Part V

On-line supplements

Coloured figures for Chapter 2

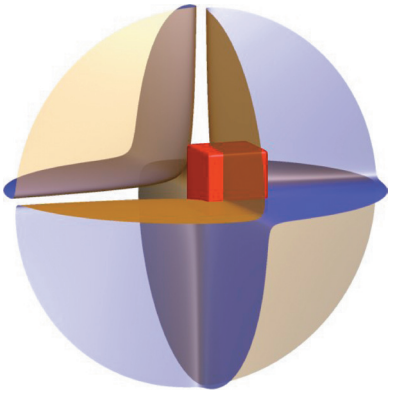


Fig. 2.2 The two-dimensional surface defined by Equation (2.12), when evaluated over the ball in $\mathbb{R}^3$ of radius 3, centred at the origin. The inner box is the unit cube $[0,1]^3$.

(a) (b)

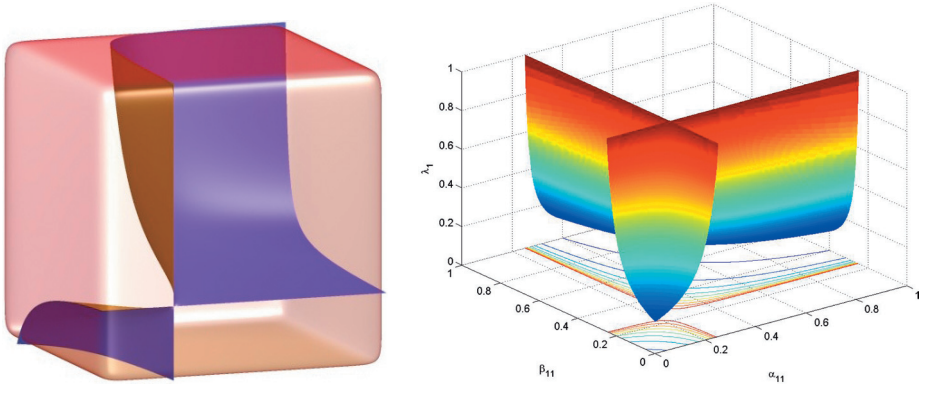


Fig. 2.3 Intersection of the surface defined by Equation (2.12) with the unit cube $[0,1]^3$, different views obtained using `surf` in (a) and `MATLAB` in (b).

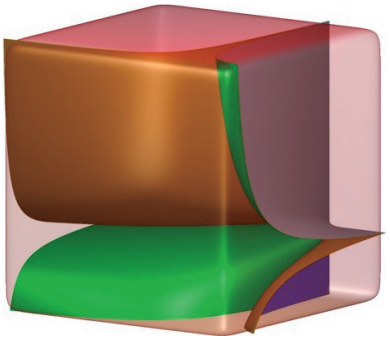


Fig. 2.4 Projection of the non-identifiable spaces corresponding to the first and second and third MLE from Table 2.2 (a) into the three-dimensional unit cube where λ_1, α_{11} and β_{21} take values.

Fig. 2.5 Projection of the non-identifiable spaces the first MLE in Table 2.2 (a), the first three local maxima and the last local maxima in Table 2.2 (b) into the three-dimensional unit cube where λ_1, α_{11} and β_{11} take values. In this coordinate system, the projection of non-identifiable subspaces for the first three local maxima in Table 2.2 (b) results in the same surface; in order to obtain distinct surfaces, it would be necessary to change the coordinates over which the projections are made.

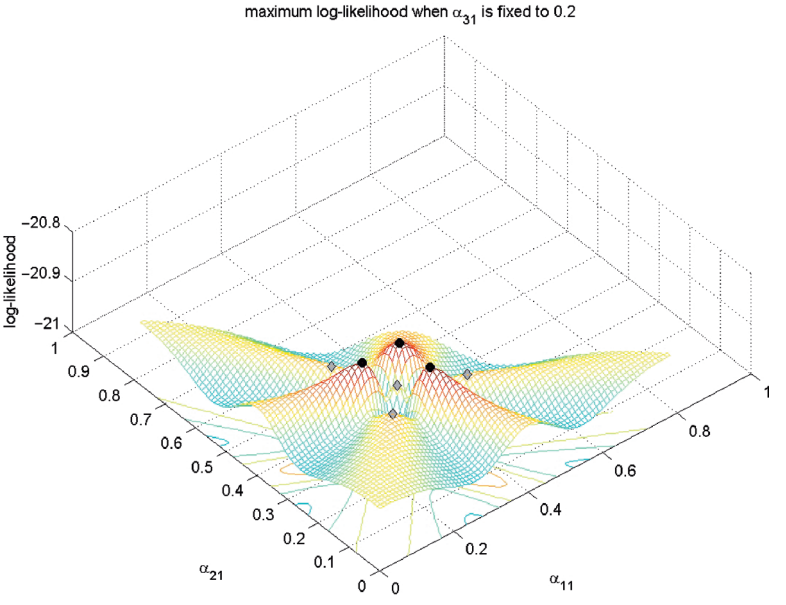


Fig. 2.6 The plot of the profile likelihood as a function of α_{11} and α_{21} when α_{31} is fixed to 0.2. There are seven peaks: the three black points are the MLEs and the four grey diamonds are the other local maxima.

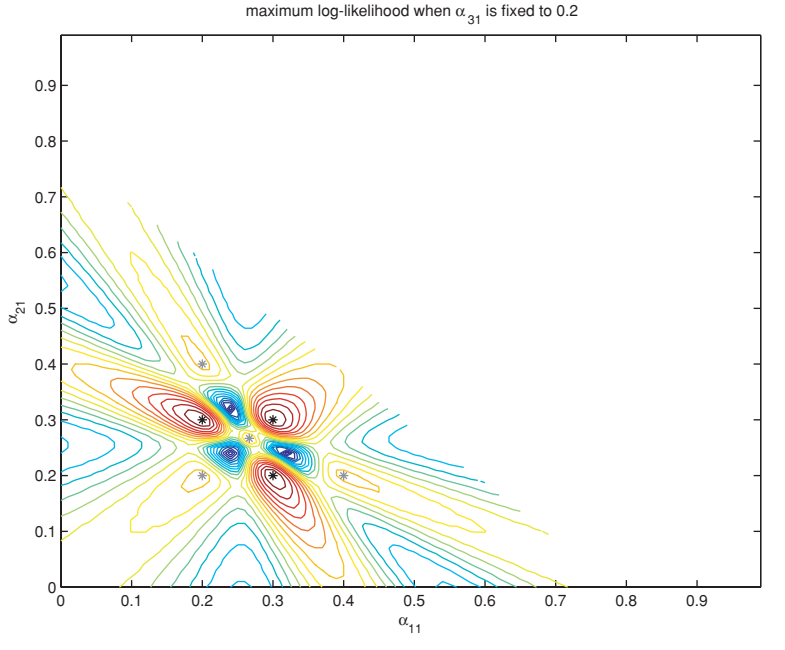


Fig. 2.7 The contour plot of the profile likelihood as a function of α_{11} and α_{21} when α_{31} is fixed. There are seven peaks: the three black points are the MLEs and the four grey points are the other local maxima.

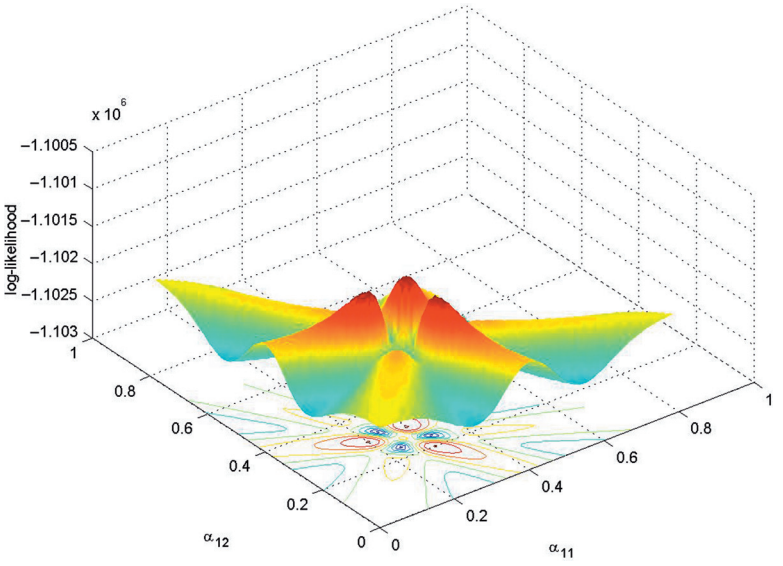


Fig. 2.8 The contour plot of the profile likelihood as a function of α_{11} and α_{21} when α_{31} is fixed for the data (2.8) multiplied by 10 000. As before, there are seven peaks: three global maxima and four identical local maxima.

22

Supplementary note to Maximum likelihood estimation in latent class models for contingency table data

Yi Zhou

22.1 Algebraic Geometry

22.1.1 Polynomial Ring, Ideal and Variety

In this section, we review some basic concepts and definitions in algebraic geometry and we draw connections between algebraic geometry and statistics. We begin with some concepts in abstract algebra. In mathematics, a *ring* is an algebraic structure in which addition and multiplication are defined and have some properties.

Definition 22.1 (Ring) *A ring is a set $\mathcal{R}$ equipped with two binary operations $+ : \mathcal{R} \times \mathcal{R} \to \mathcal{R}$ and $\cdot : \mathcal{R} \times \mathcal{R} \to \mathcal{R}$, called addition and multiplication, such that:*

- *$(\mathcal{R}, +)$ is an abelian group with identity element 0, so that $\forall a, b, c \in \mathcal{R}$, the following axiom hold:*
  - $a + b \in \mathcal{R}$
  - $(a + b) + c = a + (b + c)$
  - $0 + a = a + 0 = a$
  - $a + b = b + a$
  - $\exists -a \in \mathcal{R}$ *such that* $a + (-a) = (-a) + a = 0$
- *$(\mathcal{R}, \cdot)$ is a monoid with identity element 1, so that $\forall a, b, c \in \mathcal{R}$, the following axioms hold:*
  - $a \cdot b \in \mathcal{R}$
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
  - $1 \cdot a = a \cdot 1 = a$
- *Multiplication distributes over addition:*
  - $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$
  - $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$

The set of integer numbers $\mathbb{Z}$, the set of real numbers $\mathbb{R}$, and the set of rational numbers $\mathbb{Q}$ all are rings with the common addition and multiplication defined for numbers. Algebraic geometry is interested in polymonials and hence the polymonial rings. A *polynomial ring* is the set of polynomials in one or more unknowns with coefficients in a ring, for example, the set of polynomials with one variable in real

numbers $\mathbb{R}[x]$ or the set of polynomials with two variables in rational numbers $\mathbb{Q}[x, y]$.

An *ideal* is a special subset of a ring. The ideal concept generalizes in an appropriate way some important properties of integers like "even number" or "multiple of 3".

Definition 22.2 (Ideal, generating set) *An ideal $\mathcal{I}$ is a subset of a ring $\mathcal{R}$ satisfying:*

- *$f + g \in \mathcal{I}$ if $f \in \mathcal{I}$ and $g \in \mathcal{I}$, and*
- *$pf \in \mathcal{I}$ if $f \in \mathcal{I}$ and $p \in \mathcal{R}$ is an arbitrary element.*

In other words, an ideal is a subset of a ring which is closed under addition and multiplication by elements of the ring. Let $\mathcal{I} = \langle \mathcal{A} \rangle$ denote the ideal $\mathcal{I}$ generated by the set $\mathcal{A}$, this means any $f \in \mathcal{I}$ is of the form $f = a_1 r_1 + \cdots + a_n r_n$ where each $a_i \in \mathcal{A}$ and $r_i \in \mathcal{R}$. If $\mathcal{A}$ is finite then $\mathcal{I}$ is a finitely generated ideal and if $\mathcal{A}$ is a singleton then $\mathcal{I}$ is called a principal ideal.

From now on, we only talk about the polynomial rings and ideals in the polynomial rings. For an ideal, we can consider the *generating set* of the ideal and a particular kind of generating set is called *Gröbner basis*. Roughly speaking, a polynomial f is in the ideal if and only if the reminder of f with respect to the Gröbner basis is 0. But here, the division algorithm requires a certain type of *ordering* on the monomials. So Gröbner basis is stated relative to some monomial order in the ring and different orders will result in different bases. Later, we will give some examples of the Gröbner basis.

The following terms and notation are present in the literature of Gröbner basis and will be useful later on.

Definition 22.3 (degree, leading term, leading coefficient, power product) *A power product is a product of indeterminants $\left\{ x_1^{\beta_1} \cdots x_n^{\beta_n} : \beta_i \in \mathbb{N}, 1 \leq i \leq n \right\}$. The degree of a term of polynomial f is the sum of exponents of the term's power product. The degree of a polynomial f, denoted $\deg(f)$, is the greatest degree of terms in f. The leading term of f, denoted $\mathrm{lt}(f)$, is the term with the greatest degree. The leading coefficient of f is the coefficient of the leading term in f while the power product of the leading term is the leading power product, denoted $\mathrm{lp}(f)$.*

But sometimes there are many terms in the polynomial which all have the greatest degree, therefore to make the *leading term* well-defined, we need a well-defined *term order*. Below is one kind of term ordering.

Definition 22.4 (Degree Reverse Lexicographic Ordering) *Let $x > y > z$ be a lex ordering and $\mathbf{u}^\alpha = x^{\alpha_1} y^{\alpha_2} z^{\alpha_3}$. Then $\mathbf{u}^\alpha < \mathbf{u}^\beta$ if and only if one of the following is true:*

- *$\alpha_1 + \alpha_2 + \alpha_3 < \beta_1 + \beta_2 + \beta_3$*
- *$\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3$ and the first coordinates α_i and β_i from the right which are different satisfy $\alpha_i > \beta_i$.*

For example, consider the polynomial $f = x^3z - 2x^2y^2 + 5y^2z^2 - 7yz$. Then the degree reverse lexicographic ordering produces $x^2y^2 > x^3z > y^2z^2 > yz$. So the leading term of f is $\mathrm{lt}(f) = -2x^2y^2$ and the leading power product is $\mathrm{lp}(f) = x^2y^2$. Now we can introduce the definition of *Gröbner basis*.

Definition 22.5 (Gröbner basis) *A set of polynomials $\mathcal{G}$ contained in an ideal $\mathcal{I}$ is called a Gröbner basis for $\mathcal{I}$ if the leading term of any polynomial in $\mathcal{I}$ is divisible by some polynomial in $\mathcal{G}$.*

Equivalent definitions for Gröbner basis can be given according to the below theorem.

Theorem 22.1 *Let $\mathcal{I}$ be an ideal and $\mathcal{G}$ be a set contained in $\mathcal{I}$. Then the following statements are equivalent:*

(a) $\mathcal{G}$ is a Gröbner basis of $\mathcal{I}$.
(b) The ideal given by the leading terms of polynomials in $\mathcal{I}$ is itself generated by the leading terms of $\mathcal{G}$.
(c) The reminder of the division of any polynomial in the ideal $\mathcal{I}$ by $\mathcal{G}$ is 0.
(d) The reminder of the division of any polynomial in the ring in which the ideal $\mathcal{I}$ is defined by $\mathcal{G}$ is unique.

Now that we can obtain a Gröbner basis, we would like to obtain a simple and probably unique basis. The concept of *minimal Gröbner basis* ensures the simplicity of the basis in some sense.

Definition 22.6 (Minimal Gröbner basis) *A Gröbner basis $\mathcal{G}$ is minimal if for all $g \in \mathcal{G}$, the leading coefficient of g is 1 and for all $g_1 \neq g_2 \in \mathcal{G}$, the leading power product of g_1 does not divide the leading power product of g_2.*

A minimal Gröbner basis has the least number of polynomials among the Gröbner bases. But a minimal Gröbner basis is not unique. For example if our basis is $\{y^2+yx+x^2, y+x, y, x^2, x\}$ for the ideal $\{y^2+yx+x^2, y+x, y\}$ with the lex $y > x$ term order then both $\{y, x\}$ and $\{y+x, x\}$ are minimal Gröbner bases. To obtain a unique Gröbner basis, we need to put further restrictions on the basis.

Definition 22.7 (Reduced Gröbner basis) *A Gröbner basis is reduced if for $g \in \mathcal{G}$ the leading coefficient of g is 1 and g is reduced with respect to other polynomials in $\mathcal{G}$.*

By the definition, in our previous example $\{y, x\}$ is a reduced Gröbner basis. Every non-zero ideal $\mathcal{I}$ has a unique reduced Gröbner basis with respect to a fixed term order. In algebraic geometry, Buchberger's algorithm is the most commonly used algorithm computing the Gröbner bases and it can be viewed as a generalization of the Euclidean algorithm for univariate Greatest Common Divisor computation and of Gaussian elimination for linear systems. The basic version of Buchberger's algorithm does not guarantee the resulting basis to be minimal and reduced, but

there are many variants of the basic algorithm to produce a minimal or reduced basis.

Now let's talk about varieties. A variety is indeed a hyper-surface or a manifold in the enveloping space where it is defined. It is essentially a finite or infinite set of points where a polynomial in one or more variables attains, or a set of such polynomials all attain, a value of zero. The ideal arising from a variety is just the set of all polynomials attaining zero on the variety. For example, the surface of independence for the 2×2 table is a variety, and the ideal of this variety is generated by the set $\{p_{11}p_{22} - p_{12}p_{21}\}$ (Gröbner basis). As a geometric object, we can consider the dimension of a variety. The dimension of a variety and the dimension of its ideal is the same thing, as the ideal dimension is the dimension of the intersection of its projective topological closure with the infinite hyperplane. As we will show later the way we compute the dimension of a variety is by computing the dimension of the ideal arising from it. The dimension of a variety may be less than the dimension of its enveloping space. Again, take the surface of independence as an example. The dimension of this variety is 2 while the dimension of the enveloping space, the probability simplex, is 3.

Definition 22.8 (Variety) *A variety is the zero set of systems of polynomial equations in several unknowns.*

Definition 22.9 (Ideal of variety) *The ideal of an variety is the set of polynomials vanishing on the variety.*

Algebraic geometry studies polynomials and varieties. And the models we are working on, the traditional log-linear models and the latent class models, are all stated with polynomials! That's why concepts in statistics and concepts in algebraic geometry connects with each other. For example, in (Pachter and Sturmfels 2005), drawed the connections between some basic concepts of statistics and algebraic geometry, and we summarized them in table 22.1.

Statistics		**Algebraic Geometry**
independence	=	Segre variety
log-linear model	=	toric variety
curved exponential family	=	manifold
mixture model	=	joint of varieties
MAP estimation	=	tropicalization
.....	=	.....

Table 22.1 *A glimpse of the statistics - algeobraic geometry dictionary.*

Algebraic geometry views statistical models as varities, for example, the model of independence is related to the surface of independence. And here we like to refer to another figure in (Pachter and Sturmfels 2005), which we show here in Figure 22.1, to illustrate the connection between models and varieties. The model of interest here corresponds to the polynomial mapping f and the image of f which is a variety in the probability simplex. The observed data is a point in the probability

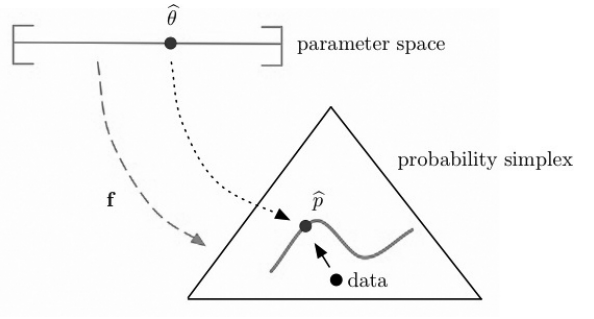


Fig. 22.1 The geometry of maximum likelihood estimation.

simplex. Thus, maximum likelihood estimation is to find a point $\hat{p}$ in the image of the mapping f, which maps back to $\hat{\theta}$ in the parameter space, *closest to* the observed data point.

In Table 22.1, we can see that specific models are corresonded to specific varieties. Here we want to talk more about the Segre variety and the secant variety because they are related to the log-linear models and the latent class models.

22.1.2 Segre Variety and Secant Variety

Let's begin by setting up the basic notations and concepts. let $\mathbb{R}^{n+1}$ be a $(n+1)$-dimensional vector space on the real field. Then the n-dimensional *projective space* $\mathbb{P}^n = \mathbb{P}(\mathbb{R}^{n+1})$ of $\mathbb{R}^{n+1}$ is a set of elements constructed from $\mathbb{R}^{n+1}$ such that a distinct element of the projective space consists of all non-zero vectors which are equal up to a multiplication by a non-zero scalar. The projective space $\mathbb{P}^n$ is isomorphic to the n-dimensional simplex.

Definition 22.10 (Segre map) *The Segre map σ is a map from the product space of two projective space $\mathbb{P}^n \times \mathbb{P}^m$ to a higher dimensional projective space $\mathbb{P}^{(n+1)(m+1)-1}$, such that for all $\mathbf{x} = (x_0, x_1, \ldots, x_n) \in \mathbb{P}^n$, all $\mathbf{y} = (y_0, y_1, \ldots, y_m) \in \mathbb{P}^m$,*

$$\sigma : \quad (\mathbf{x},\ \mathbf{y}) \mapsto \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} y_0, & y_1, & \cdots, & y_m \end{pmatrix}$$

The *Segre varieties* are the varieties $\mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_t}$ embedded in $\mathbb{P}^N$, $N = \prod(n_i+1)-1$, by Segre mapping, and the Segre embedding is based on the canonical multilinear map:

$$\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_t} \to \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_t}$$

where $\otimes$ is the *tensor product*, a.k.a. outer product. Now we denote the enveloping space $\mathbb{P}(\mathbb{R}^{n_1}\otimes\cdots\otimes\mathbb{R}^{n_t})$ by $\mathbb{P}^N$ and denote the embedded Segre variety $\mathbb{P}^{n_1}\otimes\cdots\otimes\mathbb{P}^{n_t}$ as $\mathbb{X}_n$. Then, with this point of view:

- the Segre variety $\mathbb{X}_n$ is the set of all classes of *decomposable tensors*, i.e. classes of tensors (i.e. multi-dimensional arrays) in $\mathbb{P}(\mathbb{R}^{n_1}\otimes\cdots\otimes\mathbb{R}^{n_t})$ of the form $v_1\otimes\cdots\otimes v_t$.
- the *secant variety*, $Sec_r(\mathbb{X}_n)$, is the closure of the set of classes of those tensors which can be written as the sum of $\leq r+1$ decomposable tensors.

Now let's consider the 2-dimensional tensors, which are actually matrices. In such case, $\mathbb{P}^{n_1}$ is the set of (n_1+1)-dimensional vectors, $\mathbb{P}^{n_2}$ is the set of (n_2+1)-dimensional vectors, and $\mathbb{P}^N$ is the set of $(n_1+1)\times(n_2+1)$ matrices, all under the projective equivalence. Then, the Segre variety $\mathbb{P}^{n_1}\otimes\mathbb{P}^{n_2}$ consists of all the rank 1 matrices in $\mathbb{P}^N$. And the r-secant variety $Sec_r(\mathbb{P}^{n_1}\otimes\mathbb{P}^{n_2})$ is the set of matrices having rank $\leq r+1$ because a matrix has rank $\leq r+1$ if and only if it is a sum of $\leq r+1$ matrices of rank 1.

For example, consider the embedding of $\mathbb{P}^2\otimes\mathbb{P}^2$ in $\mathbb{P}^8$, where $\mathbb{P}^8$ is the projective space of 3×3 matrices under projective equivalence. The ideal of 2×2 minors of the generic matrix of size 3×3 defines $\mathbb{P}^2\otimes\mathbb{P}^2$ and the determinant of the generic matrix gives the equation of $Sec_1(\mathbb{P}^2\otimes\mathbb{P}^2)$. *The Segre variety* $\mathbb{P}^2\otimes\mathbb{P}^2$ *corresponds to the no 2nd-effect log-linear model for the* 3×3 *table and the secant variety* $Sec_1(\mathbb{P}^1\otimes\mathbb{P}^2)$ *corresponds to the 2-level latent class model for the* 3×3 *table.*

Back to the former notations, we have $\mathbb{X}_n=\mathbb{P}^{n_1}\otimes\cdots\otimes\mathbb{P}^{n_t}$. What is the dimension of the secant variety $Sec_r(\mathbb{X}_n)$? There is an *expected dimension* by counting parameters:

$$\min\{N,(r+1)\prod_i n_i+r\}$$

which is only an upper bound of the actual dimension of $Sec_r(\mathbb{X}_n)$. If the actual dimension is different from the expected dimension, the secant variety is *deficient*. Computing the dimension of secant varieties has been a challenge problem in algebraic geometry. We summarize some results in the following theorems.

For the case of two factors, we have a complete answer for the actual dimension of the secant variety.

Theorem 22.2 (Proposition 2.3 in Catalisano etc.'s (Catalisano *et al.* 2002)) *For the case of two factors, for all* r, $1\leq r<\min(n_1,n_2)$ *the secant varieties* $Sec_r(\mathbb{X}_n)$ *all have dimension less than the expected dimension. Moreover, the least integer for which* $Sec_r(\mathbb{X}_n)$ *fills its enveloping space is* $r=n_1$.

