Research Report
Statistical Research Unit
Department of Economics
University of Gothenburg
Sweden

A Markov Chain Model for Analysing the Progression of Patient's Health States

Robert Jonsson

Research Report
2011:6
ISSN 0349-8034

Mailing address:
Statistical Research Unit
P.O. Box 640
SE 405 30 Göteborg
Sweden

Phone
Nat: 031-786 00 00
Int: +46 31 786 00 00

Home Page:
http://www.statistics.gu.se/

Research Report
2011:6
ISSN 0349-8034

Mailing address:
Statistical Research Unit
P.O. Box 640
SE 405 30 Göteborg
Sweden

Phone
Nat: 031-786 00 00
Int: +46 31 786 00 00

Home Page:
http://www.statistics.gu.se/

A Markov Chain Model for Analysing the Progression of Patient's Health States

ROBERT JONSSON

Robert.Jonsson@economics.gu.se

Department of Economics, Goteborg University, Box 640, 405 30 Goteborg, Sweden

Abstract. Markov chains (MCs) have been used to study how the health states of patients are progressing in time. With few exceptions the studies have been based on the questionable assumptions that the MC has order m=1 and is homogeneous in time. In this paper a three-state non-homogeneous MC model is introduced that allows m to vary. It is demonstrated how wrong assumptions about homogeneity and about the value of m can invalidate predictions of future health states. This can in turn seriously bias a cost-benefit analysis when costs are attached to the predicted outcomes. The present paper only considers problems connected with model construction and estimation. Problems of testing for a proper value of m and of homogeneity is treated in a subsequent paper. Data of work resumption among sick-listed women and men are used to illustrate the theory. A non-homogeneous MC with m = 2 was well fitted to data for both sexes. The essential difference between the rehabilitation processes for the two sexes was that men had a higher chance to move from the intermediate health state to the state 'healthy', while women tended to remain in the intermediate state for a longer time.

Keywords: Rehabilitation, transition probability, prediction, Maximum Likelihood

1. Introduction

The patient's health state is more or less related to earlier health states. For this reason Markov chain (MC) models can be useful to study how the health states of patients are progressing in time. In a MC one has to specify the following constituents: (1) time unit (e.g. day, month, year), (2) possible states (e.g. diseased, improved, healthy), (3) Markov order, i.e. the number of time points back in history that has to be considered when assigning a transition probability one step ahead, and finally (4) how the latter transition probabilities change with time. The Markov order will in the sequel be denoted by m. If a transition to a state is independent of earlier states then m=0, if it depends on the last reached state m=1, if it depends on the last two reached states m=2, and so on. Thus, m refers to the last history preceding a transition. More general cases where transitions depend on other sub-spaces of the history (see e.g. [14]) will not be considered here. When the transition probabilities are constant in time the MC is said to be homogeneous and otherwise non-homogeneous.

In practice, data may be insufficient in order to meet the requirements needed to specify a MC correctly. Consider for instance a homogeneous MC with m=2 and with the three states 0, 1 and 2. Given the nine possible preceding states (0,0), (0,1),…, (2,2), there is a total of 27 transition frequencies to the states 0, 1 and 2. In small samples there is clearly a risk of getting zeros in some of the 27 cells and it will be hard to accurately estimate all the (27-9=18) linearly independent transition probabilities, or even impossible if marginal cells contains zeros. For a non-homogeneous MC these problems become much more severe since the number of transition probabilities to be estimated increases rapidly. It is easily shown that for a MC of order $m \geq 1$ with $s \geq 2$ possible states that is observed at the times t=1…T, T>m, there will be $s^{m+1} - s$ non-linearly dependent transition probabilities to estimate in a homogeneous MC, while the corresponding number in a non-homogeneous MC is $[1+(T-m)(s-1)]s^m - s$. These expressions being obtained under the assumption that transitions to all states are possible and that $m \leq r$ for transitions from a state at time r to a state at time r+1.

In small samples one may be forced to use a homogeneous MC model with a small value of m and a small number of states, without having the possibility to check the validity of the model. A typical example of this is the study of Gay and Wong [3] who used a homogeneous MC model with m=1 to predict the two states 'successful' and

‘unsuccessful’ in a sample of 71 clients from private rehabilitation agencies. The many-parameter problem that arises due to a large value of *m* or due to a large number of states can be tackled in several ways. When *m* is large in a homogeneous MC model the number of parameters to estimate can be reduced by fitting autoregressive-like functions to the transition probabilities (Raftery and Tavare´ [13]). Such an approach may however be questionable if data show signs of a non-homogeneous structure (see comments in Sect. 4.3 in [13]). Sometimes it is possible to reduce the number of transition probabilities by utilizing prior information. E.g. McLean and Millard [11] used a homogeneous MC model with *m*=1 to study the progression of geriatric patients through the four states ‘Acute care’, ‘Rehabalitive’, ‘Long-Stay’ and ‘Dead’. Here, 8 transition probabilities could be put equal to either 0 or 1, so there were just 4 probabilities left to estimate. A further example of such a parsimonious MC model is presented in Section 2 of this paper.

Principles for estimation and test of Markov order in homogeneous MC models have been known since long ago (Hoel [5]). These results have mainly been applied to meteorological data [9, 10] and to DNA sequences [2], just to mention a few examples. Few attempts, if any, seem to have been made to test the Markov order of series of health states. The value of *m* gives in fact important information about how the patient’s health state depends on history. E.g. a large *m* tells us that the health state at a time point is determined by factors that were present far ago. It is furthermore important that *m* is correctly estimated if the object of a study is to make predictions of the patient’s future health. As will be seen in this paper, unrealistic assumptions about the value of *m* can lead to predictions that are seriously biased. This may in turn invalidate a cost-benefit analysis, if costs have been attached to the different health states.

With few exceptions (see e.g. [6]), the MC models used in natural sciences have been homogeneous. This may be a reasonable assumption when analysing meteorological data over relatively short periods or DNA sequences, just to take a few examples. But, it can be put in question whether this assumption holds for e.g. rehabilitation processes, since it implies that the patient at any time has the same chance to move towards the state ‘healthy’ during the whole rehabilitation period.

The two problems, to test for homogeneity and to determine the Markov order of a (possibly) non-homogeneous MC by various test procedures, require an extensive investigation that will be communicated in a subsequent paper [7]. In this paper, a non-homogeneous three-state MC model with Markov order $m \geq 0$ is introduced for the health states (Section 2). Section 3 is devoted to principles for parameter estimation and in

Section 4 an application is given that compares the different patterns between women and men in work resumption during rehabilitation. The paper ends with some final remarks (Section 5).

2. Basic properties of the three-state model for progress of health

2.1 Notations and assumptions for probabilities

Let $X_t^{(m)}$, $t = 1, 2, ..., T$ denote the health state at time *t* for a MC of order *m*, with the possible outcomes 0 (Healthy), 1 (Improved) and 2 (Acute diseased). The probabilities of these outcomes are $P(X_t^{(m)} = j)$, $j = 0, 1, 2$. At *t*=1 only the states 1 and 2 are possible and for these initial states the notations $\pi_2 = P(X_1^{(m)} = 2)$ and $1 - \pi_2 = P(X_t^{(m)} = 1)$ are used. From one state at time *t* to the following state at time *t*+1 only the following transitions are possible: $2_t \to (2_{t+1}$ or $1_{t+1})$, $1_t \to (1_{t+1}$ or $0_{t+1})$ and $0_t \to 0_{t+1}$. Here the notation j_t has been used to denote that state *j* is occupied at time *t*. Transitions to state 2 can thus only take place from state 2. Therefore, omitting the index *m* for simplicity, the outcome $(X_{t-s} = 2, \ldots, X_t = 2)$ is equivalent to the outcome $(X_t = 2)$ for any $s \geq 1$, and this in turn implies that

$$P\left(X_{t+1} = 2 \middle| X_{t-s} = 2 \ldots X_t = 2\right) = P\left(X_{t+1} = 2 \middle| X_t = 2\right) = \beta_t \tag{1}$$

$$P\left(X_{t+1} = 1 \middle| X_{t-s} = 2 \ldots X_t = 2\right) = P\left(X_{t+1} = 1 \middle| X_t = 2\right) = 1 - \beta_t$$

When the last preceding state at time *t* is 1, the earlier states can be either 1 or 2. Such transitions are denoted by α_t-probabilities as in Table 1. The latter has to be denoted in such a way that they reflect the preceding states. The following notations will be used:

$$\begin{aligned} &P\left(X_{t+1} = 1 \middle| X_{t-s+1} = 1 \ldots X_t = 1\right) = \alpha_t(1^s) \\ &P\left(X_{t+1} = 1 \middle| X_{t-s} = 2, X_{t-s+1} = 1 \ldots X_t = 1\right) = \alpha_t(2, 1^s) \end{aligned} \tag{2}$$

Table 1 Schematic illustration of probabilities for transitions from the states at time *t* to the states at time *t*+1.

		State at *t*+1		
		0	1	2
State at *t*	0	1	0	0
	1	$1-\alpha_t$	α_t	0
	2	0	$1-\beta_t$	β_t

The Markov order m of the model is defined as the Markov order of transitions to 1_{t+1}, where the transition probabilities are given in (2).

