# Stochastic Process (ENPC)
# Monday, 27th of January 2021 (2h30)

Vocabulary (english/*français*): urn=*urne*; distribution =*loi*; positive = *strictement positif*.

We shall assume that all the random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

**Exercise 1** (Characterisation of martingales)**.** Let $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ be a filtration. Let $M = (M_n, n \in \mathbb{N})$ be an $\mathbb{F}$-adapted integrable process such that for all bounded stopping time $\tau$, we have $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$. We shall prove that $M$ is then a martingale.

1. Let $m > n \in \mathbb{N}$ and $A \in \mathcal{F}_n$.
   (a) Check that $\mathbb{E}[M_m] = \mathbb{E}[M_n]$.
   (b) Prove that $\tau = n\mathbf{1}_A + m\mathbf{1}_{A^c}$ is a $\mathcal{F}$-stopping time.
   (c) Prove that $\mathbb{E}[M_m\mathbf{1}_A] = \mathbb{E}[M_n\mathbf{1}_A]$.
2. Deduce that $M$ is a martingale.

△

**Exercise 2** (Pólya's urn or the progress of an epidemic)**.** We consider an elementary model of global propagation of an epidemic from Pólya [1] (1930), where a new individual is uninfected or infected with probability depending on the proportion of already uninfected or infected people. More precisely, we consider an urn with initially $r \in \mathbb{N}^*$ red balls and $d \in \mathbb{N}^*$ deep blue balls. At each step, pick a ball at random, and put it back in the urn, together with an additional ball of the same color. At step $n \in \mathbb{N}$: there are exactly $r + d + n$ balls in the urn; we denote by $S_n$ the number of red balls in the urn; and we set $X_{n+1} = 1$ if the ball taken at next step is red and $X_{n+1} = 0$ otherwise. Notice that $S_n = r + \sum_{k=1}^n X_k$ for all $n \in \mathbb{N}^*$ and $S_0 = r$. We denote by $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ the natural filtration of the process $S = (S_n, n \in \mathbb{N})$.

1. (Properties of the process $S$.)
   (a) Prove that $\mathbb{P}(X_{n+1} = 1 | \mathcal{F}_n) = S_n/(r + d + n)$.
   (b) Is $S$ an homogeneous Markov chain?
2. (Martingales.) We define the process of proportion of red balls $M = (M_n, n \in \mathbb{N})$ by:

$$M_n = \frac{S_n}{r + d + n}.$$

   (a) Prove that $M$ is a martingale.
   (b) Prove that the sequence $M$ converges (in what sense?) to a limit, say $M_\infty$, and that $\mathbb{E}[M_\infty] = r/(r + d)$.

   For $k \in \mathbb{N}^*$, we define the processes $M^{(k)} = (M_n^{(k)}, n \in \mathbb{N})$ by:

$$M_n^{(k)} = \prod_{\ell=0}^{k-1} \frac{S_n + \ell}{r + d + n + \ell}.$$

   In particular, we have $M = M^{(1)}$.

[1] G. Pólya. Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré*, 1(2):117-161, 1930.

(c) Prove that $M^{(k)}$ is a martingale for $k \in \mathbb{N}^*$.

(d) Prove that $\lim_{n\to\infty} M_n^{(k)} = M_\infty^k$ a.s. and $\mathbb{E}[M_\infty^k] = \prod_{\ell=0}^{k-1}(r+\ell)/(r+d+\ell)$ for $k \in \mathbb{N}^*$.

3. (Law of $M_\infty$.) Let $Y_{(a,b)}$ be a random variable with $\beta(a,b)$ distribution, where $a > 0$ and $b > 0$; its density (with respect to the Lebesgue measure) is given by:

$$f_{(a,b)}(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, y^{a-1}(1-y)^{b-1}\,\mathbf{1}_{(0,1)}(y) \quad \text{with} \quad \Gamma(r) = \int_0^{+\infty} x^{r-1}\,\mathrm{e}^{-x}\,dx.$$

(a) Prove that for $k \in \mathbb{N}$:

$$\mathbb{E}\left[Y_{(a,b)}^k\right] = \frac{\Gamma(a+b)}{\Gamma(a+b+k)}\,\frac{\Gamma(a+k)}{\Gamma(a)}.$$

(b) Using that $\Gamma(x+1) = x\Gamma(x)$, deduce that $\mathbb{E}[M_\infty^k] = \mathbb{E}[Y_{(r,d)}^k]$ for all $k \in \mathbb{N}^*$.

(c) Using that $\left|\mathrm{e}^z - \sum_{k=0}^{n-1}\frac{z^k}{k!}\right| \leq \frac{|z|^n}{n!}\,\mathrm{e}^{|z|}$ for $z \in \mathbb{C}$, prove that if $Y, Z$ are $[0,1]$-valued random variables such that $\mathbb{E}[Y^k] = \mathbb{E}[Z^k]$ for all $k \in \mathbb{N}^*$, then $Y$ and $Z$ have the same distribution.

