Contingency tables with prescribed marginals

Robert Aebi

For an adjustment of contingency tables to prescribed marginal frequencies Deming and Stephan (1940) minimize a Chi-square expression. Asymptotically equivalently, Ireland and Kullback (1968) minimize a Leibler-Kullback divergence, where the probabilistical arguments for both methods remain vague.

Here we deduce a probabilistical model based on observed contingency tables. It shows that the two above methods and the maximum likelihood approach in Smith (1947) yield asymptotically the 'most probable' adjustment under prescribed marginal frequencies.

The fundamental hypothesis of statistical mechanics relates observations to 'most probable' realizations. 'Most probable' is going to be used in the sense of so-called large deviations. The proposed adjustment has a significant product form and will be generalized to contingency tables with infinitely many cells.

1 Introduction

Let us consider the problem of adjusting contingency tables to prescribed marginal frequencies introduced by Deming and Stephan (1940). Their method is to minimize a Chi-square expression. Alternatively, Ireland and Kullback (1968) propose to minimize a Kullback-Leibler divergence, i.e., a relative entropy, where the two resulting adjustments coincide asymptotically as the sample size tends to infinity.

This paper intends to show that these methods are optimal from a probabilistical point of view. Actually, in a universe of infinitely many independent individuals the Ireland-Kullback and asymptotically the Deming-Stephan adjustment as well as the maximum likelihood method in Smith (1947) yield the relatively to the observed sample 'most probable' contingency table with relative marginal frequencies prescribed by the universe. An overview of developments in the adjustment of contingency tables and bivariate distributions to prescribed marginals can be found in the introductions of Ireland and Kullback (1968), Fienberg (1970) and more recently of Bickel, Ritov and Wellner (1991).

Deming and Stephan (1940) explain that, motivated from the census of population, there are situations in sampling from a universe wherein the data furnished by the sample must be adjusted for consistency to the universe with data obtained from other sources. Let us investigate here the crossing of two characteristics $\mathcal{A}$ and $\mathcal{B}$ in a universe of N independent individuals. It is assumed to be known that character $\mathcal{A}$ and $\mathcal{B}$ appear with relative frequencies $(A_k)_{k=1,...,m_a}$ and $(B_l)_{l=1,...,m_b}$, respectively. In order to describe the crossing of the two characteristics $\mathcal{A}$ and $\mathcal{B}$, we consequently have to deal with $m_a \times m_b$-contingency tables. In case of the universe however, we cannot say anything about the cell entries a priorily. Hence we evaluate samples $\gamma^{(n)}$, $n \leq N$, of n randomly selected individuals from the universe. Tabulation yields $m_a \times m_b$-contingency tables $p^{(n)} = (p_{kl}^{(n)})$ where we agree for convenience with Ireland and Kullback (1968) that all cell entries are relative frequencies, i.e., $p_{kl}^{(n)} = \#(\mathcal{A}_k, \mathcal{B}_l)/n$. As a matter of fact, the relative marginal frequencies $\sum_{l=1}^{m_b} p_{kl}^{(n)}$ and $\sum_{k=1}^{m_a} p_{kl}^{(n)}$ differ from A_k and B_l, respectively, and an adjustment to the universal relative marginal frequencies seems appropriate.

The basic idea behind our new approach is the fundamental hypothesis of statistical mechanics. This general experience in particle physics (cf Schrödinger (1931), Lanford (1973)) claims that observations are asymptotically realized by 'most probable' microscopical systems, e.g., N-samples $\gamma^{(N)}$ for N tending to infinity. Section 2 is devoted to a setting for the adjustment of contingency tables to prescribed marginals. In reality we might only be able to afford watching one n-sample, $n \leq N$, which

provides the observed $m_a \times m_b$-contingency table $p^{(n)}$. It allows us to establish a multinomial distribution P_n with weights $p_{kl}^{(n)}$, $k = 1, \ldots, m_a$, $l = 1, \ldots, m_b$, on the set of universal $m_a \times m_b$-contingency tables $t^{(N)}$. $P_n(t^{(N)})$ indicates the probability of the $m_a \times m_b$-contingency table $t^{(N)}$ relative to the $m_a \times m_b$-contingency table $p^{(n)}$. However, an application of Stirling's formula to $P_n(t^{(N)})$ shows that these probabilities vanish for every $t^{(N)}$ as N tends to infinity.