2.2 Some expressions for transition- and state probabilities

The t-step transition probabilities $P\left(X_{1+t}^{(m)}=j\middle|X_1^{(m)}=i\right)$ can be used for predicting the future state t steps ahead. In this section some expressions for the latter are given that will be used in subsequent sections. For transitions from state 2 to state 1, results are only presented for m=1, 2 and 3, since the expressions are quite extensive. Results for chains of higher order can easily be deducted from the latter. For simplicity the following notation is introduced:

$$\text{For } j\geq 1\text{, put } \Phi_j=\begin{cases}(1-\beta_1), & \text{for } j=1\\ (1-\beta_j)\prod_{i=1}^{j-1}\beta_i, & \text{for } j\geq 2\end{cases}.$$

2.2.1 Transition probabilities in the non-homogeneous case

Transitions from state 2:

$$P\left(X_{1+t}^{(m)}=2\middle|X_1^{(m)}=2\right)=\prod_{j=1}^{t}\beta_j$$

$$P\left(X_{1+t}^{(1)}=1\middle|X_1^{(1)}=2\right)=\begin{cases}\Phi_1, & t=1\\ \Phi_t+\sum_{j=1}^{t-1}\Phi_j\prod_{i=j+1}^{t}\alpha_i(1), & t\geq 2\end{cases}$$

$$P\left(X_{1+t}^{(2)}=1\middle|X_1^{(2)}=2\right)=\begin{cases}P\left(X_{1+t}^{(1)}=1\middle|X_1^{(1)}=2\right), & t=1\\ \Phi_2+\Phi_1\alpha_2(2,1), & t=2\\ \Phi_t+\Phi_{t-1}\alpha_t(2,1)+\sum_{j=1}^{t-2}\Phi_j\alpha_{j+1}(2,1)\prod_{i=j+2}^{t}\alpha_i(1^2), & t\geq 3\end{cases}$$

$$P\left(X_{1+t}^{(3)}=1\middle|X_1^{(3)}=2\right)=\begin{cases}P\left(X_{1+t}^{(2)}=1\middle|X_1^{(2)}=2\right), & t=1,2\\ \Phi_3+\Phi_2\alpha_3(2,1)+\Phi_1\alpha_2(2,1)\alpha_3(2,1^2), & t=3\\ \Phi_t+\Phi_{t-1}\alpha_t(2,1)+\Phi_{t-2}\alpha_{t-1}(2,1)\alpha_t(2,1^2)+\sum_{j=1}^{t-3}\Phi_j\alpha_{j+1}(2,1)\alpha_{j+2}(2,1^2)\prod_{i=j+3}^{t}\alpha_i(1^3), & \\ t\geq 4 & \end{cases}$$

$$P\left(X_{1+t}^{(m)}=0\middle|X_1^{(m)}=2\right)=1-P\left(X_{1+t}^{(m)}=1\middle|X_1^{(m)}=2\right)-P\left(X_{1+t}^{(m)}=2\middle|X_1^{(m)}=2\right)\ \text{ for } t\geq 1.$$

Transitions from state 1:

$$P\left(X_{1+t}^{(m)}=1\middle|X_1^{(m)}=1\right)=\begin{cases}\prod_{i=1}^{t}\alpha_i(1^i), & t\le m\\ \prod_{i=1}^{m}\alpha_i(1^i)\cdot\prod_{i=m+1}^{t}\alpha_i(1^m), & t\ge m+1\end{cases}$$

$$P\left(X_{1+t}^{(m)}=0\middle|X_1^{(m)}=1\right)=1-P\left(X_{1+t}^{(m)}=1\middle|X_1^{(m)}=1\right)$$

Notice that $P\left(X_{1+t}^{(m)}=1\middle|X_1^{(m)}=2\right)=P\left(X_{1+t}^{(t)}=1\middle|X_1^{(t)}=2\right)$ for $t\le m-1$. This is simply because the *m*:th order Markov property can not be applied on transitions that are smaller than *m*.

Proofs of the above expressions are straightforward but tedious. Consider e.g. the expression for $P\left(X_{1+t}^{(m)}=1\middle|X_1^{(m)}=2\right)$ with *m*=*t*=3. $P\left(X_4^{(3)}=1\middle|X_1^{(3)}=2\right)=S/P(2_1)$, where $S=P(2_1,2_2,2_3,1_4)+P(2_1,2_2,1_3,1_4)+P(2_1,1_2,1_3,1_4)$. Here,

$$P(2_1,2_2,2_3,1_4)=P\left(1_4\middle|2_3\right)P\left(2_3\middle|2_2\right)P\left(2_2\middle|2_1\right)P\left(2_1\right)=(1-\beta_3)\prod_{i=1}^{2}\beta_i\cdot P\left(2_1\right),$$

$$P\left(2_1,2_2,1_3,1_4\right)=P\left(1_4\middle|2_2,1_3\right)P\left(1_3\middle|2_2\right)P\left(2_2\middle|2_1\right)P\left(2_1\right)=\alpha_3(2,1)(1-\beta_2)\beta_1P\left(2_1\right),$$

$$P\left(2_1,1_2,1_3,1_4\right)=P\left(1_4\middle|2_1,1_2,1_3\right)P\left(1_3\middle|2_1,1_2\right)P\left(1_2\middle|2_1\right)P\left(2_1\right)=\alpha_3(2,1^2)\alpha_2(2,1)(1-\beta_1)P\left(2_1\right),$$

and from this the result follows.

2.2.2 Transition probabilities in the homogeneous case

Various degree of homogeneity occurs when the α and β transition probabilities do not change with time. Here, results will only be given for the case when all transition probabilities are constant at all the times $t=1\ldots T-1$. The corresponding expressions for various cases with partial homogeneity are easily obtained from the results in Section 2.2.1. Introduce the ratios $r_1=\alpha(1)/\beta$, $r_2=\alpha(1^2)/\beta$, $r_3=\alpha(2,1)/\beta$, $r_4=\alpha(1^3)/\beta$ and $r_5=\alpha(2,1^2)/\beta$. Then one gets the following.

For $t\ge 1$:

$$P\left(X_{1+t}^{(m)}=2\middle|X_1^{(m)}=2\right)=\beta^t$$

$$P\left(X_{1+t}^{(1)}=1\middle|X_1^{(1)}=2\right)=\begin{cases}(1-\beta)\beta^{t-1}\dfrac{(1-r_1^t)}{(1-r_1)}, & r_1\ne 1\\ t(1-\beta)\beta^{t-1}, & r_1=1\end{cases}$$

For $t \geq 2$:

$$P\left(X_{1+t}^{(2)}=1 \middle| X_1^{(2)}=2\right)=\begin{cases}(1-\beta)\beta^{t-1}\left[1+\dfrac{r_3(1-r_2^{t-1})}{(1-r_2)}\right], & r_2 \neq 1\\ (1-\beta)\beta^{t-1}\left[1+(t-1)r_3\right], & r_2=1\end{cases}$$

For $t \geq 3$:

$$P\left(X_{1+t}^{(3)}=1 \middle| X_1^{(3)}=2\right)=\begin{cases}(1-\beta)\beta^{t-1}\left[1+r_3+r_3r_5\dfrac{(1-r_4^{t-2})}{(1-r_4)}\right], & r_4 \neq 1\\ (1-\beta)\beta^{t-1}\left[1+r_3+(t-2)r_3r_5\right], & r_4=1\end{cases}$$

Finally,

$$P\left(X_{1+t}^{(m)}=1 \middle| X_1^{(m)}=1\right)=\begin{cases}\prod_{i=1}^{t}\alpha(1^i), & t \leq m\\ \prod_{i=1}^{m}\alpha(1^i)\cdot\left[\alpha(1^m)\right]^{t-m}, & t \geq m+1\end{cases}$$

2.2.3 Probabilities of the states 0, 1, 2

One easily gets the following relations,

$$P\left(X_{1+t}^{(m)}=2\right)=\pi_2 P\left(X_{1+t}^{(m)}=2 \middle| X_1^{(m)}=2\right) \text{ and}$$

$$P\left(X_{1+t}^{(m)}=1\right)=(1-\pi_2)P\left(X_{1+t}^{(m)}=1 \middle| X_1^{(m)}=1\right)+\pi_2 P\left(X_{1+t}^{(m)}=1 \middle| X_1^{(m)}=2\right),$$

where the t-step transition probabilities are given in 2.2.1 above. Also, $P\left(X_{1+t}^{(m)}=0\right)=1-P\left(X_{1+t}^{(m)}=1\right)-P\left(X_{1+t}^{(m)}=2\right)$.

2.3 Effects of miss-specifying homogeneity and Markov order

Non-homogeneous high order MCs may contain many parameters. Even for the parsimonious model considered in this paper, the number of parameters can be large. From the expressions in Section 2.2 it is seen that with m=1 and t transitions there are 2t parameters to estimate in the non-homogeneous case, compared to just two parameters in the homogeneous case. Since the parameters in a many-parameter model are estimated with less accuracy, it may be tempting to deliberately specifying a homogeneous model of low Markov order. However, this can lead to seriously biased results.

To illustrate the effect of wrongly assuming a homogeneous model, consider the following simple example with m=1. For t=1…7, let the true β-parameters change

according to $\beta_t = 0.5 \cdot t^{1/3}$ so that the β parameters increas from $\beta_1 = 0.50$ to $\beta_7 = 0.96$, and let $\alpha_t(1) = 0.50$, t=1…7. From the expressions in Section 2.2.1 for the non-homogeneous case one gets $P\left(X_{1+7}^{(1)} = 1 \middle| X_1^{(1)} = 2\right) = 0.04$. On the other hand, by using the same values of $\alpha_t(1)$ but with $\beta_t = 0.77$ (assuming homogeneity and taking the average of $\beta_1 ... \beta_7$) one gets $P\left(X_{1+7}^{(1)} = 1 \middle| X_1^{(1)} = 2\right) = 0.13$ which is more than three times larger than the former value.

MC models with different values of m can give rise to large differences between the t-step transition probabilities. It is extremely complicated to compare the latter probabilities for various m in general, because many parameters are involved. Here only the simple cases $P\left(X_3^{(m)} = 0 \middle| X_1^{(m)} = 2\right)$ with m=1 and m=2 are compared. To make the comparison fair one also has to impose the restriction that the one-step transition probabilities at time 2 to time 3 are the same. This is achieved by following the advice in Appendix A1.

