Contingency tables from the algebraic statistics view point *

Gérard Letac, Université Paul Sabatier, Toulouse

College Park, October 14 th, 2010.

*. Joint work with Hélène Massam

Contingency tables

Let $I = I_1 \times \ldots \times I_k = \prod_{v \in V} I_v$ a product of finite sets and a probability p on I such that $p(i) > 0$ for all $i \in I$. The Bernoulli distribution B_p is the distribution of X valued in $\mathbb{R}^I$ such that $\Pr(X = e_i) = p_i$ where $(e_i)_{i \in I}$ is the basis of $\mathbb{R}^I$. If $X_1, \ldots, X_N$ are iid with distribution B_p then $S_N = X_1 + \cdots + X_N$ follows a multinomial distribution $B_{p,N}$. An observation S_N is a contingency table. In general p is unknown and we are willing to estimate it under some reasonable constraints like

$$p(i_1, i_2) = p(i_1, .)p(., i_2) = f(i_1)g(i_2) \text{ for } k = 2$$

$$p(i_1, i_2, i_3) = f(i_1, i_2)g(i_2, i_3) \quad \text{for } k = 3$$

or in a more linear way

$$\log p(i_1, i_2, i_3) = \lambda_{12}(i_1, i_2) + \lambda_{23}(i_2, i_3).$$

Hierarchical model

We formalize this type of constraints under the name of hierarchical model : given a family $\mathcal{D}$ of subsets of $V = \{1, \ldots, k\}$ such that $D_1 \subset D_2$ for $D_1, D_2 \in \mathcal{D}$ is impossible (unless $D_1 = D_2$) the hierarchical model $M_{\mathcal{D}}$ governed by $\mathcal{D}$ is the set of probabilities on I such that

$$\log p(i_1, \ldots, i_k) = \sum_{D \in \mathcal{D}} \lambda_D(i)$$

where $i \mapsto \lambda_D(i)$ is a function depending only on the coordinates i_j of i such that $j \in D$. In the sequel $\mathcal{D}_<$ is the family of non empty subsets of the elements of $\mathcal{D}$. An important point for performing the estimation of p under these constraints is the fact that $M_{\mathcal{D}}$ is an exponential family.

Examples of hierarchical models

(1)The most famous hierarchical model are the one given by an undirected graph with V as set of vertices. In this case $\mathcal{D}$ is the family of the cliques of the graph and the hierarchical model is called a graphical model.

We mention two trivial models (2) $\mathcal{D} = \{V\}$: no constraints at all, called the saturated model.

(3) $\mathcal{D}$ is the set of singletons : that means that $p(i) = p_1(i_1) \dots p_k(i_k)$ and each components are independent.

(4) $V = \{a, b, c\}$ and $\mathcal{D} = \{ab, bc, ac\}$ is an example of hierarchical but not graphical model.

From now on we select a point called zero in each component $I_v = \{0, 1, \dots, c_v\}$. $0 \in I$ is the point with all coordinates 0.

If $\mathcal{D} = \{V\}$ (saturated model), the exponential family is generated by the counting measure on the basis $(e_i)_{I \setminus 0}$ of $\mathbb{R}^{I \setminus 0}$ and we get for $i \neq 0$

$$p(i) = \frac{e^{\langle \xi, e_i \rangle}}{1 + \sum_{j \in I \setminus 0} e^{\langle \xi, e_j \rangle}} = \frac{e^{\xi_i}}{1 + \sum_{j \in I \setminus 0} e^{\xi_j}}.$$

For a general hierarchical model we define the support of $i = (i_1, \dots, i_k) \in I$ as the subset $S(i)$ of V of the $v \in V$ such that $i_v \neq 0$. We define the set $J \subset I$ of the $j \in I$ such that $S(j)$ is in $\mathcal{D}_<$ we write $j \triangleleft i$ if $j \in J$ is such that $S(j) \subset S(i)$ and the restriction of i to $S(j)$ coincide with j. The exponential family is now concentrated on $\mathbb{R}^J$ and is generated by the counting measure on the vectors $f_i = \sum_{j \triangleleft i} e_j$ for $i \in I$. The parameterization is now

$$\theta_j = \sum_{j \triangleleft i} (-1)^{|S(i) \setminus S(j)|} \log p(i).$$

Therefore the exponential family is concentrated on the polytope of $\mathbb{R}^J$ with extreme points $(f_i)_{i\in I}$.

