Statistica Sinica **10**(2000), 999-1010

ON THE EXACT DISTRIBUTION OF SECON AND ITS APPLICATION

James C. Fu and W. Y. Wendy Lou

University of Manitoba and Mount Sinai School of Medicine

Abstract: The index for SEquential CONtinuity of care (*SECON*, Steinwachs (1979)) can be defined as the average of a sequence of random variables $\{Y_t\}$ which measure the sequential continuity of stationary Markov-dependent m-state trials $\{X_t\}$, where Y_t is defined as 1 if $X_{t-1} = X_t$ and as 0 otherwise. In the health care sector, *SECON* is usually applied as the fraction of sequential patient-visit pairs at which the same provider was seen, and represents the standard estimate of the sequential nature of continuity of care, an important health policy aim that drives many of the changes underway in the current US health care market. After almost two decades of application, however, the exact distribution of *SECON* is still unknown except for the case where the X_t are i.i.d. with equal probabilities for each state. In this article, the distribution problem is cast into a finite Markov chain setting via the imbedding technique developed by Fu and Koutras (1994), and the exact probabilities under one-step Markov dependence can be obtained either directly or via recursive equations. It is also shown that *SECON* is the minimum variance unbiased estimator, and the maximum likelihood estimator, for the sequential continuity measure. Numerical and real-data examples are given to illustrate the theoretical results.

Key words and phrases: Ergodic distribution, Markov dependence, sequential continuity.

1. Introduction

Continuity of care, and in particular provider continuity, is increasingly a central aim of health care policy in a majority of clinical settings. The benefits of broader information exchange between doctor and patient attributed to provider continuity include earlier recognition of health problems and psycho-social effects such as greater patient satisfaction, leading to an overall improvement in the quality of care as well as a reduction in cost.

Continuity of care describes the extent to which information about the diagnosis and management of health problems is conveyed from one visit to the next. This definition, in its broadest sense, includes not only provider (*e.g.* physician, health team, HMO) continuity, but also other dimensions such as continuity of medical records and of geographical location of treatment site (Wall

(1981)). Although existing measures for the quantification of continuity can be applied to each of these dimensions in a similar manner, corresponding literature has focused mainly on provider continuity, which is expected to have the greatest impact upon treatment outcome. Over the past two decades, more than a dozen measurement indices of continuity have been proposed and, in general, they can be categorized into visit-based (*e.g.* Eriksson and Mattsson (1983)) and individual-based measures (*e.g.* Breslau and Reeb (1975) and Steinwachs (1979)). Previous literature has focused mainly upon the latter, and this article will examine one of the most commonly-used individual-based statistics: the sequential continuity measure $SECON$ (Steinwachs (1979)), which is the fraction of sequential visit pairs at which the same provider is seen. The $SECON$ statistic is a basic yet non-trivial measure of sequential continuity in a sequence of observations, and although its application has thus far been confined to the field of health care and policy research, an improved understanding of its theoretical properties might lead to applications in other areas as well.

The exact distribution of $SECON$ under the assumption of random assignment (i.i.d. with equal probabilities) at each visit was given by Steinwachs (1979). However, under the more realistic model of dependence between visits, the exact distribution of $SECON$ has never been investigated, and traditional combinatorial approaches, such as the urn model used by Eriksson (1990) to study random assignment, cannot easily be extended to non-random cases. In this article, the exact distribution of $SECON$ will be derived based on the method of finite Markov chain imbedding (see Fu (1996), Fu and Koutras (1994) and Lou (1996)), under the assumption of random assignment as well as under one-step Markov dependence. This dependence will be allowed to vary between providers, and hence these distributions will effectively capture variations in provider accessibility.

Let $\{X_t, t = 0, 1, \ldots\}$ be a stationary sequence of random variables forming a homogeneous Markov chain on a finite state space $S = \{1, \ldots, m\}$. Assume the chain is irreducible and aperiodic with transition probability matrix $\boldsymbol{A} = (p_{ij})_{m\times m}$. Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ be the ergodic distribution associated with the Markov chain, *i.e.*, for $i = 1, \ldots, m$, $\pi_i = \lim_{n\to\infty} P(X_n = i)$. Define a parameter $\theta = \sum_{i=1}^{m} \pi_i p_{ii}$, the sequential continuity measure associated with the Markov chain, and define a sequence of index random variables Y_t such that $Y_t = 1$ if $X_t = X_{t-1}$, $t = 1, 2, \ldots$ Thus $Y_t = 1$ if the same health provider is seen at consecutive visits, and zero otherwise, where the random variable X_0 corresponds to the outcome of the visit of the patient at time $t = 0$ with initial probability $\boldsymbol{\pi}_0 = (P(X_0 = 1), \ldots, P(X_0 = m))$. Throughout, we assume that the initial probability $\boldsymbol{\pi}_0$ is the ergodic distribution $\boldsymbol{\pi}$ of the Markov chain, unless otherwise specified.