Consider $P\left(X_3^{(m)} = 0 \middle| X_1^{(m)} = 2\right) = 1 - P\left(X_3^{(m)} = 1 \middle| X_1^{(m)} = 2\right) - P\left(X_3^{(m)} = 2 \middle| X_1^{(m)} = 2\right)$.

Here $P\left(X_3^{(m)} = 2 \middle| X_1^{(m)} = 2\right) = \beta_1 \beta_2$ for m=1,2 (cf. Section 2.2.1), while

$$P\left(X_3^{(1)} = 1 \middle| X_1^{(1)} = 2\right) = (1-\beta_2)\beta_1 + (1-\beta_1)\alpha_2(1)$$
$$P\left(X_3^{(2)} = 1 \middle| X_1^{(2)} = 2\right) = (1-\beta_2)\beta_1 + (1-\beta_1)\alpha_2(2,1)$$

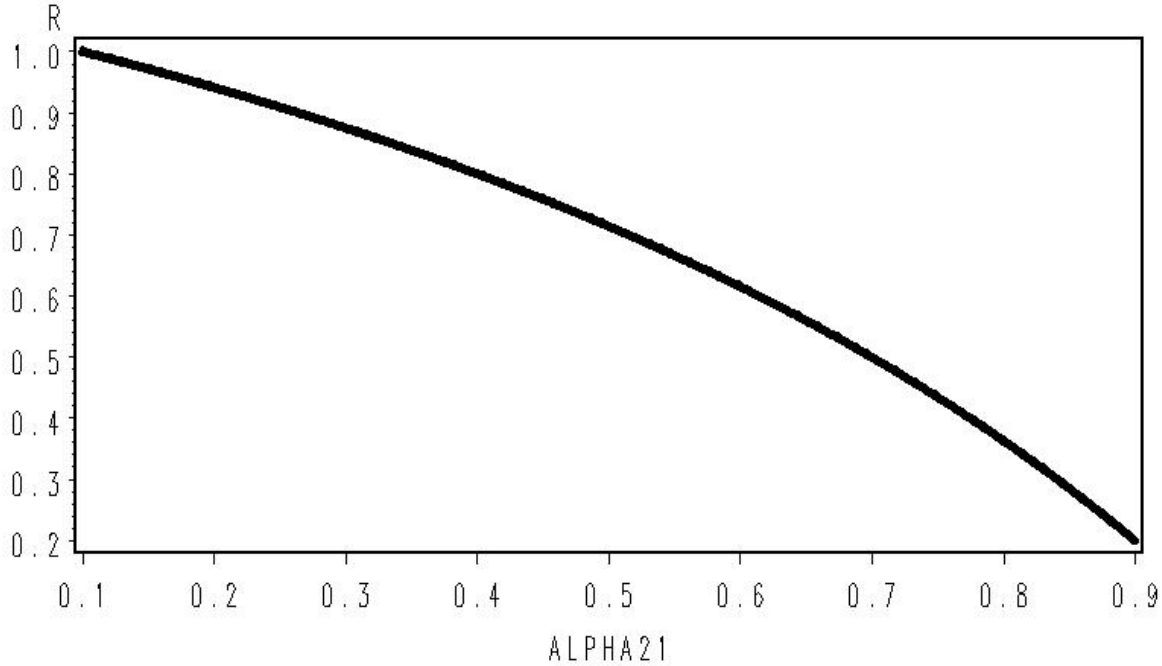
From this one gets the ratio

$$R = \frac{P\left(X_3^{(2)} = 0 \middle| X_1^{(2)} = 2\right)}{P\left(X_3^{(1)} = 0 \middle| X_1^{(1)} = 2\right)} = \frac{1-\alpha_2(2,1)}{1-\alpha_2(1)} = \frac{1-\alpha_2(2,1)}{1-\left[\alpha_2(1^2)W_2 + \alpha_2(2,1)(1-W_2)\right]},$$

where $W_2 = \dfrac{\alpha_1(1)(1-\pi_2)}{\alpha_1(1)(1-\pi_2) + (1-\beta_1)\pi_2}$. It is easily seen that R=1 only if $\alpha_2(1^2) = \alpha_2(2,1)$.

When the latter probabilities are different, R can be much smaller or larger than 1. This is illustrated in Figure 1a and Figure 1b where $\alpha_2(2,1)$ varies between 0.1 and 0.9. In both figures the ratio R is five times larger at the beginning than at the end. The conclusion is that it is important that m is correctly specified when t-step transition probabilities are to be computed, even when t is small. For larger values of t and for larger differences between the Markov orders, the ratio R can be much larger.

1a



1b

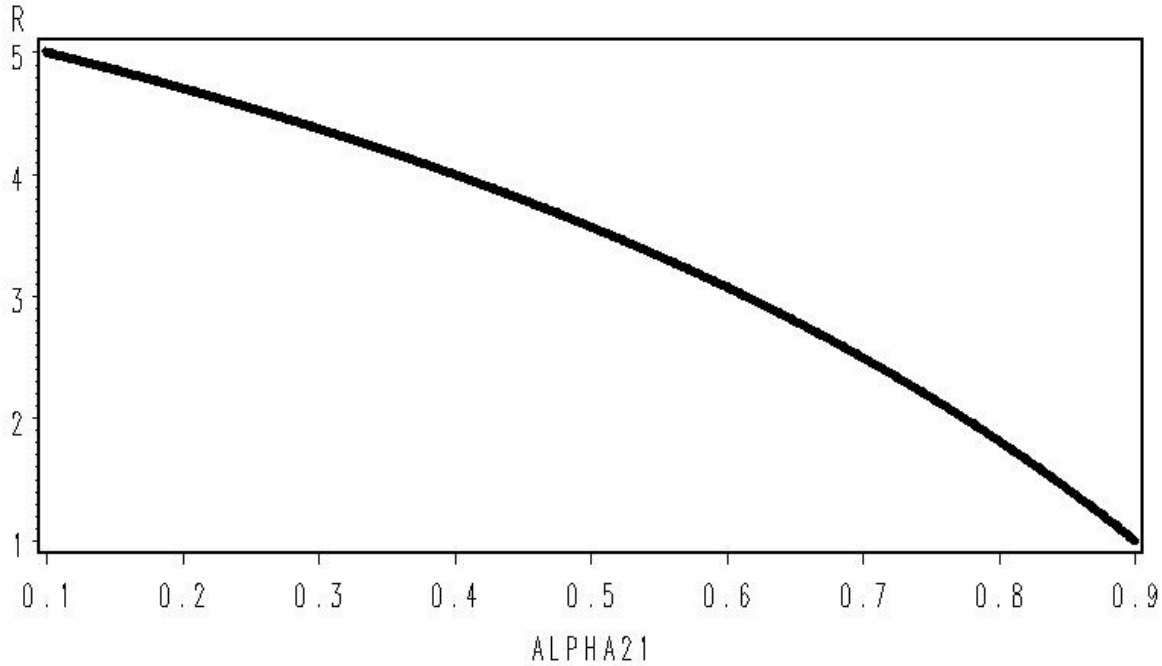


Figure 1 The ratio $R = P\left(X_3^{(2)} = 0 \middle| X_1^{(2)} = 2\right) / P\left(X_3^{(1)} = 0 \middle| X_1^{(1)} = 2\right)$ as a function of $\alpha_2(2,1)$.

(a) When $W_2 = 0.5$ and $\alpha(1^2) = 0.1$. (b) When $W_2 = 0.9$ and $\alpha(1^2) = 0.9$.

3. Estimation of parameters

3.1 Notations for frequencies and an expression for the Likelihood

In general, let $(i_s \ldots j_t)$ denote the event that a sequence of states is occupied, from state *i* at time *s* to state *j* at time *t*. In analogy with the notations for the transition probabilities in (1) and (2) which were designated by Greek symbols, the following notations are used for the transition frequencies:

$$\begin{aligned} B_t &= \text{ Number of transitions from } (2_t) \text{ to } (2_{t+1}). \\ A_t(1^s) &= \quad \text{-"-} \quad (1_{t-s+1} \ldots 1_t) \text{ to } (1_{t+1}) \\ A_t(2,1^s) &= \quad \text{-"-} \quad (2,1_{t-s+1} \ldots 1_t) \text{ to } (1_{t+1}) \end{aligned} \tag{3}$$

The state frequencies or risk masses, i.e. the number of persons being in a state just before transitions occur, are denoted in the following way:

$$\begin{aligned} N_t(2) &= \text{ Number of persons in state } 2_t \\ N_t(1^s) &= \quad \text{-"-} \quad (1_{t-s+1} \ldots 1_t) \\ N_t(2,1^s) &= \quad \text{-"-} \quad (2,1_{t-s+1} \ldots 1_t) \end{aligned} \tag{4}$$

The quantities in (3) and (4) are related. If no subjects in the sample disappear between the transitions, then e.g. $B_t = N_{t+1}(2)$ and $A_t(1^s) = N_{t+1}(1^{s+1})$. Since withdrawals may occur in practise both notations in (3) and (4) will be used. The total fixed sample size is denoted by *n*. An illustration of these frequencies is shown in Table 2 for a Markov chain of order *m*=2. Here one may notice that $A_t(1^2) + A_t(2,1) = A_t(1)$ and $N_t(1^2) + N_t(2,1) = N_t(1)$.

Table 2 Transition- and state frequencies in the three-state model for progress of health when the Markov order is *m*=2.