When it comes to the case of three factors, the dimension of the secant variety is still an open problem in general. But for some special varieties, there are beautiful results. The below two theorems are for $n=(n_1,n_2,n_3)$.

Theorem 22.3 (Proposition 2.3 in Catalisano etc.'s (Catalisano *et al.* 2002)) *If* $n=(n_1,n_2,n_3)$ *and* $r\leq min(n_1,n_2,n_3)$, *then* $Sec_r(\mathbb{X}_n)$ *has the expected dimension.*

As a direct proposition from theorem 22.3, we have a complete answer for 2-level latent class model for 3×3 tables.

Theorem 22.4 *When $n = (n_1, n_2, n_3)$, the secant line variety for any Segre variety has the expected dimension.*

Remark 22.1 *Theorem 22.3 and 22.4 says that 2-level and "small" latent class models for 3×3 tables have the dimension*

$$\min\{(n_1+1)(n_2+1)(n_3+1)-1, (r+1)(n_1+n_2+n_3)+r\}$$

Note that the first term is the free dimension of the observed table and the second term is the dimension of underlining parameter space. And obviously, Theorem 22.4 can be directly applied to our conjecture about $2\times2\times K$ models.

For more factors, the dimension of some special varieties can still be derived.

Theorem 22.5 (Proposition 3.7 in (Catalisano *et al.* 2002)) *Let $n = (n_1, \ldots, n_t)$ and let $t \geq 3$, $n_1 \leq n_2 \leq \cdots \leq n_t$,*

$$\left[\frac{n_1 + n_2 + \cdots + n_t + 1}{2}\right] \geq \max(n_t + 1, r + 1)$$

Then $\dim Sec_r(\mathbb{X}_n) = (r+1)(n_1 + n_2 + \cdots + n_t) + r$.

Another result concerning about higher secant varieties is from coding theory when the dimensions of the Segre varieties are equal, that is, $n_1 = n_2 = \cdots = n_t = q - 1$.

Theorem 22.6 (Example 2.4 in (Catalisano *et al.* 2002))

(i) Let k be any positive integer, $q = 2$, $t = 2^k - 1$, $r = 2^{t-k}$. For these numbers the Segre embedding

$$\mathbb{X}_t = \underbrace{\mathbb{P}^1 \times \cdots \times \mathbb{P}^1}_{t} \rightarrow \mathbb{P}^{2^t-1}$$

we have $Sec_{r-1}(\mathbb{X}_t) = \mathbb{P}^{2^t-1}$ and these secant varieties fit "exactly" into their enveloping space.

(ii) We can make families of similar examples for products of $\mathbb{P}^2$, $\mathbb{P}^3$, $\mathbb{P}^4$, $\mathbb{P}^7$, $\mathbb{P}^8$, …, $\mathbb{P}^{q-1}$ where q is a prime power. Given such a q, for any integer $k \geq 1$ we take $t = (q^k - 1)/(q - 1)$ copies of $\mathbb{P}^{q-1}$, which gets embedded in $\mathbb{P}^{q^t-1}$. Then for $r = q^{t-k}$ we get

$$Sec_{r-1}(\underbrace{\mathbb{P}^{q-1} \times \cdots \times \mathbb{P}^{q-1}}_{t\text{-times}}) = \mathbb{P}^{q^t-1}$$

22.2 Symbolic Software of Computational Algebra

Unlike many numerical softwares we use in machine learning, by which we get the answer for a particular set of values of the variables of interest, symbolic softwares provide us an algebraic answer for all possible values of the variables. The symbolic computation can fill up the machine very quickly. So current symbolic softwares can only deal with limited-scale problems. Here we use some examples to show some symbolic computations relevant to the problems we have been discussed so far. We have been using various symbolic softwares for different purposes and here we will talk about the software SINGULAR because it is the software we need to do the computations related to our problems in this paper.

22.2.1 Computing the dimension of the image variety

Let's take the $2 \times 2 \times 3$ table with 2 latent classes as an example, to see how to compute the dimension of the image variety defined by the polynomial mapping f:

$$\begin{array}{rccc} f: & \Delta_1 \times \Delta_1 \times \Delta_1 \times \Delta_2 & \to & \Delta_{11} \\ & (a_t, x_{it}, y_{jt}, z_{kt}) & \mapsto & p_{ijk} = \sum_t a_t x_{it} y_{jt} z_{kt} \end{array}$$

where Δ_n is the n-dimensional probability simplex. The first step is to get the ideal arising from the model that is only defined on the probabilities $\{p_{ijk}\}$. In SINGLUAR, we define a polynomial ring r on the unknowns p_{ijk} which stand for cell probabilities and the unknowns $a_t, x_{it}, y_{jt}, z_{kt}$ which stand for the conditional probabilities. The ideal I on the ring r is defined by the model equalities (the first 12 polynomials) and sum 1 constraints of the probabilties (the last 7 polynomials).

```
ring r=0, (a1,x11,x21,y11,y21,z11,z21,z31,a2,x12,x22,
y12,y22,z12,z22,z32,p111,p112,p113,p121,p122,p123,p211,
p212,p213,p221,p222,p223), lp;
ideal I=p111-a1*x11*y11*z11-a2*x12*y12*z12,
p112-a1*x11*y11*z21-a2*x12*y12*z22,
p113-a1*x11*y11*z31-a2*x12*y12*z32,
p121-a1*x11*y21*z11-a2*x12*y22*z12,
p122-a1*x11*y21*z21-a2*x12*y22*z22,
p123-a1*x11*y21*z31-a2*x12*y22*z32,
p211-a1*x21*y11*z11-a2*x22*y12*z12,
p212-a1*x21*y11*z21-a2*x22*y12*z22,
p213-a1*x21*y11*z31-a2*x22*y12*z32,
p221-a1*x21*y21*z11-a2*x22*y22*z12,
p222-a1*x21*y21*z21-a2*x22*y22*z22,
p223-a1*x21*y21*z31-a2*x22*y22*z32,
a1+a2-1,
x11+x21-1,
x12+x22-1,
y11+y21-1,
y12+y22-1,
z11+z21+z31-1,
z12+z22+z32-1;
```

But the ideal I defined as above is on all the unknowns, including both the cell probabilities and the conditional probabilities. So the next step is to eliminate the unknowns $a_t, x_{it}, y_{jt}, z_{kt}$ and then to get the image variety where p_{ijk} lies. To use the elimination functions in SINGULAR, we need to include the library "ELIM.LIB".

```
LIB "elim.lib";
ideal J=elim1(I, a1*x11*x21*y11*y21*z11*z21*z31*a2*x12*x22
*y12*y22*z12*z22*z32);
J;
===>
J[1]=p121*p212*p223-p121*p213*p222-....;
J[2]=p112*p211*p223+p112*p212*p223-p112*p213*p221-....;
J[3]=p112*p121*p223+p112*p122*p223-p112*p123*p221-....;
J[4]=p112*p121*p213+p112*p121*p223+p112*p122*p213+....;
J[5]=p111+p112+p113+p121+p122+p123+p211+p212+p213+p221+p222+p223-1;
```

Now we can see the image variety is defined by five polynomials of ideal J. And the first four polynomials are the determinants in Equation (22.1) and the last one corresponds to the sum 1 constant. We can also get the five polynomials by computing Gröbner basis.

$$\begin{vmatrix} p_{121} & p_{211} & p_{221} \\ p_{122} & p_{212} & p_{222} \\ p_{123} & p_{213} & p_{223} \end{vmatrix} \quad \begin{vmatrix} p_{1+1} & p_{211} & p_{221} \\ p_{1+2} & p_{212} & p_{222} \\ p_{1+3} & p_{213} & p_{223} \end{vmatrix} \quad \begin{vmatrix} p_{+11} & p_{121} & p_{221} \\ p_{+12} & p_{122} & p_{222} \\ p_{+13} & p_{123} & p_{223} \end{vmatrix}$$

$$\begin{vmatrix} p_{111} & p_{121}+p_{211} & p_{221} \\ p_{112} & p_{122}+p_{212} & p_{222} \\ p_{113} & p_{123}+p_{213} & p_{223} \end{vmatrix} \tag{22.1}$$

```
ideal J=groebner(I);
```

Using the above command "GROEBNER", we will get an ideal J defined by 184 polynomials. Among them, the first five polynomials only involve the variable p_{ijk} and they are the five polynomials we have got before. When using the "GROEBNER" command, please be aware that the resulting basis is subject to the monomial ordering you choose for defining the ring.

To compute the dimension of the ideal, we need to define another ring r_1 only with unknowns p_{ijk} and then an ideal (which we also call J) defined by the above five polynomials. Note that the dimension of the ideal and the size of the Gröbner basis for the ideal are different things.

```
ring r1=0, (p111,p112,p113,p121,p122,p123,p211,p212,p213,p221,p222,
p223), lp;
ideal J;
J[1]=p121*p212*p223-p121*p213*p222-....;
J[2]=p112*p211*p223+p112*p212*p223-p112*p213*p221-....;
J[3]=p112*p121*p223+p112*p122*p223-p112*p123*p221-....;
J[4]=p112*p121*p213+p112*p121*p223+p112*p122*p213+....;
J[5]=p111+p112+p113+p121+p122+p123+p211+p212+p213+p221+p222+p223-1;
dim(groebner(J));
===> 7
```

Table 22.2 lists the effective dimenions of some latent class models which have been considered so far. (Kocka and Zhang 2002) have showed that the maximal numerical rank of the Jacobian of polynomial mapping equals the symbolic rank and the numerical rank reaches the maximal rank almost surely. Therefore, although it is impossible to compute the symbolic rank of the Jacobian or to compute the

dimension of the image variety, we can calculate the numerical rank of the Jacobian at many points to find the possible maximal rank.

Latent class model		Effective dimension	
dim of table	num of latent class	dim of image variety	max numerical rank of Jacobi
2×2	$r = 2$	3	3
3×3	$r = 2$	7	7
4×5	$r = 3$	17	17
$2 \times 2 \times 2$	$r = 2$	7	7
$2 \times 2 \times 2$	$r = 3$	7	7
$2 \times 2 \times 2$	$r = 4$	7	7
$3 \times 3 \times 3$	$r = 2$	N/A	13
$3 \times 3 \times 3$	$r = 3$	N/A	20
$3 \times 3 \times 3$	$r = 4$	N/A	25
$3 \times 3 \times 3$	$r = 5$	N/A	26
$3 \times 3 \times 3$	$r = 6$	N/A	26
$5 \times 2 \times 2$	$r = 3$	N/A	17
$4 \times 2 \times 2$	$r = 3$	N/A	14
$3 \times 3 \times 2$	$r = 5$	N/A	17
$6 \times 3 \times 2$	$r = 5$	N/A	34
$10 \times 3 \times 2$	$r = 5$	N/A	54
$2 \times 2 \times 2 \times 2$	$r = 2$	N/A	9
$2 \times 2 \times 2 \times 2$	$r = 3$	N/A	13
$2 \times 2 \times 2 \times 2$	$r = 4$	N/A	15
$2 \times 2 \times 2 \times 2$	$r = 5$	N/A	15
$2 \times 2 \times 2 \times 2$	$r = 6$	N/A	15

Table 22.2 *Effective dimensions of some latent class models. 'N/A' means it is computationally infeasible.*

22.2.2 Solving Polynomial Equations

SINGULAR can also be used to solve polynomial equations. For example, in the 100 Swiss Franks Problem, we need to solve the optimization problem in Equation (22.2).

$$\ell(\mathbf{p}) = \sum_{i,j} n_{ij} \log p_{ij}, \qquad \mathbf{p} \in \Delta_{15}, \ \det(\mathbf{p}^*_{ij}) = 0 \text{ all } i, j \in [4], \tag{22.2}$$

where $\mathbf{p}^*_{ij}$ is the 3×3 sub-matrix of $\mathbf{p}$ obtained by erasing the ith row and the jth column. Using Lagrange multipliers method, the objective becomes finding all the local extrema of the below function $H(\cdot)$

$$H(p_{ij}, h_0, h_{ij}) = \sum_{i,j} n_{ij} \log p_{ij} + h_0 \left(\sum_{i,j} p_{ij} - 1 \right) + h_{ij} \det \mathbf{p}^*_{ij} \tag{22.3}$$

Taking the derivative of $H(\cdot)$ with respect to p_{ij}, h_0 and h_{ij}, we get a system of 33 polynomial functions. In SINGULAR, we can define the ideal generated by these 33 polynomials.

```
ring r=0, (p11,p21,p31,p41,p12,p22,p32,p42,p13,p23,p33,p43,p14,p24,p34,p44,
h11,h21,h31,h41,h12,h22,h32,h42,h13,h23,h33,h43,h14,h24,h34,h44,h0), lp;
ideal I=4+h0*p11+h23*p11*p32*p44-h23*p11*p34*p42+h24*p11*p32*p43 ...,
2+h0*p21+h13*p21*p32*p44-h13*p21*p34*p42+h14*p21*p32*p43 ...,
2+h0*p31-h13*p31*p22*p44+h13*p31*p24*p42-h14*p31*p22*p43 ...,
2+h0*p41+h13*p41*p22*p34-h13*p41*p24*p32+h14*p41*p22*p33 ...,
2+h0*p12-h23*p31*p12*p44+h23*p41*p12*p34-h24*p31*p12*p43 ...,
4+h0*p22-h13*p22*p31*p44+h13*p41*p22*p34-h14*p22*p31*p43 ...,
2+h0*p32+h13*p32*p21*p44-h13*p41*p24*p32+h14*p32*p21*p43 ...,
2+h0*p42-h13*p42*p21*p34+h13*p42*p31*p24-h14*p42*p21*p33 ...,
2+h0*p13+h24*p42*p31*p13-h24*p41*p13*p32-h21*p32*p13*p44 ...,
2+h0*p23+h14*p42*p31*p23-h14*p41*p23*p32-h11*p32*p23*p44 ...,
4+h0*p33-h14*p42*p21*p33+h14*p41*p22*p33+h11*p22*p33*p44 ...,
2+h0*p43+h14*p32*p21*p43-h14*p22*p31*p43-h11*p22*p34*p43 ...,
2+h0*p14+h23*p31*p14*p42-h23*p41*p14*p32+h21*p32*p14*p43 ...,
2+h0*p24+h13*p42*p31*p24-h13*p41*p24*p32+h11*p32*p24*p43 ...,
2+h0*p34-h13*p42*p21*p34+h13*p41*p22*p34-h11*p22*p34*p43 ...,
4+h0*p44+h13*p32*p21*p44-h13*p22*p31*p44+h11*p22*p33*p44 ...,
p22*p33*p44-p22*p34*p43-p32*p23*p44+p32*p24*p43+p42*p23*p34-p42*p24*p33,
p12*p33*p44-p12*p34*p43-p32*p13*p44+p32*p14*p43+p42*p13*p34-p42*p14*p33,
p12*p23*p44-p12*p24*p43-p22*p13*p44+p22*p14*p43+p42*p13*p24-p42*p14*p23,
p12*p23*p34-p12*p24*p33-p22*p13*p34+p22*p14*p33+p32*p13*p24-p32*p14*p23,
p21*p33*p44-p21*p34*p43-p31*p23*p44+p31*p24*p43+p41*p23*p34-p41*p24*p33,
p11*p33*p44-p11*p34*p43-p31*p13*p44+p31*p14*p43+p41*p13*p34-p41*p14*p33,
p11*p23*p44-p11*p24*p43-p21*p13*p44+p21*p14*p43+p41*p13*p24-p41*p14*p23,
p11*p23*p34-p11*p24*p33-p21*p13*p34+p21*p14*p33+p31*p13*p24-p31*p14*p23,
p21*p32*p44-p21*p34*p42-p31*p22*p44+p31*p24*p42+p41*p22*p34-p41*p24*p32,
p11*p32*p44-p11*p34*p42-p31*p12*p44+p31*p14*p42+p41*p12*p34-p41*p14*p32,
p11*p22*p44-p11*p24*p42-p21*p12*p44+p21*p14*p42+p41*p12*p24-p41*p14*p22,
p11*p22*p34-p11*p24*p32-p21*p12*p34+p21*p14*p32+p31*p12*p24-p31*p14*p22,
p21*p32*p43-p21*p33*p42-p31*p22*p43+p31*p23*p42+p41*p22*p33-p41*p23*p32,
p11*p32*p43-p11*p33*p42-p31*p12*p43+p31*p13*p42+p41*p12*p33-p41*p13*p32,
p11*p22*p43-p11*p23*p42-p21*p12*p43+p21*p13*p42+p41*p12*p23-p41*p13*p22,
p11*p22*p33-p11*p23*p32-p21*p12*p33+p21*p13*p32+p31*p12*p23-p31*p13*p22,
p11+p21+p31+p41+p12+p22+p32+p42+p13+p23+p33+p43+p14+p24+p34+p44-1;
```

By using the routine 'SOLVE' in SINGULAR we can find the numerical solutions to the system of polynomial equations.

```
LIB 'solve.lib';
solve(I, 6, 0 , 'nodisplay');
```

Unfortunately, the system we want to solve is beyond what SINGULAR can handle. But we can check whether a given table $\{p_{ij}\}$ is a solution to the system or not, by substituting the values of p_{ij} into the ideal I. And if the resulting ideal is not an empty set, then $\{p_{ij}\}$ is a solution to the system.

```
LIB "poly.lib"
ideal v=p11,p21,p31,p41,p12,p22,p32,p42,p13,p23,p33,p43,p14,p24,p34,p44;
ideal p=3/40,3/40,2/40,2/40,3/40,3/40,2/40,2/40,2/40,2/40,3/40,3/40,
2/40,2/40,3/40,3/40;
ideal J=substitute(I,v,p);
dim(std(J));
===> 28
```

It should be noted that the reason we get a dimension 28 is that the ideal v and p are defined on the ring r which has additional 17 unknowns other than p_{ij}. No matter what the number is, the positiveness of the number means p is a solution for p_{ij}. Otherwise, if it is zero, p is not a solution for p_{ij}.

22.2.3 Plotting Unidentifiable Space

For the 100 Swiss Franks problem, we know that

$$\frac{1}{40}\begin{pmatrix} 3 & 3 & 2 & 2 \\ 3 & 3 & 2 & 2 \\ 2 & 2 & 3 & 3 \\ 2 & 2 & 3 & 3 \end{pmatrix}$$

is one MLE for the 2-level latent class model, that is, the MLE maximizing Equation (22.2). And we also know there is a 2-dimensional subspace in the parameter space of conditional probabilities corresponding to this MLE. Now we show how to find the equations defining this unidentifiable space. In the below code, w_t's are the marginal probabilities of the latent variable, a_{it}'s and b_{jt}'s are the conditional probabilities of the observed variables given the latent variable. Then we define an ideal I, in which the first 5 polynomials corresponds to the sum 1 constraints and the last 16 polynomials corresponds to the model equalities $p_{ij} = \sum_t w_t a_{it} b_{jt}$ for the MLE.

```
ring r=0, (w1,a11,a21,a31,a41,b11,b21,b31,b41,
w2,a12,a22,a32,a42,b12,b22,b32,b42), lp;
ideal I=w1+w2-1,
a11+a21+a31+a41-1,
a12+a22+a32+a42-1,
b11+b21+b31+b41-1,
b12+b22+b32+b42-1,
w1*a11*b11+w2*a12*b12-3/40,
w1*a11*b21+w2*a12*b22-3/40,
w1*a11*b31+w2*a12*b32-2/40,
w1*a11*b41+w2*a12*b42-2/40,
w1*a21*b11+w2*a22*b12-3/40,
w1*a21*b21+w2*a22*b22-3/40,
w1*a21*b31+w2*a22*b32-2/40,
w1*a21*b41+w2*a22*b42-2/40,
w1*a31*b11+w2*a32*b12-2/40,
w1*a31*b21+w2*a32*b22-2/40,
w1*a31*b31+w2*a32*b32-3/40,
w1*a31*b41+w2*a32*b42-3/40,
w1*a41*b11+w2*a42*b12-2/40,
w1*a41*b21+w2*a42*b22-2/40,
w1*a41*b31+w2*a42*b32-3/40,
w1*a41*b41+w2*a42*b42-3/40;
dim(std(I));
===> 2
```

Now we can see the dimension of the ideal I is really 2. Then we can eliminate the unknowns other than w_1, a_{11}, b_{11} from the ideal I, thus we get the equation for the projection of the 2-dimensional unidentifiable subspace in (w_1, a_{11}, b_{11}) coordinates.

```
ideal J=elim1(I, a21*a31*a41*b21*b31*b41*w2*a12*a22*a32*a42
*b12*b22*b32*b42);
J;
===> J[1]=80*w1*a11*b11-20*w1*a11-20*w1*b11+6*w1-1;
```

The resulting ideal J has a one-to-one correspondence to the identifiable space. This is because the unidentifiable space is 2-dimensional, thus once the values of w_1, a_{11} and b_{11} are known so do the other paramters.

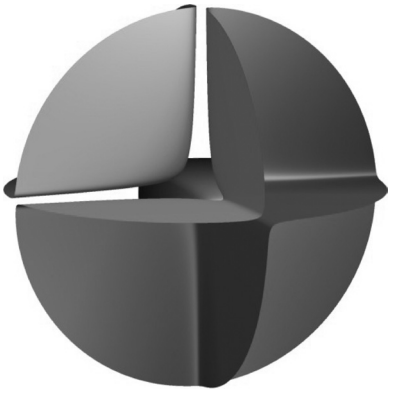
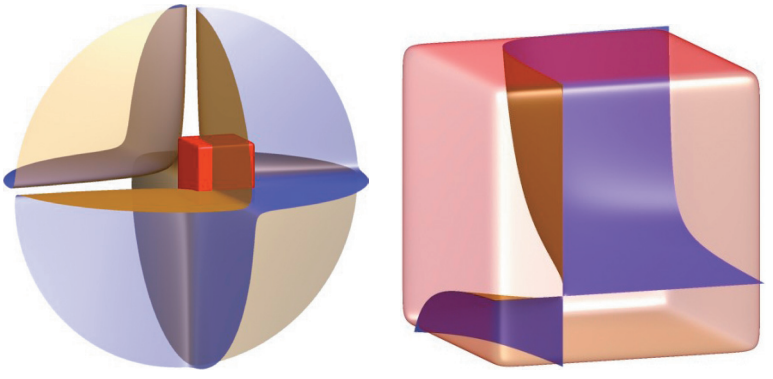


Fig. 22.2 The surface that the ideal J is vanishing.



the vanishing surface
(a) intersected with the unit cube (b) inside the unit cube

Fig. 22.3 The intersection of the vanishing surface for ideal J and the $[0,1]^3$ cube.

```
LIB "surf.lib";
ring r2=0, (w1, a11, b11), lp;
ideal J=80*w1*a11*b11-20*w1*a11-20*w1*b11+6*w1-1;
plot(J);
```

SINGULAR calls the programme `surf` to draw real pictures of plane curves and surfaces in 3-D space. If you load library "SURF.LIB" in SINGULAR and execute the "PLOT" command to show the vanishing surface of the ideal J, you will get a picture in Figure (22.2).

But the surface showed in figure 22.2 doesn't guarantee w_1, a_{11}, b_{11} to be within 0 and 1. If we want to plot more sophisticated surfaces, we can use the stand-alone programme `surf`. The unidentifiable space is the intersection of the vanishing surface and the $[0,1]^3$ cube, which is shown in Figure (22.3). We include the script used in `surf` to draw the pictures in the next section.

22.2.4 `Surf` *Script*

Below is the script used in `surf` to draw the pictures in figure 22.3-(b).

```
width = 500; height = 500; double pi = 3.1415926; double ss = 0.15;
origin_x = -0.5; origin_y = -0.5; origin_z = 0;
clip = cube;radius = 0.5;center_x = 0.5;center_y = 0.5;center_z = 0.5;
scale_x = ss;scale_y = ss;scale_z = ss;
rot_x = pi / 180 * 10;rot_y = - pi / 180 * 20;rot_z =  pi / 180 * 0;
antialiasing = 4;antialiasing_threshold = 0.05;antialiasing_radius = 1.5;
surface2_red = 255;surface2_green = 0;surface2_blue = 0;
inside2_red = 255;inside2_green = 0;inside2_blue = 0;
transparence = 0;transparence2 = 70;
illumination = ambient_light + diffuse_light + reflected_light + transmitted_light;
surface = 80*x*y*z - 20*x*z - 20*y*z + 6*z -1;
surface2 = (x-0.500)^30 + (y-0.500)^30+(z-0.500)^30 - (0.499)^30;
clear_screen;
draw_surface;
```

22.3 Proof of the Fixed Points for 100 Swiss Franks Problem

In this section, we show that when maximizing the log-likelihood function of 2-level latent class model for the 100 Swiss Franks problem, the table

$$f = \frac{1}{40}\begin{pmatrix} 3 & 3 & 2 & 2 \\ 3 & 3 & 2 & 2 \\ 2 & 2 & 3 & 3 \\ 2 & 2 & 3 & 3 \end{pmatrix} \tag{22.4}$$

is a fixed point in the Expectation Maximization algorithm. Here the observed table is

$$p = \frac{1}{40}\begin{pmatrix} 4 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 4 \end{pmatrix}$$

Under the conditional independence of the latent structure model, we have

$$f_{ij} = \sum_{t\in\{0,1\}} \lambda_t \alpha_{it} \beta_{jt}$$

where $\sum_t \lambda_t = \sum_i \alpha_{it} = \sum_j \beta_{jt} = 1$, $\lambda_t \geq 0$, $\alpha_{it} \geq 0$ and $\beta_{jt} \geq 0$.