Let $v(T)$, a positive integer random variable defined on $J^+ = \{1, 2, \ldots\}$ with probability mass function $\mu(\cdot)$, be the number of visits, excluding the initial

visit, that have occurred up to time T. Steinwachs (1979) introduced the statistic $SECON$ as the average of the sequence $\{Y_t\}_{t=1}^{v(T)}$ for a given $v(T) = n$, or in a practical sense, as the fraction of sequential visit pairs at which the same health provider is seen. In our context, $SECON$ estimaies θ, with

$$SECON = S_{v(T)}/v(T) \tag{1}$$

and $S_{v(T)} = Y_1 + Y_2 + \cdots + Y_{v(T)}$. If the integer random variable $v(T)$ is independent of the Markov chain $\{X_t\}$, and the initial distribution $\boldsymbol{\pi}_0$ is the ergodic distribution $\boldsymbol{\pi}$ of the Markov chain, then $SECON$ is an unbiased estimator for θ. This follows from the fact that for every given $v(T) = n$ with $n \in J^+$, $E(SECON|v(T) = n) = E(S_n/n) = \theta$.

It is easy to see that the sequence of random variables $\{Y_t\}$ induced by the Markov chain $\{X_t\}$ is stationary but not always a Markov chain. Hence the usual Markov chain techniques cannot be applied directly to study the exact distributions of $S_{v(T)}$ and $SECON$. This is probably the foremost reason why their exact distributions are still unknown for non-random cases. The goal of this article is to find the exact distributions of $S_{v(T)}$ and $SECON$ under one-step Markov dependence between two consecutive events/visits, and the main results, based on the finite Markov chain imbedding technique, are presented in Section 2. In Section 3, we show that $SECON$ is the minimum variance unbiased estimator, and also the maximum likelihood estimator, for the sequential continuity measure θ. Two numerical examples and one data set from the Mount Sinai AIDS Center are analyzed in Section 4, followed by some remarks in the concluding section.

2. The Exact Distribution

For a given positive integer n, let $\boldsymbol{\Gamma}_n = \{0, \ldots, n\}$. Consider the homogeneous Markov chain $\{X_t : t \in \boldsymbol{\Gamma}_n\}$ defined in Section 1 with transition probability matrix $\boldsymbol{A}$, and decompose $\boldsymbol{A}$ into two matrices $\boldsymbol{B}$ and $\boldsymbol{D}$: $\boldsymbol{A}_{m\times m} = (p_{ij})_{m\times m} = \boldsymbol{B}_{m\times m} + \boldsymbol{D}_{m\times m}$, where

$$\boldsymbol{B}_{m\times m} = \begin{pmatrix} 0 & & p_{ij} \\ & \ddots & \\ p_{ij} & & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{D}_{m\times m} = \begin{pmatrix} p_{11} & & 0 \\ & \ddots & \\ 0 & & p_{mm} \end{pmatrix}.$$

Let $\boldsymbol{\Omega}_n = \{(u, v) : u = 0, \ldots, n, \text{ and } v = 1, \ldots, m\}$ be the state space containing a total of $(n+1)m$ states. Define a homogeneous Markov chain $\{Z_t : t \in \boldsymbol{\Gamma}_n\}$ on $\boldsymbol{\Omega}_n$ as

$$Z_t(X) = \begin{cases} (\sum_{i=1}^{t} Y_i,\ X_t), & 1 \le t \le n, \\ (0, X_0), & t = 0, \end{cases} \tag{2}$$

with transition probability matrices for, $t = 1, \ldots, n$,

$$\boldsymbol{M}_t = \boldsymbol{M}_{(n+1)m\times(n+1)m} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{D} & \boldsymbol{O} & \cdots & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B} & \boldsymbol{D} & \cdots & \boldsymbol{O} & \boldsymbol{O} \\ & \cdots & \ddots & & \cdots & \\ \boldsymbol{O} & & \cdots & & \boldsymbol{B} & \boldsymbol{D} \\ \boldsymbol{O} & & \cdots & & \boldsymbol{O} & \boldsymbol{B}^\star \end{pmatrix}, \tag{3}$$

where $\boldsymbol{O}$ and $\boldsymbol{B}^\star$ are the $m \times m$ zero and identity matrices, respectively.