		State at *t*+1			
		0	1	2	Total
State at (*t*-1,*t*)	(0,0)	$N_t(0)$	0	0	$N_t(0)$
	(1,1)	$N_t(1^2) - A_t(1^2)$	$A_t(1^2)$	0	$N_t(1^2)$
	(2,1)	$N_t(2,1) - A_t(2,1)$	$A_t(2,1)$	0	$N_t(2,1)$
	(2,2)	0	$N_t(2) - B_t$	B_t	$N_t(2)$
					n

$N_t(1^s)$, $N_t(2,1^s)$ and $N_t(2)$ will in the sequel be termed working sample sizes and these will be used for estimating the α-and β-parameters. The working sample sizes will gradually become smaller as persons move to state 0.

The likelihood $L^{(m)}$ of all observations in the three-state model of Markov order *m* can be expressed as

$$L^{(m)} = C \cdot \prod_{t=1}^{T-1} \beta_t^{B_t} (1-\beta_t)^{N_t(2)-B_t} \cdot F^{(m)} \cdot G^{(m)} \tag{5}$$

Here *C* is a constant that does not depend on the α- and β-parameters, and

$$F^{(m)} = \begin{cases} \prod_{t=1}^{T-1} [\alpha_t(1)]^{A_t(1)} [1-\alpha_t(1)]^{N_t(1)-A_t(1)}, & \text{if } m=1 \\ \prod_{t=1}^{m-1} \left[\alpha_t(1^t)\right]^{A_t(1^t)} \left[1-\alpha_t(1^t)\right]^{N_t(1^t)-A_t(1^t)} \prod_{t=m}^{T-1} \left[\alpha_t(1^m)\right]^{A_t(1^m)} \left[1-\alpha_t(1^m)\right]^{N_t(1^m)-A_t(1^m)}, & \text{if } m \geq 2 \end{cases}$$

$$G^{(m)} = \begin{cases} 1, \text{ if } m=1 \\ \prod_{i=2}^{m} \prod_{t=i}^{T-1} \left[\alpha_t(2,1^{i-1})\right]^{A_t(2,1^{i-1})} \left[1-\alpha_t(2,1^{i-1})\right]^{N_t(2,1^{i-1})-A_t(2,1^{i-1})}, & \text{if } m \geq 2 \end{cases}$$

A proof of the above relations is outlined in the Appendix A1.

Some special cases are, for *m*=1, 2:

$$L^{(1)} = C \cdot \prod_{t=1}^{T-1} \beta_t^{B_t} (1-\beta_t)^{N_t(2)-B_t} [\alpha_t(1)]^{A_t(1)} \left[1-\alpha_t(1)\right]^{N_t(1)-A_t(1)} = \text{(Homogeneous case)}$$

$$= C \cdot \beta^{\sum B_t} (1-\beta)^{\sum N_t(2)-\sum B_t} [\alpha(1)]^{\sum A_t(1)} \left[1-\alpha(1)\right]^{\sum N_t(1)-\sum A_t(1)}$$

$L^{(2)} = C \cdot F^{(2)} \cdot G^{(2)}$, where $F^{(2)} = \alpha_1(1)^{A_1(1)} \left[1-\alpha_1(1)\right]^{N_1(1)-A_1(1)}$ (the same in the homogeneous case) and

$$G^{(2)} = \prod_{t=2}^{T-1} [\alpha_t(1^2)]^{A_t(1^2)} \left[1-\alpha_t(1^2)\right]^{N_t(1^2)-A_t(1^2)} [\alpha_t(2,1)]^{A_t(2,1)} \left[1-\alpha_t(2,1)\right]^{N_t(2,1)-A_t(2,1)} = \text{(Homogeneous case)}$$

$$= [\alpha(1^2)]^{\sum A_t(1^2)} \left[1-\alpha(1^2)\right]^{\sum N_t(1^2)-\sum A_t(1^2)} [\alpha(2,1)]^{\sum A_t(2,1)} \left[1-\alpha_t(2,1)\right]^{\sum N_t(2,1)-\sum A_t(2,1)}$$

3.2 Maximum Likelihood estimation of the parameters

By taking the derivatives of the logarithm of $L^{(m)}$ in (5) with respect to the unknown parameters and equating to zero, one easily finds the following Maximum Likelihood (ML) estimators of the parameters.

$$\hat{\beta}_t = \frac{B_t}{N_t(2)}, \ \hat{\alpha}_t(1^s) = \frac{A_t(1^s)}{N_t(1^s)} \ , \ \hat{\alpha}_t(2,1^s) = \frac{A_t(2,1^s)}{N_t(2,1^s)} \tag{6}$$

In the homogeneous case,

$$\hat{\beta} = \frac{\sum B_t}{\sum N_t(2)} \ , \ \hat{\alpha}(1^s) = \frac{\sum_{t=s}^{T-1} A_t(1^s)}{\sum_{t=s}^{T-1} N_t(1^s)}, \ \hat{\alpha}(2,1^s) = \frac{\sum_{t=s}^{T-1} A_t(2,1^s)}{\sum_{t=s}^{T-1} N_t(2,1^s)} \tag{7}$$

In the special case when m=0, one has the restriction that $\alpha_t(1) + \beta_t = 1$. (This is because $\beta_t = P(2_{t+1}|2_t) = 1 - P(1_{t+1}|2_t) = 1 - P(1_{t+1}|1_t) = 1 - \alpha_t(1)$.) The ML estimator of the single linearly independent parameter $\alpha_t(1)$ is

$$\hat{\alpha}_t(t) = \frac{A_t(1) + N_t(2) - B_t}{N_t(1) + N_t(2)} \tag{8}$$

In the homogeneous case all terms in (8) are preceded by summation signs, as in (7).

Special cases of the α -estimators for m=1,2,3

<u>m=1</u>: $\hat{\alpha}_t(1) = \frac{A_t(1)}{N_t(1)}$ $(1 \le t \le T-1)$, and $\hat{\alpha}(1) = \frac{\sum A_t(1)}{\sum N_t(1)}$ in the homogeneous case.

<u>m=2</u>: $\hat{\alpha}_1(1) = \frac{A_1(1)}{N_1(1)}$, $\hat{\alpha}_t(1^2) = \frac{A_t(1^2)}{N_t(1^2)}$ and $\hat{\alpha}_t(2,1) = \frac{A_t(2,1)}{N_t(2,1)}$ $(2 \le t \le T-1)$. In the homogeneous case the estimator of $\alpha_1(1)$ is the same, while $\hat{\alpha}(1^2) = \frac{\sum A_t(1^2)}{\sum N_t(1^2)}$ and $\hat{\alpha}(2,1) = \frac{\sum A_t(2,1)}{\sum N_t(2,1)}$, where the summation is from t=2 to t=T-1.

<u>m=3:</u> $\hat{\alpha}_1(1) = \frac{A_1(1)}{N_1(1)}$, $\hat{\alpha}_2(1^2) = \frac{A_2(1^2)}{N_2(1^2)}$, $\hat{\alpha}_t(2,1) = \frac{A_2(2,1)}{N_2(2,1)}$ $(2 \le t \le T-1)$, ,

$$\hat{\alpha}_t(1^3) = \frac{A_t(1^3)}{N_t(1^3)} \text{ and } \hat{\alpha}_t(2,1^2) = \frac{A_t(2,1^2)}{N_t(2,1^2)} \quad (3 \le t \le T-1)$$

In the homogeneous case the estimators of $\alpha_1(1)$ and $\alpha_2(1^2)$ remain the same, while

$\hat{\alpha}(1^3) = \frac{\sum A_t(1^3)}{\sum N_t(1^3)}$ and $\hat{\alpha}(2,1^2) = \frac{\sum A_t(2,1^2)}{\sum N_t(2,1^2)}$, $\hat{\alpha}(2,1) = \frac{\sum A_t(2,1)}{\sum N_t(2,1)}$. The latter sums are computed from t=3 to t=T-1.

3.3 Properties of ML estimators

To study some properties of the ML estimators, a lemma will be used that is a generalization of well known results for proportions based on fixed sample sizes. Introduce the notation $X \sim B(n,\theta)$ for a random variable X that has a binomial distribution with integer parameter n and proportion θ.

Lemma. If $X \sim B(n, p_N)$ and the conditional random variable $(X|N) \sim B(N, p_X)$ then

(a) $X \sim B(n, p_N p_X)$.

(b) $\hat{p}_X = \frac{X}{N}$ is unbiased for p_X with $V(\hat{p}_X) = p_X(1-p_X)E\left(N^{-1}\right)$. Furthermore,

$\hat{V}(\hat{p}_X) = \frac{\hat{p}_X(1-\hat{p}_X)}{N-1}$ is an unbiased estimator of $V(\hat{p}_X)$.

Also, if $N_i \sim B(n, p_{N_i})$ and $(X_i|N_i) \sim B(N_i, p_{X_i})$, i=1,2, where $(X_1|N_1)$ and $(X_2|N_2)$ are independent, then

(c) $\hat{p}_{X_1} - \hat{p}_{X_2} = \frac{X_1}{N_1} - \frac{X_2}{N_2}$ is unbiased for $p_{X_1} - p_{X_2}$. Furthermore, if $p_{X_1} = p_{X_2} = p_X$ say,

then $\hat{p}_X = \frac{X_1 + X_2}{N_1 + N_2}$ is unbiased for p_X and $\hat{p}_X(1-\hat{p}_X)\left(N_1^{-1} + N_2^{-1}\right)$ is an approximately unbiased estimator of $V\left(\hat{p}_{X_1} - \hat{p}_{X_2}\right)$.

For a proof of the above relations, see Appendix A2.