(d) Deduce that $M_\infty$ has the $\beta(r,d)$ distribution.

This result could be easily generalized: instead of adding one ball of the same color, we add $k_0 \in \mathbb{N}^*$ balls of the same color, then the distribution of the limiting proportion of red balls has the $\beta(r/k_0, d/k_0)$ distribution. In the Friedman's urn model [2] [3] (1949), one adds $k_0$ balls of the same color and $\ell_0 \in \mathbb{N}^*$ balls of the other color. On can then prove that the proportion of red balls has a very different behavior as it converges a.s. to $1/2$: the limit is no more random and does not depend on the positive parameters $r, d, k_0, \ell_0$ of the model! △

**Exercise 3** (Markov property for Gaussian processes)**.** We say a random process $W = (W_t, t \in \mathbb{R}_+)$ has the Markov property if for all $t \in (0,+\infty)$, conditionally on $W_t$ the processes $(W_u, u \in [0,t])$ and $(W_v, v \in [t,+\infty))$ are independent. We shall describe the centered Gaussian processes which enjoy the Markov property.

1. Let $(Y, Z, G)$ be a $\mathbb{R}^3$-valued centered Gaussian vector such that $\mathrm{Var}(G) > 0$.

   (a) Determine $\alpha, \beta \in \mathbb{R}$ such that $G_1 = Y - \alpha G$ is independent of $G$ and $G_2 = Z - \beta G$ is independent of $G$.

   (b) Compute the covariance matrix of random vector $(G_1, G_2)$.

   (c) Give a necessary and sufficient condition for $Y$ and $Z$ to be independent conditionally on $G$.

2. Let $X = (X_t, t \in \mathbb{R}_+)$ be a centered Gaussian process with covariance kernel $K = (K(s,t) = \mathrm{Cov}(X_s, X_t);\ s,t \in \mathbb{R}_+)$. We assume that $K(t,t) > 0$ for all $t > 0$.

   (a) Prove that $X$ has the Markov property if and only if:

   $$K(u,v) = \frac{K(u,t)K(t,v)}{K(t,t)} \quad \text{for all } 0 \leq u < t < v.$$

   (b) Deduce that a standard Brownian motion has the Markov property.

△

[2] D. Freedman. Bernard Friedman's urn. *Ann. Math. Statist.*, 36:956-970, 1965.

[3] R. Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1-79, 2007.

## Correction

*Exercise 1* 1. (a) Clear.

(b) Let $k \in \mathbb{N}$. We have $\{\tau = k\} = \emptyset \in \mathcal{F}_k$ if $k \notin \{n, m\}$, $\{\tau = n\} = A \in \mathcal{F}_n$, and $\{\tau = m\} = A^c \in \mathcal{F}_n \subset \mathcal{F}_m$. This implies that $\tau$ is a stopping time with respect to the filtration $\mathbb{F}$.

(c) Since $\tau$ is a bounded stopping time, we get $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$ by hypothesis on $M$. This gives:

$$\mathbb{E}[M_0] = \mathbb{E}[M_n \mathbf{1}_A + M_m \mathbf{1}_{A^c}] = \mathbb{E}[M_n \mathbf{1}_A] + \mathbb{E}[M_m] - \mathbb{E}[M_m \mathbf{1}_A].$$

Then, use that $\mathbb{E}[M_m] = \mathbb{E}[M_0]$ to conclude.

2. Since $\mathbb{E}[M_m \mathbf{1}_A] = \mathbb{E}[M_n \mathbf{1}_A]$ for all $A \in \mathcal{F}_n$, we deduce that $\mathbb{E}[M_m | \mathcal{F}_n] = M_n$. Since this holds for $m = n+1$ and all $n \in \mathbb{N}$, we get that $M$ is a martingale.

*Exercise 2* 1. (Properties of the process $S$.)

(a) By construction, we have $\mathbb{P}(X_{n+1} = 1 | \mathcal{F}_n) = S_n/(r+d+n)$.

(b) We deduce that $\mathbb{P}(S_{n+1} = S_n + 1 | \mathcal{F}_n) = S_n/(r+d+n)$ and $\mathbb{P}(S_{n+1} = S_n | \mathcal{F}_n) = 1 - S_n/(r+d+n)$. This gives that $\mathbb{P}(S_{n+1} = \bullet | \mathcal{F}_n) = \mathbb{P}(S_{n+1} = \bullet | S_n)$. This implies that $S$ is an in-homogeneous Markov chain on $\mathbb{N}^*$ with transition matrices given by $P_{n+1}(S_n, z) = \mathbb{P}(S_{n+1} = z | \mathcal{F}_n)$, that is for $s, z \in \mathbb{N}^*$ and $n \in \mathbb{N}$:

$$P_{n+1}(s,z) = \begin{cases} \min\Big(1, s/(r+d+n)\Big) & \text{if } z = s+1, \\ \max\Big(0, (r+d+n-s)/(r+d+n)\Big) & \text{if } z = s, \\ 0 & \text{otherwise.} \end{cases}$$

But $S$ is not an homogeneous Markov chain.