Now, we show that if we start with the values such that

$$\begin{array}{l} \alpha_{1t} = \alpha_{2t},\ \alpha_{3t} = \alpha_{4t} \\ \beta_{1t} = \beta_{2t},\ \beta_{3t} = \beta_{4t} \\ \sum_t \lambda_t \alpha_{1t} \beta_{1t} = \sum_t \lambda_t \alpha_{3t} \beta_{3t} = 3/40 \\ \sum_t \lambda_t \alpha_{1t} \beta_{3t} = \sum_t \lambda_t \alpha_{3t} \beta_{1t} = 2/40 \end{array} \tag{22.5}$$

then the EM will stay in these values and the fitted table is right the one in Equation (22.4). In fact, in the E step, the posterior probability is updated by

$$\pi_{ijt}^{AB\bar{X}} = P(X = t | A = i, B = j) = \frac{\lambda_t \alpha_{it} \beta_{jt}}{f_{ij}}$$

Then in the M step, the parameters are updated by

$$\begin{aligned}
\hat{\lambda}_t &= \textstyle\sum_{i,j} p_{ij}\pi_{ijt}^{AB\bar{X}} \\
&= \textstyle\sum_{i,j} p_{ij}\frac{\lambda_t\alpha_{it}\beta_{jt}}{f_{ij}} \\
&= \lambda_t + \tfrac{1}{3}[\alpha_{1t}\beta_{1t} + \alpha_{2t}\beta_{2t} + \alpha_{3t}\beta_{3t} + \alpha_{4t}\beta_{4t}] \\
&\quad -\tfrac{1}{3}[\alpha_{1t}\beta_{2t} + \alpha_{2t}\beta_{1t} + \alpha_{3t}\beta_{4t} + \alpha_{4t}\beta_{3t}] = \lambda_t \\
\hat{\alpha}_{it} &= \textstyle\sum_j p_{ij}\pi_{ijt}^{AB\bar{X}}/\hat{\lambda}_t \\
&= \alpha_{it}\textstyle\sum_j p_{ij}\beta_{jt}/f_{ij} \\
&= \begin{cases}
\alpha_{it}[1 + \frac{1}{3}\beta_{1t} - \frac{1}{3}\beta_{2t}], & i = 1 \\
\alpha_{it}[1 + \frac{1}{3}\beta_{2t} - \frac{1}{3}\beta_{1t}], & i = 2 \\
\alpha_{it}[1 + \frac{1}{3}\beta_{3t} - \frac{1}{3}\beta_{4t}], & i = 3 \\
\alpha_{it}[1 + \frac{1}{3}\beta_{4t} - \frac{1}{3}\beta_{3t}], & i = 4
\end{cases} = \alpha_{it} \\
\hat{\beta}_{jt} &= \textstyle\sum_i p_{ij}\pi_{ijt}^{AB\bar{X}}/\hat{\lambda}_t \\
&= \beta_{jt}\textstyle\sum_i p_{ij}\alpha_{it}/f_{ij} \\
&= \begin{cases}
\beta_{jt}[1 + \frac{1}{3}\alpha_{1t} - \frac{1}{3}\alpha_{2t}], & j = 1 \\
\beta_{jt}[1 + \frac{1}{3}\alpha_{2t} - \frac{1}{3}\alpha_{1t}], & j = 2 \\
\beta_{jt}[1 + \frac{1}{3}\alpha_{3t} - \frac{1}{3}\alpha_{4t}], & j = 3 \\
\beta_{jt}[1 + \frac{1}{3}\alpha_{4t} - \frac{1}{3}\alpha_{3t}], & j = 4
\end{cases} = \beta_{jt}
\end{aligned}$$

Thus, we have proved that the starting point given by Equation (22.5) is a fixed point in the EM algorithm. And this fixed point will give us the fitted table f in Equation (22.4). However, this is not the only fixed points for the EM. In fact, according to the above, we can also show that the points

$$\alpha_{1t} = \alpha_{3t},\ \alpha_{2t} = \alpha_{4t},\ \ \beta_{1t} = \beta_{3t},\ \beta_{2t} = \beta_{4t}$$

and

$$\alpha_{1t} = \alpha_{4t},\ \alpha_{2t} = \alpha_{3t},\ \ \beta_{1t} = \beta_{4t},\ \beta_{2t} = \beta_{3t}$$

are fixed points too. And the two points will lead to the tables

$$\tfrac{1}{40}\begin{pmatrix} 3 & 2 & 3 & 2 \\ 2 & 3 & 2 & 3 \\ 3 & 2 & 3 & 2 \\ 2 & 3 & 2 & 3 \end{pmatrix} \quad \text{and} \quad \tfrac{1}{40}\begin{pmatrix} 3 & 2 & 2 & 3 \\ 2 & 3 & 3 & 2 \\ 2 & 3 & 3 & 2 \\ 3 & 2 & 2 & 3 \end{pmatrix}$$

Similarly, we can show that the table

$$\frac{1}{40}\begin{pmatrix} 4 & 2 & 2 & 2 \\ 2 & 8/3 & 8/3 & 8/3 \\ 2 & 8/3 & 8/3 & 8/3 \\ 2 & 8/3 & 8/3 & 8/3 \end{pmatrix}$$

and its permutations are also the fixed points in the EM algorithm.

22.4 Matlab Codes

Here we include the two matlab subroutines which are used to compute the Jacobian of the polynomial mapping $f\colon \Delta_{d_1-1} \times \cdots \times \Delta_{d_k-1} \times \Delta_{r-1} \to \Delta_{d-1}$ $(d = \prod_i d_i)$ in

Equation (22.6) and its numerical rank for latent class models

$$(p_1(i_1)\dots p_k(i_k),\lambda_h)\mapsto\sum_{h\in[r]}p_1(i_1)\dots p_k(i_k)\lambda_h. \tag{22.6}$$

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [J,f,x,w,a] = jacob_lcm(T, I)
% ------------------------------------------------------------------------
% JACOB_LCM computes the Jacobian of the latent class model.
% For example:
%       [J, f, x, w, a] = jacob_lcm(2, [3,3,3]);
%
w = sym('', 'real');
a = sym('', 'real');
for t=1:T
    w(end+1) = sym(['w', int2str(t)], 'real');
    for k=1:length(I)
        for i=1:I(k)
            a{k}(i,t) = sym(['a', int2str(i), int2str(t), int2str(k)], 'real');
        end
    end
end
w(end) = 1 - sum(w(1:end-1));
x = w(1:end-1);
for k=1:length(I)
    for t=1:T
        a{k}(end,t) = 1 - sum(a{k}(1:end-1,t));
        x = [x, a{k}(1:end-1,t)'];
    end
end
% get the mapping from parameters to table
f = sym('', 'real');
for idx=1:prod(I)
    subv = ind2subv(I, idx);
    val = sym('0');
    for t=1:T
        temp = w(t);
        for k=1:length(I)
            temp = temp * a{k}(subv(k),t);
        end
        val = val + temp;
    end
    f(end+1) = val;
end
% get the Jacobian
J = jacobian(f, x);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function r = rank_lcm(J, w, a)
% ------------------------------------------------------------------------
% RANK_LCM computes the numberical rank of the sybotical matri 'J', which
% is a function of 'w' and 'a'. It is used after calling the funtion JACOB_LCM.
% For example,
%       [J,f,x,w,a] = jacob_lcm(2, [2,2,2,2]);
%       rank_lcm(J,w,a);
%
T = length(w);
I = zeros(1, length(a));
for k=1:length(a)
    I(k) = size(a{k},1);
end
% compute the numberical rank
v = unifrnd(0,1,1,T);
v = v ./ sum(v);
```

```
for t=1:T
    for k=1:length(I)
        b{k}(:,t) = unifrnd(0,1,I(k),1);
        b{k}(:,t) = b{k}(:,t) ./ sum(b{k}(:,t));
    end
end
JJ = zeros(size(J));
for  i=1:size(J,1)
    for j=1:size(J,2)
        cc = char(J(i,j));
        for t=1:T
            cc = strrep(cc, char(w(t)), num2str(v(t)));
            for k=1:length(I)
                for p=1:I(k)
                    cc = strrep(cc, char(a{k}(p,t)), num2str(b{k}(p,t)));
                end
            end
        end
        JJ(i,j) = eval(cc);
    end
end
r = rank(JJ);
```

Here are the EM and Newton-Raphson codes for maximum likelihood estimation in latent class models.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [nhat,m,b,se,llk,retcode,X] = LCM_newton(n,T,maxiter,eps,m,X,verbose)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% INPUT:
%       n(required):          observed table, a multi-dimensional array
%       T(required):          number of latent classes
%       maxiter(required):    maximum number of iterations
%       eps(required):        converge threshold
%       m(optional):          initial value for the mean vector
%       X(optional):          design matrix
%       verbose(optional):    display results if true
% OUTPUT:
%       nhat:       estimated observed table
%       m:          estimated probability for the full table
%       b:          estimated parameter
%       se:         standard error of mle
%       llk:        log-likelihood values in iterations
%       retcode:    1, if the algorithm terminates normally; 0, otherwise
%       X:          design matrix
%
dbstop if warning;
dbstop if error;
%
% 1. initialize
y = n(:);                                   % observed table
k = length(y);                              % number of cells
dim = size(n);                              % dimensions of observed table
s = catrep(2, T, [1:k]);
S = zeros(T*k, k);                          % scatter matrix ===> S'm = nhat
for i=1:k
    idx = find( s==i );
    S(idx, i) = 1;
end
z = S * inv(S'*S) * y;                      % observed full table ===> S'z = y
fulldim = [dim, T];                         % dimensions of full table
if nargin < 7     verbose = 1; end
if nargin < 6     X = [];  end
if nargin < 5    m = [];  end
```

```
if isempty(X)
    X = zeros(T*k, 1+(T-1)+sum(dim-1)+sum((T-1)*(dim-1)));  % design matrix
    for idx=1:prod(fulldim)
        % for main effect
        xrow = 1;
        % for first order effect
        G = {};
        subv = ind2subv(fulldim, idx);
        for i=1:length(subv)
            if subv(i)==fulldim(i)
                G{i} = - ones(fulldim(i)-1, 1);
            else
                G{i} = zeros(fulldim(i)-1, 1);
                G{i}(subv(i)) = 1;
            end
            xrow = [xrow, G{i}'];
        end
        % for second order effect
        for i=1:length(subv)-1
            temp = G{end} * G{i}';
            xrow = [xrow, temp(:)'];
        end
        %
        if length(xrow)~=size(X,2)
            keyboard;
        end
        X(idx,:) = xrow;
    end
end
if isempty(m)
    b = unifrnd(-1, 1, size(X,2), 1);           % initial value of the parameter
    m = exp(X*b);                               % estimated mean counts
else
    b = inv(X'*X) * (X' * log(m));
    m = exp(X*b);
end
%
% 2. newton-raphson
llk = sum(y .* log(S' * m ./ sum(m)));
retcode = 1;
for i=1:maxiter
    % Jacobi
    A = S'*diag(m)*S;
    if min(diag(A))<eps     % A is diagonal
        disp('maxtrix A for the Jacobi is singular.');
        disp('the algorithm stops without converging.');
        retcode = 0;
        break;
    end
    A = inv(A);
    P = S * A * S';
    J = (z-m)' * P * diag(m) * X;
% Hessian
    C = X' * (diag(z' * P) * diag(m) - diag(m) *
           (S * diag(y) * (A^2) * S') * diag(m)) * X;
    D = X' * diag(m) * X;
    H = C - D;
    if  max(eig(H)) >= 0
       H = -D;
    end
    [eigvec, eigval] = eig(H);
    eigval = diag(eigval);
    if  min(eigval) >= 0
        disp('the hessian matrix is non-negative definite.');
```

```
        retcode = 0;
        break;
    end
    eigval(find(eigval<0)) = 1 ./ eigval(find(eigval<0));
    eigval(find(eigval>=0)) = 0;
    db = eigvec * diag(eigval) * eigvec' * J';
    ss = 1;
    b = b - ss * db;
    m = exp(X*b);
    % log-likelihood
    llk(end+1) = sum(y .* log(S' * m ./ sum(m)));
    %if abs(llk(end)-llk(end-1))<eps
    if max(abs(J)) < eps
        disp(['algorithm convergs in ', int2str(i), ' steps.']);
        break;
    end
end
% log-likelihood
llk = llk;
% fitted table
nhat = S'* (m ./ sum(m)) * sum(n(:));
% standard errors
se = sqrt(-diag(inv(H)));
%
% 3. show results
if verbose
    disp('the fitted and observed counts:');
    disp([nhat, n(:)]);
    disp('mle and stand error of the parameter:');
    disp([b, se]);
    plot(llk);
    axis tight;
    xlabel('iteration');
    ylabel('log-likelihood');
end


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [f,m,llk,llr,df,c,p,devbuf,c00,p00]=em_lsm(n,T,maxiter,eps,c0,p0)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% EM algorithm for latent class model
%
% input:
%       n(required):        obserbed table. multi-dimensional array
%       T(required):        number of latent classes
%       maxiter(required):  maximum number of iterations
%       eps(required):      converge threshold
%       c0(optional):       initial value for class probabilities
%       p0(optional):       initial value for conditional probabilities
% output:
%       f:          fitted table
%       m:          expected mean vector
%       llk:        log-likelihoods
%       llr:        likelihood ratio statistic
%       df:         degree of freedoms
%       c:          class probabilities
%       p:          conditional probabilities
%       devbuf:     maximum deviations of the estimates in iterations
%       c00:        initial class probabilties
%       p00:        initial conditional probabilities
%
dbstop if warning;
f0 = n;
n = n / sum(n(:));
sz = size(n);
```

```
if nargin < 6
    p0 = cell(1, length(sz));
    for i=1:length(p0)
        A = rand(sz(i), T);
        A = A ./ kron(ones(sz(i),1), sum(A, 1));
        p0{i} = A;
    end
end
if nargin < 5
    c0 = rand(1,T);
    c0 = c0 ./ sum(c0);
end
c00 = c0;
p00 = p0;
nn = zeros([sz, T]);
c = c0;
p = p0;
iter = 0;
devbuf = [];
llk = 0;
while iter < maxiter
    % E step
    for idx=1:prod(size(nn))
        subv = ind2subv(size(nn), idx);
        nn(idx) = c(subv(end));
        for i=1:length(sz)
            nn(idx) = nn(idx) * p{i}(subv(i), subv(end));
        end
    end
    nnhat = sum(nn, length(sz)+1);
    nnhat = catrep(length(sz)+1, T, nnhat);
    nnhat = nn ./ nnhat;
% M step
    for t=1:T
        A = subarray(length(sz)+1, t, nnhat);
        A = n .* A;
        c(t) = sum(A(:));
        for i=1:length(sz)
            for k=1:sz(i)
                B = subarray(i, k, A);
                p{i}(k, t) = sum(B(:)) / c(t);
            end
        end
    end
% mle of counts
    f = zeros([sz, T]);
    for idx=1:prod(size(f))
        subv = ind2subv(size(f), idx);
        f(idx) = c(subv(end));
        for i=1:length(sz)
            f(idx) = f(idx) * p{i}(subv(i), subv(end));
        end
    end
    f = sum(f, length(sz)+1);
    llk(end+1) = sum( f0(:) .* log(f(:)) );
% if converged
    maxdev = max(abs(c-c0));
    for i=1:length(p)
        A = abs(p{i}-p0{i});
        maxdev = max(maxdev, max(A(:)));
    end
    devbuf = [devbuf, maxdev];
    if maxdev < eps
        disp(['algorithm converges in ', int2str(iter), ' steps.']);
```

```
        break;
    end
    c0 = c;
    p0 = p;
    iter = iter + 1;
end
% frequencies estimation
f = zeros([sz, T]);
for idx=1:prod(size(f))
    subv = ind2subv(size(f), idx);
    f(idx) = c(subv(end));
    for i=1:length(sz)
        f(idx) = f(idx) * p{i}(subv(i), subv(end));
    end
end
m = f;      % full table
f = sum(f, length(sz)+1);
f = f .* sum(f0(:));
% likelihood ratio test statistics
f0 = f0(:);
f1 = f(:);
llr = f0./f1;
llr( find(llr==0) ) = 1;
llr = 2 * sum( f0.*log(llr) );
% degree of freedom
df = (prod(size(n))-1) - (T-1+T*sum(size(n)-1));
llk = llk(2:end);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function C = catrep(dim, n, A) str = ['C = cat(', int2str(dim), ',']; for i=1:n
    str = [str, 'A,'];
end
str = [str(1:end-1), ');'];
eval(str);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function subv = ind2subv(siz, idx) fn = '[';
for k=1:length(siz)
    fn = [fn, 'subv(', num2str(k), '),'];
end
fn = [fn(1:length(fn)-1), '] = ind2sub(siz, idx);'];
eval(fn);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function ind = subv2ind(siz, subv)
fn = 'ind = sub2ind(siz, ';
for k=1:length(siz)
    fn = [fn, 'subv(', num2str(k), '),'];
end
fn = [fn(1:length(fn)-1), ');'];
eval(fn);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function C = subarray(dim, idx, A)
str = 'C = A(';
for i=1:length(size(A))
    if i==dim
        str = [str, int2str(idx), ','];
    else
        str = [str, ':,'];
    end
end
str = [str(1:end-1), ');'];
eval(str);
squeeze(C);
```

Bibliography

Catalisano, M.V., Geramita, A.V. and Gimigliano, A. (2002). Ranks of tensors, secant varieties of Segre varieties and fat points, *Linear Algebra and Its Applications* **355**, 263–285. Corrigendum **367**, 347–348 (2003).

Kocka, T. and Zhang, N. L. (2002). Dimension correction for hierarchical latent class models, *Proceeding of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-02)*, 267–274, Morgan Kaufmann.

Pachter, L. and Sturmfels, B., eds. (2005). *Algebraic Statistics for Computational Biology* (New York, Cambridge University Press).

23

On-line Supplement to The generalized shuttle algorithm

Adrian Dobra

Stephen E. Fienberg

23.1 Proofs

Proposition 8.1 *Let n^* be the count in the $(1,1,\ldots,1)$ cell. Consider an index $i^0 = (i_1^0, i_2^0, \ldots, i_k^0) \in \mathcal{I}$. Let $\{q_1, q_2, \ldots, q_l\} \subset K$ such that, for $r \in K$, we have*

$$i_r^0 = \begin{cases} 1, & \text{if } r \in K \setminus \{q_1, q_2, \ldots, q_l\}, \\ 2, & \text{if } r \in \{q_1, q_2, \ldots, q_l\}. \end{cases}$$

For $s = 1, 2, \ldots, l$, denote $C_s := K \setminus \{q_s\}$. Then

$$n(i^0) = (-1)^l \cdot n^* - \sum_{s=0}^{l-1} (-1)^{l+s} \cdot n_{C_{(l-s)}}(1, \ldots, 1, i^0_{q_{(l-s)}+1}, \ldots, i_k^0). \tag{23.1}$$

Proof We start from the $(1,1,\ldots,1)$ cell and go through a sequence of cells $n(i)$ until we reach $n(i^0)$. We can write

$$\begin{aligned} n^* &= n_{C_l}(1, \ldots, 1, i^0_{q_l+1}, \ldots, i_k^0) - n(1, \ldots, 1, i^0_{q_l}, \ldots, i_k^0), \\ n(1, \ldots, 1, i^0_{q_l}, \ldots, i_k^0) &= n_{C_{(l-1)}}(1, \ldots, 1, i^0_{q_{(l-1)}+1}, \ldots, i_k^0) - n(1, \ldots, 1, i^0_{q_{(l-1)}}, \ldots, i_k^0), \\ &\vdots \\ n(1, \ldots, 1, i^0_{q_2}, \ldots, i_k^0) &= n_{C_1}(1, \ldots, 1, i^0_{q_1+1}, \ldots, i_k^0) - n(i^0). \end{aligned}$$

We add the above equalities to obtain Equation (23.1). □

Proposition 8.2 *The generalized shuttle algorithm converges to the bounds in equations*

$$L(n^*) = \max\left\{ \sum_{s=0}^{l-1} (-1)^s \cdot n_{C_{(l-s)}}(1, \ldots, 1, i^0_{q_{(l-s)}+1}, \ldots, i_k^0) : l \text{ even} \right\}, \tag{23.2}$$

and

$$U(n^*) = \min\left\{ \sum_{s=0}^{l-1} (-1)^s \cdot n_{C_{(l-s)}}(1, \ldots, 1, i^0_{q_{(l-s)}+1}, \ldots, i_k^0) : l \text{ odd} \right\}. \tag{23.3}$$

Proof We write the equalities in the proof of Proposition 8.1 as

$$t_{\{1\}\ldots\{1\}} \oplus t_{\{1\}\ldots\{1\}\{i^0_{q_l}\}\ldots\{i^0_k\}} = t_{\{1\}\ldots\{i^0_{q_l-1}\}\{1,2\}\{i^0_{q_l+1}\}\ldots\{1\}},$$

and

$$t_{\{1\}\ldots\{1\}\{i^0_{q_{(s+1)}}\}\ldots\{i^0_k\}} \oplus t_{\{1\}\ldots\{1\}\{i^0_{q_s}\}\ldots\{i^0_k\}} = t_{\{1\}\ldots\{1\}\{i^0_{q_s-1}\}\{1,2\}\{i^0_{q_s+1}\}\ldots\{i^0_k\}},$$

for $s = 1, 2, \ldots, l-1$. Hence

$$\left(t_{\{1\}\ldots\{1\}}, t_{\{1\}\ldots\{i^0_{q_l-1}\}\{1,2\}\{i^0_{q_l+1}\}\ldots\{1\}}, t_{\{1\}\ldots\{1\}\{i^0_{q_l}\}\ldots\{i^0_k\}}\right) \in \mathcal{Q}(\mathbf{T}),$$

and

$$\left(t_{\{1\}\ldots\{1\}\{i^0_{q_{(s+1)}}\}\ldots\{i^0_k\}}, t_{\{1\}\ldots\{1\}\{i^0_{q_s-1}\}\{1,2\}\{i^0_{q_s+1}\}\ldots\{i^0_k\}}, t_{\{1\}\ldots\{1\}\{i^0_{q_s}\}\ldots\{i^0_k\}}\right) \in \mathcal{Q}(\mathbf{T}),$$

for $s = 1, 2, \ldots, l-1$. Since

$$\mathbf{T}'_0 := \left\{t_{\{1\}\ldots\{1\}\{i^0_{q_s-1}\}\{1,2\}\{i^0_{q_s+1}\}\ldots\{i^0_k\}} : s = 1, 2, \ldots, l\right\} \subset \mathbf{T}_0,$$

the cells in $\mathbf{T}'_0$ have a fixed value

$$V\left(t_{\{1\}\ldots\{1\}\{i^0_{q_s-1}\}\{1,2\}\{i^0_{q_s+1}\}\ldots\{i^0_k\}}\right) n_{C_s}(1, \ldots, 1, i^0_{q_s-1}, i^0_{q_s+1}, \ldots, i^0_k),$$

for $s = 1, 2, \ldots, l$. GSA sequentially updates the bounds for the cells in $\mathbf{T}'_0$ in the following way:

$$\begin{aligned}
L\left(t_{\{1\}\ldots\{1\}}\right) &= \max\left\{0, n_{C_l}(1, \ldots, 1) - U\left(t_{\{1\}\ldots\{1\}\{i^0_{q_l}\}\ldots\{i^0_k\}}\right)\right\},\\
U\left(t_{\ldots\{1\}\{i^0_{q_l}\}\ldots}\right) &= \min\left\{n_\phi, n_{C_{(l-1)}}(\ldots, 1, i^0_{q_{(l-1)}+1}, \ldots) - L\left(t_{\ldots\{1\}\{i^0_{q_{(l-1)}}\}\ldots}\right)\right\},\\
&\vdots
\end{aligned}$$

We set the non-negativity constraints

$$L\left(t_{\{1\}\ldots\{1\}\{i^0_{q_s}\}\ldots\{i^0_k\}}\right) \geq 0, \text{ for } s = 1, 2, \ldots, l, \tag{23.4}$$

then combine the above equalities to obtain Equation (23.2). In an analogous manner we obtain the upper bounds in Equation (23.3) from the identities:

$$\begin{aligned}
U\left(t_{\{1\}\{1\}\ldots\{1\}}\right) &= \min\left\{n_\phi, n_{C_l}(1, \ldots, 1) - L\left(t_{\{1\}\ldots\{1\}\{i^0_{q_l}\}\ldots\{i^0_k\}}\right)\right\},\\
L\left(t_{\ldots\{1\}\{i^0_{q_l}\}\ldots}\right) &= \max\left\{0, n_{C_{(l-1)}}(\ldots, 1, i^0_{q_{(l-1)}+1}, \ldots) - U\left(t_{\ldots\{1\}\{i^0_{q_{(l-1)}}\}\ldots}\right)\right\},\\
&\vdots
\end{aligned}$$

Once GSA reaches the bounds in Equations (23.2) and (23.3), no further changes are possible. □

Theorem 8.1 *Equations 23.5 below are sharp bounds given the marginals* $\mathbf{n}_{C_1}$, ..., $\mathbf{n}_{C_p}$:

$$\min\left\{n_{C_1}\left(i_{C_1}\right),\ldots,n_{C_p}\left(i_{C_p}\right)\right\} \geq n(i) \geq \max\left\{\sum_{j=1}^{p} n_{C_j}\left(i_{C_j}\right) - \sum_{j=2}^{p} n_{S_j}\left(i_{S_j}\right), 0\right\} \tag{23.5}$$

Proposition 8.3 *For a subset* $D_0 \subset K$ *and an index* $i^0_{D_0} \in \mathcal{I}_{D_0}$, *the following inequalities hold:*

$$\begin{aligned}\min\left\{n_{C\cap D_0}\left(i^0_{C\cap D_0}\right)|C\in\mathcal{C}(\mathcal{G})\right\} &\geq n_{D_0}(i^0_{D_0}) \\ &\geq \max\left\{0, \sum_{C\in\mathcal{C}(\mathcal{G})} n_{C\cap D_0}\left(i^0_{C\cap D_0}\right) - \sum_{S\in\mathcal{S}(\mathcal{G})} n_{S\cap D_0}\left(i^0_{S\cap D_0}\right)\right\}. \end{aligned} \tag{23.6}$$

The upper and lower bounds in Equation (23.6) are defined to be the Fréchet bounds for the cell entry $n_{D_0}(i^0_{D_0})$ *given* $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$.