Theorem 2.1. (a) *The distribution of S_n is given as*

$$P(S_n = s) = \boldsymbol{\xi}(\prod_{t=1}^{n} \boldsymbol{M}_t)\boldsymbol{U}'(C_s) = \boldsymbol{\xi}\boldsymbol{M}^n\boldsymbol{U}'(C_s), \;\; 0 \le s \le n, \tag{4}$$

where $\boldsymbol{\xi} = (\boldsymbol{\pi}, \mathbf{0}, \ldots, \mathbf{0}, \ldots, \mathbf{0})_{1\times(n+1)m}$ is the initial probability vector of Z_0, $\mathbf{0} = (0, \ldots, 0)_{1\times m}$, $\boldsymbol{U}(C_s) = (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)$ is a $1 \times (n+1)m$ row vector with 1 at the coordinates associated with states in C_s and 0 elsewhere, and the $C_s = [(s, 1), \ldots, (s, m)], s = 0, \ldots, n$, form a partition of $\boldsymbol{\Omega}_n$.

(b) *The probability that $S_n = s$ satisfies the following recursive equation:*

$$P(S_n = s) = P(S_{n-1} = s) + \sum_{j=1}^{m} p_{jj}\boldsymbol{\xi}\boldsymbol{M}^{n-1}(\boldsymbol{U}'(s-1, j) - \boldsymbol{U}'(s, j)), \tag{5}$$

where $\boldsymbol{U}(s, j) = (0, \ldots, 0, 1, 0, \ldots, 0)_{1\times(n+1)m}$ is a unit vector associated with the state (s, j).

Proof. Since $\{X_t\}$ is a Markov chain with initial probability $\boldsymbol{\pi}$ and transition probability matrix $\boldsymbol{A}$, it follows from (2) that $Z_t(X)$ is also a Markov chain with initial probability $\boldsymbol{\xi}$ and transition probabilities determined by the following: for $0 \le s \le n-1$ and $i = 1, \ldots, m$,

$$P(Z_t(X) = (u, v)|Z_{t-1}(X) = (s, i)) = \begin{cases} p_{ii}, \text{ if } u = s+1 \text{ and } v = i, \\ p_{ij}, \text{ if } u = s \text{ and } v = j, \; i \ne j, \\ 0, \;\; \text{otherwise.} \end{cases} \tag{6}$$

For states with $s = n$, by convention, $P(Z_t(X) = (u, v)|Z_{t-1}(X) = (n, i)) = 1$ if $u = n$ and $v = i$, and 0 otherwise. The transition probability matrices $\boldsymbol{M}_t = \boldsymbol{M}$ given at (3) are hence defined. From the definition of $\{Z_t\}$, $P(S_n = s) = P(Z_n \in C_s)$, so (4) is a direct consequence of the Chapman-Kolmogorov equation. The recursive Equation (5) follows immediately from (4) and

$$\boldsymbol{M}\boldsymbol{U}'(C_s) = \boldsymbol{U}'(C_s) + \sum_{j=1}^{m} p_{jj}(\boldsymbol{U}'(s-1, j) - \boldsymbol{U}'(s, j)).$$

Note that $\boldsymbol{BD} = \boldsymbol{DB}$ if and only if $p_{11} = p_{22} = \cdots = p_{mm} = \theta > 0$. Given $v(T) = n$, if $\boldsymbol{BD} = \boldsymbol{DB}$, it follows from (4) and

$$\boldsymbol{\xi M}^n = \left[\binom{n}{0}\boldsymbol{\pi B}^n, \binom{n}{1}\boldsymbol{\pi B}^{n-1}\boldsymbol{D}, \ldots, \binom{n}{n}\boldsymbol{\pi D}^n\right] \tag{7}$$

that S_n has a binomial distribution

$$P(S_n = s) = \binom{n}{s}\boldsymbol{\pi B}^{n-s}\boldsymbol{D}^s\mathbf{1}' = \binom{n}{s}\theta^s(1-\theta)^{n-s}. \tag{8}$$

Thus in this special case, the result is independent of the number of available providers m and the initial probability $\boldsymbol{\pi}$.

The Markov chain $Z_t(X)$ defined at (2) can be regarded as a Markov random walk (Pyke (1961)). Our construction of $\{Z_t\}$, Theorem 2.1, and Corollary 2.1 show that S_n is finite Markov chain imbeddable in the sense of Fu and Koutras (1994). In addition, if $\boldsymbol{BD} = \boldsymbol{DB}$, S_n is binomial finite Markov chain imbeddable in the sense of Koutras and Alexandrou (1995). If $v(T)$ has a probability measure $\mu(\cdot)$ on J^+ and is independent of $\{X_t\}$, then the distributions of $S_{v(T)}$ and $SECON$ can be specified as follows.