2. (Martingales.)

(a) The process $M$ is $\mathbb{F}$-adapted as $\mathbb{F}$ is the natural filtration of $S$. We have that $M_n \in [0,1]$ for $n \in \mathbb{N}$, and thus the process $M$ is integrable. For $n \in \mathbb{N}$, we have:

$$\mathbb{E}\left[S_{n+1} | \mathcal{F}_n\right] = (S_n + 1)\mathbb{P}(S_{n+1} = S_n + 1 | \mathcal{F}_n) + S_n \mathbb{P}(S_{n+1} = S_n | \mathcal{F}_n) = S_n + \frac{S_n}{r+d+n}\cdot$$

This gives that $\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n$ for $n \in \mathbb{N}$. Thus the process $M$ is a martingale.

(b) The martingale $M$ is non-negative and thus it converges a.s. to a limit, say $M_\infty$. Since $M_n \in [0,1]$ for all $n \in \mathbb{N}$, we deduce by dominated convergence that $M$ converges also in $L^1$ and thus $\lim_{n\to\infty} \mathbb{E}[M_n] = \mathbb{E}[M_\infty]$. We deduce that:

$$\mathbb{E}[M_\infty] = \mathbb{E}[M_0] = \frac{r}{r+d}\cdot$$

(c) The process $M^{(k)}$ is $\mathbb{F}$-adapted as $\mathbb{F}$ is the natural filtration of $S$. We have that

$M_n^{(k)} \in [0,1]$ for $n \in \mathbb{N}$, and thus the process $M^{(k)}$ is integrable. For $n \in \mathbb{N}$, we have:

$$\begin{aligned}
\mathbb{E}\left[\prod_{\ell=0}^{k-1}(S_{n+1}+\ell)\middle|\, \mathcal{F}_n\right] & \\
&= \prod_{\ell=0}^{k-1}(S_n+\ell+1)\mathbb{P}(S_{n+1}=S_n+1|\, \mathcal{F}_n) + \prod_{\ell=0}^{k-1}(S_n+\ell)\mathbb{P}(S_{n+1}=S_n|\, \mathcal{F}_n) \\
&= \frac{S_n+k}{r+d+n}\prod_{\ell=0}^{k-1}(S_n+\ell) + \frac{r+d+n-S_n}{r+d+n}\prod_{\ell=0}^{k-1}(S_n+\ell) \\
&= \frac{r+d+n+k}{r+d+n}\prod_{\ell=0}^{k-1}(S_n+\ell).
\end{aligned}$$

This gives that $\mathbb{E}[M_{n+1}^{(k)}|\, \mathcal{F}_n] = M_n^{(k)}$ for $n \in \mathbb{N}$. Thus the process $M^{(k)}$ is a martingale.

(d) Since

$$M_n^{(k)} = \prod_{\ell=0}^{k-1}\left(M_n + \frac{\ell}{r+d+n}\right)\frac{r+d+n}{r+d+n+\ell},$$

We deduce that $\lim_{n\to\infty} M_n^{(k)} = M_\infty^k$. Since $M_n^{(k)}$ is bounded by 1 for all $n \in \mathbb{N}$ and $k \geq 2$, we get by dominated convergence that $\lim_{n\to\infty} \mathbb{E}[M_n^{(k)}] = \mathbb{E}[M_\infty^k]$ and thus:

$$\mathbb{E}\left[M_\infty^k\right] = \mathbb{E}\left[M_0^{(k)}\right] = \prod_{\ell=0}^{k-1}\frac{r+\ell}{r+d+\ell}.$$

3. (Law of $M_\infty$.)