Proof The subgraph $\mathcal{G}(D)$ is decomposable since $\mathcal{G}$ is decomposable. Equation (23.6) follows directly from Theorem 8.1 applied for table $\mathbf{n}_D$ which has a fixed set of marginals $\mathbf{n}_{C_1\cap D}$, $\mathbf{n}_{C_2\cap D}$, ..., $\mathbf{n}_{C_p\cap D}$. We clearly have $\mathcal{C}(\mathcal{G}(D)) = \{C_1\cap D, C_2\cap D, \ldots, C_p\cap D\}$ and $\mathcal{S}(\mathcal{G}(D)) = \{S_2\cap D, \ldots, S_p\cap D\}$. □

Lemma 8.1 *Let* $\mathcal{G} = (K, E)$ *be a decomposable independence graph induced by the marginals* $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$. *Consider a subset* $D_0 \subset K$ *and let* $v \in K \setminus D_0$ *be a simplicial vertex of* $\mathcal{G}$. *It is known that a simplicial vertex belongs to precisely one clique, say* $v \in C_1$. *Then finding bounds for a cell* $n_{D_0}(i^0_{D_0})$, $i^0_{D_0} \in \mathcal{I}_{D_0}$, *given* $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$ *is equivalent to finding bounds for* $n_{D_0}(i^0_{D_0})$ *given* $\mathbf{n}_{C_1\setminus\{v\}}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$.

Proof If $\mathcal{G}$ is complete, i.e. $p = 1$, we have $D_0 \subset K = C_1$, hence every entry $n_{D_0}(i^0_{D_0})$ will be fixed. Otherwise, it is known that $(\{v\}, bd(v), V\setminus cl(v))$ is a proper decomposition of $\mathcal{G}$. Since $bd(v)$ is a separator of $\mathcal{G}$, X_v is independent of $X_{V\setminus(v)}$ given $X_{bd(v)}$. Therefore no information is lost if we think about $\mathbf{n}_{D_0}$ as being the marginal of $\mathbf{n}_{V\setminus\{v\}}$. The table $\mathbf{n}_{V\setminus\{v\}}$ has fixed marginals $\mathbf{n}_{C_1\setminus\{v\}}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$. □

Lemma 8.2 *Assume there are two fixed marginals* $\mathbf{n}_{C_1}$ *and* $\mathbf{n}_{C_2}$ *such that* $C_1\cup C_2 = K$, *but* $C_1\cap C_2 = \emptyset$. *Consider* $D_0 \subset K$. *The Fréchet bounds for* $n_{D_0}(i^0_{D_0})$ *given* $\mathbf{n}_{C_1}$ *and* $\mathbf{n}_{C_2}$

$$\begin{aligned}\min\left\{n_{C_1\cap D_0}(i^0_{C_1\cap D_0}), n_{C_2\cap D_0}(i^0_{C_2\cap D_0})\right\} &\geq n_{D_0}(i^0_{D_0}) \\ &\geq \max\left\{0, n_{C_1\cap D_0}(i^0_{C_1\cap D_0}) + n_{C_2\cap D_0}(i^0_{C_2\cap D_0}) - n_\phi\right\} \end{aligned} \tag{23.7}$$

are sharp given $\mathbf{n}_{C_1}$ *and* $\mathbf{n}_{C_2}$.

Proof The induced independence graph is obviously decomposable, and its cliques C_1 and C_2 are separated by the empty set. Every vertex $v \in (C_1 \setminus D_0) \cup (C_2 \setminus D_0)$ is simplicial in $\mathcal{G}$, hence we could think about $\mathbf{n}_{D_0}$ as being a table with two fixed non-overlapping marginals $\mathbf{n}_{C_1 \cap D_0}$ and $\mathbf{n}_{C_2 \cap D_0}$. Lemma 8.1 implies that we do not loose any information about the cell entry $n_{D_0}(i^0_{D_0})$ when collapsing across the variables $\{X_v : v \in (C_1 \setminus D_0) \cup (C_2 \setminus D_0)\}$. Thus the bounds in Equation (23.7) are indeed sharp. □

Lemma 8.3 *Let the two fixed marginals $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$ be such that $C_1 \cup C_2 = K$. Consider $D_0 \subset K$ and denote $D_1 := (C_1 \setminus C_2) \cap D_0$, $D_2 := (C_2 \setminus C_1) \cap D_0$ and $D_{12} := (C_1 \cap C_2) \cap D_0$. In addition, we let $C_{12} := (C_1 \cap C_2) \setminus D_0$. Then an upper bound for $n_{D_0}(i^0_{D_0})$ given $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$ is:*

$$\sum_{i^1_{C_{12}} \in \mathcal{I}_{C_{12}}} \min\left\{ n_{(C_1 \cap D_0) \cup C_{12}}\left(i^0_{C_1 \cap D_0}, i^1_{C_{12}}\right), n_{(C_2 \cap D_0) \cup C_{12}}\left(i^0_{C_2 \cap D_0}, i^1_{C_{12}}\right)\right\}, \quad (23.8)$$

while a lower bound is

$$\sum_{i^1_{C_{12}} \in \mathcal{I}_{C_{12}}} \max\Big\{0, n_{(C_1 \cap D_0) \cup C_{12}}\left(i^0_{C_1 \cap D_0}, i^1_{C_{12}}\right) + n_{(C_2 \cap D_0) \cup C_{12}}\left(i^0_{C_2 \cap D_0}, i^1_{C_{12}}\right) - n_{D_{12}}\left(i^0_{D_{12}}\right)\Big\}. \quad (23.9)$$

Proof We assume that $C_{12} \neq \emptyset$. The vertices in $C_1 \setminus (C_2 \cup D_0)$ and $C_2 \setminus (C_1 \cup D_0)$ are simplicial in the independence graph $\mathcal{G} = (K, E)$ induced by $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$. From Lemma 8.1, we deduce that we can restrict our attention to the marginal $\mathbf{n}_{D_0 \cup C_{12}}$ that has two fixed marginals $\mathbf{n}_{D_1 \cup (C_1 \cap C_2)} = \mathbf{n}_{(C_1 \cap D_0) \cup C_{12}}$ and $\mathbf{n}_{D_2 \cup (C_1 \cap C_2)} = \mathbf{n}_{(C_2 \cap D_0) \cup C_{12}}$. We choose an arbitrary index $i^1_{C_{12}} \in \mathcal{I}_{C_{12}}$. Consider the hyperplane $\mathbf{n}^{i^1_{C_{12}}}_{D_0}$ of $\mathbf{n}_{D_1 \cup (C_1 \cap C_2)}$ with entries

$$\mathbf{n}^{i^1_{C_{12}}}_{D_0}(i_{D_0}) := n_{D_0 \cup C_{12}}(i_{D_0}, i^1_{C_{12}}), \text{ for } i_{D_0} \in \mathcal{I}_{D_0}.$$

This hyperplane has two fixed marginals

$$\mathbf{n}^{i^1_{C_{12}}}_{C_1 \cap D_0} = \left\{ n_{(C_1 \cap D_0) \cup C_{12}}\left(i_{C_1 \cap D_0}, i^1_{C_{12}}\right) : i_{C_1 \cap D_0} \in \mathcal{I}_{C_1 \cap D_0} \right\},$$

and

$$\mathbf{n}^{i^1_{C_{12}}}_{C_2 \cap D_0} = \left\{ n_{(C_2 \cap D_0) \cup C_{12}}\left(i_{C_2 \cap D_0}, i^1_{C_{12}}\right) : i_{C_2 \cap D_0} \in \mathcal{I}_{C_2 \cap D_0} \right\}.$$

We have $D_0 = D_1 \cup D_{12} \cup D_2$, hence it is possible to make use of Theorem 8.1 to obtain the Fréchet bounds for the cell entry $n^{i^1_{C_{12}}}_{D_0}(i^0_{D_0}) = n_{D_0 \cup C_{12}}(i^0_{D_0}, i^1_{C_{12}})$, i.e.

$$\min\left\{ n_{(C_1 \cap D_0) \cup C_{12}}\left(i^0_{C_1 \cap D_0}, i^1_{C_{12}}\right), n_{(C_2 \cap D_0) \cup C_{12}}\left(i^0_{C_2 \cap D_0}, i^1_{C_{12}}\right)\right\},$$

and

$$\max\Big\{0, n_{(C_1 \cap D_0) \cup C_{12}}\left(i^0_{C_1 \cap D_0}, i^1_{C_{12}}\right) + n_{(C_2 \cap D_0) \cup C_{12}}\left(i^0_{C_2 \cap D_0}, i^1_{C_{12}}\right) - n_{D_{12}}\left(i^0_{D_0 \cap D_{12}}\right)\Big\}. \quad (23.10)$$

Since

$$n_{D_0}(i^0_{D_0}) = \sum_{i^1_{C_{12}} \in \mathcal{I}_{C_{12}}} n_{D_0 \cup C_{12}}(i^0_{D_0}, i^1_{C_{12}}),$$

Equations (23.8) and (23.9) follow from Equation (23.10) by adding over all the indices $i^1_{C_{12}} \in \mathcal{I}_{C_{12}}$. Although the bounds in every hyperplane $\mathbf{n}_D^{i^1_{C_{12}}}$ are sharp, the bounds in Equations (23.8) and (23.9) are guaranteed to be sharp only if $C_{12} = \emptyset$. If $C_{12} \neq \emptyset$, there is no reason to believe that Equations (23.8) and (23.9) give sharp bounds for $\mathbf{n}_{D_0}(i^0_{D_0})$. We conclude that the Fréchet upper and lower bounds for $\mathbf{n}_{D_0}(i^0_{D_0})$ are not necessarily the best bounds possible if $C_{12} \neq \emptyset$. □

Proposition 8.4 *Let* $\mathbf{n}$ *be a* k*-dimensional table and consider the set of cells* $\mathbf{T} = \mathbf{T}^{(\mathbf{n})}$ *associated with* $\mathbf{n}$*. The marginals* $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$ *induce a decomposable independence graph* $\mathcal{G} = (K, E)$ *with* $\mathcal{C}(\mathcal{G}) = \{C_1, C_2, \ldots, C_p\}$ *and* $\mathcal{S}(\mathcal{G}) = \{S_2, \ldots, S_p\}$*. The set of fixed cells* $\mathbf{T}_0 \subset \mathbf{T}^{(\mathbf{n})}$ *is given by the cell entries contained in the tables*

$$\bigcup_{r=1}^{p} \bigcup_{\{C : C \subseteq C_r\}} \mathcal{RD}(\mathbf{n}_C).$$

For every cell $t \in \mathbf{T}$*, we let* $\mathbf{n}_1^{(t)}$, $\mathbf{n}_2^{(t)}$, …, $\mathbf{n}_{k_t}^{(t)}$ *be the tables in* $\mathcal{RD}$ *such that* t *is a cell entry in* $\mathbf{n}_r^{(t)}$*,* $r = 1, 2, \ldots, k_t$*. Under these conditions, GSA converges to an upper bound* $U_s(t)$ *and to a lower bound* $L_s(t)$ *such that*

$$\max\{L^r(t) : r = 1, 2, \ldots, k_t\} \leq L_s(t),\ U_s(t) \leq \min\{U^r(t) : r = 1, 2, \ldots, k_t\}, \tag{23.11}$$

where $U^r(t)$ *and* $L^r(t)$ *are the Fréchet bounds of the cell* t *in table* $\mathbf{n}_r^{(t)}$*.*

Proof We prove Proposition 8.4 by sequentially considering several particular cases. First we show that the shuttle procedure obtains the Fréchet bounds for a 2×2 table. Since any two-way table can be reduced to a number of 2×2 tables, it follows that the Fréchet bounds are also attained for a two-dimensional cross-classification with fixed one-dimensional totals. By induction on the number of fixed marginals of an arbitrary k-dimensional table $\mathbf{n}$ and by exploiting the fact that, if $\mathbf{n}$ has two marginals fixed, $\mathbf{n}$ can be split in several two-way tables with fixed one-way marginals, we are able to prove in the last subsection that the Fréchet bounds are attained for decomposable log-linear models with any number of minimal sufficient statistics.

- The 2×2 case.
  Consider a 2×2 table $\mathbf{n} = \{n_{ij} : 1 \leq i, j \leq 2\}$ with fixed row totals $\{n_{1+}, n_{2+}\}$ and column totals $\{n_{+1}, n_{+2}\}$. The grand total of the table is n_{++}. The set $\mathbf{T}$ associated with $\mathbf{n}$ is given by

$$\mathbf{T} = \{n_{11}, n_{12}, n_{21}, n_{22}, n_{1+}, n_{2+}, n_{+1}, n_{+2}, n_{++}\}, \tag{23.12}$$

while the set of cells having a fixed value is $\mathbf{T}_0 = \{n_{1+}, n_{2+}, n_{+1}, n_{+2}, n_{++}\}$. There are only six dependencies

$$\mathcal{Q}(\mathbf{T})\{(n_{1+}, n_{++}, n_{2+}), (n_{+1}, n_{++}, n_{+2}), (n_{11}, n_{1+}, n_{12}), (n_{12}, n_{+2}, n_{22}), (n_{21}, n_{2+}, n_{22}), (n_{11}, n_{21}, n_{+1})\}. \quad (23.13)$$

The first two dependencies are redundant because they involve only the cells in $\mathbf{T}_0$. We show that GSA converges to the Fréchet bounds:

$$\min\{n_{i+}, n_{+j}\} \geq n_{ij} \geq \max\{0, n_{i+} + n_{+j} - n_{++}\}, \text{ for } 1 \leq i, j \leq 2. \quad (23.14)$$

We initialize the upper and lower bounds of the four cells in $\mathbf{T} \setminus \mathbf{T}_0$:

$$L(n_{11}) = L(n_{12}) = L(n_{21}) = L(n_{22}) := 0, \quad \text{and}$$
$$U(n_{11}) = U(n_{12}) = U(n_{21}) = U(n_{22}) := n_{++}.$$

We sequentially go through the dependencies in $\mathcal{Q}(\mathbf{T})$. When we obtain a Fréchet bound, we mark it with "$\Diamond$". Since the Fréchet bounds are sharp, once GSA reaches such a bound, it stays at that bound.

First iteration, dependency: $n_{11} \oplus n_{12} = n_{1+}$.

$$\begin{aligned}
L(n_{11}) &= \max\{L(n_{11}), n_{1+} - U(n_{12})\} = \max\{0, n_{1+} - n_{++}\} = 0,\\
U(n_{11}) &= \min\{U(n_{11}), n_{1+} - L(n_{12})\} = \min\{n_{++}, n_{1+}\} = n_{1+},\\
L(n_{12}) &= \max\{L(n_{12}), n_{1+} - U(n_{11})\} = \max\{0, n_{1+} - n_{1+}\} = 0.\\
U(n_{12}) &= \min\{U(n_{12}), n_{1+} - L(n_{11})\} = \min\{n_{++}, n_{1+}\} = n_{1+}.
\end{aligned}$$

First iteration, dependency: $n_{12} \oplus n_{22} = n_{+2}$.

$$\begin{aligned}
L(n_{22}) &= \max\{L(n_{22}), n_{+2} - U(n_{12})\},\\
&= \max\{0, n_{+2} - n_{1+}\} = \max\{0, n_{+2} + n_{2+} - n_{++}\}.\Diamond\\
U(n_{22}) &= \min\{U(n_{22}), n_{+2} - L(n_{12})\},\\
&= \min\{n_{++}, n_{+2}\} = n_{+2}.\\
L(n_{12}) &= \max\{L(n_{12}), n_{+2} - U(n_{22})\} = 0.\\
U(n_{12}) &= \min\{U(n_{12}), n_{+2} - L(n_{22})\},\\
&= \min\{n_{1+}, n_{+2} + \min\{0, n_{1+} - n_{+2}\}\} = \min\{n_{+2}, n_{1+}\}.\Diamond
\end{aligned}$$

First iteration, dependency: $n_{21} \oplus n_{22} = n_{2+}$.

$$\begin{aligned}
L(n_{21}) &= \max\{L(n_{21}), n_{2+} - U(n_{22})\},\\
&= \max\{0, n_{2+} - n_{+2}\} = \max\{0, n_{2+} + n_{+1} - n_{++}\}.\Diamond\\
U(n_{21}) &= \min\{U(n_{21}), n_{2+} - L(n_{22})\},\\
&= \min\{0, n_{2+} + \min\{0, n_{1+} - n_{+2}\}\},\\
&= \min\{n_{2+}, n_{++} - n_{+2}\} = \min\{n_{2+}, n_{+1}\}.\Diamond\\
U(n_{22}) &= \min\{U(n_{22}), n_{2+} - L(n_{21})\},\\
&= \min\{n_{+2}, n_{2+} + \min\{0, n_{+2} - n_{2+}\}\} = \min\{n_{+2}, n_{2+}\}.\Diamond
\end{aligned}$$

First iteration, dependency: $n_{11} \oplus n_{21} = n_{+1}$.

$$\begin{aligned} L(n_{11}) &= \max\{L(n_{11}), n_{+1} - U(n_{21})\}, \\ &= \max\{0, n_{+1} + \max\{-n_{2+}, -n_{+1}\}\}, \\ &= \max\{0, n_{+1} - n_{2+}\} = \max\{0, n_{+1} + n_{1+} - n_{++}\}.\Diamond \\ U(n_{11}) &= \min\{U(n_{11}), n_{+1} - L(n_{21})\}, \\ &= \min\{n_{1+}, n_{+1} + \max\{-n_{2+}, -n_{+1}\}\}, \\ &= \min\{n_{+1}, n_{++} - n_{2+}\} = \min\{n_{+1}, n_{1+}\}.\Diamond \end{aligned}$$

Second iteration, dependency: $n_{11} \oplus n_{12} = n_{1+}$.

$$\begin{aligned} L(n_{12}) &= \max\{L(n_{12}), n_{1+} - U(n_{11})\}, \\ &= \max\{0, n_{1+} + \max\{-n_{+1}, -n_{1+}\}\}, \\ &= \max\{0, n_{1+} - n_{+1}\} = \max\{0, n_{1+} + n_{+2} - n_{++}\}.\Diamond \end{aligned}$$

We see that the Fréchet bounds in Equation (23.14) for all four cells in table $\mathbf{n}$ are obtained before completing the second iteration. Therefore Proposition 8.4 holds for any 2×2 table.

- The two-way case.
  The next step is to examine a two-dimensional table $\mathbf{n} = \{n_{ij} : 1 \le i \le I_1, 1 \le j \le I_2\}$ for some $I_1, I_2 \ge 2$. This table has fixed row sums and column sums:

$$\{n_{i+} : 1 \le i \le I_1\} \cup \{n_{+j} : 1 \le j \le I_2\} \cup \{n_{++}\} \subset \mathbf{T}_0. \tag{23.15}$$

More precisely, the set of fixed cells $\mathbf{T}_0 \subset \mathbf{T} = \mathbf{T}^{(\mathbf{n})}$ is

$$\begin{aligned} \mathbf{T}_0 = \left\{t_{J_1\{1,2,\ldots,I_2\}} : \emptyset \ne J_1 \subseteq \{1,2,\ldots,I_1\}\right\} \cup \\ \left\{t_{\{1,2,\ldots,I_1\}J_2} : \emptyset \ne J_2 \subseteq \{1,2,\ldots,I_2\}\right\}. \end{aligned} \tag{23.16}$$

We remark that $t \in \mathbf{T} \setminus \mathbf{T}_0$ if and only if t is a cell in a table $\mathbf{n}' \in \mathcal{RD}(\mathbf{n})$. In other words, t does not have a fixed value if and only if t is a cell in a table of dimension 2 that could be obtained from $\mathbf{n}$ by table redesign. We show that the Fréchet bounds from Theorem 8.1 are attained when running GSA for every cell n_{ij}. It is sufficient to prove it for the $(1,1)$ cell. The Fréchet inequality for n_{11} is

$$\min\{n_{1+}, n_{+1}\} \ge n_{11} \ge \max\{0, n_{1+} + n_{+1} - n_{++}\}. \tag{23.17}$$

We notice the similarity of Equation (23.17) with Equation (23.14). We define the 2×2 table $\mathbf{n}' = \{n'_{ij} : 1 \le i, j \le 2\}$, where

$$n'_{11} := n_{11},\ n'_{12} := \sum_{j>1} n_{1j},\ n'_{21} := \sum_{i>1} n_{i1},\ n'_{22} := \sum_{i>1}\sum_{j>1} n_{ij}. \tag{23.18}$$

This table has fixed row totals $\left\{n_{1+}, \sum_{i>1} n_{i+}\right\}$ as well as fixed column totals $\left\{n_{+1}, \sum_{j>1} n_{+j}\right\}$. The Fréchet bounds for the $(1,1)$ count in table $\mathbf{n}'$ coincide

with the Fréchet bounds for the $(1,1)$ count in table $\mathbf{n}$. Since the four cells in table $\mathbf{n}'$ are also cells in the set $\mathbf{T}$ associated with $\mathbf{n}$, the generalized shuttle algorithm employed for the table $\mathbf{n}$ is equivalent to the shuttle procedure employed for the table $\mathbf{n}'$ from the perspective of finding sharp bounds for $\{n'_{11}, n'_{12}, n'_{21}, n'_{22}\}$. We proved before that the generalized shuttle algorithm will converge to the Fréchet bounds for any 2×2 table, hence GSA finds the Fréchet bounds for the $(1,1)$ cell in table $\mathbf{n}$.

Now take an arbitrary cell $t = t_{\{i_1,i_2,\ldots,i_l\}\{j_1,j_2,\ldots,j_s\}} \in \mathbf{T} \setminus \mathbf{T}_0$. Consider the 2×2 table $\mathbf{n}^{(t)}$ with entries

$$\begin{aligned}\Big\{ & t_{\{i_1,i_2,\ldots,i_l\}\{j_1,j_2,\ldots,j_s\}}, t_{\{i_1,i_2,\ldots,i_l\}(\mathcal{I}_2 \setminus \{j_1,j_2,\ldots,j_s\})}\ , \\ & t_{(\mathcal{I}_1 \setminus \{i_1,i_2,\ldots,i_l\})\{j_1,j_2,\ldots,j_s\}}, t_{(\mathcal{I}_1 \setminus \{i_1,i_2,\ldots,i_l\})(\mathcal{I}_2 \setminus \{j_1,j_2,\ldots,j_s\})} \Big\}\ .\end{aligned}$$

The Fréchet bounds for the value $V(t)$ of cell t in the above table are

$$\min\left\{V(t_{\{i_1,i_2,\ldots,i_l\}\{1,2,\ldots,I_2\}}), V(t_{\{1,2,\ldots,I_1\}\{j_1,j_2,\ldots,j_s\}})\right\}$$

and

$$\max\left\{0, V(t_{\{i_1,\ldots,i_l\}\{1,\ldots,I_2\}}) + V(t_{\{1,\ldots,I_1\}\{j_1,\ldots,j_s\}}) - V(t_{\{1,\ldots,I_1\}\{1,\ldots,I_2\}})\right\}. \tag{23.19}$$

The table $\mathbf{n}^{(t)}$ has fixed one-dimensional totals, hence we know the cell values

$$\begin{aligned} V(t_{\{i_1,i_2,\ldots,i_l\}\{1,2,\ldots,I_2\}}) &= \sum_{r=1}^{l} n_{i_r+}\,, \\ (t_{\{1,2,\ldots,I_1\}\{j_1,j_2,\ldots,j_s\}}) &= \sum_{r=1}^{s} n_{+j_r}\,, \\ V(t_{\{1,2,\ldots,I_1\}\{1,2,\ldots,I_2\}}) &= n_\phi. \end{aligned}$$

The Fréchet bounds in Equation (23.19) are the Fréchet bounds associated with cell t in every table $\mathbf{n}' \in \mathcal{RD}$ such that t is a cell in $\mathbf{n}'$. Again, for every such table $\mathbf{n}'$, it is true that $\mathbf{T}^{(\mathbf{n}')} \subset \mathbf{T}^{(\mathbf{n})}$ and $\mathcal{Q}(\mathbf{T}^{(\mathbf{n}')}) \subset \mathcal{Q}(\mathbf{T}^{(\mathbf{n})})$. When employing the shuttle procedure for $\mathbf{n}$ we also run the shuttle procedure in $\mathbf{n}'$, thus the bounds in Equation (23.19) are attained by GSA and hence Proposition 8.4 holds for an arbitrary two-dimensional table.

