STK 2130 våren 2010
Oblig

Dette er oppgavesettet for det obligatoriske innleveringsprosjektet for kurset STK 2130. Det legges ut **fredag 5. mars**, og leveringsfrist er **fredag 19. mars** s.å., til instituttkontoret ved Matematisk institutt, senest kl. 13:58. Du kan levere tekstbehandlet eller håndskrevet besvarelse, på bokmål, nynorsk, riksmål eller engelsk. Oppgavesettet består av tre oppgaver over tilsammen fem sider. Vennligst anvend «Forside til bruk ved innlevering av obligatoriske oppgaver», som du finner på kursets nettside.

Exercise 1

A SIMPLE ILLNESS MODEL is of the following type: An individual is most of the time in 'normal state', called state 0; but once in a while the individual experiences 'beginning illness', called state 1; with a consequent 'ill, but getting better again' period, called state 2; after which the person again is back in normal state 0. It is assumed that the individuals's process $X_0, X_1, X_2, \ldots$ over time points $0, 1, 2, \ldots$ follows a time-homogeneous Markov chain model with probability transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-p_0 & p_0 & 0 \\ 0 & 1-p_1 & p_1 \\ p_2 & 0 & 1-p_2 \end{pmatrix},$$

for certain parameter values p_0, p_1, p_2. (People who are largely in good shape hence have a p_0 rather smaller than p_1 and p_2.) To make this description concrete we take the framework to mean one where time is given and counted in days.

In this exercise you are allowed to use, if you need it, that if M has a geometric distribution, with point probabilities $(1-p)^{m-1}p$ for $m = 1, 2, 3, \ldots$, then its expected value and variance are respectively $1/p$ and $(1-p)/p^2$.

(a) If a man with normalcy parameter $p_0 = 0.03$ is in fact not ill 1st of April, what is the probability that he keeps himself out of illness during all of April, May and June?

(b) Assume $X_0 = 0$, and let T_0 be the time at which the individual for the first time experiences 'beginning illness' (i.e. the first time where $X_n = 1$). Find the distribution of T_0 and give its expected value.

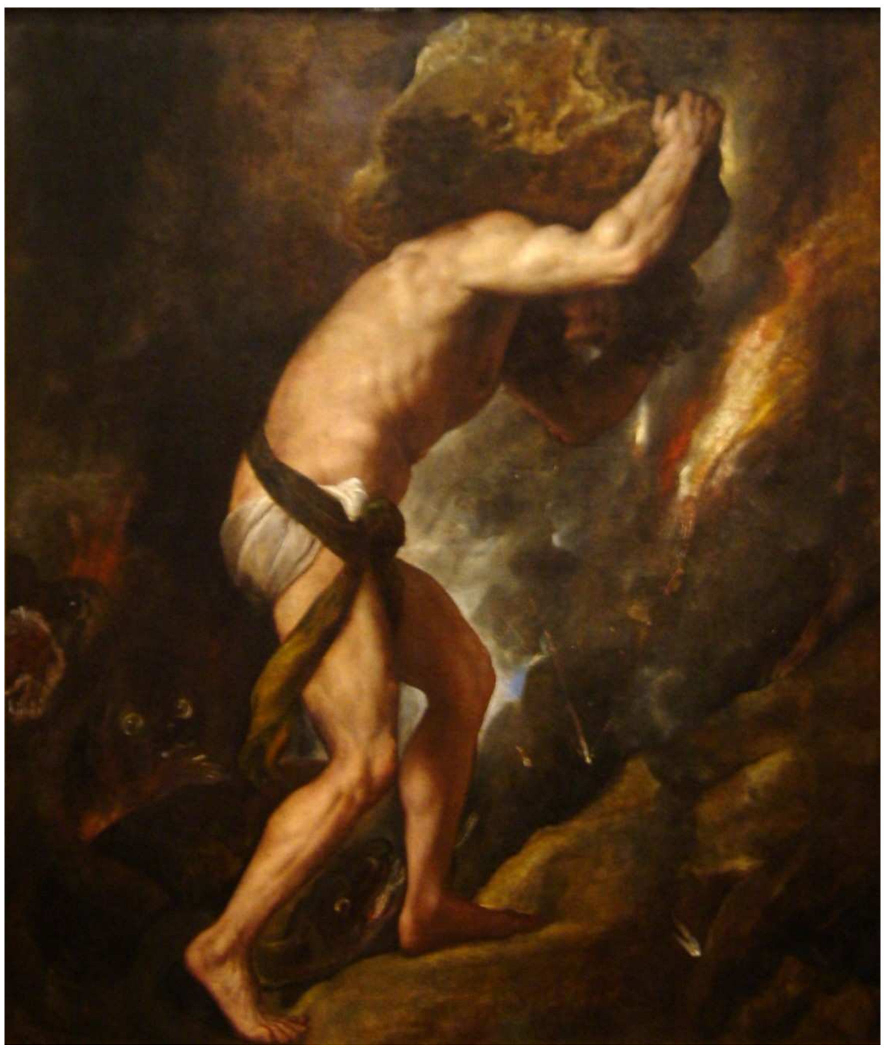
(c) Let T be the time it takes to go through a full cycle, from start in normal position at time zero until the person has gone through the illness's states 1 and 2 and then is back to normal again. Find the expected value and variance of T.

