(ii) See p. 55 in Norris

→ Application to statistics:

Objective: Estimation of P of a Markov chain $(X_n)_{n \geq 0}$ based on observations $x_0, x_1, \dots, x_M \in J$

→ tool: log-likelihood estimator of P:

For this purpose define the log-likehood function $\ell(P)$:

$$\ell(P) \stackrel{\text{def.}}{=} \log\left(P(X_0 = x_0, X_1 = x_1, \dots, X_M = x_M)\right) \qquad (3.8.2)$$

$$\stackrel{\text{Th. 3.1.3}}{=} \log\left(\lambda_{x_0} P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{M-1} x_M}\right)$$

The trans. prob. $\hat{P}_{ij} = P_{ij}, i,j \in J$ for which $\ell(P)$ attains its maximum is called maximum likelihood estimator (MLE) of the trans. matrix P

→ interpret: MLE maximizes the prob. for the occurence of the observations $x_0, \dots, x_M$

Rewrite $\ell(P)$ as (if $\lambda_{x_0} = 1$)

$$\ell(P) = \sum_{i,j \in J} N_{ij} \log(P_{ij}) = \sum_{i \in J} \sum_{j \in J} N_{ij} \log(P_{ij}) \qquad (3.8.3)$$

$N_{ij} \stackrel{\text{def.}}{=}$ number of transitions from i to j

Define

$$N_i = \sum_{j \in J} N_{ij} \qquad (3.8.4)$$

→ total number of transitions from state i

Let us maximize $\ell(P)$ by maximizing for each i

$$\sum_{j \in J} N_{ij} \log(P_{ij})$$

in $P_{ij}, j \in J$ under the constraint

$$\sum_{j \in J} P_{ij} = 1$$

Lagrange multiplier (3.8.4) $\Rightarrow$

$$N_{ik} \cdot \frac{1}{P_{ik}} - \lambda = 0 \quad \text{for all } k$$

$$\Rightarrow \lambda = N_i$$

$$\Rightarrow \hat{P}_{ik} \stackrel{\text{def.}}{=} \frac{N_{ik}}{N_i} \quad \text{MLE} \qquad (3.8.5)$$

Claim: $\hat{P}_{ij}$ consistent estimator, i.e. $\hat{P}_{ij} \longrightarrow P_{ij}$ as $M \longrightarrow \infty$ (M ← number of observations)

1. case: $T_i^{(r)} < \infty$ with prob. 1 for all r ($T_i^{(r)}$ ← r-th passage time to i)

(e.g. if $(X_n)_{n \geq 0}$ recurrent)

$$\Rightarrow N_{ij} = \sum_{r=0}^{N_i} Y_r \text{ , where}$$

$$Y_r \overset{\text{def}}{=} 1_{\{X_{T_i^{(r)}+1} = j\}}$$

$\longrightarrow$ Strong Markov prop. applied to $T = T_i^{(r)}$, $r \geq 1$ shows that

Th. 3.8.1 (SLLN) $\Rightarrow$ Y_r, $r \geq 1$ i.i.d with $E[Y_r] = p_{ij}$

$$\frac{1}{n} \sum_{r=1}^{n} Y_r \xrightarrow[n\to\infty]{} p_{ij}$$

$n = N_i \Rightarrow N_i \xrightarrow[M\to\infty]{} \infty$

$$\hat{p}_{ij} = \frac{1}{N_i} \sum_{r=0}^{N_i} Y_r \xrightarrow[M\to\infty]{} p_{ij}$$

2. case: $T_i^{(r)} = \infty$ for some r. See Norris

Ex. 3.8.6 (fire insurance)

Consider the following (rounded) Danish fire insurance data collected at Copenhagen Reinsurance between 06/01/1990 and 06/15/1990 (in millions of Danish Krone)

Date n	X_n Loss in DKM
06/03/1990	2
06/04/1990	2
06/06/1990	4
06/07/1990	2
06/11/1990	2
06/14/1990	1
06/15/1990	17 (large number)

$\longrightarrow$ X_n fire loss claim at time n

$\longrightarrow$ $S = \{0,1,2,3,\dots,L\}$, $M+1 = 15$ observations

$n = 0 \triangleq 06/01/1990$, $n = 1 \triangleq 06/02/1990$, ...