Example : we take $I = I_a \times I_b \times I_c$ where $I_a = I_b = I_c = \{0, 1\}$ (a binary model) and the hierarchical model given by the pair ab and bc. This is the graphical model

$$a \bullet -b \bullet -c\bullet$$

Here $J = \{a, b, c, ab, bc\}$ and

$$f_0 = 0,\ f_a = e_a,\ f_b = e_b,\ f_c = e_c,$$

$$f_{ab} = e_a + e_b + e_{ab},\ f_{bc} = e_b + e_c + e_{bc},$$

$$f_{ac} = e_a + e_c,\ f_{abc} = e_a + e_b + e_c + e_{ab} + e_{bc}$$

The polytope C has 8 vertices in a space of 5 dimensions.

Bayesian methods for selecting models

The estimation of p by maximum likelihood is difficult to perform and we are here actually more interested in deciding if the hierarchical model is the good one or not. Since the model is an exponential family we rather take a Bayesian approach by using the Diaconis Ylvisaker conjugate family as *a priori* distributions on the natural parameter

$$\theta = (\theta_j)_{j\in J} \in \mathbb{R}^J,$$

namely

$$\pi_{m,\alpha}(d\theta) = \frac{1}{I(m,\alpha)} \times \frac{e^{\alpha\langle\theta,m\rangle}}{(\sum_{i\in I} e^{\langle\theta,f_i\rangle})^\alpha} d\theta$$

Here, $\alpha > 0$, m is an interior point of the convex polytope C with vertices $(f_i)_{i\in I}$ and $I(m,\alpha)$ is the normalization constant such that $\pi_{m,\alpha}$ is a probability.

The characteristic function of a convex set

Introduce a new general tool. Given a convex set C in $\mathbb{R}^n$ its support function is $h_C(\theta) = \max\{\langle\theta, x\rangle \ ; \ x \in C\}$ and its characteristic function is defined on the interior of C by

$$\mathbb{J}_C(m) = \int_{\mathbb{R}^n} e^{\langle\theta,m\rangle - h_C(\theta)} d\theta$$

$$= n!\text{Vol}(C - m)^o = n! \int_{C^o} \frac{d\theta}{(1 - \langle\theta, m\rangle)^{n+1}}$$

where

$$C^o = \{\theta \ ; \ \langle\theta, x\rangle \leq 1 \ \forall x \in C\}$$

is the polar set of C. This concept is useful here since we have the following result

Theorem 1 :

$$\lim_{\alpha\to 0} \alpha^n I(m, \alpha) = \mathbb{J}_C(m)$$

(just apply $\|f\|_p \to \|f\|_\infty$ in the theory of L_p spaces to prove this).

Note that this remarkable limit depends only on the convex support of the measure μ, not on the measure itself. For example if $E = \mathbb{R}$ the measures

$$\mu = \delta_{-1} + \delta_1$$

and

$$\nu(dx) = \mathbf{1}_{(-1,1)}(x)dx$$

share the same $C = (-1, 1)$ (which satisfies $h_{(-1,1)}(\theta) = |\theta|$. Therefore if $-1 < m < 1$ the two functions

$$\alpha I_\mu(m, \alpha) = \frac{\alpha}{2^{\alpha+1}} B(\frac{\alpha}{2}(1+m), \frac{\alpha}{2}(1-m))$$

and the untractable

$$\alpha I_\nu(m, \alpha) = \frac{\alpha}{2^{2\alpha}} \int_0^\infty \left(\frac{\log u}{u-1}\right)^\alpha u^{\frac{\alpha}{2}(1+m)-1} du$$

have the same limit

$$\mathbb{J}_{(-1,1)}(m) = \int_{-\infty}^{\infty} e^{\theta m - |\theta|} d\theta = \frac{2}{1-m^2}.$$

