Orthogonal polynomials, hypergeometric functions and multivariate probability distributions

Tamás Szántai
Institute of Mathematics,
Budapest University of Technology and Economics

International Colloquium on Stochastic Modeling and Optimization,

**Dedicated to the 80th Birthday of
Professor András Prékopa**

November 30 – December 1, 2009, RUTCOR

Outline:

1. Bivariate gamma distribution and the Laguerre polynomials
2. Bivariate normal distribution and the Hermite polynomials
3. Dirichlet distribution and the Lauricella hypergeometric functions
4. Bivariate Poisson distribution and the Charlier polynomials
5. Bivariate binomial distribution and the Krawtchouk polynomials

Definition:

$$\begin{aligned}\xi_1 &= \eta_1 + \eta_2 \\ \xi_2 &= \eta_1 + \eta_3,\end{aligned}$$

where η_1, η_2, η_3 are independent, standard gamma distributed with parameters $\vartheta_1, \vartheta_2, \vartheta_3$.

The two variable characteristic function:

$$\begin{aligned}\phi(t_1, t_2) &= E\left(e^{it_1\xi_1 + it_2\xi_2}\right) \\ &= E\left(e^{i(t_1+t_2)\eta_1}\right) E\left(e^{it_1\eta_2}\right) E\left(e^{it_2\eta_3}\right) \\ &= \frac{1}{(1-i(t_1+t_2))^{\vartheta_1}} \frac{1}{(1-it_1)^{\vartheta_2}} \frac{1}{(1-it_2)^{\vartheta_3}}.\end{aligned}$$

A series expansion of the two variable characteristic function:

$$\begin{aligned}\phi(t_1, t_2) &= \frac{1}{(1-it_1)^{\vartheta_1+\vartheta_2}} \frac{1}{(1-it_2)^{\vartheta_1+\vartheta_3}} \left(1 - \frac{it_1it_2}{(1-it_1)(1-it_2)}\right)^{-\vartheta_1} \\ &= \frac{1}{(1-it_1)^{\vartheta_1+\vartheta_2}} \frac{1}{(1-it_2)^{\vartheta_1+\vartheta_3}} \sum_{k=0}^{\infty} \binom{-\vartheta_1}{k} \left(\frac{-t_1t_2}{(1-it_1)(1-it_2)}\right)^k \\ &= \sum_{k=0}^{\infty} \frac{\vartheta_1(\vartheta_1+1)\cdots(\vartheta_1+k-1)}{k!} \frac{(t_1t_2)^k}{(1-it_1)^{\vartheta_1+\vartheta_2+k}(1-it_2)^{\vartheta_1+\vartheta_3+k}}.\end{aligned}$$

Bivariate gamma distribution (2)

Using the notations:

$$f_\vartheta(x) = \frac{1}{\Gamma(\vartheta)} x^{\vartheta-1} e^{-x}, \text{ ha } x > 0,$$

$$F_\vartheta(x) = \int_0^x f_\vartheta(t)\,dt$$

the inversion formula:

$$f_\vartheta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx}(1-it)^{-\vartheta}\,dt$$

the *k*-th derivative of the inversion formula:

$$\frac{d^k f_\vartheta(x)}{dx^k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^k e^{-itx}(1-it)^{-\vartheta}\,dt$$

and the two dimensional inversion formula:

$$f(x_1, x_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-it_1 x_1} e^{-it_2 x_2} \varphi(t_1, t_2)\,dt_1\,dt_2$$

From the series expansion of the two variable characteristic function we get:

$$\begin{aligned} f(x_1, x_2) &= f_{\vartheta_1+\vartheta_2}(x_1) f_{\vartheta_1+\vartheta_3}(x_2) \\ &+ \sum_{k=1}^{\infty} \frac{\vartheta_1(\vartheta_1+1)\cdots(\vartheta_1+k-1)}{k!} \frac{d^k f_{\vartheta_1+\vartheta_2+k}(x_1)}{dx_1^k} \frac{d^k f_{\vartheta_1+\vartheta_3+k}(x_2)}{dx_2^k} \end{aligned}$$

By the generalized Laguerre polynomials

$$L_k^{\vartheta}(x) = \frac{1}{k!} x^{-\vartheta} e^x \frac{d^k}{dx^k} x^{\vartheta+k} e^{-x}, k = 0, 1, \ldots$$

we have:

$$\frac{d^k f_{\vartheta+k}(x)}{dx^k} = \frac{k!\Gamma(\vartheta)}{\Gamma(\vartheta+k)} f_{\vartheta}(x) L_k^{\vartheta-1}(x), k = 0, 1, \ldots.$$

So the final formula for the probability density function of the bivariate gamma distribution is:

$$\begin{aligned} f(x_1, x_2) &= f_{\vartheta_1+\vartheta_2}(x_1) f_{\vartheta_1+\vartheta_3}(x_2) \\ &\times \left[1 + \sum_{k=1}^{\infty} k! \frac{\Gamma(\vartheta_1+k)}{\Gamma(\vartheta_1)} \frac{\Gamma(\vartheta_1+\vartheta_2)}{\Gamma(\vartheta_1+\vartheta_2+k)} \frac{\Gamma(\vartheta_1+\vartheta_3)}{\Gamma(\vartheta_1+\vartheta_3+k)} L_k^{\vartheta_1+\vartheta_2-1}(x_1) L_k^{\vartheta_1+\vartheta_3-1}(x_2)\right]. \end{aligned}$$