In Section 3 we will introduce the notion 'most probable' in the sense of so-called large deviations dealing with the exponential decay of probabilities. Theorem 1 identifies a specific relative entropy as the exponential rate function of a large deviation principle satisfied by $m_a \times m_b$-contingency tables. We will be able to conclude that our 'most probable' adjustment of $m_a \times m_b$-contingency tables coincides with the Ireland-Kullback as well as asymptotically with the Deming-Stephan adjustment and with the maximum likelihood adjustment in Smith (1947). The identification of these adjustments as 'most probable' could be a point in Carnal (1993)'s discussion on the numbers of parliament seats representing the results of an election in a fair way.

Section 4 provides by means of a Lagrangian procedure the product form of the 'most probable' adjustment in terms of so-called adjustment multipliers. They deserve an interpretation of the adjustment in terms of prescription from each character $\mathcal{A}$ and $\mathcal{B}$ for the crossing $\mathcal{A} \times \mathcal{B}$. These adjustment multipliers appear as solutions of a system of two non-linear integral equations. Ireland and Kullback found a discrete version in 1968, being apparently not aware that Schrödinger had already discussed the general continuous version in 1931. This link gives access to further references, among others Fortet (1940), Beurling (1960), Föllmer (1988), Nagasawa (1989,1993), Aebi and Nagasawa (1992) and Aebi (1995).

2 Probabilities on contingency tables

Let us sample the crossings $(\mathcal{A}_k, \mathcal{B}_l)$, $k = 1, \ldots, m_a$, $l = 1, \ldots, m_b$, of two characteristics $\mathcal{A}$ and $\mathcal{B}$ from n randomly selected individuals in a universe of N independent candidates. We obtain n-samples $\gamma^{(n)}$, $n \leq N$, i.e., lists of n labeled lines which con-

tain the two characteristics of each individual of the sample $\gamma^{(n)}$. Every sample $\gamma^{(n)}$

yields a $m_a \times m_b$-contingency table $t^{(n)}(\gamma^{(n)})$ which cells contain the relative frequen-

cies $t_{kl}^{(n)}$ of crossings $(\mathcal{A}_k, \mathcal{B}_l)$, $k = 1, \ldots, m_a$, $l = 1, \ldots, m_b$. Considered as a mapping

from n-samples $\gamma^{(n)}$ to $m_a \times m_b$-contingency tables $t^{(n)}(\gamma^{(n)})$, this correspondence is

not one-to-one. In fact, rearranging the n individuals of a sample $\gamma^{(n)}$ according to the

$n\, t_{kl}^{(n)}$-subsets of crossings provides other samples which all yield the same $m_a \times m_b$-

contingency table $t^{(n)} = (t_{kl}^{(n)})$. The possible number of such rearrangements can easily

be computed as

$$\frac{n!}{\prod_{k,l=1}^{m_a,m_b}(n\, t_{kl}^{(n)})!} . \tag{1}$$

In reality, we might only be able to afford watching one n-sample from which we

determine our observed $m_a \times m_b$-contingency table

$$p^{(n)} = (p_{kl}^{(n)}) \begin{array}{c} k = 1, \ldots, m_a \\ l = 1, \ldots, m_b \end{array} .$$

It allows us to define a probability measure P_n on the set of universal $m_a \times m_b$-

contingency tables $t^{(N)}$. In fact, let $\gamma^{(N)}$ be any N-sample generating a given $t^{(N)}$.

By assumption, $\gamma^{(N)}$ is obtained from N independent individuals; hence it can be

assigned the probability

$$P_n(\gamma^{(N)}) = \prod_{k,l=1}^{m_a,m_b} (p_{kl}^{(n)})^{N\, t_{kl}^{(N)}} \tag{2}$$

which does not depend on $\gamma^{(N)}$ directly, but on the generated contingency table $t^{(N)}$.

As a consequence of (1), the probability of $t^{(N)}$ relative to $p^{(n)}$ can be given as

$$P_n(t^{(N)}) = N! \prod_{k,l=1}^{m_a,m_b} \frac{(p_{kl}^{(n)})^{N\, t_{kl}^{(N)}}}{(N\, t_{kl}^{(N)})!} . \tag{3}$$

Hence, if considered relative to the observation $p^{(n)}$, $m_a \times m_b$-contingency tables $t^{(N)}$

are multinomial distributed with $m_a m_b$ categories possessing the weights $p_{kl}^{(n)}$.