Comparing two exponential models

We now use this result for comparing two hierarchical models by the method of the Bayes factor. We are given two general exponential families

$$F_1 = \{e^{\langle \theta_1, t_1(w)\rangle - k_1(\theta_1)}\nu_1(dw)\ ;\ \theta_1 \in \Theta_1\}$$

and

$$F_2 = \{e^{\langle \theta_2, t_2(w)\rangle - k_2(\theta_2)}\nu_2(dw)\ ;\ \theta_1 \in \Theta_1\}$$

on the same abstract space Ω. On Ω we observe $(w_1, \ldots, w_N)$ and from this sample we have to choose between F_1 and F_2. The technique of the Bayesian factor is the following : F_1 and F_2 have natural associated exponential families with respective domains of the means M_1 and M_2 in which we choose m_1 and m_2 as close as possible. The normalisation constants for the a priori densities of D.-Y. are $I_1(m_1, \alpha)$ and $I_2(m_2, \alpha)$. We denote

$$S_N^{(1)} = t_1(\omega_1) + \ldots + t_1(\omega_N)$$

et

$$S_N^{(2)} = t_2(\omega_1) + \ldots + t_2(\omega_N)$$

Here comes the Bayesian factor

$$f = \frac{I_1(\frac{\alpha m_1 + S_N^{(1)}}{\alpha + N}, \alpha + N)}{I_1(m_1, \alpha)} \times \frac{I_2(m_2, \alpha)}{I_2(\frac{\alpha m_2 + S_N^{(2)}}{\alpha + N}, \alpha + N)}$$

If f is big we choose the model F_1. If f is close to zero we choose the model F_2.

Suppose now that we are comparing two hierarchical models : for instance the graphical model $a\bullet - b\bullet - c\bullet$ (say M_1) against the saturated one, say M_2. The observed contigency table has entries $Nt_i^{(2)}$ and its projection to model M_1 has entries $Nt_j^{(1)}$. Thus the Bayesian factor f is

$$f = \frac{I_1(\frac{\alpha m_1 + Nt^{(1)}}{\alpha+N}, \alpha + N)}{I_1(m_1, \alpha)} \times \frac{I_2(m_2, \alpha)}{I_2(\frac{\alpha m_2 + Nt^{(2)}}{\alpha+N}, \alpha + N)}$$

Now suppose that $t^{(1)}$ and $t^{(2)}$ are interior points of the corresponding polytopes C_1 and C_2. In this case the limiting behavior of f when α goes to zero is

$$f \sim \frac{I_1(t^{(1)}, N)}{I_2(t^{(2)}, N)} \times \frac{\mathbb{J}_2(m_2)}{\mathbb{J}_1(m_1)} \alpha^{|J_1| - |J_2|}.$$

The real first difficulty arises when the observations t are on the boundary of the polytope since $I_1(t^{(1)}, N)$ is now infinite. We use the following delicate result : let y belonging to a face $F(y)$ of C of codimension $M(y)$ and let m be in interior of the polytope C. Then

Theorem 2 :

$\lambda^{M(y)}\mathbb{J}_C(\lambda m+(1-\lambda)y) \to_{\lambda\to 0} \mathbb{J}_{F(y)}(y)\mathbb{J}_{C_2}(m_2)$

where C_2 is the cone generated by $(C-y)\cap F(y)^{\perp}$

The consequence is that if the observation t is on the boundary of C we have, with the simpler notation $\lambda = \frac{\alpha}{\alpha+N}$

$$\lambda^{M(t)} I(\lambda m + (1-\lambda)t, N) \to_{\lambda\to 0} Cte$$

which implies that the Bayesian factor behaves as $f \sim \alpha^{D(t^{(1)})-D(t^{(2)})} \times Cte$ where $D(t)$ is the dimension of the face containing t.

Linking the facets of $\overline{C}$ and the characteristic function $\mathbb{J}_C$

How do we check that t is on the boundary ? By knowing the facets of C.

Theorem 3 : If $\mathcal{F}$ is the family of facets of the polytope $\overline{C}$ and if $g_F(m) = 0$ is the affine equation of the facet m, then

$$m \mapsto \mathbb{J}_C(m) \prod_{F \in \mathcal{F}} g_F(m)$$

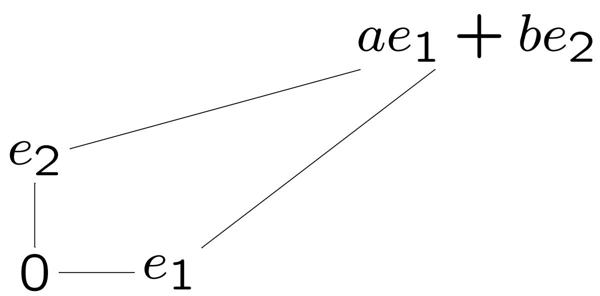
is a polynomial of degree $< |\mathcal{F}|$.