Corollary 2.1. *If $v(T)$ and $\{X_t\}$ are independent, then*

$$P(S_{v(T)} = s) = \sum_{n=s}^{\infty} \boldsymbol{\xi M}^n\boldsymbol{U}'(C_s)\mu(n); \tag{9}$$

if $p_{11} = \cdots = p_{mm} = \theta$ and $\alpha \in [0,1]$, then

$$P(S_{v(T)}/v(T) \le \alpha) = \sum_{n=1}^{\infty}\sum_{s=0}^{[\alpha n]}\binom{n}{s}\theta^s(1-\theta)^{n-s}\mu(n), \tag{10}$$

where $[\alpha n]$ denotes the integer part of αn.

Proof. Since $v(T)$ and $\{X_t\}$ are independent, (9) follows immediately from Theorem 2.1 and $P(S_{v(T)} = s) = \sum_{n=s}^{\infty} P(S_n = s|v(T) = n)\mu(n)$. If $p_{11} = \cdots = p_{mm} = \theta$, (10) follows directly from (8) and (9).

If $v(T)$ is independent of $\{X_t\}$ and has, for example, a truncated Poisson distribution with parameter λ such that $\mu(n) = \lambda^n e^{-\lambda}/(n!(1-e^{-\lambda}))$, $n \in J^+$, then for cases where $p_{11} = \cdots = p_{mm} = \theta$, the $SECON$ statistic has the following distribution: for $\alpha \in [0,1]$,

$$P(S_{v(T)}/v(T) \le \alpha) = \frac{e^{-\lambda}}{1-e^{-\lambda}}\sum_{n=1}^{\infty}\frac{\lambda^n}{n!}\sum_{s=0}^{[\alpha n]}\binom{n}{s}\theta^s(1-\theta)^{n-s}. \tag{11}$$

Further, under the above assumptions, the $SECON$ statistic has mean θ and variance

$$\mathrm{Var}\,(S_{v(T)}/v(T)) = \frac{\theta(1-\theta)e^{-\lambda}}{1-e^{-\lambda}} \sum_{n=1}^{\infty} \frac{\lambda^n}{n(n!)}. \tag{12}$$

The assumption of independence between the number of visits $v(T)$ and the visit sequence $\{X_t\}$ is indispensable for the above results. It seems to us that if $v(T)$ is a stopping random variable, then under certain conditions our formulation can be extended to the case of dependence between $v(T)$ and $\{X_t\}$, and the limiting distribution of $SECON$ can then be obtained with some modifications by using the method of Fuh (1997).

3. Estimation of Sequential Continuity Measure

If $v(T)$ and $\{X_t\}$ are independent, $SECON$ is an unbiased estimator for the sequential continuity measure parameter $\theta = \sum_{i=1}^{m} \pi_i p_{ii}$. It is easy to see that $\theta = 0$ if and only if $p_{ii} = 0$ for all $i = 1, \ldots, m$. If $\{X_t\}$ is a sequence of i.i.d. m-state trials with probabilities $p_{ij} = p_j > 0$, then the following theorem provides a lower bound for θ.

Theorem 3.1. *Consider a sequence of i.i.d. multi-state trials* $\{X_t = i : t \in \Gamma_n, i = 1, \ldots, m\}$. *The parameter* θ *satisfies* $\theta(p_1, \ldots, p_m) \geq 1/m$ *for all* $(p_1, \ldots, p_m)$, *with equality if and only if* $p_1 = \cdots = p_m = 1/m$.

In other words, random assignment with equal probabilities will yield the smallest value of θ amongst all i.i.d. sequences of multi-state trials.

Proof. Since the X_t are i.i.d., the ergodic distribution is $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ with $p_i = \pi_i$ and $\theta = \sum_{i=1}^{m} p_i^2$, and hence the theorem is an immediate consequence of Lagrange's method.