To get the cumulative probability distribution function of the bivariate gamma distribution we have to apply the following derivation formula of the Laguerre polynomials:

$$\frac{d}{dx}\left(x^{\vartheta}e^{-x}L_{k-1}^{\vartheta}(x)\right) = kx^{\vartheta-1}e^{-x}L_{k}^{\vartheta-1}(x),$$

integrating both sides:

$$x^{\vartheta}e^{-x}L_{k-1}^{\vartheta}(x) = k\int_0^x t^{\vartheta-1}e^{-t}L_{k}^{\vartheta-1}(t)\,dt$$

after reordering we get:

$$\int_0^x t^{\vartheta-1}e^{-t}L_{k}^{\vartheta-1}(t)\,dt = \frac{1}{k}\frac{\Gamma(\vartheta+1)}{\Gamma(\vartheta)}f_{\vartheta+1}(x)L_{k-1}^{\vartheta}(x).$$

So finally we get:

$$
\begin{aligned}
F(x_1, x_2) &= \int_0^{x_2}\int_0^{x_1} f(t_1, t_2)\,dt_1 dt_2 \\
&= F_{\vartheta_1+\vartheta_2}(x_1) F_{\vartheta_1+\vartheta_3}(x_2) \\
&+ \sum_{k=1}^{\infty} C(\vartheta_1, \vartheta_2, \vartheta_3) f_{\vartheta_1+\vartheta_2+1}(x_1) L_{k-1}^{\vartheta_1+\vartheta_2}(x_1) f_{\vartheta_1+\vartheta_3+1}(x_2) L_{k-1}^{\vartheta_1+\vartheta_3}(x_2),
\end{aligned}
$$

where:

$$
\begin{aligned}
C(\vartheta_1, \vartheta_2, \vartheta_3) &= \frac{(k-1)!}{k} \frac{\Gamma(\vartheta_1+k)}{\Gamma(\vartheta_1)} \frac{\Gamma(\vartheta_1+\vartheta_2+1)}{\Gamma(\vartheta_1+\vartheta_2+k)} \frac{\Gamma(\vartheta_1+\vartheta_3+1)}{\Gamma(\vartheta_1+\vartheta_3+k)} \\
&= \frac{(k-1)!}{k} \frac{\vartheta_1(\vartheta_1+1)\cdots(\vartheta_1+k-1)}{(\vartheta_1+\vartheta_2+1)\cdots(\vartheta_1+\vartheta_2+k-1)(\vartheta_1+\vartheta_3+1)\cdots(\vartheta_1+\vartheta_3+k-1)}
\end{aligned}
$$

and the values of the Laguerre polynomials easily can be calculated by the following recursion:

$$L_0^{\vartheta}(x) = 1, L_1^{\vartheta}(x) = \vartheta + 1 - x$$

$$(k+1)L_{k+1}^{\vartheta}(x) - (2k+\vartheta+1-x)L_k^{\vartheta}(x) + (k+\vartheta)L_{k-1}^{\vartheta}(x) = 0, k = 0, 1, \ldots$$

Program $\Gamma(z)$ neboli Γ-funkce (Gamma fcn.)

(HT. PTK-1072)

$$\Gamma(z) = \int_0^\infty e^{-t}\, t^{z-1}\, dt$$

$$\ln \Gamma(z) \sim \left(z - \tfrac{1}{2}\right)\ln z - z + \tfrac{1}{2}\ln(2\pi) + \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5} - \frac{1}{1680z^7} + \frac{1}{1188z^9} - \frac{691}{360360z^{11}} + \frac{1}{156z^{13}} - \frac{3617}{122400z^{15}} + \cdots$$

(Lásd: Budinowska matematicheskikh mashinu, Vypusk 40, Vychislitelnyi tsentr A. N. SSSR, Moskva, 1966, X. old. (8') képlet. vill. Mat. Tn. Tho-potéka: 342 f.)

M0 : 1/z
M1 : 8.3333333 -02
M2 : -2.7777778 -03
M3 : 7.9365079 -04
M4 : -5.9523810 -04
M5 : 8.4175084 -04
M6 : -1.9175269 -03
M7 : 6.4102564 -03
M8 : -2.9550654 -02
M9 : 1/z²

Program [illegible]

0:	MR	24:	5	48:	2		
1:	8	25:	=	49:	=		
2:	×	26:	×	50:	×		
3:	MR	27:	MR	51:	MR	ln	÷
4:	9	28:	9	52:	9	×	2
5:	=	29:	=	53:	=	(	)
6:	+	30:	+	54:	+	MR	=
7:	MR	31:	MR	55:	MR	C	+
8:	+	32:	4	56:	1	−	MR
9:	=	33:	=	57:	=	.	9
10:	×	34:	×	58:	×	5	=
11:	MR	35:	MR	59:	MR	)	[illegible]
12:	9	36:	9	60:	0	=	e^x
13:	=	37:	=	61:	−	−	
14:	+	38:	+	62:	M	MR	
15:	MR	39:	MR	63:	9	C	
16:	6	40:	3	64:	MR	+	
17:	=	41:	=	65:	C	(	
18:	×	42:	×	66:	F	(	
19:	MR	43:	MR	67:	1/x	π	
20:	9	44:	9	68:	M	×	
21:	=	45:	=	69:	C	2	
22:	+	46:	+	70:	R/S	)	
23:	MR	47:	MR	71:	99	ln	