Example 1 : $C = [-1, 1]$ Its facets are $\{-1, 1\}$ with equations $g_{-1}(m) = 1 + m$, $g_1(m) = 1 - m$ and $(1 - m^2)\mathbb{J}_{(-1,1)}(m) = 2$. A consequence is that $\mathbb{J}_C(m)$ is a rational function whose denominator is factorisable in affine forms which give the equations of the facets.

Example 2 : quadrangle in the plane. In $\mathbb{R}^2$ with canonical basis (e_1, e_2) consider for $a, b > 1$ the vector $f = ae_1 + be_2$ and the convex C generated by $0, e_1, e_2, f$. We get by direct calculation

$$\mathbb{J}_C(m) = \int_{\mathbb{R}^2} e^{\langle \theta, m\rangle - h_C(\theta)} d\theta =$$

$$\frac{(1 + \frac{a-1}{b} m_2)(1 + \frac{b-1}{a} m_1)}{m_1 m_2 (1 - m_1 + \frac{a-1}{b} m_2)(1 + \frac{b-1}{a} m_1 - m_2)}.$$



This illustrates the preceeding theorem : $m_1 = m_2 = 1 - m_1 + \frac{a-1}{b} m_2 = 1 + \frac{b-1}{a} m_1 - m_2 = 0$ define the 4 facets of the quadrangle C.

Example 3 : Octahedron in $\mathbb{R}^d$.

In the Euclidean space $E = \mathbb{R}^d$ denote by $(e_j)_{j=1}^d$ the canonical basis. The octahedron is the convex hull $\overline{C}$ of the $2d$ vectors $(\pm e_j)_{j=1}^d$. Thus we have $m \in C$ if and only if

$$\sum_{j=1}^{d} |m_j| < 1.$$

It is easily seen that

$$h_C(\theta) = \max\{|\theta_j|; j = 1, \dots, d\}.$$

If $\epsilon = (\epsilon_1, \dots, \epsilon_d) \in \{-1, 1\}^d = E_d$ we write $s(\epsilon) = \epsilon_1 \cdots \epsilon_d$. We define the linear form

$$f_\epsilon(m) = \sum_{j=1}^{d} \epsilon_j m_j.$$

We get the general formula

$$\mathbb{J}_C(m) = \int_{\mathbb{R}^d} e^{\langle \theta, m \rangle - h_C(\theta)} d\theta =$$

$$\frac{1}{\prod_{j=1}^d m_j} \sum_{\epsilon \in E_d} \frac{s(\epsilon)}{1 - f_\epsilon(m)}.$$

Example 3, continued : Octahedron in $\mathbb{R}^3$.

For instance in the particular case $d = 3$ we get after reduction to the same denominator

$$\mathbb{J}_C(m)\frac{1}{8}\prod_{\epsilon\in E_3}(1-f_\epsilon(m)) = 3-2(m_1^2+m_2^2+m_3^2)$$

$$+2(m_1^2m_2^2+m_2^2m_3^2+m_3^2m_1^2)-m_1^4-m_2^4-m_3^4$$

We observe that the numerator of $\mathbb{J}_C(m)$ is a polynomial with degree 4, while the denominator has eight factors describing the facets of the classical regular octahedron. One important point of the present example is the following : the numerator has in general no special properties of factorization as it occurs in the quadrangle or in the hierarchical model in the decomposable case.

Decomposable graphs

Theorem 4 : Let $(V, \mathcal{E})$ be a decomposable graph, let $\mathcal{D}$ be the family of the complete subsets of V, let $\mathcal{C}$ be the family of its cliques, let $\mathcal{S}$ be the family of its minimal separators and let $\nu(S)$ be the multiplicity of the minimal separator S. For each $D \in \mathcal{D}$ and each $j \in J$ such that $S(j) \subset D$ consider the affine forms defined by

$$\begin{aligned} g_{0,D}(m) &= 1 + \sum_{j; S(j) \subset D} (-1)^{|S(j)|} m_j \\ g_{j_0,D}(m) &= \sum_{j; S(j) \subset D,\ j_0 \triangleleft j} (-1)^{|S(j)| - |S(j_0)|} m_j \end{aligned}$$