The transition probabilities p_{ij} are often unknown in real situations, and the estimation problem has been studied by many researchers. The likelihood approach is adopted here. Given $v(T) = n$, let $N_{ij}(n)$, $i, j, = 1, \ldots, m$, be the total number of X_t such that $X_{t-1} = i$ and $X_t = j$, $t = 1, \ldots, n$. Assuming that the distribution of $v(T)$ does not depend on the parameters p_{ij} and that $v(T)$ is independent of the sequence $\{X_t\}$, the conditional likelihood function of p_{ij}, given $v(T) = n$ and $\{x_0, \ldots, x_n\}$, can be written as

$$L(p_{ij}|x_0, \ldots, x_n) \quad \propto \quad \boldsymbol{\pi}(x_0) \prod_{i,j=1}^{m} p_{ij}^{N_{ij}(n)}. \tag{13}$$

Theorem 3.2. *If the initial probability is the ergodic distribution of the Markov chain* $\{X_t\}$, *and* $v(T)$ *and* $\{X_t\}$ *are independent, then* $SECON$ *is the minimum variance unbiased estimator, and also the maximum likelihood estimator, for* θ.

Proof. Since $\{X_t\}$ and $v(T)$ are independent and, for every given $v(T) = n$, the $N_{ij}(n)$ are sufficient and complete statistics for p_{ij}, $SECON$ is the MVUE for θ since for every $v(T) = n$, $n = 1, 2, \ldots$,

$$SECON = \sum_{i=1}^{m} N_{ii}(n)/n, \tag{14}$$

and one can appeal to the the Rao-Blackwell Theorem. It follows directly from the likelihood function (13) that $SECON$ is also the maximum likelihood estimator.

In fact $SECON$ becomes intuitively obvious as an estimator for $\theta = \sum_{i=1}^{m} \pi_i p_{ii}$ once it is rewritten as

$$\begin{aligned}\hat{\theta}_{v(T)} = SECON &= \sum_{i=1}^{m} N_{ii}(v(T))/v(T) = \sum_{i=1}^{m} [N_i(v(T))/v(T)]\,[N_{ii}(v(T))/N_i(v(T))] \\ &= \sum_{i=1}^{m} \hat{\pi}_i \hat{p}_{ii},\end{aligned} \tag{15}$$

where $N_i(v(T)) = \sum_{j=1}^{m} N_{ij}(v(T))$.

4. Examples

To illustrate our method for finding the distributions of S_n and $SECON$, two numerical examples are given. A real-data example is also given to demonstrate practical application.

Example 4.1. Let $\{X_t\}$ be a two-state homogeneous Markov chain with transition matrix

$$\boldsymbol{A}_{2\times 2} = \begin{bmatrix} 0.3 & 0.7 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} 0 & 0.7 \\ 0.3 & 0 \end{bmatrix} + \begin{bmatrix} 0.3 & 0 \\ 0 & 0.7 \end{bmatrix} = \boldsymbol{B} + \boldsymbol{D}. \tag{16}$$

The distribution of S_n can be obtained via Theorem 2.1. The numerical results for $n = 10$ and $n = 100$ with initial probabilities of $\boldsymbol{\pi}_0 = (0.5, 0.5)$, (0.3,0.7), and (0.8,0.2), are presented in Figure 1, along with the exact means and variances. It is easy to see that, given $v(T) = n$, the mean of $SECON$ is θ when the initial probability $\boldsymbol{\pi}_0$ is the same as the ergodic probability of the Markov chain $\{X_t\}$, otherwise the bias depends on the initial probability, especially when n is small.

Example 4.2. Let $\{X_t\}$ be a homogeneous m-state Markov chain having symmetric transition matrix:

$$\boldsymbol{A}^{*}_{m\times m} = \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \end{array} \begin{bmatrix} & \beta & & & \\ & & \beta & & (1-\beta)/(m-1) \\ (1-\beta)/(m-1) & & \ddots & & \\ & & & & \beta \\ & & & & \end{bmatrix}. \tag{17}$$

Since $\boldsymbol{A}^*$ contains only one parameter β $(0 \le \beta \le 1)$, and $p_{11} = \cdots = p_{mm} = \beta$, the distribution of S_n can be easily obtained via Corollary 2.1. The interpretation of this model is that the probability of a patient seeing the same provider at consecutive visits is β, with probability $(1-\beta)/(m-1)$ of seeing any of the other $(m-1)$ providers. It is easy to see that the above Markov chain has the ergodic distribution $\boldsymbol{\pi} = (1/m, \ldots, 1/m)$ for all $0 < \beta < 1$, and that the parameter β is equal to the sequential continuity measure θ. Further, if $\beta = 1/m$, the model reduces to random assignment of patient to provider at each visit.