(a) Since $f_{(a+k,b)}$ is a probability density, we have for $k \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{E}\left[Y_{(a,b)}^k\right] = \int y^k\, f_{(a,b)}(y)\, dy &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\frac{\Gamma(a+k)\Gamma(b)}{\Gamma(a+b+k)}\int f_{(a+k,b)}(y)\, dy \\
&= \frac{\Gamma(a+b)}{\Gamma(a+b+k)}\,\frac{\Gamma(a+k)}{\Gamma(a)}.
\end{aligned}$$

(b) Since $\Gamma(r+d+k+1) = \Gamma(r+d)\prod_{\ell=0}^k(r+d+\ell)$ and $\Gamma(r+k+1) = \Gamma(r)\prod_{\ell=0}^k(r+\ell)$ for $k \in \mathbb{N}$, we deduce from Questions 2.(b) and 2.(d) that for $k \in \mathbb{N}$:

$$\mathbb{E}\left[Y_{(r,d)}^{k+1}\right] = \prod_{\ell=0}^{k}\frac{r+\ell}{r+d+\ell} = \mathbb{E}\left[M_\infty^{k+1}\right].$$

(c) Let $X$ be a random variable taking values in $[0,1]$ and $\psi_X$ be its characteristic function. Let $u \in \mathbb{R}$. We have:

$$\left|\psi_X(u) - \sum_{k=0}^{n-1}\frac{u^k\mathbb{E}[X^k]}{k!}\right| \leq \mathbb{E}\left[\left|\mathrm{e}^{iuX} - \sum_{k=0}^{n-1}\frac{(uX)^k}{k!}\right|\right] \leq \frac{|u|^n\mathbb{E}\left[X^n\right]}{n!} \leq \frac{|u|^n}{n!}.$$

We deduce that if $X$ and $Y$ are random variables taking values in $[0,1]$ such that $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ for all $k \in \mathbb{N}^*$, then $\psi_X = \psi_Y$ and thus $X$ and $Y$ have the same law.

(d) Using the previous questions, we get that $M_\infty$ has distribution $\beta(r, d)$.

*Exercise 3* 1. (a) Since $(Y, G)$ is a Gaussian vector, we deduce that $(Y - \alpha G, G)$ is also a Gaussian vector. Since $\mathrm{Cov}(Y - \alpha G, G) = \mathrm{Cov}(Y, G) - \alpha \mathrm{Var}(G, G)$, we get that $Y - \alpha G$ and $G$ are independent if and only if $\alpha = \mathrm{Cov}(Y, G)/\mathrm{Var}(G)$. We prove similarly that $Z - \beta G$ and $G$ are independent if and only if $\beta = \mathrm{Cov}(Z, G)/\mathrm{Var}(G)$.

(b) We have $\mathrm{Var}(G_1) = \mathrm{Var}(Y - \alpha G) = \left(\mathrm{Var}(Y)\mathrm{Var}(G) - \mathrm{Cov}(Y, G)^2\right)/\mathrm{Var}(G)$. Similarly, we have $\mathrm{Var}(G_2) = \left(\mathrm{Var}(Z)\mathrm{Var}(G) - \mathrm{Cov}(Z, G)^2\right)/\mathrm{Var}(G)$. We also have:

$$\mathrm{Cov}(G_1, G_2) = \mathrm{Cov}(Y - \alpha G, Z - \beta G) = \frac{\mathrm{Cov}(Y, Z)\mathrm{Var}(G) - \mathrm{Cov}(Y, G)\mathrm{Cov}(Z, G)}{\mathrm{Var}(G)}\cdot$$

(c) Since $(Y, Z)$ is conditionally on $G$ distributed as $(G_1 + \alpha G, G_2 + \beta G)$, we deduce that $Y$ and $Z$ are independent conditionally on $G$ if and only if $G_1$ and $G_2$ are independent. Since $(G_1, G_2)$ is a Gaussian vector, we get that $G_1$ and $G_2$ are independent if and only if $\mathrm{Cov}(G_1, G_2) = 0$ that is, according to the previous question:

$$\mathrm{Cov}(Y, Z)\mathrm{Var}(G) = \mathrm{Cov}(Y, G)\mathrm{Cov}(Z, G).$$

2. (a) Since the distribution of a process is characterised by the distribution of its finite marginals, we get that the process $X$ has the Markov property if and only if for all $m, n \in \mathbb{N}^*$, $t \in (0, +\infty)$, $u_1, \ldots, u_m \in [0, t)$ and $v_1, \ldots, v_n \in (t, +\infty)$, we have that $(X_{u_1}, \ldots, X_{u_m})$ and $(X_{v_1}, \ldots, X_{v_n})$ are independent conditionally on $X_t$. Because $(X_{u_1}, \ldots, X_{u_m}, X_t, X_{v_1}, \ldots, X_{v_n})$ is a Gaussian vector, this is equivalent to $X_u$ and $X_v$ being independent conditionally on $X_t$ for any choice of $u \in [0, t)$ and $v \in (t, +\infty)$. According to Question 1(c), this is equivalent to:

$$K(u, v)K(t, t) = K(u, t)K(t, v) \quad \text{for all } 0 \leq u < t < v. \tag{1}$$

(b) The covariance kernel $K = (K(s, t);\, s, t \in \mathbb{R}_+)$ of a standard Brownian motion is given by $K(s, t) = s \wedge t$. Notice that for this covariance kernel (1) holds trivially. Hence a standard Brownian motion has the Markov property.