Then for m in the interior C of the convex hull of the f_i's we have $I(m, \alpha) =$

$$\frac{\prod_{C \in \mathcal{C}} \Gamma(\alpha g_{0,C}(m)) \prod_{\{j \in J; S(j) \subset C\}} \Gamma(\alpha g_{j,C}(m))}{\Gamma(\alpha) \prod_{S \in \mathcal{S}} \left[\Gamma(\alpha g_{0,S}(m)) \prod_{\{j \in J; S(j) \subset S\}} \Gamma(\alpha g_{j,S}(m))\right]^{\nu(}}$$

$\mathbb{J}_C(m)$ for a decomposable graph

In particular for $m \in C$ we have

$$\lim_{\alpha\to} \alpha^{|J|} I(m,\alpha) = \mathbb{J}_C(m) =$$

$$\frac{\prod_{S\in\mathcal{S}} \left[g_{0,S}(m) \prod_{\{j\in J; S(j)\subset S\}} g_{j,S}(m)\right]^{\nu(S)}}{\prod_{C\in\mathcal{C}} g_{0,C}(m) \prod_{\{j\in J; S(j)\subset C\}} g_{j,C}(m)}$$

As a consequence of Theorem 3, the family facets of $\overline{C}$ are parameterized by the set

$$\bigcup_{C\in\mathcal{C}} \{j \in J \cup \{0\}; S(j) \subset C\}$$

and the corresponding equation for the facet F associated to the clique C and to the point $j \in J \cup \{0\}$ such that $S(j) \subset C$ is $g_{j,C}(m) = 0$.

Basic facets. Consider now a general hierarchical model governed by the family $\mathcal{D}$ of non empty subsets of V (recall such that $D \in \mathcal{D}$ and $D_1 \subset D$ implies $D \in \mathcal{D}$). Denote by $\mathcal{C} \subset \mathcal{D}$ the family of maximal elements (if the hierarchical model was a graphical one, $\mathcal{C}$ would be the family of cliques.

Theorem 5 : Let D be a maximal element of $\mathcal{D}$. The affine equations

$$\begin{aligned} g_{0,C}(m) &= 1 + \sum_{j;S(j)\subset D} (-1)^{|S(j)|} m_j \\ g_{j_0,D}(m) &= \sum_{j;S(j)\subset D,\ j_0 \triangleleft j} (-1)^{|S(j)|-|S(j_0)|} m_j \end{aligned}$$

define facets of $\overline{C}$ (called basic facets)

Clearly this is an extension of Theorem 4 which was devoted to the decomposable case. In Theorem 4 ALL facets where basic. In the present general case, there are non basic facets. We have the following conjecture : if all facets are basic, the hierarchical model is a decomposable graphical model.

Facets for a cycle The binary graphical model governed by a cycle has been much studied. It happens that the facets of the corresponding polytope $\overline{C}$ have been studied in different contexts. Since we are in the binary case and since V has n elements, there exist $4n$ basic facets. But there also exist other ones, as discovered by Deza and Laurent (1995), or Hosten and Sullivant (2002).

Theorem 6 : The non basic facets of the binary graphical model governed by a cycle $V = \{1, 2, \dots, n\}$ are indexed by the 2^{n-1} odd subsets F of the set E of the n edges of the cycle. The corresponding affine equation is $g_F(m) =$

$$\sum_{(a,b)\in F}(m_a+m_b-2m_{ab})-\Big(\sum_{v\in V}m_v-\sum_{e\in E}m_e\Big)-\frac{|F|-1}{2}$$

Hence in the case of a cycle the polytope C has $4n + 2^{n-1}$ facets : 24 for $n = 4$ and 36 for $n = 5$.

Two way interaction models They are the hierarchical models such that all maximal elements of $\mathcal{D}$ have only two elements. We can therefore represent such a model by a graph, but the interpretation of this graph should not be the interpretation of a graph in the case of a graphical model (unless the cliques of this graphical model have always two elements, like in a cycle of size ≥ 4 or a tree).

A binary two way interaction model whose graph has m edges has exactly $4m$ basic facets.