Program $I(x,m)$ számítására (Nem teljes gamma fv)

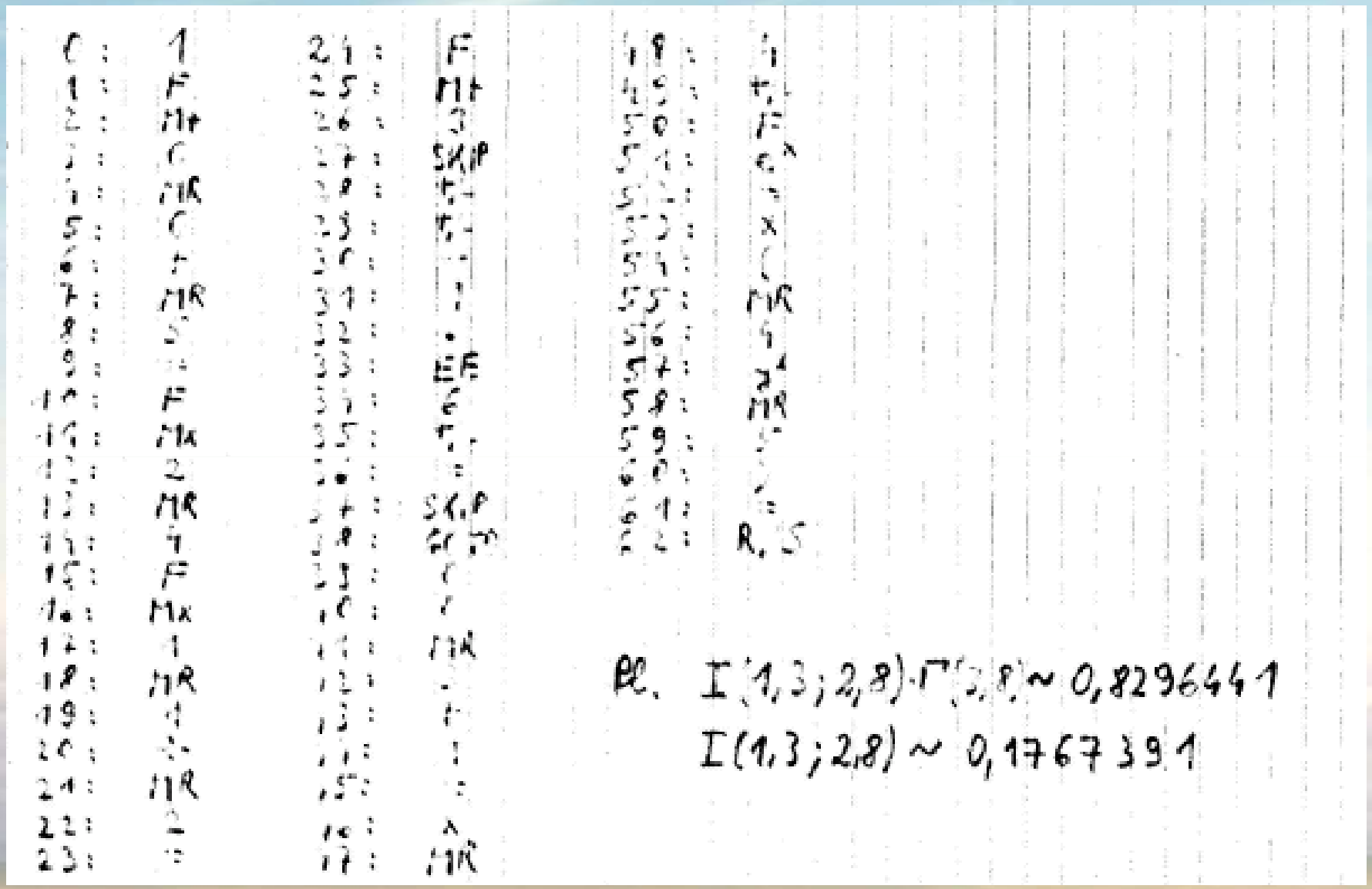
(HT PTK-1072)

$$I(x,m) = \frac{1}{\Gamma(m)} \int_0^x t^{m-1} e^{-t}\, dt \,, \text{ ahol } \Gamma(m) = \int_0^\infty e^{-t} t^{m-1}\, dt$$

$$I(x,m)\cdot\Gamma(m+1) \sim e^{-x}\cdot x^m \left\{ 1 + \sum_{k=1}^{\infty} \frac{x^k}{(m+1)\cdots(m+k)} \right\}$$