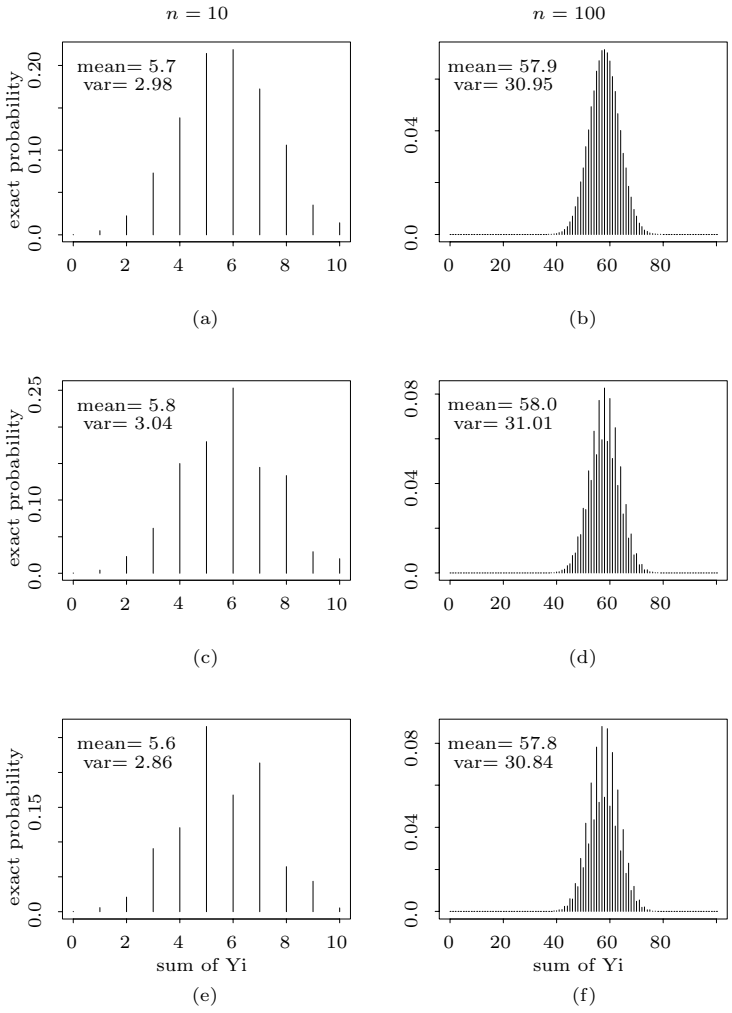

Figure 1. The exact probability of $S_n = \sum_{i=1}^{n} Y_i$ for Example 4.1: $\pi_0 = (0.5, 0.5)$ for (a) and (b), $\pi_0 = (0.3, 0.7)$ for (c) and (d), and $\pi_0 = (0.8, 0.2)$ for (e) and (f).

Some numerical results are plotted in Figure 2 for $\theta = \beta = 0.5$ and 0.8, where $v(T)$ has a truncated Poisson distribution (excluding $n = 0$) with $\lambda = 4$, 15 or 30. For fixed θ, the distribution functions become more alike as the Poisson parameter λ increases, and they have larger probability increments with α when α is near θ. Note that, for this special case, the distribution of $S_{v(T)}/v(T)$ is independent of the number of available providers m and the initial probability $\boldsymbol{\pi}_0$. In practice, this is a very useful feature when comparing the sequential continuity of care across different health care units where the number of available providers varies.

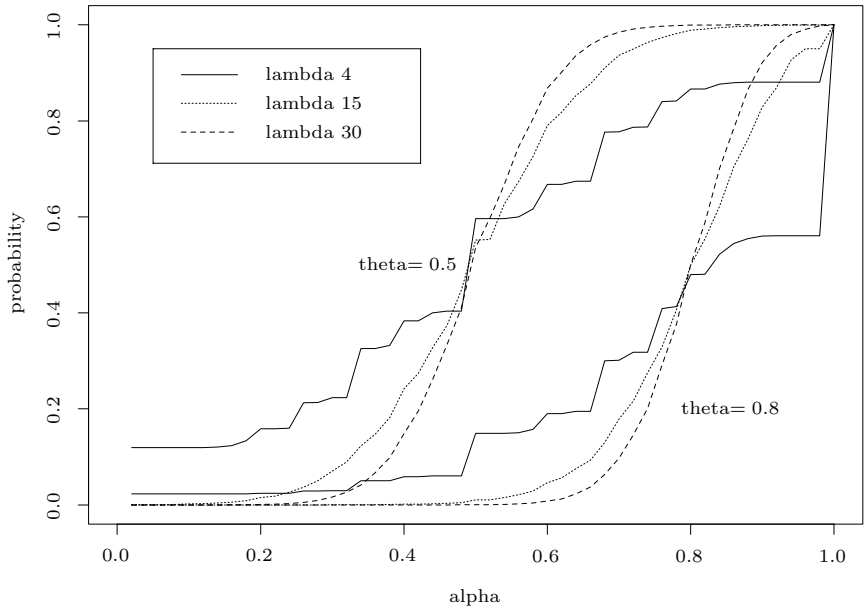

Figure 2. The distribution of $S_{v(T)}/v(T)$ when $\theta = 0.5$ and $\theta = 0.8$, where $v(T)$ follows the truncated Poisson distribution with $\lambda = 4, 15$ or 30.