The correlation polytope The binary two way interaction model obtained by considering the complete graph K_n is supported by the polytope $\overline{C_n}$. The polytope C_n has been the subject of intense studies in different contexts (see Deza and Laurent 1995) where it is called the correlation polytope. The number $|J| = n(n+1)/2$ is the dimension of the space of symmetric matrices of order n and C_n can be seen as the convex hull of the 2^n symmetric matrices f_T where $T \subset \{1, \ldots, n\}$ and $f_T(i,j) = 1$ if $i, j \in T$ and $f_T(i,j) = 0$ if not. The set of facets of C_n and their number F_n has been studied for the first values of $n \leq 7$. One gets in particular

$$F_3 = 16, \ F_4 = 56, \ F_5 = 368,$$

$$F_6 = 116{,}764, \ F_7 = 49{,}594{,}520.$$

The two way interaction model C_3

Let us insist on the fact that the two way interaction model C_3 is the hierarchical model of probabilities $(p(i))_{i \in I}$ on $I = I_1 \times I_2 \times I_3$ such that there exists 3 functions λ_j such that

$$\log p(i_1, i_2, i_3) = \lambda_3(i_1, i_2) + \lambda_2(i_1, i_3) + \lambda_1(i_2, i_3),$$

and that we consider only the binary case $I_j = \{0, 1\}$.

For the correlation polytope C_3, we now explain how to compute $I(m,\alpha)$ and $\mathbb{J}_{C_3}(m)$.

Theorem 7 : Define the positive numbers $A = m_{ab}$, $B = m_{ac}$, $C = m_a$, $A' = m_b - m_{ab}$, $B' = m_c - m_{ac}$, $C' = 1 - m_a$, $D = m_{bc}$. Then the sixth order integral $I(m,\alpha)$ is equal to the one dimensional Mellin transform

$$\int_0^\infty x^{\alpha D - 1} \times$$

$${}_2F_1(\alpha A, \alpha B; \alpha C; 1-x)\, {}_2F_1(\alpha A', \alpha B'; \alpha C'; 1-x) dx$$

multiplied by the following product of gamma functions

$$\frac{\Gamma(\alpha A)\Gamma(\alpha B)\Gamma(\alpha A')\Gamma(\alpha B')}{\Gamma(\alpha)\Gamma(\alpha C)\Gamma(\alpha C')}$$

$$\times \Gamma(\alpha(C-A))\Gamma(\alpha(C-B))\Gamma(\alpha(C'-A'))\Gamma(\alpha(C'-B'))$$

Let us indicate the sources of the Gauss hypergeometric function in this calculation. In $I(m,\alpha)$ we use the change of variable $u_a = e^{\theta},\ u_{ab} = e^{\theta_{ab}}, \dots$. Consider the polynomial

$$P = 1 + u_a + u_b + u_c + u_a u_b u_{ab} + u_b u_c u_{bc} + u_a u_c u_{ac}$$

$$+ u_a u_b u_c u_{ab} u_{bc} u_{ac}.$$

The integral $I(m,\alpha)$ becomes

$$\int_{(0,\infty)^6} \frac{u_a^{\alpha m_a - 1} u_b^{\alpha m_b - 1} u_c^{\alpha m_c - 1} u_{ab}^{\alpha m_{ab} - 1} u_{bc}^{\alpha m_{bc} - 1} u_{ac}^{\alpha m_{ac} - 1}}{P^{\alpha}}$$

$$du_a du_b du_c du_{ab} du_{bc} du_{ac}$$

We first integrate with respect to u_a. This leads to an integral on $(0,\infty)^5$ that we treat by the following observation

If a, b, c, R are positive numbers denote

$$I = \int_0^\infty \int_0^\infty \frac{x^{a-1} y^{b-1}}{(1 + x + y + Rxy)^{a+b+c}} dx dy.$$

Then $I =$

$$\frac{\Gamma(a)\Gamma(b)\Gamma(b+c)\Gamma(a+c)}{\Gamma(a+b+c)^2} \, {}_2F_1(a, b, a+b+c; 1-R).$$

Applying it twice reduces the integral on $(0, \infty)^5$ to an integral on $(0, \infty)$ involving a product of two hypergeometric functions.

Why this horrible $I(m, \alpha)$?