(d) Find the equilibrium (or stationary) distribution π_0, π_1, π_2 for the chain (expressed via p_0, p_1, p_2).

(e) Let A_n be the number of times our man has been ill (i.e. 'beginning illness' or 'ill, but getting better again') in the course of the first n days (where it is assumed that he was in normal state when the process started). Find the limit of $\mathrm{E}(A_n/n)$ as n increases. Explore also what happens to $\mathrm{Var}\,(A_n/n)$, and comment on what you find.

Exercise 2

SISYPHUS [på norsk: Sisyfos] CLIMBS A LADDER according to the following setup. The ladder has m steps, which we initially take to be equal to 4, for the sake of concrete illustration. If he at a certain time point is positioned at height i, then his efforts during the next hour takes him to height $i+1$, with probability p, or he falls down to the floor (which we label position 0), with probability $q = 1 - p$. This description is also valid when he is at the bottom position (i.e. $i = 0$). When he has finally reached the top of the ladder ($i = m$), he will during the coming hour fall down to the bottom again. Sisyphus's ladder climbing process goes on and on, in this truly Sisyphean manner.



(a) Let X_n be his position on the ladder at time point n. Explain why the assumptions above lead to this being a Markov chain on $\{0, 1, \ldots, m\}$ with transition probability matrix equal to

$$P = \begin{vmatrix} q & p & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ q & 0 & 0 & p & 0 \\ q & 0 & 0 & 0 & p \\ 1 & 0 & 0 & 0 & 0 \end{vmatrix}$$

(and as mentioned above we take $m = 4$, so far).

(b) Let $\pi_0, \pi_1, \ldots, \pi_m$ be the equilibrium distribution for Sisyphus's position on his ladder. Show that $\pi_i = p\pi_{i-1}$ for $i = 1, \ldots, m$ and use this to show that the equilibrium distribution is

$$\pi_i = \frac{1-p}{1-p^5} p^i \quad \text{for } i = 0, 1, \ldots, 4.$$

If $p = 0.15$, what proportion of his time is Sisyphus spending at the bottom, and what proportion at the very top?

(c) For $i = 0, 1, \ldots, 4$, let

$$u_i = \mathrm{E}(T \mid X_0 = i),$$

the expected time he needs to spend before he reaches the top position (i.e. $T = \min\{n \geq 0 \colon X_n = 4\}$). Argue that

$$u_i = 1 + qu_0 + pu_{i+1} \quad \text{for } i = 0, 1, 2, 3,$$

and that $u_4 = 0$.

(d) Then find the solutions to these equations, and give the numerical values of u_0, u_1, u_2, u_3, u_4 for the case of $p = 0.15$.

(e) Albert Camus (Nobel Prize winner 1957) appears to claim in his *Le Mythe de Sisyphe* that Sysiphus's quandary is the deepest and most crucially important of all human philosophical themes. So let us at least generalise Sysiphus's problem from the above setting, where $m = 4$, to the case of a general m. Find the equilibrium distribution and the expected time to reach the top (before he and his heavy rock inevitably tumble down again), given that he starts at the bottom at time zero.

Exercise 3

WILL I HAVE ANY GREAT-GREAT-GRANDCHILDREN? The theory of branching processes has been used to estimate the probability that a given surname will survive that mighty sculptor, time. We now envisage a society where women take their husbands' surname, so that only sons contribute to the survival of the surname. Data from an investigation carried out by A.J. Lotka ['The extinction of families', *Journal of Washington's Academy of Sciences* **21**, 1931, pages 377–380 and 453–459] showed that the the probability distribution

$P(Y = j) = p_j$ for the number of sons from a randomly sampled man (within a certain relevant socio-cultural stratum) gave a satisfactory fit to the distribution with generating function

$$G(s) = \mathrm{E}\, s^Y = \sum_{j=0}^{\infty} P(Y = j) s^j = p_0 + p_1 s + p_2 s^2 + \cdots = \frac{0.48 - 0.04\, s}{1 - 0.56\, s}.$$

Let X_n be the number of boys in generation n stemming from a given start individual (in particular, $X_0 = 1$).