To demonstrate the application of the sequential continuity measure, one subset of data from the Mount Sinai AIDS Center is analyzed below.

As part of a study of HIV/AIDS patients we investigated the relationship between continuity of care and overall outcomes, such as total charges and frequency of hospitalization, for patients from the Mount Sinai AIDS Center. This patient population is composed mainly of East and Central Harlem residents. These low-income residential communities with high illicit drug use have among the highest AIDS and tuberculosis prevalences of any community in the country, and it is important to quantify the patterns of care received by patients in such areas.

As an illustration here, we selected one set of 45 patients whose primary provider is Medicaid, whose ages are between 25 and 55, and who had regular visits since the initial date of receiving case management care during the period of July 1995 to June 1996. Among the 45 observed sequences, 24 had $SECON$ values equal to one, and the minimum value was 0.3. Hence these patients received much better sequential continuity of care than if they had been randomly assigned to any one of the $m = 13$ providers at each visit, in which case θ would be 1/13. To gain further insight into the measure θ for these 45 patients, the empirical distribution of $SECON$ is plotted in Figure 3, along with the theoretical distribution of $SECON$, where $\theta = 0.8$ and the number of visits $v(T)$ is assumed to follow a Poisson distribution with $\lambda = 8$. Here, as a reasonable starting point, we naively assumed the symmetric transition probability relationship given by (17) for the 13-provider Markov chain, resulting in the small discrepancies apparent in Figure 3. Note that both distributions are heavily skewed toward the left. A chi-square test with four intervals, each containing at least five observations, was carried out for goodness-of-fit of the theoretical distribution of $SECON$ with parameters $\theta = 0.8$ and $\lambda = 8$, yielding a p-value of 0.09. It's worth mentioning that when the restriction of having at least five observations in each interval is removed, as in using the interval alpha=[0, 0.34] which contains two observations, the p-value decreases to less than 0.05. The requirement of at least five observations in each interval was chosen to eliminate potential problems in the chi-square approximation due to extremely small expected values. Based on Figure 3 and the chi-square goodness-of-fit test, it appears that the theoretical distribution approximates the empirical distribution well in the center, and less well in the two tails. With proper estimation of the transition probabilities, we can then model the relationship between the distribution of $S_{v(T)}/v(T)$ and the characteristics of the patient population (health care provider, race, *etc.*), and further with overall outcomes such as total costs and quality of care.

5. Remarks

Under the one-parameter model proposed in (17), the distribution of $SECON$ is independent of the number of available providers m. This model is a very useful starting point for comparing sequential continuity of care across health care units with different numbers of providers. In fact, as long as the diagonal transition probabilities are equal, regardless of the off-diagonal transition probabilities, this advantage is still realized.

Much of the previous discussion has required that the number of visits $v(T)$ be independent of the visit sequence $\{X_t\}$. This is often a mild condition from a practical point of view. If all doctors in a study provide roughly the same quality of care and the patients' care-seeking is based solely on their own well-being and

not on their relationship to a specific provider, then it is quite reasonable to assume that the frequency of visits is not related to which provider is seen. In reality, of course, while there are factors which could possibly inter-relate the number of visits and the visit sequence, we expect this dependence to be weak in most cases, and feel that this assumption of independence is a reasonable starting point in the health care sector. An extension of our present formulation to more general cases of dependence between the number of visits and the visit sequence is very challenging from both a theoretical as well as a practical point of view, and is a topic of future research.