To apply the lemma on the estimator $\hat{\beta}_t$ in (6), notice that $N_t(2) \sim B(n, P(X_t^{(m)} = 2))$, where $P(X_t^{(m)} = 2)$ can be obtained from the results in Sections 2.2.1 and 2.2.2. Since $(B_t|N_t(2)) \sim B(N_t(2), \beta_t)$ it follows that $\hat{\beta}_t$ is unbiased for β_t and also that

$$V(\hat{\beta}_t) = \beta_t(1-\beta_t)E\left(\frac{1}{N_t(2)}\right) \text{ with } \hat{V}(\hat{\beta}_t) = \frac{\hat{\beta}_t(1-\hat{\beta}_t)}{N_t(2)-1} \tag{9}$$

In a similar way it is easily seen that $\hat{\alpha}_t(1^s)$ and $\hat{\alpha}_t(2,1^s)$ in (6) are unbiased with unbiased variance estimators

$$\hat{V}\left(\hat{\alpha}_t(1^s)\right) = \frac{\hat{\alpha}_t(1^s)\left[1-\hat{\alpha}_t(1^s)\right]}{N_t(1^r)-1} \text{ and } \hat{V}\left(\hat{\alpha}_t(2,1^s)\right) = \frac{\hat{\alpha}_t(2,1^s)\left[1-\hat{\alpha}_t(2,1^s)\right]}{N_t(2,1^s)-1} \tag{10}$$

By the same arguments it can be shown that the estimators in the homogeneous case in (7) are unbiased. In the latter case an unbiased estimator of $V\left(\hat{\beta}\right)$ is

$$\hat{V}(\hat{\beta}) = \frac{\hat{\beta}(1-\hat{\beta})}{\sum N_t(2)-1} \tag{11}$$

The variances of the other estimators in (7) are estimated analogously.

Notice that the variances of the above estimators differ from the variances of the corresponding estimators based on fixed sample sizes. Consider for instance $V(\hat{\beta}_t)$ in (9). With a fixed sample size it is well known that the latter has a maximum for $\beta_t = 1/2$. But, in the present case the variance depends on $E\left(1/N_t(2)\right)$, which in turn is a function of β_t. Here, $E\left(1/N_t(2)\right)$ can be obtained from a Taylor expansion, using the fact that $E\left(N_t(2)\right)$ $= n \cdot P(X_t^{(m)}=2)$ and $V\left(N_t(2)\right) = n \cdot P(X_t^{(m)}=2)\left[1-P(X_t^{(m)}=2)\right]$. In the last two expressions $P(X_t^{(m)}=2) = \pi_2 \prod_{j=1}^{t-1} \beta_j$ (cf. Sections 2.2.1 and 2.2.3). Assume for simplicity that all β_t's are approximately equal to β. Then $E\left(1/N_t(2)\right) \approx \left(n\pi_2\beta^{t-1}\right)^{-1}$, omitting terms of order n^{-2} and smaller. In this case it is easily seen that $V(\hat{\beta}_t)$ is a strictly decreasing function of β when $t \geq 3$ that tends to infinity as β approaches 0, and tends to 0 as β approaches 1.

3.4 Confidence intervals for parameters

Conditionally on the working sample sizes, all numerators in (6) and (7) have Binomial distributions, e.g. $B_t \sim B\left(N_t(2), \beta_t\right)$. CI's for the parameters can thus be constructed by

using the relation between the Binomial and F distributions noticed by Jowett [8]. In this way the 95 % lower and upper confidence limits for β_t, $\hat{\beta}_t^{(L)}$ and $\hat{\beta}_t^{(U)}$, respectively, are

$$\begin{aligned}\hat{\beta}_t^{(L)} &= \frac{B_t}{B_t + \left(N_t(2) - B_t + 1\right) \cdot F_{.975}\left(2(N_t(2) - B_t + 1), 2B_t\right)} \\ \hat{\beta}_t^{(U)} &= \frac{\left(B_t + 1\right) \cdot F_{.975}\left(2(B_t + 1), 2(N_t(2) - B_t)\right)}{N_t(2) - B_t + \left(B_t + 1\right) \cdot F_{.975}\left(2(B_t + 1), 2(N_t(2) - B_t)\right)}\end{aligned} \tag{12}$$

In (12) the notation $F_{.975}(n_1, n_2)$ is used for the 97.5 percentile in the *F*-distribution with n_1 and n_2 degrees of freedom. CIs for other parameters are constructed in a similar way. The CIs constructed in this way are conservative in the sense that the actual coverage rate is at least 95 %.

3.5 A simulation study

To study the performance of the proposed estimators in Section 3.2-3.4 some simulations were carried out for MCs of order $m \leq 1$. The probabilities for remaining in state 1 were $\alpha_t(1) = 1 - \beta_t + \delta, t = 1,2,3$ with $\delta = 0\,(m = 0)$ and $\delta = 0.15\,(m = 1)$. The β-parameters were gradually increasing from 0.30 at *t* = 1 to 0.40 at *t* = 3 and from 0.60 at *t* = 1 to 0.70 at *t* = 3. Frequencies of the initial state at *t* = 1were obtained by assigning the states 2 and 1 for each subject in a sample of size *n* with probabilities $\pi_2 = 0.5$ and $1 - \pi_2 = 0.5$. *n* was chosen as 100, 500 and 1000. Each simulation consisted of 10 000 replicates.

The results are shown in Table 3. Here only the figures for the case $\delta = 0.15$ are given since no essential difference between the two cases was seen. In the table one notices that the bias of the β-estimators and of the estimated variances can be neglected. As *n* increases from 100 to 1000 there is a considerably reduction of the variance of the estimators and also of the average length of the CI intervals. This is of course a result of the fact that by increasing *n*, the working sample sizes $N_t(2)$ become larger. Notice that the 95 % CIs given by (12) are conservative and that the coverage rate can be much higher than 95 % in small samples, which in terms imply wider CIs.

Results for estimators of the α-parameters show a similar pattern and are therefore not presented.

	$\beta_1 = 0.30,\ \beta_2 = 0.35,\ \beta_3 = 0.40$								
	Bias of $\hat{\beta}_t$			Variance of $\hat{\beta}_t$			Estimated variance of $\hat{\beta}_t$		
n	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
100	.00	.00	.00	.0042	.0161	.0566	.0043	.0175	.0638
500	-.00	-.00	.00	.0008	.0031	.0096	.0008	.0031	.0099
1000	-.00	-.00	.00	.0004	.0015	.0046	.0004	.0015	.0048

	$\beta_1 = 0.30,\ \beta_2 = 0.35,\ \beta_3 = 0.40$								
	Mean of $N_t(2)$			Mean CI-level for β_t			Mean of CI-length		
n	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
100	50	15	5	96.7	97.4	97.5	.27	.50	.75
500	250	75	26	95.9	96.2	96.8	.12	.23	.39
1000	500	150	52	95.3	96.0	96.6	.08	.16	.28

	$\beta_1 = 0.60,\ \beta_2 = 0.65,\ \beta_3 = 0.70$								
	Bias of $\hat{\beta}_t$			Variance of $\hat{\beta}_t$			Estimated variance of $\hat{\beta}_t$		
n	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
100	-.00	.00	.00	.0049	.0078	.0111	.0049	.0080	.0119
500	-.00	.00	-.00	.0009	.0015	.0023	.0010	.0015	.0022
1000	-.00	.00	.00	.0004	.0008	.0011	.0005	.0008	.0011

	$\beta_1 = 0.60,\ \beta_2 = 0.65,\ \beta_3 = 0.70$								
	Mean of $N_t(2)$			Mean CI-level for β_t			Mean of CI-length		
n	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
100	50	30	20	96.4	96.6	97.7	.28	.36	.42
500	250	150	98	95.8	96.1	96.1	.12	.16	.19
1000	500	300	195	95.2	95.6	96.6	.09	.11	.13

Table 3 Results from the simulation studies with two sets of the β-parameters. (See text.)

4. An application: Work resumption among sick-listed women and men

In Sweden the annual costs for sick absence increased from 15 billion Euro in 1997 to 26 billion Euro in 1999, costs for production losses not included [4]. The growing awareness of these raising costs initiated several surveys in order to deal with the problem. One of

these was a survey performed in the county of Vastra Gotaland in Sweden to study work resumption among long-termed sick-listed persons who participated in various rehabilitation programs. After start of the rehabilitation the health states 0 (Healthy), 1 (Improved) and 2 (Acute diseased) were recorded for each person at the beginning of each quarter t = 1,2,...8. The classification into the states was made by social insurance authorities. In order to illustrate the results in previous sections, only different patterns in work resumption between the sexes will be examined. The time unit was originally one month (30 days), but there were several reasons for using quarter instead. One was that the transition frequencies changed very little between months. A further reason was that the unit quarter gave rise to the structure of Table 1, where no transitions from state 2 to state 0 were obtained. This structure was violated if longer periods than quarter were used.

The process X_t with the state space (0,1,2), t=1…8, will for simplicity be called the rehabilitation process. The sample consisted of 2440 women and 1801 men and the initial probability of being in state 2 was $\pi_2 = 0.45$ for women and $\pi_2 = 0.47$ for men. The percentage of the persons being in the different states is summarized in Table 4.

		Quarter								
Sex	State	1	2	3	4	5	6	7	8	Total
Women	2	45	37	34	33	32	32	31	30	2440
	1	55	55	50	47	45	43	42	42	
	0	-	8	16	20	23	25	27	28	
Men	2	47	40	36	34	33	32	31	30	1801
	1	53	48	43	39	35	33	32	32	
	0	-	12	21	27	32	35	37	38	

Table 4 Percentage of women and men that were in the different health states at the beginning of each quarter.