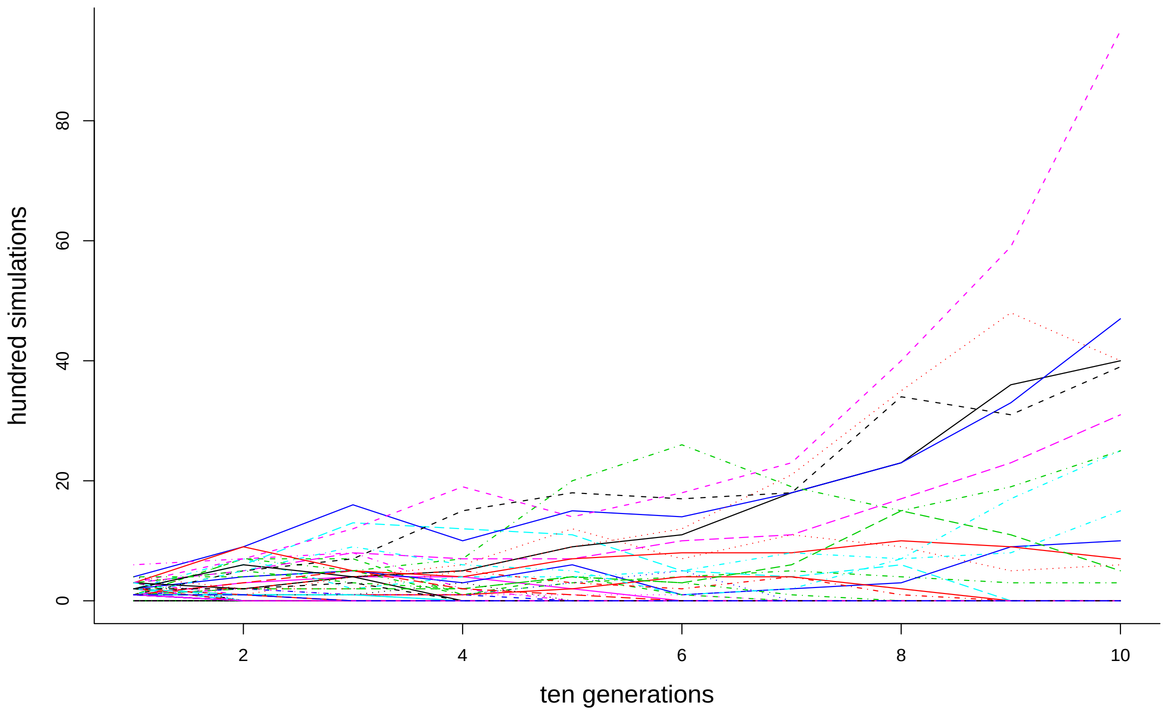


Figure 1: The number of Ffrench boys in generations $1, 2, \ldots, 10$, *in 100 independent simulations, each starting with a single Ffrench boy.*

(a) Find $p_0, p_1, p_2, \ldots$.

(b) Show that the mean of Y is 1.1818 (the exact value is $\frac{13}{11}$). Compute also the variance of Y. [Hint: For generating functions, $G'(1) = \mathrm{E}\, Y$ and $G''(1) = \mathrm{E}\, Y(Y-1)$. The variance becomes 2.7934 (exact value is $\frac{26}{11}\frac{13}{11}$).]

(c) What are the conditions required for $\{X_0, X_1, X_2, \ldots\}$ to form a time-homogeneous Markovian branching process? To what extent do you find these assumptions reasonable?

For the rest of the present exercise assume that the chain really satisfies the required conditions from (c).

(d) Find the generating function for X_2, i.e. $G_2(s) = \mathrm{E}\, s^{X_2}$. Find the probability that a man will not have any grandsons.

(e) Find the expected value and variance of X_n.

(f) Find the probability π_0 that a branching process with $X_0 = 1$ start individual will die out.

(g) Assume that there in the year 1931 were ten young and unmarried men of the Ffrench family. What is then the probability that the Ffrench name will die out?

Nils Lid Hjort