- Bounds induced by two fixed marginals.
  Let $\mathbf{n} = \{n(i)\}_{i \in \mathcal{I}}$ be a k-dimensional frequency count table having fixed marginals $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$ such that $C_1 \cup C_2 = K$. The Fréchet bounds for a cell entry $n(i^0)$ are

$$\begin{aligned}\min\left\{n_{C_1}\left(i^0_{C_1}\right), n_{C_2}\left(i^0_{C_2}\right)\right\} &\geq n(i^0), \\ n(i^0) &\geq \min\left\{n_{C_1}\left(i^0_{C_1}\right) + n_{C_2}\left(i^0_{C_2}\right) - n_{C_1 \cap C_2}\left(i^0_{C_1 \cap C_2}\right)\right\}.\end{aligned}$$

  First we study the case when the fixed marginals are non-overlapping. i.e. $C_1 \cap C_2 = \emptyset$. We attempt to reduce this case to the case of two-dimensional tables we studied before for which we know that Proposition 8.4 is true. The

above inequalities become

$$\min\left\{n_{C_1}\left(i^0_{C_1}\right), n_{K\setminus C_1}\left(i^0_{K\setminus C_1}\right)\right\} \geq n(i^0),$$
$$n(i^0) \geq \min\left\{n_{C_1}\left(i^0_{C_1}\right) + n_{K\setminus C_1}\left(i^0_{K\setminus C_1}\right) - n_\phi\right\}. \quad (23.20)$$

Without restricting the generality, we can assume that $C_1 = \{1,\ldots,l\}$ and $C_2 = \{l+1,\ldots,k\}$. To every index $i_{C_1} = (i_1,\ldots,i_l) \in \mathcal{I}_{C_1}$ we define:

$$IND_{C_1}(i_{C_1}) := \sum_{r=1}^{l}\left[\prod_{s=r+1}^{l} I_s\right]\cdot(i_r - 1) + 1 \in \{1,\ldots,I_1\cdot I_2\cdot\ldots\cdot I_l\}.$$

IND_{C_1} induces a one-to-one correspondence between the sets $\mathcal{I}_{C_1}$ and $\{1,\ldots,I_1\cdot\ldots\cdot I_l\}$. Similarly, to every $i_{C_2} = (i_{l+1},\ldots,i_k) \in \mathcal{I}_{C_2}$, we assign

$$IND_{C_2}(i_{C_2}) := \sum_{r=l+1}^{k}\left[\prod_{s=r+1}^{k} I_s\right]\cdot(i_r - 1) + 1 \in \{1,\ldots,I_{l+1}\cdot\ldots\cdot I_k\}.$$

Introduce two new compound variables Y_1 and Y_2 that take values in the sets $\{1,\ldots,I_1\cdot I_2\cdot\ldots\cdot I_l\}$ and $\{1,\ldots,I_{l+1}\cdot\ldots\cdot I_k\}$, respectively. Consider a two-way table

$$\mathbf{n}' = \{n'_{j_1 j_2} : 1 \leq j_1 \leq I_1\cdot I_2\cdot\ldots\cdot I_l, 1 \leq j_2 \leq I_{l+1}\cdot\ldots\cdot I_k\}$$

with entries given by

$$n'_{j_1 j_2} = n_K\left(IND^{-1}_{C_1}(j_1), IND^{-1}_{C_2}(j_2)\right).$$

The table $\mathbf{n}'$ has fixed row totals

$$\{n_{j_1+} : 1 \leq j_1 \leq I_1\cdot I_2\cdot\ldots\cdot I_l\},$$

where $n_{j_1+} n_{C_1}\left(IND^{-1}_{C_1}(j_1)\right)$, and column totals

$$\{n_{+j_2} : 1 \leq j_2 \leq I_{l+1}\cdot\ldots\cdot I_k\},$$

where $n_{+j_2} n_{C_2}\left(IND^{-1}_{C_2}(j_2)\right)$. Therefore there is a one-to-one correspondence between the cells in the original k-dimensional table $\mathbf{n}$ and the cells in the two-way table $\mathbf{n}'$. Moreover, there is a one-to-one correspondence between the fixed cells in $\mathbf{n}$ and the set of fixed cells in $\mathbf{n}'$. Running GSA for $\mathbf{n}$ assuming fixed marginals $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$ is the same as running the shuttle procedure for $\mathbf{n}'$ assuming fixed one-dimensional totals. This implies that the Fréchet bounds in Equation (23.20) are attained.

Consider a cell $t \in \mathbf{T}\setminus\mathcal{N}$ and let $\mathbf{n}' \in \mathcal{RD}$ such that $t = n'(i^0)$, for some $i^0 \in \mathcal{I}'_1\times\mathcal{I}'_2\times\ldots\times\mathcal{I}'_k$. If $\mathbf{n}' \in \mathcal{RD}(\mathbf{n})$, then the Fréchet bounds for $t = n'(i^0)$ in table $\mathbf{n}'$ are

$$\min\left\{n'_{C_1}\left(i^0_{C_1}\right), n'_{K\setminus C_1}\left(i^0_{K\setminus C_1}\right)\right\} \geq n'(i^0),$$
$$n'(i^0) \geq \min\left\{n'_{C_1}\left(i^0_{C_1}\right) + n'_{K\setminus C_1}\left(i^0_{K\setminus C_1}\right) - n_\phi\right\}. \quad (23.21)$$

$\mathbf{n}'_{C_1}$ and $\mathbf{n}'_{K\setminus C_1}$ are fixed marginals of $\mathbf{n}'$ obtained from $\mathbf{n}_{C_1}$ and $\mathbf{n}_{K\setminus C_1}$

by the same sequence of "category-join" operations that was necessary to transform the initial table $\mathbf{n}$ in $\mathbf{n}'$. Again, we have $\mathbf{T}^{(\mathbf{n}')} \subset \mathbf{T}^{(\mathbf{n})}$ and $\mathcal{Q}(\mathbf{T}^{(\mathbf{n}')}) \subset \mathcal{Q}(\mathbf{T}^{(\mathbf{n})})$, thus the Fréchet bounds in Equation (23.21) are obtained by employing the shuttle procedure for the same reasons the bounds in Equation (23.20) were reached.

Now assume that $\mathbf{n}' = \mathbf{n}_{D_0}$, $D_0 \subset K$, with $t = n_{D_0}(i^0_{D_0})$ for some $i^0_{D_0} \in \mathcal{I}_{D_0}$. The Fréchet bounds in $\mathbf{n}_{D_0}$ are given in Lemma 8.2. The table $\mathbf{n}_{D_0}$ has two fixed non-overlapping marginals $\mathbf{n}_{C_1 \cap D_0}$ and $\mathbf{n}_{C_2 \cap D_0}$, hence GSA reaches the Fréchet bounds in Equation (23.7) because $\mathbf{T}^{(\mathbf{n}_{D_0})} \subset \mathbf{T}^{(\mathbf{n})}$ and $\mathcal{Q}(\mathbf{T}^{(\mathbf{n}_{D_0})}) \subset \mathcal{Q}(\mathbf{T}^{(\mathbf{n})})$. If $\mathbf{n}' \in \mathcal{RD}(\mathbf{n}_{D_0})$ $\mathbf{n}'$ has two fixed marginals $\mathbf{n}'_{C_1 \cap D_0}$ and $\mathbf{n}'_{C_2 \cap D_0}$ obtained from $\mathbf{n}_{C_1 \cap D_0}$ and $\mathbf{n}_{C_2 \cap D_0}$ by joining categories associated with the variables cross-classified in $\mathbf{n}$. It is sufficient to replace $\mathbf{n}_{D_0}$ with $\mathbf{n}'$ in Equation (23.7) to calculate the Fréchet bounds for t in table $\mathbf{n}'$.

If the two fixed marginals are overlapping, we can assume that there exist q and l with $1 \leq q \leq l \leq k$, such that $C_1 = \{1, 2, \ldots, l\}$ and $C_2 = \{q, q+1, \ldots, k\}$. Then $C_1 \cap C_2 = \{q, \ldots, l\}$. We reduce the case of two fixed overlapping marginals to the case of two fixed non-overlapping marginals by decomposing the tables $\mathbf{n}$, $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$ in a number of hyperplanes. Each hyperplane of $\mathbf{n}$ has two non-overlapping marginals that are hyperplanes of $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$. Denote

$$D_1 := C_1 \setminus C_2 = \{1, 2, \ldots, q-1\}, \text{ and } D_2 := C_2 \setminus C_1 = \{l+1, l+2, \ldots, k\}.$$

Take the set of contingency tables

$$\left\{ \mathbf{n}^{i^0_q, \ldots, i^0_l} = \left\{ n^{i^0_q, \ldots, i^0_l}(i_{D_1 \cup D_2}) : i_{D_1 \cup D_2} \in \mathcal{I}_{D_1 \cup D_2} \right\} : i^0_q \in \mathcal{I}_q, \ldots, i^0_l \in \mathcal{I}_l \right\},$$

where

$$\begin{aligned} n^{i^0_q, \ldots, i^0_l}(i_{D_1 \cup D_2}) &= n^{i^0_q, \ldots, i^0_l}(i_1, \ldots, i_{q-1}, i_{l+1}, \ldots, i_k) \\ &= n(i_1, \ldots, i_{q-1}, i^0_q, \ldots, i^0_l, i_{l+1}, \ldots, i_k). \end{aligned}$$

Every table $\mathbf{n}^{i^0_q, \ldots, i^0_l}$ has two fixed non-overlapping marginals

$$\mathbf{n}^{i^0_q, \ldots, i^0_l}_{D_1} = \left\{ n^{i^0_q, \ldots, i^0_l}(i_{D_1}) : i_{D_1} \in \mathcal{I}_{D_1} \right\},$$

with entries given by

$$n^{i^0_q, \ldots, i^0_l}(i_{D_1}) = n^{i^0_q, \ldots, i^0_l}(i_1, \ldots, i_{q-1}) = n_{C_1}\left(i_1, \ldots, i_{q-1}, i^0_q, \ldots, i^0_l\right),$$

and $\mathbf{n}^{i^0_q, \ldots, i^0_l}_{D_2} = \left\{ n^{i^0_q, \ldots, i^0_l}(i_{D_2}) : i_{D_2} \in \mathcal{I}_{D_2} \right\}$, with entries given by

$$n^{i^0_q, \ldots, i^0_l}(i_{D_2})\, n^{i^0_q, \ldots, i^0_l}(i_{l+1}, \ldots, i_k) = n_{C_2}\left(i^0_q, \ldots, i^0_l, i_{l+1}, \ldots, i_k\right).$$

Notice that the table $\mathbf{n}^{i^0_q, \ldots, i^0_l}$ is a hyperplane of the original table $\mathbf{n}$, whereas $\mathbf{n}^{i^0_q, \ldots, i^0_l}_{D_1}$ is a hyperplane of $\mathbf{n}_{C_1}$, and $\mathbf{n}^{i^0_q, \ldots, i^0_l}_{D_2}$ is a hyperplane of $\mathbf{n}_{C_2}$. Employing the generalized shuttle algorithm for $\mathbf{n}$ is equivalent to employing distinct versions of the shuttle procedure for every hyperplane determined by an index $\left(i^0_q, \ldots, i^0_l\right) \in \mathcal{I}_{C_1 \cap C_2}$. We already showed that GSA for $\mathbf{n}^{i^0_q, \ldots, i^0_l}$

converges to the Fréchet bounds of the cell entry $n^{i_q^0,\ldots,i_l^0}\left(i_{D_1}^0, i_{D_2}^0\right)$ (compare with Equation (23.20)):

$$\min\left\{n_{D_1}^{i_q^0,\ldots,i_l^0}\left(i_{D_1}^0\right), n_{D_2}^{i_q^0,\ldots,i_l^0}\left(i_{D_2}^0\right)\right\} \geq$$
$$n^{i_q^0,\ldots,i_l^0}\left(i_{D_1}^0, i_{D_2}^0\right) \geq \max\left\{n_{D_1}^{i_q^0,\ldots,i_l^0}\left(i_{D_1}^0\right) + n_{D_2}^{i_q^0,\ldots,i_l^0}\left(i_{D_2}^0\right) - n_{\phi}^{i_q^0,\ldots,i_l^0}\right\}, \tag{23.22}$$

where $n_{\phi}^{i_q^0,\ldots,i_l^0} = n_{C_1\cap C_2}\left(i_q^0,\ldots,i_l^0\right)$ is the grand total of the hyperplane $\mathbf{n}^{i_q^0,\ldots,i_l^0}$. Equation (23.22) can equivalently be written as

$$\min\left\{n_{C_1}\left(i_{D_1}^0, i_q^0,\ldots,i_l^0\right), n_{C_2}\left(i_q^0,\ldots,i_l^0,i_{D_2}\right)\right\} \geq n\left(i_{D_1}^0, i_q^0,\ldots,i_l^0,i_{D_2}^0\right)$$
$$\geq \max\left\{0, n_{C_1}\left(i_{D_1}^0, i_{C_1\cap C_2}^0\right) + n_{C_2}\left(i_{C_1\cap C_2}^0, i_{D_2}^0\right) - n_{C_1\cap C_2}\left(i_{C_1\cap C_2}^0\right)\right\}.$$

These inequalities represent the Fréchet bounds for the cell count

$$n\left(i^0\right) = n\left(i_{D_1}^0, i_q^0,\ldots,i_l^0,i_{D_2}^0\right).$$

Now we show that any table $\mathbf{n}' \in \mathcal{RD} \setminus \bigcup_{r=1}^{2} \bigcup_{\{C: C\subseteq C_r\}} \mathcal{RD}(\mathbf{n}_C)$ can be separated in a number of hyperplanes such that the two fixed marginals of every hyperplane are non-overlapping. Consider an arbitrary cell in $\mathbf{n}'$ specified by the index $(J_1^0,\ldots,J_k^0) \in \mathcal{I}_1'\times\ldots\times\mathcal{I}_k'$. The hyperplane $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}$ of table $\mathbf{n}'$ has entries

$$\left\{n'(J_1,\ldots,J_{q-1},J_q^0,\ldots,J_l^0,J_{l+1},\ldots,J_k) : J_r \in \mathcal{I}_r'\right\},$$

for $r = 1,\ldots,q-1,l+1,\ldots,k$. The fixed overlapping marginals $\mathbf{n}_{C_1}$ and $\mathbf{n}_{C_2}$ induce two fixed overlapping marginals $\mathbf{n}'_{C_1}$ and $\mathbf{n}'_{C_2}$ of $\mathbf{n}'$. The index set of $\mathbf{n}'_{C_r}$, $r = 1, 2$, is $\mathcal{I}'_{1;C_r}\times\ldots\times\mathcal{I}'_{k;C_r}$, where

$$\mathcal{I}'_{s;C_r} = \begin{cases} \mathcal{I}'_s, & \text{if } s \in C_r, \\ \{\mathcal{I}_s\}, & \text{if } s \notin C_r. \end{cases}$$

We define the hyperplanes $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}_{C_1}$ of $\mathbf{n}'_{C_1}$ and $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}_{C_2}$ of $\mathbf{n}'_{C_2}$ in the same way we defined the hyperplane $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}$ of $\mathbf{n}'$. Therefore $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}$ is a table having two fixed non-overlapping marginals $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}_{C_1}$ and $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}_{C_2}$. The Fréchet bounds for $n'(J_1^0,\ldots,J_k^0)$ coincide with the Fréchet bounds for the cell entry

$$n'^{(J_q^0,\ldots,J_l^0)}(J_1^0,\ldots,J_{q-1}^0,J_{l+1}^0,\ldots,J_k^0)$$

in table $\mathbf{n}'^{(J_q^0,\ldots,J_l^0)}$. Therefore Proposition 8.4 holds for any table of counts with two fixed marginals.

- Calculating bounds in the general decomposable case.
  The set of fixed cliques defines a decomposable independence graph $\mathcal{G} = (K, E)$ with cliques $\mathcal{C}(\mathcal{G})$ and separators $\mathcal{S}(\mathcal{G})$. We prove Proposition 8.4 by

induction on the number of fixed marginals. Because the notation tends to be quite cumbersome, we will show that the Fréchet bounds for the cells in only the initial table $\mathbf{n}$ are attained. A similar argument can be made about every table in

$$\mathcal{RD} \setminus \bigcup_{r=1}^{p} \bigcup_{\{C:C\subseteq C_r\}} \mathcal{RD}(\mathbf{n}_C).$$

If $\mathcal{G}$ decomposes in $p = 2$ cliques, we already proved that GSA converges to the Fréchet bounds in Equation (23.5). We assume that Proposition 8.4 is true if $\mathbf{n}$ has at most $(p-1)$ fixed marginals that induce a decomposable independence graph. We want to prove Proposition 8.4 for an independence graph with p cliques. We take an arbitrary index $i^0 \in \mathcal{I}$ that will remain fixed for the rest of this proof.

The cliques of $\mathcal{G}$ can be numbered so that they form a perfect sequence of vertex sets. Let $H_{p-1} := C_1 \cup C_2 \cup \ldots \cup C_{p-1}$. The subgraph $\mathcal{G}(H_{p-1})$ is decomposable and its cliques are $\{C_1, \ldots, C_{p-1}\}$, while its separators are $\{S_2, \ldots, S_{p-1}\}$. As before, $\mathbf{T} = \mathbf{T}^{(\mathbf{n})}$ is the set of cells associated with $\mathbf{n}$. In an analogous manner we define the set of cells $\mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}$ associated with the marginal table $\mathbf{n}_{H_{p-1}}$. The set of fixed cells $\mathbf{T}_0 = \mathbf{T}_0^{(\mathbf{n})} \subset \mathbf{T}$ induced by fixing the cell counts in the marginals $\mathbf{n}_{C_1}$, $\mathbf{n}_{C_2}$, ..., $\mathbf{n}_{C_p}$ of the table $\mathbf{n}$ includes the set of fixed cells $\mathbf{T}_0^{\left(\mathbf{n}_{H_{p-1}}\right)} \subset \mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}$ obtained by fixing the marginals $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_{p-1}}$ of the table $\mathbf{n}_{H_{p-1}}$.

We have $\mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)} \subset \mathbf{T}^{(\mathbf{n})}$ and $\mathcal{Q}\left(\mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}\right) \subset \mathcal{Q}\left(\mathbf{T}^{(\mathbf{n})}\right)$. This implies that, when we run GSA for $\mathbf{T}^{(\mathbf{n})}$ and $\mathbf{T}_0^{(\mathbf{n})}$, it is as if we would run an instance ofGSA for $\mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}$ and $\mathbf{T}_0^{\left(\mathbf{n}_{H_{p-1}}\right)}$. Every vertex in $C_p \setminus S_p = C_p \setminus H_{p-1}$ is simplicial in the graph $\mathcal{G}$, hence Lemma 8.1 tells us that finding bounds for a cell in $t \in \mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}$ given $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_{p-1}}$ is equivalent to finding bounds for t given $\mathbf{n}_{C_1}, \mathbf{n}_{C_2}, \ldots, \mathbf{n}_{C_p}$. We do not lose any information by not considering the marginal $\mathbf{n}_{C_p}$ when computing bounds for $t \in \mathbf{T}\left(\mathbf{n}_{H_{p-1}}\right)$.

From the induction hypothesis we know that GSA employed for table $\mathbf{n}_{H_{p-1}}$ with the set of fixed cells $\mathbf{T}_0^{\left(\mathbf{n}_{H_{p-1}}\right)}$ converges to the Fréchet bounds for the cell $n_{H_{p-1}}\left(i^0_{H_{p-1}}\right)$:

$$n^U_{H_{p-1}}\left(i^0_{H_{p-1}}\right) = \min\left\{n_{C_1}\left(i^0_{C_1}\right), \ldots, n_{C_{p-1}}\left(i^0_{C_{p-1}}\right)\right\}, \quad \text{and}$$

$$n^L_{H_{p-1}}\left(i^0_{H_{p-1}}\right) = \max\left\{0, \sum_{r=1}^{p-1} n_{C_r}\left(i^0_{C_r}\right) - \sum_{r=2}^{p-1} n_{S_r}\left(i^0_{S_r}\right)\right\}.$$

The shuttle procedure generates feasibility intervals $[L_s(t), U_s(t)]$ for every $t \in \mathbf{T}\left(\mathbf{n}_{H_{p-1}}\right)$. These are the tightest feasibility intervals GSA can find given the values of the cells in $\mathbf{T}_0^{\left(\mathbf{n}_{H_{p-1}}\right)}$. Because the information about the cells in the marginal $\mathbf{n}_{C_p}$ is not relevant for computing bounds for the cells

in $\mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}$, GSA employed for table $\mathbf{n}$ converges to the same feasibility intervals $[L_s(t), U_s(t)]$ for every $t \in \mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}$.
Since the sequence $C_1, C_2, \ldots, C_p$ is perfect in $\mathcal{G}$, $(H_{p-1} \setminus S_p, S_p, C_p \setminus S_p)$ is a proper decomposition of $\mathcal{G}$. Consider the graph $\mathcal{G}' = (K, E')$, where

$$E' := \{(u, v) : \{u, v\} \subset H_{p-1} \text{ or } \{u, v\} \subset C_p\}.$$

$\mathcal{G}'$ is a decomposable graph with two cliques H_{p-1}, C_p and one separator $H_{p-1} \cap C_p = S_p$. Running GSA for table $\mathbf{n}$ and the set of fixed cells $\mathbf{T}_0^{(\mathbf{n})}$ is equivalent to running GSA for $\mathbf{n}$ given the feasibility intervals $\left\{[L_s(t), U_s(t)] : t \in \mathbf{T}^{\left(\mathbf{n}_{H_{p-1}}\right)}\right\}$ and the set of fixed cells in $\mathbf{T}^{(\mathbf{n})}$ obtained by fixing the cells in the marginal $\mathbf{n}_{C_p}$.
As a consequence, by employing the shuttle procedure for table $\mathbf{n}$, we end up with the following Fréchet bounds for the count $n(i^0)$:

$$\begin{aligned} &\min\left\{n^U_{H_{p-1}}\left(i^0_{H_{p-1}}\right), n_{C_p}\left(i^0_{C_p}\right)\right\} \geq n(i^0), \text{ and} \\ &\quad n(i^0) \geq \max\left\{0, n^L_{H_{p-1}}\left(i^0_{H_{p-1}}\right) + n_{C_p}\left(i^0_{C_p}\right) - n_{S_p}\left(i^0_{S_p}\right)\right\}. \end{aligned} \tag{23.23}$$

It is straightforward to notice that Equation (23.5) is obtained by combining Equations (23.23) and (23.23). We can conclude that Proposition 8.4 is true when the set of fixed marginals are the minimal sufficient statistics of a decomposable log-linear model.

□

24

On-line supplement to Indicator function and sudoku designs

Roberto Fontana

Maria Piera Rogantin

24.1 An example of complex coding for sudoku design

A row r of the sudoku grid is coded by the levels of the pseudo-factors R_1 and R_2

$$(\omega_{r_1}, \omega_{r_2}) \quad \text{with } r_i \in \mathbb{Z}_p \text{ and } r - 1 = p\, r_1 + r_2.$$

Similarly, for columns and symbols. Figure 24.1 gives a 9×9 partially filled sudoku grid and the array on the right gives the complex coding of the fraction. For example, for the symbol 3 in the first row and second column we have: first row $R_1 = \omega_0, R_2 = \omega_0$, second column $C_1 = \omega_0, C_2 = \omega_1$, symbol 3 $S_1 = \omega_0, S_2 = \omega_2$. The box is the first, in fact $R_1 = \omega_0, C_2 = \omega_0$.