$\Rightarrow$ $x_0 = 0$, $x_1 = 0$, $x_2 = 2$, $x_3 = 2$, ..., $x_{14} = 17$

$\longrightarrow$ e.g. $N_{20} = 3$ and $N_2 = 4$

(3.8.5) $\Rightarrow$

$$\hat{p}_{20} = \frac{N_{20}}{N_2} = \frac{3}{4} = 75\%$$

or e.g. $N_{02} = 2$ and $N_0 = 8$

(3.8.5) $\Rightarrow$

$$\hat{p}_{02} = \frac{N_{02}}{N_0} = \frac{2}{8} = \frac{1}{4} = 25\%$$

as a rough estimator

4. Applications

4.1 Branching processes

→ introduced by Galton and Watson (1870s) to study the phenomenon of disapperance of family names in populations

Assumption: Family names are passed from fathers to sons

In generation $n=0$ there is 1 male, i.e. $X_0 = 1$
This male dies and is replaced in generation $n=1$ by N_1^1 sons

→ $X_1 := N_1^1$ number of family name bearers in the 1st generation

These sons also die and leave in gen. $n=2$

$N_1^2, \dots, N_{X_1}^2$ sons

→ $X_2 = N_1^2 + \dots + N_{X_1}^2$ number of family name bearers in gen. $n=2$

⋮

More generally, we get

$$X_n = N_1^n + \dots + N_{X_{n-1}}^n \qquad (4.1.1)$$

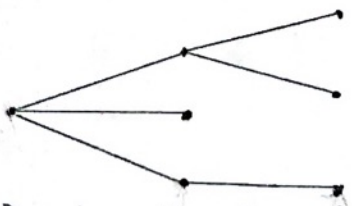
numb. of fam. name bearers in generation n

$N_K^n \triangleq$ number of male descendants of the K-th son of the $(n-1)$-th gen.

Assumption: N_K^n, $K \geq 1$, $n \geq 1$ are i.i.d random variables with common distr.

$$N \overset{\text{def}}{=} N_1^1$$

→ $X_n, n \geq 0$ Markov chain with $\mathbb{N} = \{0,1,2,\dots\}$



$X_0 = 1$, $X_1 = 3$, $X_2 = 3$

→ $X_n, n \geq 0$ is a special case of a branching process

If $X_m = 0 \overset{\text{def. of } X_{m+1}}{\Longrightarrow} X_{m+1} = 0 \Longrightarrow X_{m+2} = 0 \dots$

$\Longrightarrow \{0\}$ closed (i.e. a "trap")

(73) In order to avoid trivial cases suppose that

$$P(N=1) < 1 \qquad (4.1.2)$$

Note that

$$P(X_n=0 \mid X_{n-1}=i) = P(N_1^n+\dots+N_i^n=0 \mid X_{n-1}=i)$$

$$= P(N_1^n=0,\dots,N_i^n=0 \mid X_{n-1}=i) = P(N_1^n=0,\dots,N_i^n=0)$$

(indep.)

$$= P(N_1^n=0)\cdot\ldots\cdot P(N_i^n=0) \overset{\text{indep}}{=} P(N=0)\cdot\ldots\cdot P(N=0) = (P(N=0))^i \quad (+)$$

1. case: $P(N=0) > 0$

Assume that $P(T_i < \infty \mid X_0=i) = 1$ for some $i \geq 1$ (T_i: return time to i)

$$\Rightarrow \quad 0 < (P(N=0))^i \overset{(+)}{=} P(X_1=0 \mid X_0=i) = P(T_i<\infty, X_1=0 \mid X_0=i)$$

$= 0 \Rightarrow$ contradiction

$\Rightarrow P_i(T_i<\infty) < 1$ for all $i \geq 1$ $\overset{\text{Th. 3.4.6}}{\Rightarrow}$ all $i \geq 1$ transient