Following Deming and Stephan (1940) and others, we may realistically assume

that the relative marginal frequencies $(A_k)_{k=1,\ldots,m_a}$ of character $\mathcal{A}$ and $(B_l)_{l=1,\ldots,m_b}$ of

character $\mathcal{B}$ are prescribed by the universe. As a matter of sampling variability and

divisibility caused by the sample size n, our observation $p^{(n)}$ cannot be expected to

satisfy the resulting marginal consistency conditions

$$A_k = \sum_{l=1}^{m_b} t_{kl}^{(N)}, \quad k = 1, \ldots, m_a \tag{4}$$

$$B_l = \sum_{k=1}^{m_a} t_{kl}^{(N)}, \quad l = 1, \ldots, m_b \tag{5}$$

for universal $m_a \times m_b$-contingency tables $t^{(N)}$. Hence the question for a criterion arises which determines a relevant adjustment of the observation $p^{(n)}$ to the prescribed relative marginal frequencies. As a new approach, this paper proposes the 'fundamental hypothesis of statistical mechanics' appearing in Schrödinger (1931), Lanford (1973)'s section A1 and further references given there. It claims that

> *an observation on a macroscopical level (of e.g., contingency tables and their marginals) is realized in the limit of infinitely many individuals by that microscopical system (e.g., N-samples as N tends to infinity) which attains maximal probability given the observation.*

We notice that this hypothesis expressing a general experience in particle physics exclusively applies to asymptotical phenomena.

3 A large deviation approach

Definition 1 *For $m_a \times m_b$-contingency tables t and p the relative entropy of t with respect to p is defined as*

$$H(t \mid p) = \sum_{k,l=1}^{m_a,m_b} \log(\frac{t_{kl}}{p_{kl}}) t_{kl} \tag{6}$$

where $0 \log 0 = 0$ by convention. In case that $p_{kl} = 0 \Rightarrow t_{kl} = 0$ for $k = 1, \ldots, m_a$, $l = 1, \ldots, m_b$, is violated, $H(t \mid p)$ is defined to be infinity.

Remark 1 *Relative entropies, also known as information divergences, informations for discrimination or Kullback-Leibler divergences, are considered in statistical mechanics as natural measures for the amount of randomness in a particle system and were already discussed by Boltzmann (1896) in his lectures on the theory of gas. For problems related to those investigated in this paper, we refer beside the standard sources to Aebi and Nagasawa (1992), Föllmer (1988) and Csiszar (1984).*

Large deviation principles describe asymptotical phenomena which possess expo-
nentially fast vanishing probabilities. Under such circumstances, 'most probable' re-
alizations have to be understood as realizations which vanish at slowest exponential
rate. In case of contingency tables we find

Theorem 1 (large deviations of contingency tables)
The probabilities $(P_n(t^{(N)}))_{N \in \mathbb{N}}$ at (3) obey the large deviation principle

$$\lim_{N \nearrow \infty} [\frac{1}{N} \log P_n(t^{(N)}) + H(t^{(N)} \mid p^{(n)})] = 0 \tag{7}$$

where $H(t^{(N)} \mid p^{(n)})$ given at (6) is the so-called exponential rate function.

Proof. By means of Stirling's formula

$$N! = (\frac{N}{e})^N \sqrt{2\pi N} (1 + \varepsilon_N) \quad \text{with} \quad \varepsilon_N \to 0 \text{ as } N \nearrow \infty \tag{8}$$

the number (1) of possible N-samples $\gamma^{(N)}$ yielding a given $m_a \times m_b$-contingency ta-
ble $t^{(N)}$ can be expressed as

$$\frac{1}{(\sqrt{2\pi N})^{m_a+m_b-1}} \frac{1 + \varepsilon_N}{\prod_{k,l=1}^{m_a,m_b}(1 + \varepsilon_{t_{kl}^{(N)}})} \prod_{k,l=1}^{m_a,m_b} (t_{kl}^{(N)})^{-N t_{kl}^{(N)} - \frac{1}{2}}.$$