	00	01	02	10	11	12	20	21	22
00	5	3	4	6	7	8			
01									
02									
10									
11									
12									
20									
21									
22							1	7	9

r_1	r_2	c_1	c_2	s_1	s_2
ω_0	ω_0	ω_0	ω_0	ω_1	ω_1
ω_0	ω_0	ω_0	ω_1	ω_0	ω_2
ω_0	ω_0	ω_0	ω_2	ω_2	ω_2
ω_0	ω_0	ω_1	ω_0	ω_1	ω_2
ω_0	ω_0	ω_1	ω_1	ω_2	ω_0
ω_0	ω_0	ω_1	ω_2	ω_2	ω_1
..	..	..	..	..	..
ω_2	ω_2	ω_2	ω_0	ω_0	ω_0
ω_2	ω_2	ω_2	ω_1	ω_2	ω_0
ω_2	ω_2	ω_2	ω_2	ω_2	ω_2

Fig. 24.1 A partially filled sudoku and its complex coding.

24.2 Proofs

Proposition 12.8 *The move corresponding to the exchange of the symbol u with the symbol v is:*

$$M(F) = E_{s,uv}\ P_{g,uv}(F) = \sum_{\alpha_g \in L_g} \sum_{\beta_s \in L_s} m_{\alpha_g,\beta_s}\ X_g^{\alpha_g}\ X_s^{\beta_s}$$

where the coefficients m_{α_g,β_s} *are:*

$$m_{\alpha_g,\beta_s} = \frac{1}{p^2}(-\overline{e_{\beta_s,uv}}) \sum_{\alpha_s \in L_s} b_{(\alpha_g,\alpha_s)} e_{\alpha_s,uv}.$$

Proof First, we prove that $F_1 = F + M(F)$ is the indicator function corresponding to the grid where the symbol u has been exchanged with the symbol v. Then, we prove that $M(F)$ is a valid move, according to Corollary 2.

Step 1.

If $E_{s,hk} = 0$ (no symbol to exchange) or if $P_u = P_v = 0$ (no cell to modify) we have $F_1 = F$ on $\mathcal{D}$.

Let's now consider the points corresponding to the cells of the grid where the symbol is u. We denote by $\zeta_{\hat{g}}$ these points of $\mathcal{D}_{1234}$: $\zeta_{\hat{g}} = (\omega_{\hat{r}_1}, \omega_{\hat{r}_2}, \omega_{\hat{c}_1}, \omega_{\hat{c}_2})$.

We have: $F(\zeta_{\hat{g}}, \zeta_u) = 1$ and $F(\zeta_{\hat{g}}, \zeta_v) = 0$. On the same points the move is:

$$M(F)(\zeta_{\hat{g}}, \zeta_u) = E_{s,hk}(\zeta_u)\ P_{g,hk}(F)(\zeta_{\hat{g}}) = -1$$
$$M(F)(\zeta_{\hat{g}}, \zeta_v) = E_{s,hk}(\zeta_v)\ P_{g,hk}(F)(\zeta_{\hat{g}}) = 1$$

and, therefore: $F_1(\zeta_{\hat{g}}, \zeta_u) = 1 - 1 = 0$ and $F_1(\zeta_{\hat{g}}, \zeta_v) = 0 + 1 = 1$.

Analogously, for the replacement of the symbol v by the symbol u. We can conclude that $F_1 = F + M(F)$ is the indicator function of the grid that has been generated exchanging u with v in the original fraction.

Step 2.

As in Lemma 1, $E_{s,hk}$ depends only by S_1 and S_2, and it is the polynomial

$$E_{s,hk} = \frac{1}{p^2} \sum_{\beta_s \in L_s} (-\overline{e_{\beta_s,hk}})\ X^{\beta_s},$$

where the constant term is zero.

It follows that the move $M(F)$ can be written as

$$\begin{aligned} M(F) =& E_{s,hk}\ P_{g,hk}(F) = \\ & -\frac{1}{p^2} \sum_{\alpha_g \in L_g; \alpha_s \in L_s} \sum_{\beta_s \in L_s} b_{(\alpha_g,\alpha_s)}\ e_{\alpha_s,hk}\ \overline{e_{\beta_s,hk}}\ X_g^{\alpha_g}\ X_s^{\beta_s} = \\ & -\frac{1}{p^2} \sum_{\alpha_g \in L_g} \sum_{\beta_s \in L_s} \left(\overline{e_{\beta_s,hk}} \sum_{\alpha_s \in L_s} b_{(\alpha_g,\alpha_s)}\ e_{\alpha_s,hk} \right) X_g^{\alpha_g}\ X_s^{\beta_s}. \end{aligned}$$

We verify that the coefficients m_α of $M(F)$ meet the requirements that are stated in Corollary 2. Indeed

(a) $m_{i_1 i_2 i_3 i_4 00} = 0$ because $-\overline{e_{0,hk}} = (\overline{\omega}_{v_1}^0\, \overline{\omega}_{v_2}^0 - \overline{\omega}_{u_1}^0\, \overline{\omega}_{u_2}^0) = 0$,
(b) $m_{i_1 i_2 00 i_5 i_6} = 0$ because $b_{i_1 i_2 00 i_5 i_6} = 0$,
(c) $m_{00 i_3 i_4 i_5 i_6} = 0$ because $b_{00 i_3 i_4 i_5 i_6} = 0$,
(d) $m_{i_1 0 i_3 0 i_5 i_6} = 0$ because $b_{i_1 0 i_3 0 i_5 i_6} = 0$.

□

Example 12.7 *Consider the following* 4×4 *sudoku grid*

1	2	3	4
3	4	1	2
2	1	4	3
4	3	2	1

The corresponding indicator function is

$$F = \frac{1}{4}(1 - R_1C_2S_2)(1 - R_2C_1S_1).$$

If we exchange the second row of the grid with the third one, the coefficient m_{101010} *of* $M(F)$ *is* $1/4$ *and conditions of Corollary 12.2 are not satisfied.*

Proof The second row corresponds to the points of $\mathcal{D}_{12}$ $\zeta_u = (\omega_{u_1}, \omega_{u_2}) = (-1, 1)$ and the third one to $\zeta_v = (\omega_{v_1}, \omega_{v_2}) = (1, -1)$. Then, the move is not valid. Indeed:

$$\begin{aligned} m_{101010} = &-\frac{1}{4}\,\overline{e_{10,hk}} \sum_{\alpha_s \in L_s} b_{\alpha_g,\alpha_s}\, e_{\alpha_s,hk} = \\ &\frac{1}{4}\,(\overline{\omega}^1_{v_1}\overline{\omega}^0_{v_2} - \overline{\omega}^1_{u_1}\overline{\omega}^0_{u_2}) \sum_{\alpha_1=0}^{1}\sum_{\alpha_2=0}^{1} b_{\alpha_1\alpha_2 1010}(\omega^{\alpha_1}_{u_1}\omega^{\alpha_2}_{u_2} - \omega^{\alpha_1}_{v_1}\omega^{\alpha_2}_{v_2}) = \\ &\frac{1}{4}\,(1+1)\,\big(b_{001010}(\omega^0_{u_1}\omega^0_{u_2} - \omega^0_{v_1}\omega^0_{v_2}) \\ &+ b_{011010}(\omega^0_{u_1}\omega^1_{u_2} - \omega^0_{v_1}\omega^1_{v_2})b_{101010}(\omega^1_{u_1}\omega^0_{u_2} - \omega^1_{v_1}\omega^0_{v_2}) \\ &+ b_{111010}(\omega^1_{u_1}\omega^1_{u_2} - \omega^1_{v_1}\omega^1_{v_2})\big) = \frac{1}{2}\,(-\frac{1}{4})(1+1) = -\frac{1}{4}. \end{aligned}$$

□

Proposition 12.10 *We identify the parts of the sudoku grid where the* $\mathcal{M}_3$ *moves can be applied. Fix*

- *a stack:* $C_1 = \omega_t$,
- *two columns of this stack* $C_2 = \omega_{c_u}$ *and* $C_2 = \omega_{c_v}$,
- *two boxes of this stack:* $(R_1, C_1) = (\omega_{b_m}, \omega_t)$ *and* $(R_1, C_1) = (\omega_{b_n}, \omega_t)$.
- *a row in each box:* $(R_1, R_2, C_1) = (\omega_{b_m}, \omega_{r_p}, \omega_t)$ *and* $(R_1, R_2, C_1) = (\omega_{b_n}, \omega_{r_q}, \omega_t)$.

In this way we select two couple of cells, as shown in the following table

R_1	R_2	C_1	C_2	*symbol*
ω_{b_m}	ω_{r_p}	ω_t	ω_{c_u}	a_1
ω_{b_m}	ω_{r_p}	ω_t	ω_{c_v}	a_2
ω_{b_n}	ω_{r_q}	ω_t	ω_{c_u}	a_3
ω_{b_n}	ω_{r_q}	ω_t	ω_{c_v}	a_4

Clearly, analogue identification holds by fixing a band, and then two rows of this band, etc. Moreover, this kind of exchange can be generalised to more than two symbols, simultaneously.

The two couples of cells selected above can be exchanged only if they contain exactly two symbols a_1 and a_2 (i.e. $a_4 = a_1$ and $a_3 = a_2$).

The coefficients of the move are

$$m_{i_1 i_2 i_3 i_4 i_5 i_6} = \frac{1}{p^4}\ \overline{\omega}_t^{i_3}\ \ (-\overline{e_{i_1 i_2, hk}})\ \ n_{i_4 i_5 i_6}$$

where

$$n_{i_4 i_5 i_6} = \sum_{\alpha_s} e_{\alpha_s, hk} \sum_{\alpha_3} \omega_t^{\alpha_3} \sum_{\alpha_4} b_{\alpha_s, \alpha_3, \alpha_4, i_5, i_6} \left(\omega_{c_h}^{[\alpha_4 - i_4]} + \omega_{c_k}^{[\alpha_4 - i_4]}\right).$$

Moreover, it holds:

$$n_{0 i_5 i_6} = 0 \quad \textit{for all} \ \ (i_5, i_6) \in \{0, \cdots, p-1\}^2 \setminus \{(0,0)\}.$$

Proof The new grid has both the boxes, the rows and the columns involved in the moves that still contain all the symbols repeated exactly once.

Let $s = \{1, 2\}$, $\zeta_u = (\omega_{b_m}, \omega_{r_p})$ and $\zeta_v = (\omega_{b_n}, \omega_{r_q})$. We define the following indicator functions of specific parts of the grid:

- S identifying the cells of the stack represented by $C_1 = \omega_s$:

$$S = \frac{1}{p}\left(\sum_{i=0}^{p-1} \left(\overline{\omega}_s C_1\right)^i\right);$$

- K_1 and K_2 identifying the cells of the columns represented by $C_2 = \omega_{c_1}$ and $C_2 = \omega_{c_2}$ respectively:

$$K_1 = \frac{1}{p}\left(\sum_{i=0}^{p-1} \left(\overline{\omega}_{c_1} C_2\right)^i\right) \quad \text{and} \quad K_2 = \frac{1}{p}\left(\sum_{i=0}^{p-1} \left(\overline{\omega}_{c_2} C_2\right)^i\right);$$

- K identifying the cells of both the columns represented by $C_2 = \omega_{c_1}$ and $C_2 = \omega_{c_2}$:

$$K = K_1 + K_2.$$

It follows that the polynomial $F \cdot S \cdot K$ is the indicator function of the cells of the specific sudoku grid in the stack and in both the columns identified by S and K respectively.

The coefficients of the polynomial move can be obtained as in Proposition 12.8, where the coefficients of the indicator function are replaced by those of $F \cdot S \cdot K$. Writing ζ_g as $(\zeta_3, \zeta_4, \zeta_5, \zeta_6)$, the polynomial form of the move is:

$$M(F) = E_{s,hk} \tilde{P}_{g,hk} \tag{24.1}$$

where $E_{s,hk}$ is the usual polynomial and $\tilde{P}_{g,hk}$ is obtained using the indicator function $F \cdot S \cdot K$ in place of F

$$\tilde{P}_{g,hk}(\zeta_g) = (F \cdot S \cdot K)(\omega_{b_m}, \omega_{r_p}, \zeta_g) - (F \cdot S \cdot K)(\omega_{b_n}, \omega_{r_q}, \zeta_g).$$

The expression of the coefficients follows from Equation (24.1), observing that:

$$\begin{aligned}&(F \cdot S \cdot K)(\omega_{b_m}, \omega_{r_p}, \zeta_g) = \\ &S(\zeta_3)K_1(\zeta_4)F(\omega_{b_m}, \omega_{r_p}, \omega_s, \omega_{c_u}, \zeta_5, \zeta_6) + S(\zeta_3)K_2(\zeta_4)F(\omega_{b_m}, \omega_{r_p}, \omega_s, \omega_{c_v}, \zeta_5, \zeta_6).\end{aligned}$$

To be a valid move the coefficients $m_{i_1 i_2 i_3 i_4 i_5 i_6}$ must meet the requirements of Corollary 2. The conditions (a) and (c) are satisfied. Indeed

(a) $m_{i_1 i_2 i_3 i_4 00} = 0$ because $b_{i_1 i_2 i_3 i_4 00} = 0$
(c) $m_{00 i_3 i_4 i_5 i_6} = 0$ because $-e_{0,hk} = \left(\overline{\omega}^0_{b_n}\,\overline{\omega}^0_{r_q} - \overline{\omega}^0_{b_m}\,\overline{\omega}^0_{r_p}\right) = 0$

Both the conditions (b) and (d) become equivalent to $n_{0 i_5 i_6} = 0$. □

Proposition 12.11 *Let σ_1, σ_2 be two exchanges in $\mathcal{M}_1(F)$ and write*

$$\sigma_1(F) = F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1} \quad \text{and} \quad \sigma_2(F) = F + E_{s_2,u_2v_2} P_{g_2,u_2v_2}.$$

where E_{s_i,u_iv_i} and P_{g_i,u_iv_i}, $i = 1, 2$, are defined in Lemma 12.1. The composed move $\sigma_1 \circ \sigma_2$ equals to $\sigma_2 \circ \sigma_1$ if one of the two following conditions holds:

- *$s_1 \cap s_2 = \emptyset$, i.e. the moves act on different factors,*
- *$s_1 = s_2$ and $\{u_1, v_1\} \cap \{u_2, v_2\} = \emptyset$, i.e. the moves act on the same factors and on different bands/rows/stacks/columns/symbols.*

Proof We remind that E_{s_i,u_iv_i}, $i = 1, 2$ depend on the set of variables whose exponents are in L_{s_1} and L_{s_2} respectively. Let's consider the composition of the moves $\sigma_2 \circ \sigma_1$:

$$\begin{aligned}(\sigma_2 \circ \sigma_1)(F) = \sigma_2(\sigma_1(F)) = \sigma_2(F_1) &= F_1 + E_{s_2,u_2v_2} P_{g_2,u_2v_2}(F_1) \\ &= F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1} + E_{s_2,u_2v_2} P_{g_2,u_2v_2}(F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1}).\end{aligned}$$

We focus on $P_{g_2,u_2v_2}(F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1})$.

- If $s_1 \cap s_2 = \emptyset$, then

$$P_{g_2,u_2v_2}(F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1}) = P_{g_2,u_2v_2} + E_{s_1,u_1v_1} P_{g_2,u_2v_2}(P_{g_1,u_1v_1}).$$

The polynomial $P_{g_2,u_2v_2}(P_{g_1,u_1v_1})$ is

$$\begin{aligned}&P_{g_2,u_2v_2}(F(u_1, \zeta_{g_1}) - F(v_1, \zeta_{g_1}) = \\ &\quad F(u_1, u_2, \zeta_{g_{1,2}}) - F(v_1, u_2, \zeta_{g_{1,2}}) - F(u_1, v_2, \zeta_{g_{1,2}}) + F(v_1, v_2, \zeta_{g_{1,2}})\end{aligned}$$

with $g_{1,2} = g_1 \cap g_2$. It follows that

$$\begin{aligned}&\sigma_2 \circ \sigma_1 = F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1} + E_{s_2,u_2v_2} P_{g_2,u_2v_2} + E_{s_2,u_2v_2} E_{s_1,u_1v_1} \times \\ &(F(u_1, u_2, \zeta_{g_{1,2}}) - F(v_1, u_2, \zeta_{g_{1,2}}) - F(u_1, v_2, \zeta_{g_{1,2}}) + F(v_1, v_2, \zeta_{g_{1,2}})) = \sigma_1 \circ \sigma_2.\end{aligned}$$

- If $s_1 = s_2 = s$ and $\{u_1, v_1\} \cap \{u_2, v_2\} = \emptyset$, then

$$P_{g,u_2v_2}(F + E_{s,u_1v_1}\ P_{g,u_1v_1}) = \\ P_{g,u_2v_2} + (E_{s,u_1v_1}(u_2) - E_{s,u_1v_1}(v_2))P_{g,u_1v_1} = P_{g,u_2v_2}$$

being $E_{s,u_1v_1}(u_2) = E_{s,u_1v_1}(v_2) = 0$. It follows that

$$\sigma_2 \circ \sigma_1 = F + E_{s_1,u_1v_1}\ P_{g_1,u_1v_1} + E_{s_2,u_2v_2}\ P_{g_2,u_2v_2} = \sigma_1 \circ \sigma_2.$$

□

Proposition 12.14 *Let $\mathcal{F}$ be a 4×4-sudoku regular fraction. A move in $\mathcal{M}_3(F)$ must satisfy the equation system:*

$$(\omega_{r_p} - \omega_{r_q})b_{0110i_5i_6} - (\omega_{r_p} + \omega_{r_q})b_{1110i_5i_6} = 0 \quad \forall \quad i_5, i_6 \in \{0,1\}.$$

It leads to a non regular fraction.

Proof We proved the system of conditions in the Example 12.10. We observe that only one of the b's is different from 0. If not, also $b_{1000[i_5+j_5][i_6+j_6]}$ must be different from 0 and it does not meet the requirements of Proposition 12.5. It follows that there always exists a solution for each regular fraction: the exchange must be made either on the same row within the band or in two different rows.

The new fraction is non regular. Indeed, referring to the proof of Proposition 12.10, the expression of the move is

$$M(F) = E_{s,hk}\tilde{P}_{g,hk}.$$

Keeping into account that the 2nd roots of unity are ± 1 and that $\omega_{b_n} = -\omega_{b_m}$, we derive the expressions of the polynomials $E_{s,hk}$ and $\tilde{P}_{g,hk}$. For $E_{s,hk}$ we get

$$E_{s,hk} = \frac{1}{4}\left((1 + \omega_{b_n}R_1)(1 + \omega_{b_q}R_2) - (1 + \omega_{b_m}R_1)(1 + \omega_{b_p}R_2)\right) = \\ -\frac{1}{4}\left(2\omega_{b_m}R_1 + (\omega_{r_p} - \omega_{r_q})R_2 + \omega_{b_m}(\omega_{r_p} + \omega_{r_q})R_1R_2\right)$$

We observe that all the three coefficients of $E_{s,hk}$ are equal to 0 or $\pm\frac{1}{2}$ and that the coefficient of R_1 is different from 0 and one of the remaining is different from 0. The expression of $\tilde{P}_{g,hk}$ is

$$\tilde{P}_{g,hk}(\zeta_g) = (F \cdot S \cdot K)(\omega_{b_m}, \omega_{r_p}, \zeta_g) - (F \cdot S \cdot K)(\omega_{b_n}, \omega_{r_q}, \zeta_g)$$

In this case there are only two columns within a stack and so $K = K_1 + K_2 = 1$ and S is $\frac{1}{2}(1 + \omega_s C_1)$. We obtain

$$\tilde{P}_{g,hk}(\zeta_g) = \frac{1}{2}(1 + \omega_s\zeta_3)\left(F(\omega_{b_m}, \omega_{r_p}, \omega_s, \zeta_4, \zeta_5, \zeta_6) - F(-\omega_{b_m}, \omega_{r_q}, \omega_s, \zeta_4, \zeta_5, \zeta_6)\right)$$

and considering the polynomial expression of the indicator function:

$$\tilde{P}_{g,hk} = \frac{1}{2}(1+\omega_s C_1)\times \sum_{\alpha_4,\alpha_5,\alpha_6}\left(\sum_{\alpha_1,\alpha_2,\alpha_3} b_{\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6}\,\omega_{b_m}^{\alpha_1}\,(\omega_{r_p}^{\alpha_2}-(-1)^{\alpha_1}\omega_{r_q}^{\alpha_2})\omega_s^{\alpha_3}\right) C_2^{\alpha_4}\,S_1^{\alpha_5}\,S_2^{\alpha_6}\,.$$

F is the indicator function of a sudoku regular fraction so all its non null coefficients are equal to $\pm\frac{1}{4}$. In particular one of the non null coefficients has $\alpha_1 = 1$ and $\alpha_2 = 0$, by definition of regular fraction and Remark 1. If we indicate with $b_{10\tilde{\alpha_3}\tilde{\alpha_4}\tilde{\alpha_5}\tilde{\alpha_6}}$ such coefficient, the coefficient of $\tilde{P}_{g,hk}$ corresponding to the monomial $C_2^{\tilde{\alpha_4}}\,S_1^{\tilde{\alpha_5}}\,S_2^{\tilde{\alpha_6}}$ is

$$\frac{1}{2}\sum_{\alpha_1,\alpha_2,\alpha_3} b_{\alpha_1\alpha_2\alpha_3\tilde{\alpha_4}\tilde{\alpha_5}\tilde{\alpha_6}}\,\omega_{b_m}^{\alpha_1}\,(\omega_{r_p}^{\alpha_2}-(-1)^{\alpha_1}\omega_{r_q}^{\alpha_2})\omega_s^{\alpha_3}\,.$$

We observe that, in this summation, only $b_{10\tilde{\alpha_3}\tilde{\alpha_4}\tilde{\alpha_5}\tilde{\alpha_6}}$ can be different from 0 in order to satisfy the requirements of Proposition 12.5 and so the coefficient of $\tilde{P}_{g,hk}$ corresponding to the monomial $C_2^{\tilde{\alpha_4}}\,S_1^{\tilde{\alpha_5}}\,S_2^{\tilde{\alpha_6}}$ reduces to

$$\frac{1}{2}b_{10\tilde{\alpha_3}\tilde{\alpha_4}\tilde{\alpha_5}\tilde{\alpha_6}}\,\omega_{b_m}\,(1+1)\omega_s^{\tilde{\alpha_3}} = b_{10\tilde{\alpha_3}\tilde{\alpha_4}\tilde{\alpha_5}\tilde{\alpha_6}}\,\omega_{b_m}\,\omega_s^{\tilde{\alpha_3}}\,.$$

It follows that the coefficient of $M(F)$ corresponding to the monomial $R_1 C_2^{\tilde{\alpha_4}}\,S_1^{\tilde{\alpha_5}}\,S_2^{\tilde{\alpha_6}}$ is equal to $\pm\frac{1}{8}$ and therefore $F_e = F + M(F)$ is an indicator function of a non regular design. □

24.3 Generation and classification of all the 4×4 sudoku

Using CoCoA software all the 288 possible 4×4 sudoku have been found. In order to simplify the presentation we consider only the grids with the symbol 4 in position $(4,4)$. In the Appendix the CoCoA code and the list of obtained sudoku grids and their indicator functions are provided. Among the 72 sudoku grids, 24 correspond to regular fractions and the other 48 correspond to non regular fractions.

There are no 4×4 symmetrical sudoku.

Removing one or two of three symmetry conditions (a)-(c) of Proposition 12.6 there are 6 sudoku in each case; all of them correspond to regular fractions.

We list below some characteristics of the obtained sudoku fractions.

Among the 24 regular fractions:

- 6 fractions which are symmetric with respect to broken rows and broken columns,
- 6 fractions which are symmetric with respect to broken rows and locations,
- 6 fractions which are symmetric with respect to broken columns and locations,
- 6 fractions which are symmetric with respect to symbols only.

All the indicator functions of non regular fractions have 10 terms: the constant (1/4), one interaction with coefficient 1/4, two interactions with coefficients -1/8 and six with coefficients 1/8. We can classify them using the word length pattern of

the indicator function. We denote by i and j the indices of the factors, $i, j \in \{1, 2\}$, and we consider $i \neq j$.

- 16 fractions have the word length pattern (0,0,2,3,4,1) and the term whose coefficient is $1/4$ is either $R_i C_j S_i$ or $R_i C_j S_j$,
- 24 fractions have the word length pattern is (0,0,2,5,2,0) and the term whose coefficient is $1/4$ is either $R_i C_j S_1 S_2$ or $R_1 C_1 C_2 S_{i,j}$ or $R_1 R_2 C_1 S_{i,j}$,
- 8 fractions have the word length pattern is (0,0,4,4,1,0) and the term whose coefficient is $1/4$ is either $R_1 C_1 C_2 S_1 S_2$ or $R_1 R_2 C_1 S_1 S_2$.

Proposition 12.7 allows us also to know how many and which solutions has a partially filled puzzle. It is enough to add to the system on the coefficients the conditions $F(x_j) = 1$, where x_j are the points of $\mathcal{F}$ already known.

For instance, among the 72 previous sudoku with the symbol 4 the position $(4, 4)$ of the sudoku grid, there are 54 sudoku grids with the symbol 3 in position $(1, 1)$ and, among them, there are 45 sudoku with the symbol 2 in position $(2, 3)$. In the Appendix the CoCoA code is provided.