(Lásd: В.И. Пагурова: Таблицы неполной гамма функции, Вычислительный центр А.Н. СССР, Москва 1963, VI. lap. (7) formula szerint, Vill. Mat. Tu. Kompjutera: 23/15)

M0 : 0
M1 : 1
M2 : 1
M3 : c
M4 : x
M5 : m


Pc. I(1,3;2,8)·Γ(2,8) ~ 0,8296441
I(1,3;2,8) ~ 0,1767391

```
 subroutine mdbgam(x,y,a,b,c,p,ier)
c                                specifications for arguments
      integer          ier
      real            x,y,a,b,c,p
c                                specifications for local variables
      integer          k
      real            cnew,cold,coldr,g1,g2,px,py,s,sfx,sfy,xpnew,
     1               xpold,xpoldr,ypnew,ypold,ypoldr
c                                first executable statement
      xpoldr=1.
      xpold=-x+a+b+1.
      ypoldr=1.
      ypold=-y+b+c+1.
      coldr=b
      cold=b*(b+1)/(2*(a+b+1)*(b+c+1))
      s=coldr*xpoldr*ypoldr+cold*xpold*ypold
c calculate the bilinear series expansion of the laguerre polinomials
      do k=3,40
c update the coefficient
         cnew=cold*(k-1)*(k-1)*(b+k-1)/(k*(a+b+k-1)*(b+c+k-1))
         cold=cnew
c update the laguerre polinom of x
         xpnew=((2*(k-1)+(a+b+1)-x-2)*xpold-((k-1)+(a+b+1)-2)*xpoldr)/(k-1)
         xpoldr=xpold
         xpold=xpnew
c update the laguerre polinom of y
         ypnew=((2*(k-1)+(b+c+1)-y-2)*ypold-((k-1)+(b+c+1)-2)*ypoldr)/(k-1)
         ypoldr=ypold
         ypold=ypnew
c sum up the product of the coefficient and the laguerre polinomials
         s=s+cnew*xpnew*ypnew
      enddo
c produce the final value of the bivariate gamma distribution function
      g1=gamma(a+b+1)
      g2=gamma(b+c+1)
      sfx=exp(-x)*x**(a+b)/g1
      sfy=exp(-y)*y**(b+c)/g2
      call mdgam(x,a+b,px,ier)
      call mdgam(y,b+c,py,ier)
      p=px*py+s*sfx*sfy
      return
      end
```

The two variable standard normal probability density function:

$$\varphi(x,y;r) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}$$

where r is the correlation coefficient of the two standard normally distributed random variables.

We know the series expansion formula:

$$\varphi(x,y;r) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \sum_{k=0}^{\infty} \frac{r^k}{k!} H_k(x) H_k(y),$$

where:

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}}, k = 0, 1, \ldots$$

is the k-th order Hermite polynomial. Its values can be calculated by the recursion:

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x), n = 1, 2, \ldots, H_0(x) = 1, H_1(x) = x$$

Integrating the probability density function $\varphi(x, y; r)$ we get the cumulative probability distribution function :

$$\Phi(x, y; r) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \sum_{k=0}^{\infty} \frac{r^k}{k!} H_{k-1}(x) H_{k-1}(y),$$

where, by definition:

$$H_{-1}(x) = -\sqrt{2\pi} e^{-\frac{x^2}{2}} \Phi(x).$$

Introducing

$$\tau_k(x) = \frac{1}{\sqrt{2\pi k!}} e^{-\frac{x^2}{2}} H_{k-1}(x), k = 0, 1, \ldots$$

we obtain the well known tetrachoric series expansion:

$$\Phi(x, y; r) = \sum_{k=0}^{\infty} r^k \tau_k(x) \tau_k(y)$$

where the τ_k tetrachoric functions similarly can be calculated by recursion as the Hermite polynomials.

We can proceed in the following way, too:

$$\begin{aligned}\Phi(x,y;r) &= \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}\sum_{k=0}^{\infty}\frac{r^k}{k!}H_{k-1}(x)H_{k-1}(y)\\ &= \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}\left[H_{-1}(x)H_{-1}(y)+r\sum_{k=1}^{\infty}\frac{r^{k-1}}{k!}H_{k-1}(x)H_{k-1}(y)\right]\\ &= \Phi(x)\Phi(y)+\frac{r}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}\sum_{k=0}^{\infty}\frac{1}{k+1}\frac{r^k}{k!}H_k(x)H_k(y).\end{aligned}$$

Here we can introduce the functions:

$$\mathcal{H}_k^{(i)}(x)=\begin{cases}\frac{r}{1}\frac{r}{3}\cdots\frac{r}{k-1}H_k(x), & \text{if } k \text{ even}\\ \frac{r}{1}\frac{r}{3}\cdots\frac{r}{k}H_k(x), & \text{if } k \text{ odd}\end{cases}$$

$$\mathcal{H}_k^{(ii)}(y)=\begin{cases}\frac{r}{2}\frac{r}{4}\cdots\frac{r}{k}H_k(y), & \text{if } k \text{ even}\\ \frac{r}{2}\frac{r}{4}\cdots\frac{r}{k-1}H_k(y), & \text{if } k \text{ odd}\end{cases}$$

and substitute the $\frac{r^k}{k!}H_k(x)H_k(y)$ term by $\mathcal{H}_k^{(i)}(x)\mathcal{H}_k^{(ii)}(y)$:

$$\Phi(x,y;r)=\Phi(x)\Phi(y)+\frac{r}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}\sum_{k=0}^{\infty}\frac{1}{k+1}\mathcal{H}_k(x)\mathcal{H}_k(y)$$