4.1 Transition probabilities

Tests of homogeneity and of Markov order m are considered in [7] and these suggested a non-homogeneous MC of order m = 2 for both sexes, possibly with a shift to m = 1 at the quarters 6 and 7. In a non-homogeneous MC of order 2 there are three transition probabilities, β_t, $\alpha_t(1^2)$ and $\alpha_t(2,1)$ to estimate at each $t \geq 2$. Figure 2a shows the estimates

of β_t, the probability of remaining in the state 2 at time *t*. The differences between the estimates for the two sexes were very small. The probability of remaining in state 2 was smallest during the first quarter and increased during the first year until it reached a stable level of about 0.98, disregarding a small decrease at the last quarter. Transitions from the most acute disease phase 2 to a less severe state 1 was thus most likely to take place at the start of the rehabilitation period, and if no transition has taken place during the first year the chance of such an event during the second year was small.

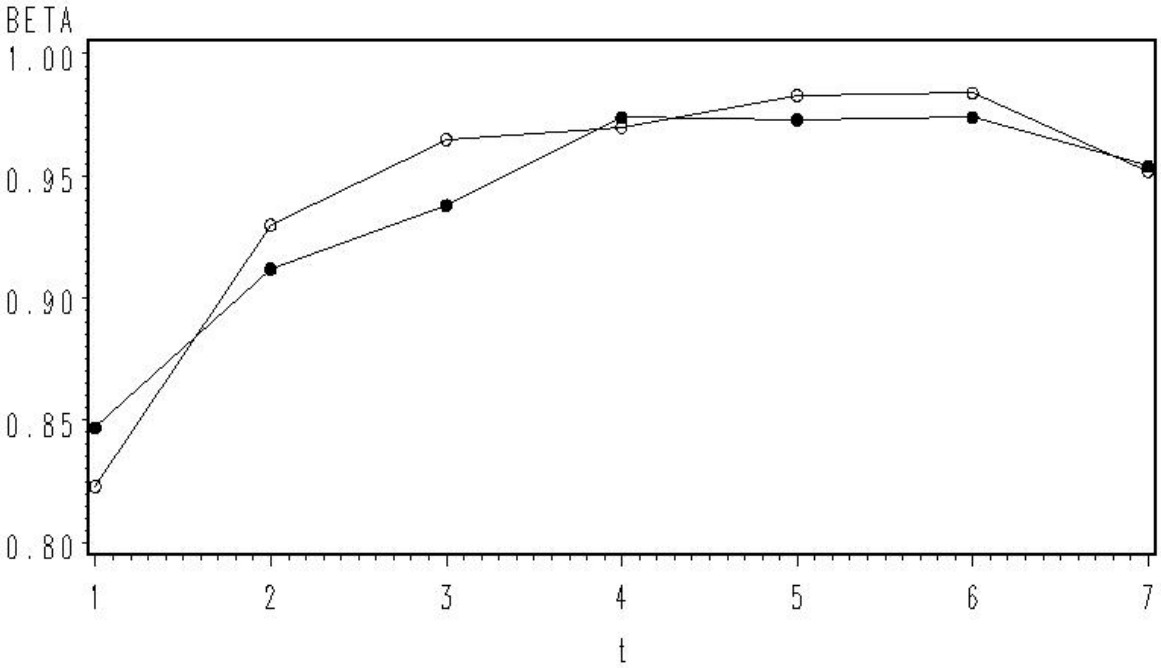
The two probabilities of remaining in state 1, $\alpha_t(1^2)$ and $\alpha_t(2,1)$, are shown in Figure 2b. The two probabilities differed markedly for both sexes, and the probability of remaining in state 1 was larger if the subject had previously remained in state 1, compared to if the subject had moved from state 2 to state 1. The two probabilities of remaining in state 1 were also constantly larger for women, except for the last two quarters. The estimates of $\alpha_t(2,1)$ are however more unreliable than the estimates of $\alpha_t(1^2)$, since $N_t(2,1)$ were considerably smaller than $N_t(1^2)$. E.g. for women $N_t(1^2)$ decreased from 1158 at *t* =2 to 1012 at *t* = 7, while $N_t(2,1)$ decreased from 177 at *t* =2 to just 10 at *t* =7.

From expressions like the one in (11) one may calculate the 95 % CI's for the parameters β_t, $\alpha_t(1^2)$ nd $\alpha_t(2,1)$. At *t*=4 the latter are (0.956, 0.981), (.918, 0.947) and (0.537, 0.889) for women, and (0.958, 0.985), (0.855, 0.906) and (0.513, 0.825) for men. Here one may notice that all CIs overlap except for the CIs for $\alpha_t(1^2)$. In general it turned out that the lengths of the confidence intervals for $\alpha_t(2,1)$ were 5-15 times wider than those for β_t and $\alpha_t(1^2)$.

4.2 r-step transition probabilities (*predictions*) *computed under various assumptions*

Given that a subject is in a state at some time, one may compute the probability that the subject is in a certain state at a future time, *r* steps ahead. Consider the probability of a transition from state 2 at the initial time 1 to state 0 at the times *r*=3…8, denoted by $P\left(X_r^{(m)}=0 \middle| X_1^{(m)}=2\right)$. The latter reflects how the probability of moving from the most acute disease state 2 at the start of the rehabilitation to the healthy state 0 develops in time,

(a)



(b)

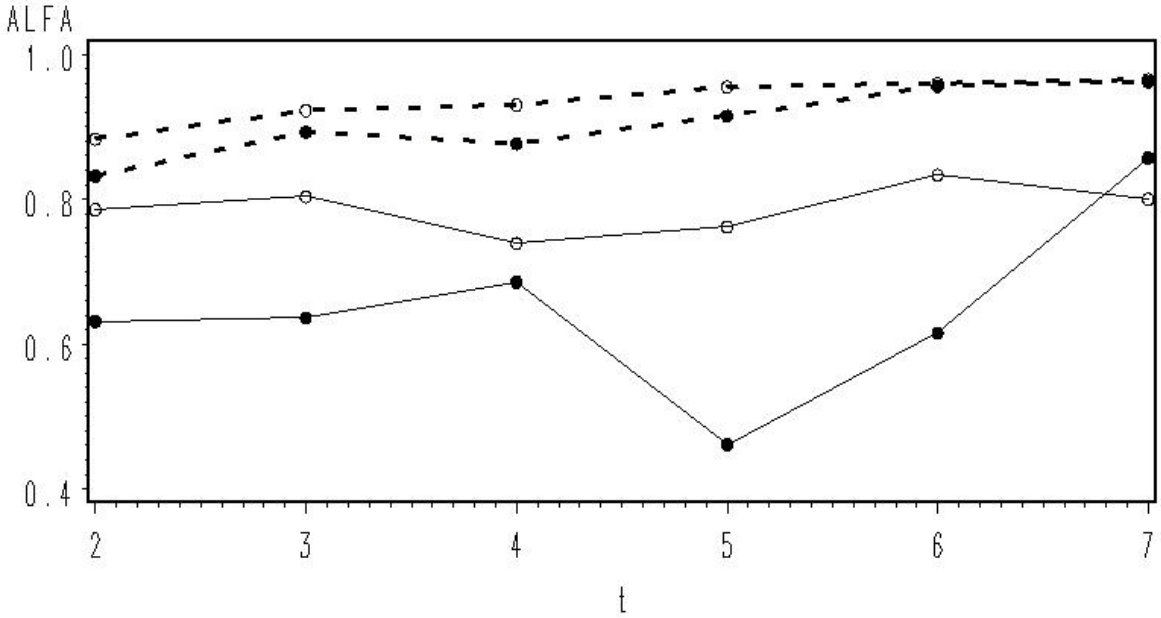


Figure 2. (a) Estimates of β_t, the probability of remaining in state 2 at t = 1...7, for women (unfilled circles) and men (filled circles). **(b)** Estimates of $\alpha_t(1^2)$ (upper two curves) and of $\alpha_t(2,1)$ (lower two curves), the two probabilities of remaining in state 1 depending on previous history, for women (unfilled circles) and men (filled circles).

from the first quarter of the rehabilitation to the *r*:th quarter. These probabilities can be estimated from the observed relative frequencies, but also by inserting the estimates for the α- and β-parameters into the expressions for $P\left(X_r^{(m)}=0 \middle| X_1^{(m)}=2\right)$ given in Section 2.2.1. The former estimates will be called model-free and the latter model-dependent.

Figure 3 shows the model-free estimates together with the model-dependent estimates for women when *m*=1 and *m*=2. The agreement between the model-free estimates and the model-dependent estimates was poor when *m*=1 but quite good when *m*=2. Given that a woman started in state 2, the probability of reaching state 0 at the last quarter *r*=8 was 0.114 (model-free estimate). The corresponding probability for men was about 50 % higher, 0.173. For both sexes these figures were considerably smaller than the probability of moving from the intermediate state 1 at quarter 1 to the state 0 at quarter 8, 0.409 for females and 0.552 for males.