24.3.1 CoCoA code for 4×4 *sudoku*

(A-1) Generation of all the indicator functions with given symmetries.

```
Use R::=Q[b[0..1,0..1,0..1,0..1,0..1,0..1]];
D:=6;   L1:=Tuples([0,1],D);   L2:=L1;   Le:=2^D;
-- LABEL A
L3:=[I | I In 1..Le];
T:=[[[Mod(L1[I,K]+L2[J,K],2)|K In 1..D]|J In 1..Le]|I In 1..Le];
Tab:=[[b[B[1],B[2],B[3],B[4],B[5],B[6]] |B In T[J]]|J In 1..Le];
Coe:=[b[B[1],B[2],B[3],B[4],B[5],B[6]] |B In L1];
LF:=[-Coe[J]+Sum([Coe[I]*Tab[I,J] | I In 1..Le])| J In 1..Le];
LOrth:=[];
For K:=2 To Le Do
 If   (L1[K][1]= 0 And L1[K][2]= 0) -- columns and symbols
 Or (L1[K][3]= 0 And L1[K][4]= 0) -- rows and symbols
 Or (L1[K][5]= 0 And L1[K][6]= 0) -- rows and columns
 Or (L1[K][2]= 0 And L1[K][4]= 0) -- boxes and symbols
 Or (L1[K][1]= 0 And L1[K][4]= 0) -- broken rows and symbols
 Or (L1[K][2]= 0 And L1[K][3]= 0) -- broken columns and symbols
-- Or (L1[K][1]= 0 And L1[K][3]= 0) -- locations and symbols
 Then Append(LOrth, L1[K]);  EndIf;
EndFor;
CoeOrth:=[b[B[1],B[2],B[3],B[4],B[5],B[6]] |B In LOrth];
EvCoeOrth:=[[C,0]|C In CoeOrth];
Append(LF,Sum(Coe)-1);  ---- 4 in  position (4,4)
Fin:=Subst(LF,EvCoeOrth);
Append(Fin,CoeOrth);Fin:=Flatten(Fin);
-- LABEL B
Define BCond(FinCond,B,V);
 FinCond:=Subst(FinCond,B,V);
 Append(FinCond,B-V);
 Return FinCond;
EndDefine;
Define Ord(L);
   L2:=[LT(L[I])-L[I] | I In 1..Len(L)];     K:=L;
   For I:=1 To Len(L) Do K[IndetIndex(LT(L[I]))]:= L2[I]; End;
   L:=K;  Return L;
EndDefine;
```

```
FinCond:=BCond(Fin,b[0,0,0,0,0,0],1/4);
G :=ReducedGBasis(Ideal(FinCond));
E:=QuotientBasis(Ideal(G));Len(E);
-- 6 solutions for symmetry w.r.t. broken rows and broken columns
Define Sol(G,C,V);
 LL:=BCond(G,C,V);
 LL:=ReducedGBasis(Ideal(LL));
 PrintLn C,' = ',V;
 E:=QuotientBasis(Ideal(LL));
 PrintLn 'Number of solution  ',Len(E);
 If Len(E)=1 Then Append(MEMORY.CT,Ord(LL));Else PrintLn LL;EndIf;
 Return LL;
EndDefine;
MEMORY.CT:=[];
```

Solutions for symmetric sudoku w.r.t. broken rows and broken columns

```
G01:=Sol(G,b[1,0,1,1,1,0],1/4);   -- 2 sol
G02:=Sol(G01,b[1,1,1,0,1,1],0);   -- 1 sol
G03:=Sol(G01,b[1,1,1,0,1,1],1/4); -- 1 sol
G04:=Sol(G,b[1,0,1,1,1,0],0);     -- 4 sol
G05:=Sol(G04,b[1,0,1,1,1,1],1/4); -- 2 sol
G06:=Sol(G05,b[1,1,1,0,1,0],1/4); -- 1 sol
G07:=Sol(G05,b[1,1,1,0,1,0],0);   -- 1 sol
G08:=Sol(G04,b[1,0,1,1,1,1],0);   -- 2 sol
G09:=Sol(G08,b[1,1,1,0,1,1],1/4); -- 1 sol
G010:=Sol(G08,b[1,1,1,0,1,1],0);  -- 1 sol
UnSet Indentation;
Len(MEMORY.CT);MEMORY.CT;
```

(A-2) Computation of sudoku grids

```
Use R::=Q[x[1..6]];
CT:=BringIn(MEMORY.CT);
D:=6;
L1:=Tuples([0,1],D);L2:=[[2*L1[I,J]-1|J In 1..6]|I IN 1..64]
SK:=NewMat(4,4); Define Sudo(ZZ,L1,SK);
 For I:= 1 To 64 Do
  If ZZ[I]=1 Then
   R:=2*L1[I,1]+L1[I,2]+1;
   C:=2*L1[I,3]+L1[I,4]+1;
   S:=2*L1[I,5]+L1[I,6]+1;
   SK[R,C]:=S;
  EndIf;
 EndFor;
Return SK; End;
F:=CT;
For J:=1 To Len(CT) Do
 F[J]:=Sum([CT[J,I]*LogToTerm(L1[I])|I In 1..64]);PrintLn(F[J]);
 ZZ:=[Eval(F[J],L2[I])|I In 1..64]; PrintLn(Sudo(ZZ,L1,SK));
EndFor;
```

(A-3) Computation of solutions of incomplete sudoku grids

```
Use S::=Q[ x[1..6]];
L1:=Tuples([0,1],6);
Le:=2^6;
X:=  [LogToTerm(L1[I]) |I In 1..Le];
Use R::=Q[b[0..1,0..1,0..1,0..1,0..1,0..1], x[1..6]];
X:=BringIn(X);
L1:=BringIn(L1);
```

Continue from Label A to Label B of Item (i)

```
 MEMORY.EvCoe:=EvCoeOrth;
 Define PS(F,S,Fin);
  P:=Subst(F,S);
  Point:=Subst(P,MEMORY.EvCoe);
  Append(Fin,P);Fin:=Flatten(Fin);
 Return Fin;
 EndDefine;
Fin:=PS(F,[[x[1],-1],[x[2],-1],[x[3],-1],[x[4],-1],[x[5],1],
    [x[6],-1]],Fin);
Fin:=PS(F,[[x[1],-1],[x[2],1],[x[3],1],[x[4],-1],[x[5],-1],
    [x[6],1]],Fin);
Use RR::=Q[b[0..1,0..1,0..1,0..1,0..1,0..1]];
Fin:=BringIn(Fin);
```

Continue from Label B of Item (i)

24.3.2 4×4 *sudoku regular fractions*

There are 96 regular fractions. Among them, 24 are symmetric for broken rows and broken columns, 24 are symmetric for broken rows and locations, 24 are symmetric for broken columns and locations, 24 are symmetric for symbols only. There are no 4×4 symmetrical sudoku.

We list only the sudoku with the symbol 4 in the position $(16, 16)$ of the grid.

After the grids we show the terms of the indicator functions; all the coefficients are $1/4$.

(A-1) Symmetric fractions for broken rows and broken columns, non symmetric for locations:

3	2	4	1
1	4	2	3
4	1	3	2
2	3	1	4

2	3	4	1
1	4	3	2
4	1	2	3
3	2	1	4

3	1	4	2
2	4	1	3
4	2	3	1
1	3	2	4

$$
\begin{array}{lll}
R_1R_2C_1S_1S_2 & R_1C_1C_2S_2 & R_2C_2S_1 \quad 1 \\
R_1R_2C_1S_1S_2 & R_1C_1C_2S_1 & R_2C_2S_2 \quad 1 \\
(R_1C_1C_2S_1S_2 & R_1R_2C_1S_2 & R_2C_2S_1 \quad 1
\end{array}
$$

2	1	4	3
3	4	1	2
4	3	2	1
1	2	3	4

1	2	4	3
3	4	2	1
4	3	1	2
2	1	3	4

1	3	4	2
2	4	3	1
4	2	1	3
3	1	2	4

$$
\begin{array}{lll}
R_1C_1C_2S_1S_2 & R_1R_2C_1S_1 & R_2C_2S_2 \quad 1 \\
R_1R_2C_1S_1 & R_1C_1C_2S_2 & R_2C_2S_1S_2 \quad 1 \\
R_1C_1C_2S_1 & R_1R_2C_1S_2 & R_2C_2S_1S_2 \quad 1
\end{array}
$$

(A-2) Symmetric fractions for broken rows and locations, non symmetric for broken columns:

2	3	4	1
4	1	2	3
1	4	3	2
3	2	1	4

2	1	4	3
4	3	2	1
3	4	1	2
1	2	3	4

3	1	4	2
4	2	3	1
2	4	1	3
1	3	2	4

$R_1R_2C_1S_1S_2$ $R_2C_1C_2S_1$ $R_1C_2S_2$ 1
$R_2C_1C_2S_1S_2$ $R_1R_2C_1S_1$ $R_1C_2S_2$ 1
$R_2C_1C_2S_1S_2$ $R_1R_2C_1S_2$ $R_1C_2S_1$ 1

3	2	4	1
4	1	3	2
1	4	2	3
2	3	1	4

1	2	4	3
4	3	1	2
3	4	2	1
2	1	3	4

1	3	4	2
4	2	1	3
2	4	3	1
3	1	2	4

$R_1R_2C_1S_1S_2$ $R_2C_1C_2S_2$ $R_1C_2S_1$ 1
$R_1R_2C_1S_1$ $R_2C_1C_2S_2$ $R_1C_2S_1S_2$ 1
$R_2C_1C_2S_1$ $R_1R_2C_1S_2$ $R_1C_2S_1S_2$ 1

(A-3) Symmetric fractions for broken columns and locations, non symmetric for broken rows:

3	4	2	1
1	2	4	3
4	3	1	2
2	1	3	4

2	4	3	1
1	3	4	2
4	2	1	3
3	1	2	4

3	4	1	2
2	1	4	3
4	3	2	1
1	2	3	4

$R_1R_2C_2S_1S_2$ $R_1C_1C_2S_2$ $R_2C_1S_1$ 1
$R_1R_2C_2S_1S_2$ $R_1C_1C_2S_1$ $R_2C_1S_2$ 1
$R_1C_1C_2S_1S_2$ $R_1R_2C_2S_2$ $R_2C_1S_1$ 1

2	4	1	3
3	1	4	2
4	2	3	1
1	3	2	4

1	4	3	2
2	3	4	1
4	1	2	3
3	2	1	4

1	4	2	3
3	2	4	1
4	1	3	2
2	3	1	4

$R_1C_1C_2S_1S_2$ $R_1R_2C_2S_1$ $R_2C_1S_2$ 1
$R_1C_1C_2S_1$ $R_1R_2C_2S_2$ $R_2C_1S_1S_2$ 1
$R_1R_2C_2S_1$ $R_1C_1C_2S_2$ $R_2C_1S_1S_2$ 1

(A-4) Symmetric fractions for locations only:

4	2	3	1
3	1	4	2
2	4	1	3
1	3	2	4

4	3	2	1
2	1	4	3
3	4	1	2
1	2	3	4

4	3	1	2
1	2	4	3
3	4	2	1
2	1	3	4

$R_1R_2C_1C_2S_1S_2$ $R_1C_2S_1$ $R_2C_1S_2$ 1
$R_1R_2C_1C_2S_1S_2$ $R_2C_1S_1$ $R_1C_2S_2$ 1
$R_1R_2C_1C_2S_2$ $R_1C_2S_1S_2$ $R_2C_1S_1$ 1

4	1	3	2
3	2	4	1
1	4	2	3
2	3	1	4

4	2	1	3
1	3	4	2
2	4	3	1
3	1	2	4

4	1	2	3
2	3	4	1
1	4	3	2
3	2	1	4

$R_1R_2C_1C_2S_2$ $R_2C_1S_1S_2$ $R_1C_2S_1$ 1
$R_1R_2C_1C_2S_1$ $R_1C_2S_1S_2$ $R_2C_1S_2$ 1
$R_1R_2C_1C_2S_1$ $R_2C_1S_1S_2$ $R_1C_2S_2$ 1

24.3.3 4×4 *non-regular sudoku fractions*

There are 192 non regular fractions. We list only the sudoku with the symbol 4 in the position $(16, 16)$ of the grid.

All the indicator functions have 10 terms: the constant (1/4), one interaction with coefficient 1/4, two interactions with coefficients -1/8 and six with coefficients 1/8.

After the grids we show the terms of the indicator functions; in bold the interactions whose coefficients are 1/4.

We can classify the fractions using the word length pattern of the indicator function. We denote by i and j the indices of the factors $i, j \in \{1, 2\}$ and we consider $i \neq j$.

(A-1) The word length pattern of the indicator function is (0,0,2,3,4,1).

The interactions whose coefficients are 1/4 are of the form:

$$R_i C_j S_i \quad \text{or} \quad R_i C_j S_j$$

(a)

2	4	1	3
1	3	4	2
4	2	3	1
3	1	2	4

4	2	1	3
3	1	4	2
2	4	3	1
1	3	2	4

2	4	3	1
3	1	4	2
4	2	1	3
1	3	2	4

4	2	3	1
1	3	4	2
2	4	1	3
3	1	2	4

b_{100110}	b_{100111}	b_{101110}	b_{101111}	b_{110110}	b_{110111}	b_{111110}	b_{111111}
-	+	+	+	+	+	+	-
+	+	-	+	+	-	+	+
+	-	+	+	+	+	-	+
+	+	+	-	-	+	+	+

$R_1R_2C_1C_2S_1S_2$ $R_1R_2C_1C_2S_1$ $R_1R_2C_2S_1S_2$ $R_1C_1C_2S_1S_2$
$R_1R_2C_2S_1$ $R_1C_1C_2S_1$ $R_1C_2S_1S_2$ $R_1C_2S_1$
$\mathbf{R_2C_1S_2}$ $\mathbf{1}$

(b)

3	4	1	2
1	2	4	3
4	3	2	1
2	1	3	4

4	3	2	1
1	2	4	3
3	4	1	2
2	1	3	4

4	3	1	2
2	1	4	3
3	4	2	1
1	2	3	4

3	4	2	1
2	1	4	3
4	3	1	2
1	2	3	4

b_{100101}	b_{100111}	b_{101101}	b_{101111}	b_{110101}	b_{110111}	b_{111101}	b_{111111}
-	+	+	+	+	+	+	-
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+
+	-	+	+	+	+	-	+

$R_1R_2C_1C_2S_1S_2$ $R_1R_2C_1C_2S_2$ $R_1R_2C_2S_1S_2$ $R_1C_1C_2S_1S_2$
$R_1R_2C_2S_2$ $R_1C_1C_2S_2$ $R_1C_2S_1S_2$ $\mathbf{R_2C_1S_1}$
$R_1C_2S_2$ $\mathbf{1}$

(c)

2	1	4	3
4	3	2	1
1	4	3	2
3	2	1	4

4	1	2	3
2	3	4	1
3	4	1	2
1	2	3	4

4	3	2	1
2	1	4	3
1	4	3	2
3	2	1	4

2	3	4	1
4	1	2	3
3	4	1	2
1	2	3	4

b_{011010}	b_{011011}	b_{011110}	b_{011111}	b_{111010}	b_{111011}	b_{111110}	b_{111111}
+	-	+	+	+	+	-	+
-	+	+	+	+	+	+	-
+	+	-	+	+	-	+	+
+	+	+	-	-	+	+	+

$R_1R_2C_1C_2S_1S_2$ $R_1R_2C_1C_2S_1$ $R_1R_2C_1S_1S_2$ $R_2C_1C_2S_1S_2$
$R_1R_2C_1S_1$ $R_2C_1C_2S_1$ $R_2C_1S_1S_2$ $R_2C_1S_1$
$\mathbf{R_1C_2S_2}$ $\mathbf{1}$

(d)

3	1	4	2
4	2	3	1
1	4	2	3
2	3	1	4

4	2	3	1
3	1	4	2
1	4	2	3
2	3	1	4

4	1	3	2
3	2	4	1
2	4	1	3
1	3	2	4

3	2	4	1
4	1	3	2
2	4	1	3
1	3	2	4

b_{011001}	b_{011011}	b_{011101}	b_{011111}	b_{111001}	b_{111011}	b_{111101}	b_{111111}
+	-	+	+	+	+	-	+
-	+	+	+	+	+	+	-
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+

$R_1R_2C_1C_2S_1S_2$ $R_1R_2C_1C_2S_2$ $R_1R_2C_1S_1S_2$ $R_2C_1C_2S_1S_2$
$R_1R_2C_1S_2$ $R_2C_1C_2S_2$ $R_2C_1S_1S_2$ $\mathbf{R_1C_2S_1}$
$R_2C_1S_2$ $\mathbf{1}$

(A-2) The word length pattern of the indicator function is (0,0,2,5,2,0).
The interactions whose coefficients are 1/4 are of the form:

$$R_iC_jS_1S_2 \quad \text{or} \quad R_1C_1C_2S_{i,j} \quad \text{or} \quad R_1R_2C_1S_{i,j}$$

(a)

4	1	2	3
3	2	4	1
1	4	3	2
2	3	1	4

4	1	3	2
2	3	4	1
1	4	2	3
3	2	1	4

1	4	2	3
2	3	4	1
4	1	3	2
3	2	1	4

1	4	3	2
3	2	4	1
4	1	2	3
2	3	1	4

b_{100101}	b_{100110}	b_{101101}	b_{101110}	b_{110101}	b_{110110}	b_{111101}	b_{111110}
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+
+	-	+	+	+	+	-	+
-	+	+	+	+	+	+	-

$R_1R_2C_1C_2S_1$ $R_1R_2C_1C_2S_2$ $R_1R_2C_2S_1$ $R_1C_1C_2S_1$
$R_1R_2C_2S_2$ $R_1C_1C_2S_2$ $\mathbf{R_2C_1S_1S_2}$ $R_1C_2S_1$
$R_1C_2S_2$ $\mathbf{1}$

(b)

1	2	4	3
4	3	1	2
2	4	3	1
3	1	2	4

1	3	4	2
4	2	1	3
3	4	2	1
2	1	3	4

4	2	1	3
1	3	4	2
3	4	2	1
2	1	3	4

4	3	1	2
1	2	4	3
2	4	3	1
3	1	2	4

b_{011001}	b_{011010}	b_{011101}	b_{011110}	b_{111001}	b_{111010}	b_{111101}	b_{111110}
+	-	+	+	+	+	-	+
-	+	+	+	+	+	+	-
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+

$R_1R_2C_1C_2S_1$ $R_1R_2C_1C_2S_2$ $R_1R_2C_1S_1$ $R_2C_1C_2S_1$
$R_1R_2C_1S_2$ $R_2C_1C_2S_2$ $\mathbf{R_1C_2S_1S_2}$ $R_2C_1S_1$
$R_2C_1S_2$ $\mathbf{1}$

(c)

1	2	4	3
3	4	2	1
4	1	3	2
2	3	1	4

3	2	4	1
1	4	2	3
4	3	1	2
2	1	3	4

3	4	2	1
1	2	4	3
4	1	3	2
2	3	1	4

1	4	2	3
3	2	4	1
4	3	1	2
2	1	3	4

b_{010110}	b_{010111}	b_{011010}	b_{011011}	b_{110110}	b_{110111}	b_{111010}	b_{111011}
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+
+	-	+	+	+	+	-	+
-	+	+	+	+	+	+	-

$R_1R_2C_1S_1S_2$ $R_1R_2C_2S_1S_2$ $R_1R_2C_1S_1$ $R_1R_2C_2S_1$
$\mathbf{R_1C_1C_2S_2}$ $R_2C_1S_1S_2$ $R_2C_2S_1S_2$ $R_2C_1S_1$
$R_2C_2S_1$ $\mathbf{1}$

(d)

1	3	4	2
2	4	3	1
4	1	2	3
3	2	1	4

2	3	4	1
1	4	3	2
4	2	1	3
3	1	2	4

1	4	3	2
2	3	4	1
4	2	1	3
3	1	2	4

2	4	3	1
1	3	4	2
4	1	2	3
3	2	1	4

b_{010101}	b_{010111}	b_{011001}	b_{011011}	b_{110101}	b_{110111}	b_{111001}	b_{111011}
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+
-	+	+	+	+	+	+	-
(+	-	+	+	+	+	-	+

$R_1R_2C_1S_1S_2$ $R_1R_2C_2S_1S_2$ $\mathbf{R_1C_1C_2S_1}$ $R_1R_2C_1S_2$
$R_1R_2C_2S_2$ $R_2C_1S_1S_2$ $R_2C_2S_1S_2$ $R_2C_1S_2$
$R_2C_2S_2$ $\mathbf{1}$

(e)

3	1	4	2
2	4	3	1
4	2	1	3
1	3	2	4

1	3	4	2
2	4	1	3
4	2	3	1
3	1	2	4

3	1	4	2
4	2	1	3
2	4	3	1
1	3	2	4

1	3	4	2
4	2	3	1
2	4	1	3
3	1	2	4

b_{010110}	b_{010111}	b_{011110}	b_{011111}	b_{100110}	b_{100111}	b_{101110}	b_{101111}
+	-	+	+	+	+	-	+
+	+	-	+	+	-	+	+
+	+	+	-	-	+	+	+
-	+	+	+	+	+	+	-

$R_1C_1C_2S_1S_2$ $R_2C_1C_2S_1S_2$ $R_1C_1C_2S_1$ $R_2C_1C_2S_1$
$\mathbf{R_1R_2C_1S_2}$ $R_1C_2S_1S_2$ $R_2C_2S_1S_2$ $R_1C_2S_1$
$R_2C_2S_1$ $\mathbf{1}$

(f)

2	1	4	3
4	3	1	2
3	4	2	1
1	2	3	4

2	1	4	3
3	4	2	1
4	3	1	2
1	2	3	4

1	2	4	3
3	4	1	2
4	3	2	1
2	1	3	4

1	2	4	3
4	3	2	1
3	4	1	2
2	1	3	4

b_{010101}	b_{010111}	b_{011101}	b_{011111}	b_{100101}	b_{100111}	b_{101101}	b_{101111}
+	-	+	+	+	+	-	+
+	+	-	+	+	-	+	+
+	+	+	-	-	+	+	+
-	+	+	+	+	+	+	-

$R_1C_1C_2S_1S_2$ $R_2C_1C_2S_1S_2$ $\mathbf{R_1R_2C_1S_1}$ $R_1C_1C_2S_2$
$R_2C_1C_2S_2$ $R_1C_2S_1S_2$ $R_2C_2S_1S_2$ $R_1C_2S_2$
$R_2C_2S_2$ $\mathbf{1}$

(A-3) The word length pattern of the indicator function is (0,0,4,4,1,0). The interactions whose coefficients are 1/4 are of the form:

$$R_1C_1C_2S_1S_2 \quad \text{or} \quad R_1R_2C_1S_1S_2$$

(a)

2	1	4	3
3	4	1	2
4	2	3	1
1	3	2	4

3	1	4	2
2	4	1	3
4	3	2	1
1	2	3	4

3	4	1	2
2	1	4	3
4	2	3	1
1	3	2	4

2	4	1	3
3	1	4	2
4	3	2	1
1	2	3	4

b_{010101}	b_{010110}	b_{011001}	b_{011010}	b_{110101}	b_{110110}	b_{111001}	b_{111010}
+	+	+	-	-	+	+	+
+	+	-	+	+	-	+	+
-	+	+	+	+	+	+	-
+	-	+	+	+	+	-	+

$\mathbf{R_1C_1C_2S_1S_2}$ $R_1R_2C_1S_1$ $R_1R_2C_2S_1$ $R_1R_2C_1S_2$
$R_1R_2C_2S_2$ $R_2C_1S_1$ $R_2C_2S_1$ $R_2C_1S_2$
$R_2C_2S_2$ $\mathbf{1}$

(b)

2	3	4	1
4	1	3	2
1	4	2	3
3	2	1	4

2	3	4	1
1	4	2	3
4	1	3	2
3	2	1	4

3	2	4	1
1	4	3	2
4	1	2	3
2	3	1	4

3	2	4	1
4	1	2	3
1	4	3	2
2	3	1	4

b_{010101}	b_{010110}	b_{011101}	b_{011110}	b_{100101}	b_{100110}	b_{101101}	b_{101110}
-	+	+	+	+	+	+	-
+	-	+	+	+	+	-	+
+	+	-	+	+	-	+	+
+	+	+	-	-	+	+	+

$\mathbf{R_1R_2C_1S_1S_2}$ $R_1C_1C_2S_1$ $R_2C_1C_2S_1$ $R_1C_1C_2S_2$
$R_2C_1C_2S_2$ $R_1C_2S_1$ $R_2C_2S_1$ $R_1C_2S_2$
$R_2C_2S_2$ $\mathbf{1}$

25

On-line Supplement to Replicated measurements and algebraic statistics

Roberto Notari and Eva Riccomagno

25.1 Proofs

Theorem 11.3 *Consider n distinct points $P_1, \ldots, P_n \in \mathbb{A}^k$ with P_i of coordinates $(a_{i1}, \ldots, a_{ik})$, and let $X = \{P_1, \ldots, P_n\}$. Then $J = \bigcap_{i=1}^{n} \langle x_1 - t a_{a1}, \ldots, x_k - t a_{ik} \rangle \subset S = \mathcal{K}[x_1, \ldots, x_k, t]$ is a flat family. Its special fibre is the origin with multiplicity n and it is defined by the ideal $I_0 = \{F \in R : F$ is homogeneous and there exists $f \in I(X)$ such that $F = \mathrm{LF}(f)\}$. Moreover, the Hilbert function does not depend on t.*

Proof At first, we prove that the ideal $J \subset S$ is homogeneous, that is to say, if $f \in J$ and $f = f_0 + \cdots + f_s$ with f_i homogeneous of degree j, then $f_i \in J$ for every $i = 0, \ldots, s$.