Recursive formulae for the calculation the values of the $\mathcal{H}_k^{(i)}$ and $\mathcal{H}_k^{(ii)}$ functions:

$$\mathcal{H}_{k+1}^{(i)}(x) = \begin{cases} \frac{r}{k+1} x \mathcal{H}_k^{(i)}(x) - \frac{r}{k+1} k \mathcal{H}_{k-1}^{(i)}(x), & \text{if } k \text{ even} \\ x \mathcal{H}_k^{(i)}(x) - r \mathcal{H}_{k-1}^{(i)}(x), & \text{if } k \text{ odd} \end{cases}$$

$$\mathcal{H}_{k+1}^{(ii)}(y) = \begin{cases} \frac{r}{k+1} y \mathcal{H}_k^{(ii)}(y) - \frac{r}{k+1} k \mathcal{H}_{k-1}^{(ii)}(y), & \text{if } k \text{ even} \\ y \mathcal{H}_k^{(ii)}(y) - r \mathcal{H}_{k-1}^{(ii)}(y), & \text{if } k \text{ odd.} \end{cases}$$

x	y	r	integration	IMSL subroutine	new Series expansion	
3.0	3.0	0.0	0.997270	0.997302	0.997302	2
3.0	3.0	0.5	0.997350	0.997382	0.997382	14
3.0	3.0	0.9	0.997879	0.997911	0.997910	46
3.0	3.0	0.95	0.998078	0.998109	0.998108	62
3.0	3.0	0.99	0.998370	0.998402	0.998372	62
3.0	3.0	0.999	0.998539	0.998571	0.998460	62
0.0	0.0	0.999	0.492850	0.492882	0.492881	5688

Proof of Relations Connected with the Tetrachoric Series and its Generalization
Author(s): M. G. Kendall
Source: *Biometrika*, Vol. 32, No. 2 (Oct., 1941), pp. 196-198

The generalized tetrachoric expansion can be obtained by an expansion of (9) in terms of the ρ's and the application of the foregoing procedure. For instance, with three variates we have

$$\exp \Sigma(-\rho_{jk} t_j t_k) = \Sigma \frac{(-1)^r}{r!} (\rho_{12} t_1 t_2 + \rho_{23} t_2 t_3 + \rho_{13} t_1 t_3)^r$$

$$= \Sigma(-1)^r \frac{\rho_{12}^j \rho_{23}^k \rho_{13}^l}{j!k!l!} t_1^{j+k} t_2^{j+l} t_3^{k+l},$$

and on integration

$$d = \int_{h_1}^{\infty} \int_{h_2}^{\infty} \int_{h_3}^{\infty} dF$$

$$= \Sigma \frac{\rho_{12}^j \rho_{23}^k \rho_{13}^l}{j!k!l!} H_{j+k-1}(h_1) f(h_1) H_{j+l-1}(h_2) f(h_2) H_{k+l-1}(h_3) f(h_3),$$

which will also be found to be convergent.

[*Note.* Since writing this paper, under the impression that the results were new, I am indebted to Dr A. C. Aitken for pointing out that similar results have been given by him in lectures for a number of years. Among published work, reference may be made to:

(*a*) P. 175 of Aitken & Turnbull's *Theory of Canonical Matrices* (Blackie, 1931), where a more direct method of deriving the characteristic function of the multivariate normal distribution has been given;

(*b*) A paper on 'Fourfold sampling with and without replacement' by A. C. Aitken & H. T. Gonin (1935, *Proc. Roy. Soc. Edinb.* **55**, 114), where the tetrachoric expansions are discussed and new series associated with the correlated binomial and the correlated hypergeometric distributions are derived.

As the work of the Edinburgh school on this subject may not, however, be generally familiar, the Editor has suggested that this short paper should be published, together with the foregoing references, in order to bring the recent developments before a wider statistical audience. M. G. K.]

The Dirichlet probability density function is:

$$f(x_1,\ldots,x_n)=\frac{\Gamma(\vartheta_1+\cdots+\vartheta_{n+1})}{\Gamma(\vartheta_1)\cdots\Gamma(\vartheta_{n+1})}x_1^{\vartheta_1-1}\cdots x_n^{\vartheta_n-1}(1-x_1-\cdots-x_n)^{\vartheta_{n+1}-1}$$

on the

$$S_n=\left\{(x_1,\ldots,x_n)^T \;\middle|\; x_i\geq 0,\, i=1,\ldots n,\, \sum_{i=1}^{n}x_i\leq 1\right\}$$

n-dimensional unit simplex and zero outside,

$$\vartheta_1>0,\ldots,\vartheta_n>0,\vartheta_{n+1}>0$$

the parameters of the distribution.