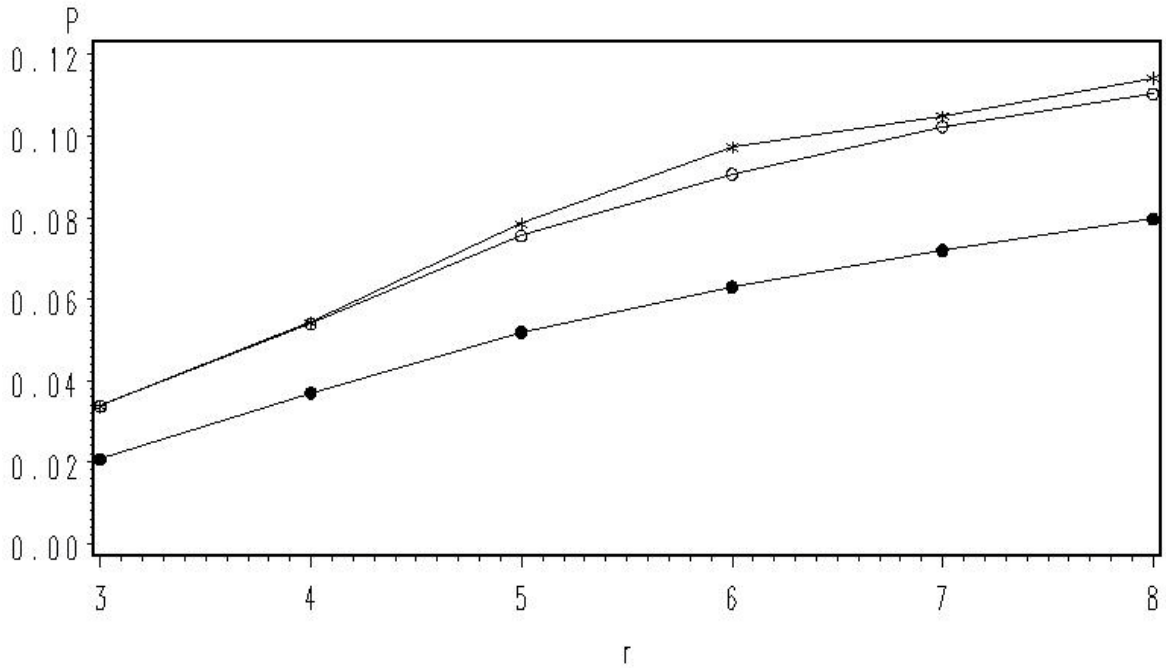


Figure 3. Estimates of the *r*-step probabilities $p = P\left(X_r^{(m)}=0 \middle| X_1^{(m)}=2\right)$, *r* =3...8, for women. Empirical model-free estimates (*), estimates assuming *m* = 2 (○) and estimates assuming *m* = 1 (●).

To compare the *r*-step transition probabilities between women and men, introduce for the moment the notations $p_{10}^{(W)}(1,r)$, $p_{20}^{(W)}(1,r)$ and $p_{10}^{(M)}(1,r)$, $p_{20}^{(M)}(1,r)$ for the probabilities

for women and men, respectively, to move from 0_1 to 1_r and from 0_1 to 2_r. Let $n_i^{(W)}$ and $n_i^{(M)}$, i=1,2, be the number women and men, respectively, who are in state i at the first quarter and let all estimates be model-free. Then it follows from large-sample theory that the statistic

$$\chi^2 = \frac{\left(\hat{p}_{10}^{(W)}(1,r) - \hat{p}_{10}^{(M)}(1,r)\right)^2}{\hat{p}_{10}(1,r)\left(1-\hat{p}_{10}(1,r)\right)\left((n_1^{(W)})^{-1} + (n_1^{(M)})^{-1}\right)},$$

$$\text{where } \hat{p}_{10}(1,r) = \frac{n_1^{(W)}\hat{p}_{10}^{(W)}(1,r) + n_1^{(M)}\hat{p}_{10}^{(M)}(1,r)\hat{p}_{10}^{(M)}(1,r)}{n_1^{(W)} + n_1^{(M)}}$$

can be used for testing the hypothesis $p_{10}^{(W)}(1,r) = p_{10}^{(M)}(1,r)$. The latter is rejected for large values of the test statistic which in large samples can be treated as a chi-square variable with 1 degree of freedom. A test of the hypothesis $\hat{p}_{20}^{(W)}(1,r) = \hat{p}_{20}^{(M)}(1,r)$ is treated similarly.

In the present data the hypothesis of equal r-step transition probabilities for women and men was rejected at the 5 % level for all r =3…8 regarding transitions from 1 to 0, and for all r=4…8 regarding transitions from 2 to 0. In all cases men had higher probabilities. The only non-significant difference was obtained for transitions from state 2 at time 1 to state 0 at time 3, where the value of the chi-square statistic was 3.45, corresponding to the p-value 0.063. In the present data, consisting of persons that had participated in rehabilitation programs, there is thus massive evidence for the fact that men have a higher chance than women to reach the final healthy state 0 within 2 years.

A similar analysis of the difference between the sexes regarding the transition probabilities from state 2 at time 1 to state 1 at r=2…8, showed no significant differences. Possible explanations of these different patterns are discussed below.

Finally, it may be instructive to demonstrate what would happen if the data were analysed by assuming a homogeneous MC of order m=1. For men one gets the estimates $\hat{\beta} = 0.936$ and $\hat{\alpha}(1) = 0.868$, and from the results in Section 2.2.2 it is now possible to compute the probabilities

$$P\left(X_{1+t}^{(1)} = 0 \middle| X_1^{(1)} = 2\right) = 1 - P\left(X_{1+t}^{(1)} = 1 \middle| X_1^{(1)} = 2\right) - P\left(X_{1+t}^{(1)} = 2 \middle| X_1^{(1)} = 2\right).$$

With t=3 the latter probability is 0.023 implying that only 2.3 % of the men who started in the most acute phase 2 could be expected to become healthy during the first year. The corresponding probability based on a non-homogeneous MC of order m=2 was 0.088, which is close to the empirical value 0.091.

5. Some final remarks

This paper has considered some aspects of model building when using Markov Chains for analyzing the progression of patient's health states.

First, that the specification of time unit, state space, Markov order and how the transition probabilities change with time, are closely connected. In the example of Section 4 the time unit was chosen as quarter which resulted in a non-homogeneous MC of order 2. Other choices of time unit would have given other options.

Second, there is a problem with high order MCs since they require large samples in order to estimate all transition probabilities accurately, especially when they are non-homogeneous. Very few rehabilitation processes are in fact homogeneous and this argues for that parsimonious models with few states should be used, at least in a first step. In the example of Section 4 a significant difference between the sexes was noticed for the parameter $\alpha_t(1^2)$ but not for $\alpha_t(2,1)$. However, this was likely caused by the fact that the working sample size for estimating the former parameter was much larger. When planning a MC study of this kind one should not just focus on the total sample size *n*, but also try to get information about the magnitude of the working sample sizes.

The present study should be viewed as a first step for analyzing progression of patient's health states. In a second step one may go further and model how the transition probabilities depends on a number of covariates, such as age and diagnosis. This approach is of importance if predictions of future health are at the individual level and not just for groups. The three-state MC model with transition probabilities that are schematically illustrated in Figure 1 can be used also in other contexts, e.g. when studying transition from HIV infection to AIDS and further to death or transition from healthy to diabetes and finally to death, just to take a few examples.

Acknowledgements

The research was supported by the National Social Insurance Board in Sweden (RFV), Dnr 3124/99-UFU.

References

1. Arber, S. and Cooper, H., “Gender and inequalities in women’s health across the life course”, in *Gender Inequalities in Health* (Annandale, E. and Hunt, K., eds.), Open University Press, Buckingham, 123-149, 2000.
2. Avery, P. J. and Henderson, D. A., “Fitting Markov chain models to discrete state series such as DNA sequences”, *Applied Statistics* 48, 53-61, 1999.
3. Gay, D. A. and Wong, D. W., “Predicting rehabilitation outcomes from clinical and statistical data: a probability model”, *International Journal of Rehabilitation Research* 11, 11-19, 1988.
4. Hansson, K., *Valfardsbulletinen* 2001:2 (in Swedish), 2001.
5. Hoel, P. G., “A test for Markoff chains”, *Biometrika* 41, 430-433, 1954.
6. Hughes, J. P., Guttorp, P. and Charles, S. P., “A non-homogeneous hidden Markov model for precipitation occurrence”, *Applied Statistics* 48, 15-30, 1999.
7. Jonsson, R. “Tests of Markov Order and Homogeneity in a Markov Chain Model for Analysing Rehabilitation”, *Research Report 2011:7*, Statistical Research Unit, Department of Economics, Göteborg University, Sweden, 2011.
8. Jowett, G. H.,”The relationship between the binomial and F distributions, *The Statistician* 13, 55-57, 1963.
9. Katz, R. W., ”On some criteria for estimating order of a Markov chain, *Technometrics* 23, 243-249, 1981.
10. Lowry, W. P. and Guthrie, D., “Markov chains of order greater than one”, *Monthly Weather Review* 96, 798-801, 1968.
11. Mc Lean, S. I. and Millard, P. H., “A three compartment model of the patient flows in a geriatric department: a decision support approach”, *Health Care Management Science* 1, 159-163, 1998.
12. Nilsson, G. and Jonsson, R., “Rapport fran projekt Regionala analyser av forsaekringsutnyttjande” (in Swedish), *Forsaekringskassan i Sodermanlands län*, 2003.

13. Raftery, A. and Tavare´, S., “Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution model”, *Applied Statistics* 43, 179-199, 1994.
14. Tikhomirova, M. I. And Chistyakov, V. P., “On two statistics of chi-square type based on frequencies of tuples of states of a high-order Markov chain”, *Discrete Mathematical Applications* 13, 319-329, 2003.

Appendix

(A1) Derivation of the expressions that follows from the requirement that Markov chains of order 1 and 2 have the same one-step transition probabilities,

$P\left(X_{t+1}^{(1)}=1\middle|X_t^{(1)}=1\right)=P\left(X_{t+1}^{(2)}=1\middle|X_t^{(2)}=1\right),\ t=2,3$.