2. case: $P(N=0) = 0$

$$\Rightarrow \quad P(X_n=j \mid X_{n-1}=i) = 0 \quad \text{if } j < i \qquad (++)$$

i.e. $i \not\to j$ if $j < i$

On the other hand it follows from $(+)$, $(++)$ that

$$P(X_1 \geq i \mid X_0=i) = 1 - (P(N=0))^i = 1 \qquad (+++)$$

Note that $P(N \geq 2) > 0$ because of (4.1.2)

$\overset{(+++)}{\Rightarrow}$ there ex. $m > i$ s.t.

$$P(X_1=m \mid X_0=i) > 0$$

i.e. $i \to m$

Suppose that

$$P(T_i<\infty \mid X_0=i) = 1 \text{ for some } i \geq 1$$

$$\Rightarrow \quad 0 < P(X_1=m \mid X_0=i) = P(T_i<\infty, X_1=m \mid X_0=i)$$

$\overset{(++),\, m>i}{=} 0 \Rightarrow$ contradiction

$\Rightarrow P(T_i<\infty \mid X_0=i) < 1$ for all $i \geq 1$

$\Rightarrow$ all $i \geq 1$ transient

Suppose ("smallest" limes, "biggest" limes)

$$\underline{\lim_{n\to\infty}} X_n \text{ or } \overline{\lim_{n\to\infty}} X_n \in (0,\infty)$$

for a positive probability.

$\Rightarrow$ some states in $\{1,2,3,\dots\}$ are infinitely often revisited for a positive probability.

(74) $\overset{1}{\Rightarrow}$ contradiction (since states $i=1,2,3,\dots$ transient)

countable $\lim_{n\to\infty} X_n$, $\overline{\lim_{n\to\infty}} X_n = 0$ or ∞

$\overset{3.2}{\Rightarrow}$ "trap" with prob. 1

either $\lim_{n\to\infty} X_n = 0$ or $\lim_{n\to\infty} X_n = \infty$

$\longrightarrow$ either the number of family names disappears or explodes

Problem: $P(X_n = 0 \text{ for some } n)$?

Consider the **probability generating function**:

$$\mathcal{P}_Y(t) \overset{\text{def}}{=} E[t^Y] = \sum_{k\geq 0} t^k P(Y=k), \quad 0\leq t\leq 1 \qquad (4.1.3)$$

$\longrightarrow$ can be used to compute moments of r.v.'s $Y \in \{0,1,2,\dots\}$

$$E[t^{X_2} \mid X_1 = k] \overset{\text{def of}}{=} E[t^{N_1^2+\dots+N_{X_1}^2} \mid X_1 = k] \overset{X_n}{=} E[t^{N_1^2+\dots+N_k^2} \mid X_1 = k]$$

$$= E[t^{N_1^2+\dots+N_k^2} \mid X_1 = k] = E[\underbrace{t^{N_1^2}\cdots}_{\text{indep.}}\underbrace{t^{N_k^2}}_{\text{indep.}}]$$

$$= E[t^{N_1^2}]\cdots E[t^{N_k^2}] \overset{N_i^n \text{ i.i.d}}{=} E[t^N]\cdots\underbrace{E[t^N]}_{=\mathcal{P}_N(t)\overset{\text{def}}{=}\phi(t)} \overset{\text{def}}{=} (\phi(t))^k \quad (*)$$

$$\Rightarrow E[t^{X_2}] = \sum_{k\geq 0} E[t^{X_2}\cdot \mathbb{1}_{\{X_1=k\}}] =$$

$$\sum_{k\geq 0} E[t^{X_2} \mid X_1 = k]\cdot P(X_1=k) \overset{(*)}{=} \sum_{k\geq 0} (\phi(t))^k \cdot P(X_1=k)$$

$$\overset{\text{def. of expect.}}{=} E[(\phi(t))^{X_1}] \overset{X_1=N}{=} E[(\phi(t))^N] \overset{\text{def}}{=} \phi(u)\big|_{u=\phi(t)} = \phi(\phi(t))$$

$$= \phi^{(2)}(t)$$

Or more generally

$$\mathcal{P}_{X_n}(t) \overset{\text{def}}{=} E[t^{X_n}] = \phi^{(n)}(t) \quad (= \underbrace{\phi\circ\dots\circ\phi}_{n \text{ times}}(t)) \qquad (4.1.4)$$