By definition, if $f \in J$ then $f \in \langle x_1 - t a_{i1}, \ldots, x_k - t a_{ik} \rangle$ for $i = 1, \ldots, n$, that is to say, $f(t, t a_{i1}, \ldots, t a_{ik})$ is the null polynomial in the variable t. Let $t^m x_1^{m_1} \ldots x_n^{m_k}$ be a term of degree $M = m + m_1 + \ldots m_n$. If we evaluate it at $(t, t a_{i1}, \ldots, t a_{ik})$ we obtain $(a_{i1}^{m_1} \ldots a_{in}^{m_k}) t^M$. Hence, if $f = f_0 + \cdots + f_s$ with f_j homogeneous of degree j, then $f(t, t a_{i1}, \ldots, t a_{ik}) = c_0 t^0 + \cdots + c_s t^s$ where $c_j = f_j(1, a_{i1}, \ldots, a_{ik})$. The polynomial $f(t, t a_{i1}, \ldots, t a_{ik})$ is the null polynomial and thus, for every j and every i, we have $f_j(1, a_{i1}, \ldots, a_{ik}) = 0$. The homogeneity of f_j guarantees that $f_j(t, t a_{i1}, \ldots t a_{ik}) = 0$ as well, and so $f_j \in \langle x_1 - t a_{i1}, \ldots, x_k - t a_{ik} \rangle$ for every j and i. The first claim then follows.

A remarkable property of homogeneous ideals in polynomial rings is that they can be generated by homogeneous polynomials. Secondly, we prove that $J = \langle t^s f_0 + \cdots + t^0 f_s : f = f_0 + \cdots + f_s \in I(X), f_j$ homogeneous of degree $j \rangle$. Let $F = t^s f_0 + \ldots t^0 f_s \in S$ with $f = f_0 + \cdots + f_s \in I(X)$. Then F is homogeneous of degree s, $f(a_{i1}, \ldots, a_{ik}) = 0$ and $F(t, t a_{i1}, \ldots, t a_{ik}) = t^s f(a_{i1}, \ldots a_{ik}) = 0$. Hence, $F \in \langle x_1 - t a_{i1}, \ldots, x_k - t a_{ik} \rangle$ for every i and so $F \in J$. Conversely, if $F \in J$ is homogeneous, then $f(a_{i1}, \ldots, a_{ik}) = F(1, a_{i1}, \ldots, a_{ik}) = 0$ for every i and so $f \in I(X)$.

To simplify notation, set $h(f, t) = t^s f_0 + \cdots + t^0 f_s$ where $f = f_0 + \cdots + f_s$ and f_j is homogeneous of degree j.

Now, we prove that there exists a monomial ideal $L \subset R$ such that $\mathrm{LT}(J) = L$ with respect a term-ordering $\preccurlyeq$ which satisfies the following properties:

(A-1) $t \preccurlyeq x_1 \preccurlyeq \cdots \preccurlyeq x_k$;

(A-2) over R, $\preccurlyeq$ is graded;
(A-3) $t^l x_1^{l_1} \dots x_k^{l_k} \prec t^m x_1^{m_1} \dots x_k^{m_k}$ if $x_1^{l_1} \dots x_k^{l_k} \prec x_1^{m_1} \dots x_k^{m_k}$ or $x_1^{l_1} \dots x_k^{l_k} = x_1^{m_1} \dots x_l^{m_k}$ and $l < m$.

With respect to $\preccurlyeq$, $\mathrm{LT}(F) = \mathrm{LT}(f_s) \in R$ for every $F = h(f,t) \in J$, with $f \in I(X)$. Furthermore, if $\mathcal{G} = \{g_1, \dots, g_m\}$ is a Gröbner base of $I(X)$ with respect to $\preccurlyeq$, then $\{h(g_1,t), \dots, h(g_m,t)\}$ is a Gröbner base of J with respect to $\preccurlyeq$. Hence, $\mathrm{LT}(J) = \mathrm{LT}(I(X)) \subset R$ and the claim follows.

For every $t_0 \in \mathcal{K}$, a Gröbner base of $\langle J, t - t_0 \rangle$ is then

$$\{h(g_1,t), \dots, h(g_m,t), t - t_0\}$$

because $\mathrm{GCD}(\mathrm{LT}(h(g_i,t)),t) = 1$, for every $i = 1, \dots, m$ and GCD stands for greatest common divisor. It follows that the Hilbert function of $S/\langle J, t - t_0 \rangle$ is equal to the Hilbert function of X and so it does not depend on $t_0 \in \mathcal{K}$. The family J is then flat and the claim follows. In particular,

$$\langle J, t \rangle = \langle \mathrm{LF}(f) : f \in I(X) \rangle.$$

□

Theorem 11.4 *Let $X = \{P_1, \dots, P_r\}, Y = \{Q_1, \dots, Q_s\}$ be sets of points in $\mathbb{A}^k$, and assume that $Z = X \cup Y$ has degree $n = r + s$; that is, n distinct points. If P_i has coordinates $(a_{i1}, \dots, a_{ik})$ then the family*

$$J = \bigcap_{i=1}^{r} \langle x_1 - t a_{i1}, \dots, x_k - t a_{ik} \rangle \cap I(Q_1) \cap \dots \cap I(Q_s)$$

is flat, with fibers of dimension 0 and degree $r + s$.

Proof Assume first that $P_i \neq O$ and $Q_i \notin l_j$ for each i, j where l_j is the line through P_j and the origin O. Then, for each $t_0 \neq 0$, the points $P_1(t_0), \dots, P_r(t_0), Q_1, \dots, Q_s$ are distinct. We have to check that J is flat also for $t_0 = 0$. If $t^a g \in J$ for some $g \in S$, then $t^a g \in I(Q_j)$ for every j and $t^a g \in J' = \bigcap_{i=1}^{r} \langle x_1 - t a_{i1}, \dots, x_k - t a_{ik} \rangle$. The ideal $I(Q_j)$ is prime and $t \notin I(Q_j)$. Then $I(Q_j) \ni g$. From the proof of Theorem 11.3 it follows that $g \in J'$ and so $g \in J$. Hence, J is flat also for $t_0 = 0$ and the claim follows.

If one or more points among the Q_j's belong to some lines among $l_1, \dots, l_r$ then for some values we obtain some double points, but the family is still flat as a straightforward computation shows.

If one point among the Q_j's or one among the P_i's is the origin, then again the family is flat for the same reasons as before. □

Theorem 11.8 *In the hypotheses and notation of Theorem 11.7, for every $i = 1, \dots, r$ it holds*

$$c_i(0) = \frac{\det(D_{i,m_i})}{\det(A(1))}.$$

Proof The hypotheses guarantee that the polynomial c_i is equal to

$$c_i = \frac{\sum_{h=m_i}^{b} t^{m+h-m_i} \det(D_{ih})}{t^m \det(A(1))} = \sum_{h=m_i} t^{h-m_i} \frac{\det(D_{ih})}{\det(A(1))}.$$

Hence, $c_i(0) = \det(D_{i,m_i})/\det(A(1))$. □

Theorem 11.9 *Let $Y = \{A_1, \ldots, A_m\} \subset \mathbb{A}^k$ be a set of distinct points, and let $X_i = \{P_{i1}, \ldots, P_{ir_i}\}$ be a set of r_i distinct points such that $Z = X_1 \cup \cdots \cup X_m$ has degree $r = r_1 + \cdots + r_m$. Let J_i be the $I(A_i)-$primary ideal of degree r_i obtained by collapsing X_i to A_i as in previous Theorem 11.6, and let $J = J_1 \cap \cdots \cap J_m$. Let $F_i \in \frac{R}{J_i}$ be the limit interpolating polynomial computed as in Theorem 11.7. Then there exists a unique polynomial $F \in \frac{R}{J}$ such that $F \mod J_i = F_i$.*

Proof The existence and uniqueness of F is a consequence of the isomorphism between $\frac{R}{J}$ and $\frac{R}{J_1} \oplus \cdots \oplus \frac{R}{J_m}$ because $J_i + J_j = R$ for every $i \neq j$. In fact, the sum of ideals correspond to the intersection of the algebraic sets associated, but $A_i \neq A_j$ and so the intersection is empty.

Now we want to describe an algorithm to get F starting from $F_1, \ldots, F_m$, from a monomial base of R/J, and from Gröbner bases $\mathcal{G}_i$ of J_i. To fix ideas, assume that $\mathcal{G}_1 = \{g_1, \ldots, g_t\}$.

Let $M_1 = 1, M_2, \ldots, M_r$ be a monomial basis of $\frac{R}{J}$, and assume that $M_1, M_2, \ldots, M_{r_1}$ is a monomial base of $\frac{R}{J_1}$. Then, for $j = r_1 + 1, \ldots, r$, there exists $\sigma(j)$ such that $M_j = LT(g_{\sigma(j)})N_j$ for a suitable monomial N_j. From the fact that $M_1, M_2, \ldots, M_r$ is a base of R/J, it follows that also $M_1, \ldots, M_{r_1}, N_{r_1+1}g_{\sigma(r_1+1)}, \ldots, N_r g_{\sigma(r)}$ is a base of R/J. The second base has the property that $N_j g_{\sigma(j)} = 0$ in R/J_1 and so their cosets are a base of $R/(J_2 \cap \cdots \cap J_m) \cong R/J_2 \oplus \cdots \oplus R/J_m$. Hence, every interpolation problem has a unique solution as linear combination of the $N_j g_{\sigma(j)}$'s.

Let $H = \sum_{j=r_1+1}^{r} a_j N_j g_{\sigma(j)} \in J_1$, and let $F = F_1 + H \in R/J$.

By its properties, we have that $F - F_i \in J_i$, for $i = 1, \ldots, m$. Then, we impose that $NF(F_1 + H - F_i) = 0$ in R/J_i. By rewriting the polynomial $F_1 + H - F_i$ modulo $\mathcal{G}_i$ we get a polynomial with coefficients that are linear polynomials in the variables $a_{r_1+1}, \ldots, a_r$. The coefficients must be zero because the normal form is 0 and so we get a linear system in the variables a_i's. The only solution gives the only H and so we get F as claimed. □

26

On-line Supplement to Geometry of extended exponential models

Daniele Imparato and Barbara Trivellato

26.1 Proofs

Proposition 19.2 *Suppose that $(\Omega, \mathcal{F}, \mu)$ is not atomic with a finite number of atoms.*

(A-1) $L_0^{\Phi_1}(p)$ *is a non-separable space.*
(A-2) $C_p = \overline{L^\infty \cap L_0^{\Phi_1}(p)} \neq L_0^{\Phi_1}(p)$.
(A-3) $\mathcal{K}_p$ *is neither a closed nor an open set.*
(A-4) $\mathcal{S}_p$ *satisfies a cylindrical property, that is, if* $v \in \mathcal{S}_p$ *then* $v + C_p \in \mathcal{S}_p$.

Proof For Items (A-1) and (A-2), see (Rao and Ren 2002). For Item (A-3), consider the Lebesgue measure on $[0,1]$ and let

$$u_n(x) = \log\left(\frac{1}{x^{1-\frac{1}{n}}}\right) - \mathrm{E}_p\left[\log\left(\frac{1}{x^{1-\frac{1}{n}}}\right)\right].$$

It should be noted that, for each $n \in \mathbb{N}$, $u_n \in \mathcal{K}_p$. More precisely, $u_n \in \mathcal{S}_p$. In fact, let $\alpha_n = 1 + 1/n$, β_n its conjugate exponent and $t_n = 1/\beta_n$. Then from Hölder's inequality one obtains that, for each $v \in B_p$, $\|v\|_{B_p} < 1$

$$\mathrm{E}_p\left[\mathrm{e}^{u_n + t_n v}\right] < \left(\mathrm{E}_p\left[\mathrm{e}^{\alpha_n u_n}\right]\right)^{1/\alpha_n} \left(\mathrm{E}_p\left[\mathrm{e}^{v}\right]\right)^{1/\beta_n} < \infty.$$

However, the sequence $(u_n)_n$ tends in norm to $u(x) = -\log(x) + \mathrm{E}_p[\log(x)]$, which does not belong to $\mathcal{K}_p$. This proves that $\mathcal{K}_p$ is not a closed set. In order to prove that $\mathcal{K}_p$ is not an open set in general, let μ be the uniform distribution on $[0, 1/2]$ and let $u(x) = -\log(x\log^2(x)) + \mathrm{E}_p\left[\log(x\log^2(x))\right]$. It is straightforward to see that $u(x)$ belongs to $\mathcal{K}_p \setminus \mathcal{S}_p$. For Item (A-4), let $v \in \mathcal{S}_p$, so that $\alpha v \in \mathcal{S}_p$ for some $\alpha > 0$, and let $u \in C_p$. Then, if $\lambda = 1/\alpha$ and $t = 1/(1-\lambda)$, it holds that

$$\lambda\alpha v + (1-\lambda)tu = u + v,$$

that is, $u + v \in \mathcal{S}_p$ as a convex combination of elements which belong to $\mathcal{S}_p$. □

Proposition 19.4 *The following statements are equivalent.*

(A-1) $q \in \widehat{\mathcal{E}}(p)$.
(A-2) $\log(q/p) \in L^{\Phi_1}(p)$.

(A-3) $p/q \in L^a(p)$ *for some* $a > 0$.
(A-4) $q = \mathrm{e}^{u - K_p(u)} \cdot p$ *for some* $u \in \mathcal{K}_p$.
(A-5) *A sequence* $q_n = e_p(u_n)$, $u_n \in \mathcal{S}_p$, $n = 1, 2, \ldots$, *exists so that* $\lim_{n\to\infty} u_n = u$ μ*-a.s. and in* $L^{\Phi_1}(p)$, $\lim K_p(u_n) = K_p(u)$, *and* $q = \mathrm{e}^{u - K_p(u)} \cdot p$.

Proof The equivalence between (A-1) and (A-2) easily follows from the definition of the exponential arc. Let $p(t)$ be a left open exponential arc connecting q to p; namely, $p(t) = \mathrm{e}^{tu - K_p(tu)} p$, $t \in (-\alpha, 1]$, $\alpha > 0$, with $p(0) = p$ and $p(1) = q$. For $p(t)$ to be an exponential model, it is necessary and sufficient that $u = \log(q/p)$ belongs to $L^{\Phi_1}(p)$.

It is trivial to say that if q satisfies (A-4), then $q \in \widehat{\mathcal{E}}(p)$. Conversely, let us suppose that $\log(q/p) \in L^{\Phi_1}(p)$; namely, $q = \mathrm{e}^v p$, where $v \in L^{\Phi_1}(p)$. Then, by centring v, we obtain

$$q = \mathrm{e}^{u - K_p(u)} p,$$

where $u = v - \mathrm{E}_p[v]$ and $K_p[u] = -\mathrm{E}_p[v]$, which is finite since $L^{\Phi_1}(p) \subset L^1(p)$. Therefore, $q \in \widehat{\mathcal{E}}(p)$.

In order to prove the equivalence between (A-1) and (A-5), let $q \propto \mathrm{e}^u p$, $q \in \widehat{\mathcal{E}}(p)$, $(t_n)_n$ be an increasing real sequence converging to 1 and define the sequence $(u_n)_n = (t_n u)_n$. By definition, $u_n \to u$ a.e. and in $L^{\Phi_1}(p)$; furthermore, $u_n \in \mathcal{S}_p$ since $\mathcal{S}_p$ is a solid convex set, see (Cena and Pistone 2007, Theorem 21). Hence, $q_n = e_p(u_n) \in \mathcal{E}(p)$. Moreover, since $\mathrm{e}^{u_n} < \mathrm{e}^u$ for $u > 0$ and $\mathrm{e}^{u_n} < 1$ for $u < 0$ from the Lebesgue dominated convergence theorem $K_p(u_n) \to K_p(u)$. □

Proposition 19.5 *Let* $p \in \mathcal{M}_>$; *then*

(A-1) $q \in \widehat{\mathcal{E}}(p)$ *if, and only if, a left open right closed exponential arc exists that connects* p *to* q. *In particular,* $q \in \partial\mathcal{E}(p)$ *if, and only if, such an arc cannot be right open.*
(A-2) $\widehat{\mathcal{E}}(p)$ *is a convex set.*

Proof Item (A-1) is straightforward from the definition of $\partial\mathcal{E}(p)$. In order to prove Item (A-2), let $q_1, q_2 \in \widehat{\mathcal{E}}(p)$ and $\lambda \in [0, 1]$; then, for some $\alpha > 0$, because of the convexity of the function $x^{-\alpha}$ for $x > 0$, it holds that

$$\mathrm{E}_p\left[\left(\lambda\frac{q_1}{p} + (1-\lambda)\frac{q_2}{p}\right)^{-\alpha}\right] \le \lambda\mathbb{E}_p\left[\left(\frac{q_1}{p}\right)^{-\alpha}\right] + (1-\lambda)\mathbb{E}_p\left[\left(\frac{q_2}{p}\right)^{-\alpha}\right] < \infty,$$

since, by hypotheses, both p/q_1 and p/q_2 belong to $L^\alpha(p)$. □

Theorem 19.2 *Let* $p \in \mathcal{M}_>$ *and* $q \in \mathcal{M}_\ge = \overline{\mathcal{E}(p)}$. *Let us consider sequences* $u_n \in \mathcal{S}_p$ *and* $q_n = \mathrm{e}^{u_n - K_p(u_n)} \cdot p \in \mathcal{E}(p)$, $n = 1, 2, \ldots$, *such that* $q_n \to q$ *in* $L^1(\mu)$ *as* $n \to \infty$.

(A-1) *The sequence* $v_n = u_n - K_p(u_n)$ *converges in* $p \cdot \mu$*-probability, as* $n \to \infty$, *to a* $[-\infty, +\infty[$*-valued random variable* v *and* $\{v \neq -\infty\} = \operatorname{Supp} q$.

(A-2) $\liminf_{n\to\infty} v_n \le \liminf_{n\to\infty} u_n$. *If the sequence* $(v_n)_n$ *is* μ*-a.s. convergent, then* $v \le \liminf_{n\to\infty} u_n$.

(A-3) If $\operatorname{Supp} q = \Omega$, *then either*

(a) $\limsup_{n\to\infty} K_p(u_n) < +\infty$ *and for each sub-sequence* $n(k)$ *such that* $u_{n(k)}$ *is* $p\cdot\mu$*-convergent, it holds that*

$$-\infty < v + \liminf_{n\to\infty} K_p(u_n) \le \lim_{k\to\infty} u_{n(k)} \le v + \limsup_{n\to\infty} K_p(u_n) < +\infty,$$

μ*-a.s., or*

(b) $\limsup_{n\to\infty} K_p(u_n) = +\infty$ *and for each sub-sequence* $n(k)$ *such that* $u_{n(k)}$ *is* $p\cdot\mu$*-convergent, it holds that* $\lim_{k\to\infty} u_{n(k)} = +\infty$.

(A-4) If $\operatorname{Supp} q \ne \Omega$, *then* $\lim_{n\to\infty} K_p(u_n) = +\infty$ *and* $\lim_{n\to\infty} u_n = +\infty$ $p\cdot\mu$*-a.s on* $\operatorname{Supp} q$. *Moreover,* $\lim_{n\to\infty} u_n - K_p(u_n) = -\infty$ *on* $\{q = 0\}$.

Proof The function $\log : [0,+\infty[\to [-\infty,+\infty[$ is continuous and $v = \log(q_n/p)$, therefore Item (A-1) holds true. Item (A-2) follows from the inequality $v_n = u_n - K_p(u_n) < u_n$ and $\limsup_{n\to\infty} v_n = \lim_{n\to\infty} v_n$ in the case of a.s. convergence.

For Item (A-3), it should first be noted that the convergence of the real sequence $(K_p(u_{n(k)}))_k$ is equivalent to the $p\cdot\mu$-convergence of the sequence of real random variables $(u_{n(k)})_k$. Therefore, the first part follows by letting $k\to\infty$ in $v_{n(k)} < u_{n(k)} = v_{n(k)} + K_p(u_{n(k)})$. On the other hand, if $\limsup_{n\to\infty} K_p(u_n) = +\infty$ then $\lim_{k\to\infty} K_p(u_{n(k)}) = +\infty$, therefore $\lim_{k\to\infty} u_{n(k)} = +\infty$, since $(v_{n(k)})_k$ converges to a finite v.

Now, let us suppose that $\operatorname{Supp} q \ne \Omega$ as in Item (A-4). Reasoning by contradiction, let $(n(k))_k$ be a subsequence such that $\lim_{k\to\infty} K_p(u_{n(k)}) = \kappa < \infty$. By Jensen inequality we obtain

$$0 = \lim_{k\to\infty} \int_{\{q=0\}} \mathrm{e}^{u_{n(k)} - K_p(u_{n(k)})}\, p d\mu = \mathrm{e}^{-\kappa} \lim_{k\to\infty} \int_{\{q=0\}} \mathrm{e}^{u_{n(k)}}\, p d\mu$$
$$\ge \mathrm{e}^{-\kappa} \exp\left(\lim_{k\to\infty} \int_{\{q=0\}} u_{n(k)}\, p d\mu \right),$$

therefore $\lim_{k\to\infty} \int_{\{q=0\}} u_{n(k)}\, p d\mu = -\infty$. Because each $u_{n(k)}$ has zero expectation, it follows that $\lim_{k\to\infty} \int_{\operatorname{Supp} q} u_{n(k)}\, p d\mu = +\infty$. This is in contradiction with

$$1 = \lim_{k\to\infty} \int_{\operatorname{Supp} q} \mathrm{e}^{u_{n(k)} - K_p(u_{n(k)})}\, p d\mu = \mathrm{e}^{-\kappa} \lim_{k\to\infty} \int_{\operatorname{Supp} q} \mathrm{e}^{u_{n(k)}}\, p d\mu$$
$$\ge \mathrm{e}^{-\kappa} \exp\left(\lim_{k\to\infty} \int_{\operatorname{Supp} q} u_{n(k)}\, p d\mu \right).$$

As $\lim_{n\to\infty} K_p(u_n) = +\infty$, then the sequence $u_n = v_n + K_p(u_n)$ is convergent to $+\infty$ where $v = \lim_{n\to\infty} \sigma_n$ is finite. □

Ttheorem 19.3 *Let $q_n = e_p(u_n) \in \mathcal{E}(p)$, and suppose that $u_n \to u$ in μ-probability. Then, possibly for a sub-sequence, the following statements are equivalent.*

(A-1) $u_n^*(q_n) \to u^*(q)$ weakly, where $q = \mathrm{e}^{u-k_p(u)}p$.
(A-2) $u_n \to u$ a.e. and $K_p(u_n) \to K_p(u) < \infty$.
(A-3) $q_n \to q$ in $L^1(\mu)$, where $q = \mathrm{e}^{u-k_p(u)}p$.

Proof If $u_n \to u$ in μ-probability, then $u_n \to u$ a.e., possibly for a sub-sequence, and $u_n^*(\mu) \to u^*(\mu)$ weakly. Hence, if $u_n^*(q_n) \to u^*(q)$ weakly, due to Proposition 19.7, $K_p(u_n) \to K_p(u) < \infty$, so that (A-1) implies (A-2). An application of Scheffé's Lemma shows that (A-2) implies (A-3), since, possibly for a sub-sequence, $q_n \to q$ a.e. and both q_n and q are densities. Finally, (A-3) implies (A-1) since by hypotheses and due to (19.3), possibly for a sub-sequence, $u_n^*(q_n) \to u^*(q)$ a.e. and hence weakly. □

Corollary 19.3 Let $q \in \mathcal{M}_{\geq} = \overline{\mathcal{E}(p)}$, i.e. sequences $(u_n)_n$, $u_n \in \mathcal{S}_p$ and $q_n = e_p(u_n)$, $q_n \to q$ in $L^1(\mu)$, exist and suppose that $u_n \to u$ in μ-probability. Then, $q = \mathrm{e}^{u-K_p(u)}p$ and, possibly for a sub-sequence, $K_p(u_n) \to K_p(u)$.

Proof Since possibly for a sub-sequence $u_n \to u$ a.e, Proposition 19.3 implies that for such a sub-sequence $\lim K_p(u_n) < \infty$; furthermore, through the lower semi-continuity of $K_p(u)$ it holds that

$$K_p(u) \leq \liminf_n K_p(u_n) = \lim K_p(u_n) < \infty,$$

so that $q = \mathrm{e}^{u-K_p(u)}p$ and eventually for a sub-sequence $\lim K_p(u_n) = K_p(u)$. □