Calculation of the Dirichlet cdf (H. Exton, 1976):

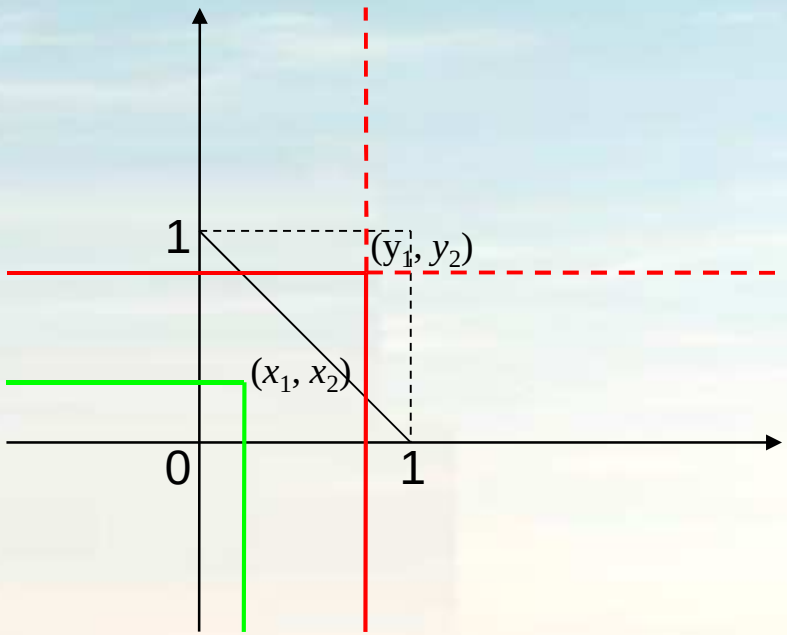
$$F(x_1,\ldots,x_n)=\frac{\Gamma(\vartheta_1+\cdots+\vartheta_{n+1})}{\Gamma(\vartheta_1)\cdots\Gamma(\vartheta_{n+1})}\frac{x_1^{\vartheta_1}}{\vartheta_1}\cdots\frac{x_n^{\vartheta_n}}{\vartheta_n}\cdot$$
$$\cdot F_A^{(n)}(1-\vartheta_{n+1},\vartheta_1,\ldots,\vartheta_n;\vartheta_1+1,\ldots,\vartheta_n+1;x_1,\ldots,x_n),$$

if $x_1\geq 0,\ldots,x_n\geq 0,\ x_1+\cdots+x_n\leq 1,$ where

$$F_A^{(n)}(a,b_1,\ldots,b_n;c_1,\ldots,c_n;x_1,\ldots,x_n)=$$
$$\sum_{m_1=0}^{\infty}\cdots\sum_{m_n=0}^{\infty}\frac{(a,m_1+\cdots+m_n)(b_1,m_1)\cdots(b_n,m_n)}{(c_1,m_1)\cdots(c_n,m_n)}\frac{x_1^{m_1}\cdots x_n^{m_n}}{m_1!\cdots m_n!}$$

is the so called first order Lauricella function and $(\lambda,k)=\dfrac{\Gamma(\lambda+k)}{\Gamma(\lambda)},$ k integer.

The Lauricella function method is not appropriate for all possible arguments of the cdf:



$$
\begin{aligned}
P(\xi_1 < y_1, \xi_2 < y_2) &= 1 - P(\{\xi_1 \geq y_1\} \cup \{\xi_2 \geq y_2\}) \\
&= 1 - P(\xi_1 \geq y_1) - P(\xi_1 \geq y_1) + P(\xi_1 \geq y_1, \xi_2 \geq y_2) \\
&= 1 - (1 - P(\xi_1 < y_1)) - (1 - P(\xi_1 < y_1)) + 0 \\
&= -1 + P(\xi_1 < y_1) + P(\xi_1 < y_1),
\end{aligned}
$$

The possibility of exact calculation using one- and two dimensional marginals only:

Let us suppose on the arguments of the Dirichlet cdf, that $x_1 \le \cdots \le x_n$.

Case a): If $x_1 + x_2 \ge 1$, then

$$F(x_1,\ldots,x_n) = 1 - n + \sum_{i=1}^{n} F(x_i)$$

Case b): If $x_1 + x_2 + x_3 \ge 1$, then

$$F(x_1,\ldots,x_n) = \frac{1}{2}(n-1)(n-2) - (n-2)\sum_{i=1}^{n} F(x_i) + \sum_{1\le i<j\le n} F(x_i, x_j)$$

Calculation by implicite recursion:

$$F(x_1) = P(\xi_1 \le x_1) = \begin{cases} 0, & \text{if } x_1 < 0, \\ F_A^{(1)}(x_1), & \text{if } 0 \le x_1 < 1, \\ 1, & \text{if } x_1 \ge 1, \end{cases}$$

$$\overline{F}(x_1) = P(\xi_1 > x_1) = 1 - P(\xi_1 \le x_1) = 1 - F(x_1),$$

$$F(x_1, x_2) = P(\xi_1 \le x_1, \xi_2 \le x_2) = \begin{cases} F_A^{(2)}(x_1, x_2), & \text{if } x_1 + x_2 < 1, \\ 1 - \overline{F}(x_1) - \overline{F}(x_2), & \text{if } x_1 + x_2 \ge 1, \end{cases}$$

$$\overline{F}(x_1, x_2) = P(\xi_1 > x_1, \xi_2 > x_2) = \begin{cases} 1 - F(x_1) - F(x_2) + \\ + F(x_1, x_2), & \text{if } x_1 + x_2 < 1, \\ 0, & \text{if } x_1 + x_2 \ge 1, \end{cases}$$