Since (8) is designed to be applied to large numbers, $t_{kl}^{(N)} > 0$ is required. In case of
vanishing relative frequencies, we put $(N t_{kl}^{(N)})! = 1$ in advance. A straightforward
computation of (7) yields

$$\lim_{N \nearrow \infty} [\frac{1}{N} \log P_n(t^{(N)}) + H(t^{(N)} \mid p^{(n)})]$$

$$= \lim_{N \nearrow \infty} [\frac{1}{N} (\sum_{k,l=1}^{m_a,m_b} \{ \log p_{kl}^{(n)} - \log t_{kl}^{(N)} \} N t_{kl}^{(N)} - \frac{1}{2} \sum_{k,l=1}^{m_a,m_b} \log t_{kl}^{(N)}$$

$$- (m_a + m_b - 1) \log \sqrt{2\pi N} + \log(1 + \varepsilon_N) - \sum_{k,l=1}^{m_a,m_b} \log(1 + \varepsilon_{t_{kl}^{(N)}}))$$

$$+ H(t^{(N)} \mid p^{(n)})] = 0.$$

Definition 2 *Provided P_n at (3), the 'most probable' $m_a \times m_b$-contingency table $\hat{t}^{(N)}$
which satisfies the marginal consistency conditions (4) and (5) is defined as the mini-
mum point of the exponential rate function of the associated large deviation principle
at (7) under the side-conditions (4) and (5).*

Corollary 1 *The 'most probable' contingency table $\hat{t}^{(N)}$ in Definition 2 approaches asymptotically as N tends to infinity the relevant, i.e., observable, realization postulated by the fundamental hypothesis of statistical mechanics.*

Proof. The limit (7) implies that $\hat{t}^{(N)}$ maximizes (3) as N tends to infinity since it minimizes the relative entropy (6) with respect to $p^{(n)}$ for $N \in \mathbb{N}$.

Corollary 2 *The 'most probable' adjustment $\hat{t}^{(N)}$ of $p^{(n)}$ coincides with the Ireland-Kullback adjustment of $m_a \times m_b$-contingency tables to prescribed marginal frequencies. Hence it is best asymptotically normal and $2n\,H(\hat{t}^{(N)} \mid p^{(n)})$ is asymptotically as n, $n \leq N$, tends to infinity Chi square-distributed with $m_a + m_b - 2$ degrees of freedom.*

Moreover, $\hat{t}^{(N)}$ corresponds to the Deming-Stephan adjustment approximately for large sample size n and it coincides asymptotically in N with the maximum likelihood adjustment in Smith (1947).

Proof. Theorem 1 identifies $H(t^{(N)} \mid p^{(n)})$ at (6) as the adequate rate function for large deviations of contingency tables. Following Definition 2, $\hat{t}^{(N)}$ minimizes (6) which is the Leibler-Kullback divergence treated in Ireland and Kullback (1968). The asymptotical statistical properties of $\hat{t}^{(N)}$ are found in their theorem 1.3. The approximate equivalence of the Deming-Stephan and the Ireland-Kullback adjustment for large sample size n is discussed in Ireland and Kullback (1968), Fienberg (1970), Csiszar (1975) and Haberman (1984). Finally, (7) provides the asymptotical behavior of (3) which determines the maximum likelihood adjustment in Smith (1947).

4   Adjustment multipliers

Theorem 1 explains in terms of asymptotically 'most probable' the traditional adjustment of contingency tables due to Deming and Stephan (1940) and Ireland and Kullback (1968) on the basis of the fundamental hypothesis of statistical mechanics. Now we would like to specify the characteristical structures of $\hat{t}^{(N)}$, i.e., the 'most probable' contingency table relative to the observed contingency table $p^{(n)}$ which satisfies the marginal consistency conditions (4) and (5).

Theorem 2 (Schrödinger 1931, Ireland and Kullback 1968)

The 'most probable' contingency table $\hat{t}^{(N)}$ in Definition 2 can be represented as

$$\hat{t}_{ij}^{(N)} = \varphi_{ai}^{(n,N)}\, p_{ij}^{(n)}\, \varphi_{bj}^{(n,N)} \tag{9}$$

for $1 \leq i \leq m_a$, $1 \leq j \leq m_b$, where $(\varphi_{ai}^{(n,N)})_{i=1,\ldots,m_a}$ and $(\varphi_{bj}^{(n,N)})_{j=1,\ldots,m_b}$ depend exclusively on character $\mathcal{A}$ and $\mathcal{B}$, respectively. These so-called adjustment multipliers of $p^{(n)}$ are non-negative and have a uniquely determined product $(\varphi_{ai}^{(n,N)}\, \varphi_{bj}^{(n,N)})_{i=1,\ldots,m_a,\, j=1,\ldots,m_b}$.

Schrödinger (1931) deduces the characteristical product structures at (9) by maximizing the probability (3) of a N-particle journey with prescribed departure and arrival according to (4) and (5), respectively. Ireland and Kullback (1968) minimize the relative entropy (6) under the marginal consistency conditions (4) and (5).