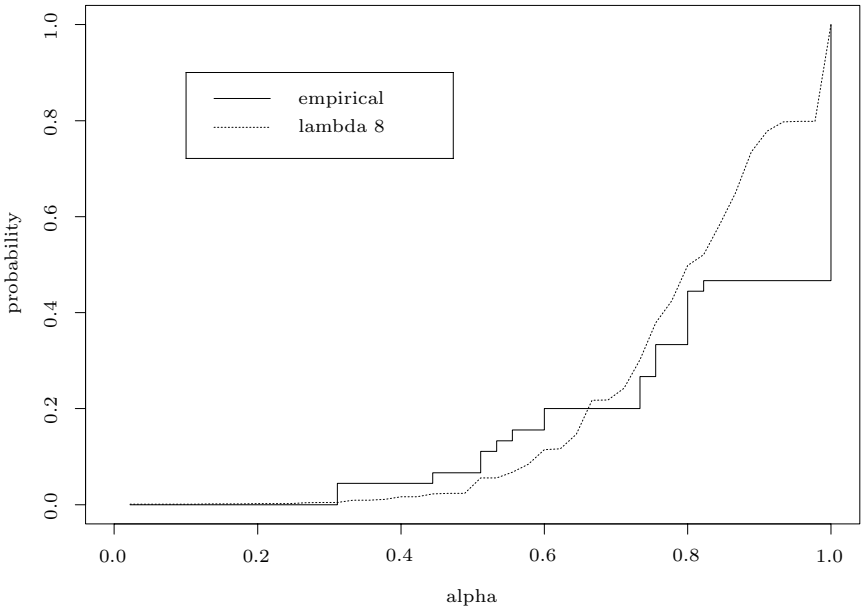

Figure 3. The empirical distribution of $S_{v(T)}/v(T)$ for the 45 patients, along with the exact distribution of $S_{v(T)}/v(T)$ where $\theta = 0.8$ and $v(T)$ follows a truncated Poisson distribution with $\lambda = 8$.

In practice, the ergodic distribution $\boldsymbol{\pi}$ is often unknown. Strictly speaking, assigning new patients to doctors based on the ergodic distribution cannot be implemented. The numerical examples depicted in Figure 1 show that the initial distribution has minimal effect on the mean and variance of $SECON$. From a practical point of view, we feel that the initial distribution should not be an issue and, with the current state of medical record information systems, the rule of assigning new patients to doctors based on past medical records provides a good and robust procedure.

In summary, the exact distribution of the $SECON$ statistic can be obtained under the assumption of independence between $v(T)$ and $\{X_t\}$, for any arbitrary sequence of stationary one-step Markov-dependent m-state trials, and the one-parameter model at (17) is a reasonable starting point in the analysis of sequential continuity data.

Acknowledgements

This work was supported in part by the NATO Collaborative Research Grants Programme CRG-970278, the National Science and Engineering Research Council of Canada Grant A-9216, and the Agency for Health Care Policy and Research Grant R01 HS09474. The authors thank Dr. M. M. Zen of the National Cheng Kung University, Taiwan for introducing them to this research topic, and also thank the referees for their helpful comments.

References

Breslau, N. and Reeb, K. (1975). Continuity of care in a university-based practice. *J. Medical Education* **50**, 965-969.

Eriksson, E. A. (1990). Continuity-of-care measures: random assignment of patients to providers and the impact of utilization level. *Medical Care* **28**, 180-190.

Eriksson, E. A. and Mattsson, L. G. (1983). Quantitative measurement of continuity of care. *Medical Care* **21**, 858-875.

Fu, J. C. (1996). Distribution theory of runs and patterns associated with a sequence of multi-state trials. *Statist. Sinica* **6**, 957-974.

Fu, J. C. and Koutras, M. V. (1994). Distribution theory of runs: a Markov chain approach. *J. Amer. Statist. Assoc.* **89**, 1050-1058.

Fuh, C. D. (1997). Corrected diffusion approximations for ruin probabilities in a Markov random walk. *Adv. Appl. Probab.* **29**, 695-712.

Koutras, M. V. and Alexandrou, V. A. (1995). Runs, scans, and urn model distributions: a unified Markov chain approach. *Ann. Inst. Statist. Math.* **47**, 743-766.

Lou, W. Y. W. (1996). On runs and longest run tests: a method of finite Markov chain imbedding. *J. Amer. Statist. Assoc.* **91**, 1595-1601.

Pyke, R. (1961). Markov renewal processes: definitions and preliminary properties. *Ann. Math. Statist.* **32**, 1231-1242.

Steinwachs, D. M. (1979). Measuring provider continuity in ambulatory care. *Medical Care* **17**, 551-565.

Wall, E. M. (1981). Continuity of care and family medicine: definition, determinants, and relationship to outcome. *J. Family Practice* **13**, 655-664.

Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.
E-mail: fu@ccu.umanitoba.ca

Department of Biomathematical Sciences, Mount Sinai School of Medicine, New York University, New York, NY 10029, U.S.A.
E-mail: lou@stat.biomath.mssm.edu

(Received March 1998; accepted October 1999)