The general recursion formula:

$$F(x_1,\ldots,x_k)=\begin{cases} F_A^{(k)}(x_1,\ldots,x_k), & \text{if } x_1+\cdots+x_k<1, \\ 1-\sum\limits_{1\le i_1\le k}\overline{F}(x_{i_1})+\sum\limits_{1\le i_1<i_2\le k}\overline{F}(x_{i_1},x_{i_2})- & \\ -\cdots+(-1)^{k-1}\sum\limits_{1\le i_1<\cdots<i_{k-1}\le k}\overline{F}(x_{i_1},\ldots,x_{i_{k-1}}), & \text{if } x_1+\cdots+x_k\ge 1, \end{cases}$$

$$\overline{F}(x_1,\ldots,x_k)=\begin{cases} 1-\sum\limits_{1\le i_1\le k}F(x_{i_1})+\sum\limits_{1\le i_1<i_2\le k}F(x_{i_1},x_{i_2}) & \\ -\cdots+(-1)^{k}\sum\limits_{1\le i_1<\cdots<i_k\le k}F(x_{i_1},\ldots,x_{i_k}), & \text{if } x_1+\cdots+x_k<1, \\ 0, & \text{if } x_1+\cdots+x_k\ge 1. \end{cases}$$

Partial derivatives of a cdf in general:

If $F(x_1,\ldots,x_n)=P(\xi_1\le x_1,\ldots,\xi_n\le x_n)$, then we have

$$\frac{\partial F(x_1,\ldots,x_n)}{\partial x_k}=F(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n|x_k)f(x_k),$$

where $F(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n|x_k)$ is the conditional cdf of the random variables $\xi_1,\ldots,\xi_{k-1},\xi_{k+1},\ldots,\xi_n$ under the condition $\xi_k=x_k$, while $f(x_k)$ is the marginal pdf of the random variable ξ_k.

First order partial derivatives of the Dirichlet cdf:

In the case of Dirichlet distribution the conditional probability distributions are slightly modified Dirichlet distributions and the one dimensional marginal distributions are beta distributions.

So by the notation $\Theta = \{\vartheta_1, \ldots, \vartheta_n\}$ we have:

$$\frac{\partial F(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n; \Theta, \vartheta_{n+1})}{\partial x_k} =$$

$$= F\left(\frac{x_1}{1-x_k}, \ldots, \frac{x_{k-1}}{1-x_k}, \frac{x_{k+1}}{1-x_k}, \ldots, \frac{x_n}{1-x_k}; \Theta \setminus \{\vartheta_k\}, \vartheta_{n+1}\right)$$

$$\cdot f\left(x_k; \vartheta_k, \sum_{\vartheta_i \in \Theta \setminus \{\vartheta_k\}} \vartheta_i + \vartheta_{n+1}\right).$$

Mixed second order partial derivatives:

The mixed second order partial dervative according to $x_k, x_l, l \neq k$ can be calculated in the following way:

$$\frac{\partial^2 F(x_1,\ldots,x_{k-1},x_k,x_{k+1},\ldots,x_{l-1},x_l,x_{l+1},\ldots,x_n;\Theta,\vartheta_{n+1})}{\partial x_k \partial x_l} = \frac{1}{1-x_k}\cdot$$

$$\cdot F\left(\frac{x_1}{1-x_k-x_l},\ldots,\frac{x_{k-1}}{1-x_k-x_l},\frac{x_{k+1}}{1-x_k-x_l},\ldots,\frac{x_{l-1}}{1-x_k-x_l},\frac{x_{l+1}}{1-x_k-x_l},\ldots,\frac{x_n}{1-x_k-x_l};\Theta\setminus\{\vartheta_k,\vartheta_l\},\vartheta_{n+1}\right)\cdot$$

$$\cdot f\left(\frac{x_l}{1-x_k};\vartheta_l,\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k,\vartheta_l\}}\vartheta_i+\vartheta_{n+1}\right)\cdot f\left(x_k;\vartheta_k,\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k\}}\vartheta_i+\vartheta_{n+1}\right).$$

Pure second order partial derivatives:

