernoulli random variables :  $p_0 = p, p_1 = 1 - p$  for a fixed value  $p$ , and  $p_k = 0$  if  $k \geq 2$ . Let us denote  $T = \min\{n \geq 0 \mid Z_n = 0\}$  the extinction time. Give a closed formula for  $G_n(x)$ , for  $\mathbb{P}(Z_n = 0)$  and  $\mathbb{P}(T = n)$ .

**Exercise 5.***Server and mirror (4 pts)*

Consider two web servers : the main one and a mirror site. Requests arrive first at a router which distributes the load between both servers. The router does not know the current load of each server, it chooses to route each request either to the main server with probability  $p$  or to the mirror server with probability  $1 - p$ . We wish to optimize the choice of  $p$ .

We assume that arrivals follow a Poisson process of intensity  $\lambda$ . Service times in the servers are exponential with parameter  $\mu$  for the main server and  $\nu$  for the mirror server.

1. Show the arrivals in the main server and in the mirror server are both Poisson processes with respective parameters  $p\lambda$  and  $(1 - p)\lambda$ .
2. What are the criteria for system stability, that is the existence of an invariant distribution ?
3. What is the mean response time in the global system under the stationary regime ?
4. Find the optimal value  $p$  that minimizes this mean response time.