Because it provides the only known way to compute $\mathbb{J}_{C_3}(m) = \lim_{\alpha \to 0} \alpha^6 I(m, \alpha)$. Recall that

$$ {}_2F_1(\alpha A, \alpha B; \alpha C; 1 - e^{\frac{v}{\alpha}}) $$

$$ = \sum_{n=0}^{\infty} \frac{(\alpha A)_n (\alpha B)_n}{n! (\alpha C)_n} (1 - e^{\frac{v}{\alpha}})^n. $$

We need two difficult limits. If $v > 0$:

$$ \lim_{\alpha \to 0} {}_2F_1(\alpha A, \alpha B; \alpha C; 1 - e^{\frac{v}{\alpha}}) = $$

$$ \frac{B(C-A)}{C(B-A)} e^{-vA} - \frac{A(C-B)}{C(B-A)} e^{-vB} $$

and if $v < 0$:

$$ \lim_{\alpha \to 0} {}_2F_1(\alpha A, \alpha B; \alpha C; 1 - e^{\frac{v}{\alpha}}) = $$

$$ \frac{(C-A)(C-B) - ABe^{v(C-A-B)}}{C(C-A-B)} $$

Among other things, the proof of the last statement uses the following undergraduate exercise :

Let $a, b > 0$. If X_α has beta distribution

$$\beta_{\alpha a, \alpha b}$$

what is the limiting distribution of X_α^α when $\alpha \to 0$?

Answer :

$$\frac{a}{a+b}\delta_1 + \frac{b}{a+b}\beta_{a,1}$$

(use Mellin transform).

Consequence : if $X_\alpha \sim \beta_{\alpha a, \alpha b}$ then

$$\Pr(X_\alpha < x^{1/\alpha}) \to_{\alpha \to 0} \frac{b}{a+b} x^a.$$

Here is $\mathbb{J}_{C_3}(m) =$

$$\frac{1}{ABA'B'(C-A)(C-B)(C'-A')(C'-B')}$$
$$\times \frac{1}{(B-A)(B'-A')}$$
$$\{ \frac{(C-A)(C'-A')BB'}{(A+A'-D)}$$
$$- \frac{(C-B)(C'-A')AB'}{(B+A'-D)}$$
$$- \frac{(C-A)(C'-B')BA'}{(A+B'-D)}$$
$$+ \frac{(C-B)(C'-B')AA'}{(B+B'-D)} \}$$
$$+ \quad \frac{1}{ABA'B'(C-A)(C-B)(C'-A')(C'-B')}$$
$$\times \frac{1}{(C-A-B)(C'-A'-B')}$$
$$\{ \frac{(C-A)(C-B)(C'-A')(C'-B')}{D}$$
$$- \frac{(C-A)(C-B)A'B'}{(C'+D-A'-B')}$$
$$- \frac{AB(C'-A')(C'-B')}{(C+D-A-B)}$$
$$+ \frac{ABA'B'}{(C+C'+D-A-A'-B-B')} \}$$

Recall that

$$A = m_{ab},\ B = m_{ac},\ C = m_a,\ A' = m_b - m_{ab},$$

$$B' = m_c - m_{ac},\ C' = 1 - m_a,\ D = m_{bc}.$$

After simplification, the four undesirable denominators $B-A, B'-A', C-A-B, C'-A'-B'$ disappear (but the numerator according to Maple is the sum of four thousand monomials). The sixteen remaining denominators give the sixteen facets. The 12 basic facets are given by the annulation of the 12 forms

$$A, B, C-A, C-B, A', B', C'-A', C'-B', D,$$

$$D-A-A', D-B-B', D+C+C'-A-A'-B-B'.$$

The 4 non basic facets are given by

$$D+C-A-B, D-A'-B, D-A-B', D+C'-A'-B'.$$

$A = m_{ab}$, $B = m_{ac}$, $D = m_{bc}$, $A' = m_b - m_{ab}$, $C - A = m_a - m_{ab}$, $B' = m_c - m_{ac}$, $C - B = m_a - m_{ac}$, $A + A' - D = m_b - m_{bc}$, $B + B' - D = m_c - m_{bc}$, $C' - A' = 1 - m_a - m_b + m_{ab}$, $C' - B' = 1 - m_a - m_c + m_{ac}$, $C + C' + D - A - A' - B - B' = 1 - m_b - m_c + m_{bc}$, $A + B' - D = m_c + m_{ab} - m_{ac} - m_{bc}$, $A' + B - D = m_b + m_{ac} - m_{ab} - m_{bc}$, $C + D - A - B = m_a + m_{bc} - m_{ac} - m_{ab}$, $C' + D - A' - B' = 1 - m_a - m_b - m_c + m_{ac} + m_{ab} + m_{bc}$.