Proof. In view of Remark 2 it is illustrative to employ a Lagrangian procedure. Setting

$$
\begin{aligned}
L_{p^{(n)},\mathcal{A},\mathcal{B}}(t^{(N)}, \lambda^{(n,N)}) \;=\;& \sum_{k,l=1}^{m_a,m_b} \log\!\left(\frac{t_{kl}^{(N)}}{p_{kl}^{(n)}}\right) t_{kl}^{(N)} \\
&- \sum_{k=1}^{m_a} \lambda_{k\cdot}^{(n,N)} \Big(\sum_{l=1}^{m_b} t_{kl}^{(N)} - A_k\Big) \\
&- \sum_{l=1}^{m_b} \lambda_{\cdot l}^{(n,N)} \Big(\sum_{k=1}^{m_a} t_{kl}^{(N)} - B_l\Big)
\end{aligned}
\tag{10}
$$

for $m_a + m_b$ Lagrangian multipliers $\lambda_{k\cdot}^{(n,N)}$ and $\lambda_{\cdot l}^{(n,N)}$, the first $m_a m_b$ Lagrangian equations

$$\frac{\partial L_{p^{(n)},\mathcal{A},\mathcal{B}}}{\partial t_{ij}^{(N)}} = 1 + \log\!\left(\frac{t_{ij}^{(N)}}{p_{ij}^{(n)}}\right) - \lambda_{i\cdot}^{(n,N)} - \lambda_{\cdot j}^{(n,N)} = 0$$

$1 \leq i \leq m_a$, $1 \leq j \leq m_b$, imply

$$\varphi_{ai}^{(n,N)} = \exp\{\lambda_{i\cdot}^{(n,N)}\}, \quad i = 1,\ldots,m_a$$

and

$$\varphi_{bj}^{(n,N)} = \exp\{\lambda_{\cdot j}^{(n,N)} - 1\}, \quad j = 1,\ldots,m_b.$$

The $m_a + m_b$ Lagrangian equations $\partial L_{p^{(n)},\mathcal{A},\mathcal{B}}/\partial \lambda_{k\cdot}^{(n,N)} = 0$ and $\partial L_{p^{(n)},\mathcal{A},\mathcal{B}}/\partial \lambda_{\cdot l}^{(n,N)} = 0$ reproduce the consistency condition (4) required by character $\mathcal{A}$ and the consistency condition (5) required by character $\mathcal{B}$, respectively. For the problem of uniqueness we have to refer to Remark 3.

Remark 2 *The structures of the representation (9) of the 'most probable' contin-gency table $\hat{t}^{(N)}$ in Definition 2 deserve an interpretation. The adjustment multipliers $(\varphi_{ai}^{(n,N)})_{i=1,\ldots,m_a}$ and $(\varphi_{bj}^{(n,N)})_{j=1,\ldots,m_b}$ can be considered as the prescriptions from each character $\mathcal{A}$ and $\mathcal{B}$, respectively. They tell us how to adjust in the 'most probable' way the observation $p^{(n)}$ to the required consistency with the universal relative frequencies of the characters $\mathcal{A}$ and $\mathcal{B}$.*

Corollary 3 (Csiszar 1975, Aebi and Nagasawa 1992)

There exists a $m_a \times m_b$-contingency table $t^{(N)}$ with $t_{kl}^{(N)} = 0 \Leftrightarrow p_{kl}^{(n)} = 0$, $k = 1,\ldots m_a$, $l = 1,\ldots,m_b$, and $H(t^{(N)} \mid p^{(n)}) < \infty$ which satisfies (4) and (5) if and only if the system (4) & (5) expressed in terms of (9) is solved by adjustment multipliers $(\varphi_{ai}^{(n,N)})_{i=1,\ldots,m_a}$ and $(\varphi_{bj}^{(n,N)})_{j=1,\ldots,m_b}$ which satisfy

$$\sum_{k=1}^{m_a} | \log \varphi_{ak}^{(n,N)} | \, A_k < \infty$$

$$\sum_{l=1}^{m_b} | \log \varphi_{bl}^{(n,N)} | \, B_l < \infty.$$

In this sense, minimizing the relative entropy with respect to $p^{(n)}$ at (6) under the marginal consistency conditions (4) and (5) is equivalent to solving (4) & (5) expressed in terms of (9). This equivalence also holds in the continuous situation of a bivariate distribution density $p(x,z)$ and prescribed marginal densities $A(x)$ and $B(z)$. The system of two non-linear integral equations corresponding to (4) & (5) in terms of (9) is a so-called Schrödinger system