Let j_t denote the event that state *j* is occupied at time *t*. For a Markov chain with *m*=2, $P\left(1_3\middle|1_2\right)=P\left(1_2,1_3\right)/P\left(1_2\right)$, where

$P\left(1_2\right)=P\left(1_1,1_2\right)+P\left(2_1,1_2\right)=P\left(1_2\middle|1_1\right)P\left(1_1\right)+P\left(1_2\middle|2_1\right)P\left(2_1\right)=\alpha_1(1)(1-\pi_2)+(1-\beta_1)\pi_2$,

$P\left(1_2,1_3\right)=P\left(1_1,1_2,1_3\right)+P\left(2_1,1_2,1_3\right)=P\left(1_3\middle|1_1,1_2\right)P\left(1_2\middle|1_1\right)P\left(1_1\right)+P\left(1_3\middle|2_1,1_2\right)P\left(1_2\middle|2_1\right)P\left(2_1\right)$ $=\alpha_2(1^2)\alpha_1(1)(1-\pi_2)+\alpha_2(2,1)(1-\beta_1)\pi_2$. Thus,

$$P\left(X_3^{(2)}=1\middle|X_2^{(2)}=1\right)=\alpha_2(1^2)W_2+\alpha_2(2,1)\left(1-W_2\right), \tag{i}$$

$$\text{where } W_2=\frac{\alpha_1(1)(1-\pi_2)}{\alpha_1(1)(1-\pi_2)+(1-\beta_1)\pi_2}$$

In the same way it is easily shown that

$$P\left(X_4^{(2)}=1\middle|X_3^{(2)}=1\right)=\alpha_3(1^2)W_3+\alpha_3(2,1)\left(1-W_3\right), \tag{ii}$$

$$\text{where } W_3=\frac{\alpha_2(1^2)\alpha_1(1)(1-\pi_2)+\alpha_2(2,1)(1-\beta_1)\pi_2}{\alpha_2(1^2)\alpha_1(1)(1-\pi_2)+\alpha_2(2,1)(1-\beta_1)\pi_2+(1-\beta_2)\beta_1\pi_2}$$

If the two expressions in (i) and (ii) are put equal to $\alpha_2(1)$ and $\alpha_3(1)$, respectively, it is guaranteed that the one-step transition probabilities are the same when *m*=1 and *m*=2.

(A2) Derivation of the expression for the Likelihood in (5)

Let $p(x_1\ldots x_T)$ be the joint probability function of the states $x_1\ldots x_T$. Then the likelihood of a Markov chain of order *m* can be written

$$L^{(m)}=p(x_1\ldots x_T)=$$
$$p(x_1)p\left(x_2\middle|x_1\right)p\left(x_3\middle|x_1x_2\right)\ldots p\left(x_m\middle|x_1\ldots x_{m-1}\right)p\left(x_{m+1}\middle|x_1\ldots x_m\right)\ldots p\left(x_T\middle|x_{T-m}\ldots x_{T-1}\right)=$$
$$p(x_1)\prod_{t=1}^{m-1}p\left(x_{t+1}\middle|x_1\ldots x_t\right)\cdot\prod_{t=m}^{T-1}p\left(x_{t+1}\middle|x_{t-m+1}\ldots x_t\right)$$

Here $p(x_1)=\pi_2^{N_2(1)}(1-\pi_2)^{N_1(1)}$, $p\left(x_2\left|x_1\right.\right)=\beta_1^{B_1}(1-\beta_1)^{N_1(2)-B_1}\alpha_1(1)^{A_1(1)}\left[1-\alpha_1(1)\right]^{N_1(1)-A_1(1)}$, and

$p\left(x_3\left|x_1x_2\right.\right)=$

$$\beta_2^{B_2}(1-\beta_2)^{N_2(2)-B_2}\alpha_2(1^2)^{A_2(1^2)}\left[1-\alpha_2(1^2)\right]^{N_2(1^2)-A_2(1^2)}\alpha_2(2,1)^{A_2(2,1)}\left[1-\alpha_2(2,1)\right]^{N_2(2,1)-A_2(2,1)}.$$

The general form of $F^{(m)}$ and $G^{(m)}$ follows by repeating the argument.

(A3) Proof of (a)-(c) of the lemma in Section 3.2

To prove (a), notice that the probability generating function (pgf) of X is

$$E\left(z^X\right)=E_N\left(E\left(z^X\left|N\right.\right)\right)=E_N\left((zp_X+1-p_X)^N\right)=\left((zp_X+1-p_X)p_N+1-p_N\right)^n=$$
$$\left(zp_Np_X+1-p_Np_X\right)$$

and the latter is the pgf for $X\sim B\left(n,p_Np_X\right)$.

$\hat{p}_X$ in (b) is unbiased because $E(\hat{p}_X)=E_N\left(E\left(\frac{X}{N}\left|N\right.\right)\right)=E_N\left(\frac{Np_X}{N}\right)=p_X$. The variance of $\hat{p}_X$ is

$$E_N\left(V\left(\hat{p}_X\left|N\right.\right)\right)+V_N\left(E\left(\hat{p}_X\left|N\right.\right)\right)=E_N\left(\frac{p_X(1-p_X)}{N}\right)+V_N\left(p_X\right)=p_X(1-p_X)E(N^{-1})+0.$$

Finally, to show that $\hat{V}(\hat{p}_X)$ is unbiased for $V(\hat{p}_X)$, consider $E\left(\frac{\hat{p}_X(1-\hat{p}_X)}{N-1}\right)=$

$$=E_N\left(E\left(\frac{X}{N(N-1)}-\frac{X^2}{N^2(N-1)}\left|N\right.\right)\right)=E_N\left(\frac{Np_X}{N(N-1)}-\frac{(Np_X(1-p_X)+N^2p_X^2)}{N^2(N-1)}\right)=$$

$$=p_X(1-p_X)E\left(N^{-1}\right)=V(\hat{p}_X).$$

(c) is proved in the same way as (b). The suggested estimator of $V\left(\hat{p}_{X_1}-\hat{p}_{X_2}\right)$ presupposes that N_1 and N_2 are so large that $\frac{N_1+N_2-1}{N_1+N_2}\approx 1$.

(A4) Proof of the proposition about the difference between the estimates in Section 3.3.2

The ML estimator of $\alpha_t(1^m)$ is

$$\frac{A_t(1^m)}{N_t(1^m)}=\frac{A_t(1^{m+1})+A_t(2,1^m)}{N_t(1^{m+1})+N_t(2,1^m)}=\frac{\hat{\alpha}_t(1^{m+1})N_t(1^{m+1})+\hat{\alpha}_t(2,1^m)N_t(2,1^m)}{N_t(1^{m+1})+N_t(2,1^m)},$$

and from this it follows that

$$\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(1^m) = [\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2,1^m)] \cdot \frac{N_t(2,1^m)}{N_t(1^m)}$$

In the same way it is easily shown that

$$\hat{\alpha}_t(2,1^m) - \hat{\alpha}_t(1^m) = [\hat{\alpha}_t(2,1^m) - \hat{\alpha}_t(1^{m+1})] \cdot \frac{N_t(1^{m+1})}{N_t(1^m)}$$

Research Report

2007:1	Andersson, E.:	Effect of dependency in systems for multivariate surveillance.
2007:2	Frisén, M.:	Optimal Sequential Surveillance for Finance, Public Health and other areas.
2007:3	Bock, D.:	Consequences of using the probability of a false alarm as the false alarm measure.
2007:4	Frisén, M.:	Principles for Multivariate Surveillance.
2007:5	Andersson, E., Bock, D. & Frisén, M.:	Modeling influenza incidence for the purpose of on-line monitoring.
2007:6	Bock, D., Andersson, E. & Frisén, M.:	Statistical Surveillance of Epidemics: Peak Detection of Influenza in Sweden.
2007:7	Andersson, E., Kuhlmann, S., Linde., A &Frisén, M.:	Predictions by early indicators of the progress of the influenza in Sweden.
2007:8	Bock, D., Andersson, E. & Frisén, M.:	Similarities and differences between statistical surveillance and certain decision rules in finance.
2007:9	Bock, D.:	Evaluations of likelihood based surveillance of volatility.
2007:10	Bock, D. & Pettersson, K.:	Explorative analysis of spatial aspects on the Swedish influenza data.
2007:11	Frisén, M. & Andersson, E.:	On-line detection of outbreaks.
2007:12	Frisén, M., Andersson, E. & Schiöler, L.:	A non-parametric system for on-line outbreak detection of epidemics.
2007:13	Frisén, M., Andersson, E. & Pettersson, K.:	Estimation of outbreak regression.
2007:14	Pettersson, K.:	Unimodal regression in the two-parameter exponential family with constant dispersion parameter.
2007:15		

Research Report

2008:1	Frisén, M.	Introduction to financial surveillance.
2008:2	Jonsson, R.	When does Heckman's two-step procedure for censored data work and when does it not?
2008:3	Andersson, E.	Hotelling´s T2 Method in Multivariate On-Line Surveillance. On the Delay of an Alarm.
2008:4	Schiöler, L. & Frisén, M.	On statistical surveillance of the performance of fund managers.
2008:5	Schiöler, L.	Explorative analysis of spatial patterns of influenza incidences in Sweden 1999—2008.
2008:6	Schiöler, L.	Aspects of Surveillance of Outbreaks.
2008:7	Andersson, E & Frisén, M.	Statistiska varningssystem för hälsorisker
2009:1	Frisén, M., Andersson, E. & Schiöler, L.	Evaluation of Multivariate Surveillance
2009:2	Frisén, M., Andersson, E. & Schiöler, L.	Sufficient Reduction in Multivariate Surveillance
2010:1	Schiöler, L	Modelling the spatial patterns of influenza incidence in Sweden
2010:2	Schiöler, L. & Frisén, M.	Multivariate outbreak detection
2010:3	Jonsson, R.	Relative Efficiency of a Quantile Method for Estimating Parameters in Censored Two-Parameter Weibull Distributions
2010:4	Jonsson, R.	A CUSUM procedure for detection of outbreaks in Poisson distributed medical health events
2011:1	Jonsson, R.	Simple conservative confidence intervals for comparing matched proportions
2011:2	Frisén, M	On multivariate control charts
2011:3	Frisén, M	Methods and evaluations for surveillance in industry, business, finance, and public health
2011:4	Knoth, S &Frisén, M	Minimax Optimality of CUSUM for an Autoregressive Model
2011:5	Marianne Frisén	Inference principles for multivariate surveillance