$$\varphi_a(x) \int p(x,z)\, dz\, \varphi_b(z) = A(x) \tag{11}$$

$$\int \varphi_a(x)\, dx\, p(x,z)\, \varphi_b(z) = B(z) \tag{12}$$

for Lebesgue-almost all $x, z \in \mathbb{R}^d$. Accordingly, the 'most probable' adjustment of $p(x,z)$ to prescribed marginal densities $A(x)$, $B(z)$ is given as

$$\hat{t}(x,z) = \varphi_a(x)\, p(x,z)\, \varphi_b(z), \quad x, z \in \mathbb{R}^d.$$

Proof. The equivalence in the general continuous situation is a consequence of Csiszar (1975). A class of singular kernels $p(x,z)$ is investigated in Aebi and Naga-sawa (1992).

Remark 3 *At (2.2) in Ireland and Kullback (1968) we find (4) & (5) expressed in terms of (9). However, Schrödinger gave the system (11) & (12) already in 1931. He pointed out that existence and uniqueness of solutions are difficult to discuss due to the non-linearity of this system. These questions have been investigated by Fortet (1940), Beurling (1960), Föllmer (1988), Nagasawa (1989,1993), Aebi and Nagasawa (1992) and Aebi (1995) where the latter three deal with more general kernels p.*

References

Aebi, R., Nagasawa, M. (1992) Large Deviations and the Propagation of Chaos for Schrödinger Processes. Probab. Theory Relat. Fields **94**: 53-68

Aebi, R. (1995) A Solution to Schrödinger's Problem of Non-Linear Integral Equations. Z. angew. Math. Phys. (ZAMP) **46**: 772-792

Beurling, A. (1960) An Automorphism of Product Measures. Ann. Math. **72**: 189-200

Bickel, P.J., Ritov, Y., Wellner, J.A. (1991) Efficient Estimation of Linear Functionals of a Probability Measure P with Known Marginal Distributions. Ann. of Stat. **19**: 1316-1346

Boltzmann, L. (1896) Vorlesungen über Gastheorie. J.A. Barth Verlag, Leipzig

Carnal, H. (1993) Mathématiques et politique. El. Math. **48**: 27-32

Csiszar, I. (1975) I-Divergence Geometry of Probability Distributions and Minimization Problems. Ann. Probab. **3**: 146-158

Csiszar, I. (1984) Sanov Property, Generalized I-Projection and a Conditional Limit Theorem. Ann. Probab. **12**: 768-793

Deming, W.E., Stephan, F.F. (1940) On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known. Ann. of Math. Stat. **11**: 427-444

Fienberg, S.E. (1970) An Iterative Procedure for Estimation in Contingency Tables. Ann. Math. Stat. **41**: 907-917

Föllmer, H. (1988) Random Fields and Diffusion Processes. École d'Été de Saint Flour XV-XVII (1985-87). Lecture Notes Math. 1362, Springer-Verlag, Berlin

Fortet, R. (1940) Résolution d'un Système d'Équation de M. Schrödinger. J. Math. Pures et Appl. IX: 83-95

Haberman, S.J. (1984) Adjustment by Minimum Discriminant Information. Ann. of Stat. **12**: 971-988

Ireland, C.T., Kullback, S. (1968) Contingency Tables with Given Marginals. Biometrika **55**: 179-188

Lanford, O.E. (1973) Entropy and Equilibrium States in Classical Statistical Mechanics. 'Statistical Mechanics and Mathematical Problems' (ed. Lenard A.). Lecture Notes in Phys. **20**, 1-113, Springer-Verlag, Berlin

Nagasawa, M. (1989) Transformations of Diffusion and Schrödinger Processes. Probab. Th. Rel. Fields **82**: 109-136

Nagasawa, M. (1993) Schrödinger Equations and Diffusion Theory. Monographs in Mathematics vol. 86, Birkhäuser-Verlag, Basel

Schrödinger, E. (1931) Über die Umkehrung der Naturgesetze. Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse: 144-153

Smith, J.H. (1947) Estimations of Linear Functions of Cell Proportions. Ann. Math. Statist. **18**: 231-254

Robert Aebi
Institute of Mathematical Statistics
University of Berne
Sidlerstrasse 5
CH-3012 Berne, Switzerland