The pure second order partial dervative according two times to x_k can be calculated in the following way:

$$\frac{\partial^2 F(x_1,\ldots,x_{k-1},x_k,x_{k+1},\ldots,x_n;\Theta,\vartheta_{n+1})}{\partial^2 x_k}=$$

$$=\left[\sum_{\substack{i=1\\ i\neq k}}^{n}\frac{x_i}{(1-x_k)^2}F\left(\frac{x_1}{1-x_k-x_i},\ldots,\frac{x_{k-1}}{1-x_k-x_i},\frac{x_{k+1}}{1-x_k-x_i},\ldots,\frac{x_{i-1}}{1-x_k-x_i},\frac{x_{i+1}}{1-x_k-x_i},\ldots,\frac{x_n}{1-x_k-x_i};\Theta\setminus\{\vartheta_k,\vartheta_i\},\vartheta_{n+1}\right)\cdot\right.$$

$$\left.\cdot f\left(\frac{x_i}{1-x_k};\vartheta_i,\sum_{\vartheta_j\in\Theta\setminus\{\vartheta_k,\vartheta_i\}}\vartheta_j+\vartheta_{n+1}\right)\right]f\left(x_k;\vartheta_k,\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k\}}\vartheta_i+\vartheta_{n+1}\right)+$$

$$+F\left(\frac{x_1}{1-x_k},\ldots,\frac{x_{k-1}}{1-x_k},\frac{x_{k+1}}{1-x_k},\ldots,\frac{x_n}{1-x_k};\Theta\setminus\{\vartheta_k\},\vartheta_{n+1}\right)\cdot$$

$$\cdot\frac{\Gamma\left(\sum_{\vartheta_i\in\Theta}\vartheta_i+\vartheta_{n+1}\right)}{\Gamma(\vartheta_k)\Gamma\left(\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k\}}\vartheta_i+\vartheta_{n+1}\right)}\cdot\left[(\vartheta_k-1)x_k^{\vartheta_k-2}(1-x_k)^{\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k\}}\vartheta_i+\vartheta_{n+1}-1}-x_k^{\vartheta_k-1}\left(\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k\}}\vartheta_i+\vartheta_{n+1}-1\right)(1-x_k)^{\sum_{\vartheta_i\in\Theta\setminus\{\vartheta_k\}}\vartheta_i+\vartheta_{n+1}-2}\right].$$

The bivariate Poisson distribution is defined as:

$$\begin{aligned} \xi_1 &= \eta_1 + \eta_2 \\ \xi_2 &= \eta_2 + \eta_3 \end{aligned}$$

where η_1, η_2, η_3 are independent, Poisson distributed with parameters $\lambda_1, \lambda_2, \lambda_3$.

Denote the one and two variate Poisson probability functions by:

$$f_P^{\lambda_i}(x) = P(\eta_i = x), i = 1, 2, 3$$ and $$f_P^{\lambda_1+\lambda_2, \lambda_2+\lambda_3}(x, y) = P(\xi_1 = x, \xi_2 = y)$$

Then the probability function of the bivariate Poisson distribution can be written as:

$$f_P^{\lambda_1+\lambda_2, \lambda_2+\lambda_3}(x, y) = f_P^{\lambda_1+\lambda_2}(x) f_P^{\lambda_2+\lambda_3}(y) \left\{ 1 + \sum_{r=1}^{\infty} \frac{\lambda_2^r}{r!} C_r^{\lambda_1+\lambda_2}(x) C_r^{\lambda_2+\lambda_3}(y) \right\}$$

where $C_r^{\lambda_1+\lambda_2}(x)$ **and** $C_r^{\lambda_2+\lambda_3}(y)$ **are the so called Charlier polynomials.**

The Charlier polynomials are defined as:

$$f_P^{\lambda}(x)\, C_r^{\lambda}(x) = (-\nabla)^r\, f_P^{\lambda - r}(x),$$

where $\nabla f(x) = f(x) - f(x-1).$

They have the explicit expression:

$$C_r^{\lambda}(x) = \sum_{k=0}^{r} \binom{r}{k}\binom{x}{k} k!(-\lambda)^{r-k}$$

and satisfy the orthogonal relation:

$$\int_0^{\infty} C_r^{\lambda}(x) C_s^{\lambda}(x)\, d\alpha(x) = \lambda^s s!\, \delta_r^s,$$

where $\alpha(x)$ **is a step function with the jump**

$$d\alpha(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\ldots, \quad \lambda > 0.$$

Applying the summation formula:

$$\sum_{x=0}^{h} f_P^{\lambda}(x) C_r^{\lambda}(x) = -f_P^{\lambda}(h) C_{r-1}^{\lambda}(h)$$

for the bivariate Poisson cumulative probability distribution function we get:

$$F_P(h,k) = \sum_{x=0}^{h} \sum_{y=0}^{k} f_P(x,y) = F_P^{\lambda_1+\lambda_2}(h) F_P^{\lambda_2+\lambda_3}(k) +$$

$$\sum_{r=1}^{\infty} \frac{\lambda_2^r}{r!} f_P^{\lambda_1+\lambda_2}(h)\, C_{r-1}^{\lambda_1+\lambda_2}(h)\, f_P^{\lambda_2+\lambda_3}(k)\, C_{r-1}^{\lambda_2+\lambda_3}(k),$$

where: $$F_P^{\lambda}(h) = \sum_{x=0}^{h} \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!}.$$

By the three term recurrence formula we can easily calculate the values of the Charlier polynomials:

$$C_{r+1}^{\lambda}(x) = (x - r - \lambda) C_r^{\lambda}(x) - \lambda r C_{r-1}^{\lambda}(x), \quad C_0^{\lambda}(x) \equiv 1, \quad C_1^{\lambda}(x) = -\lambda + x.$$

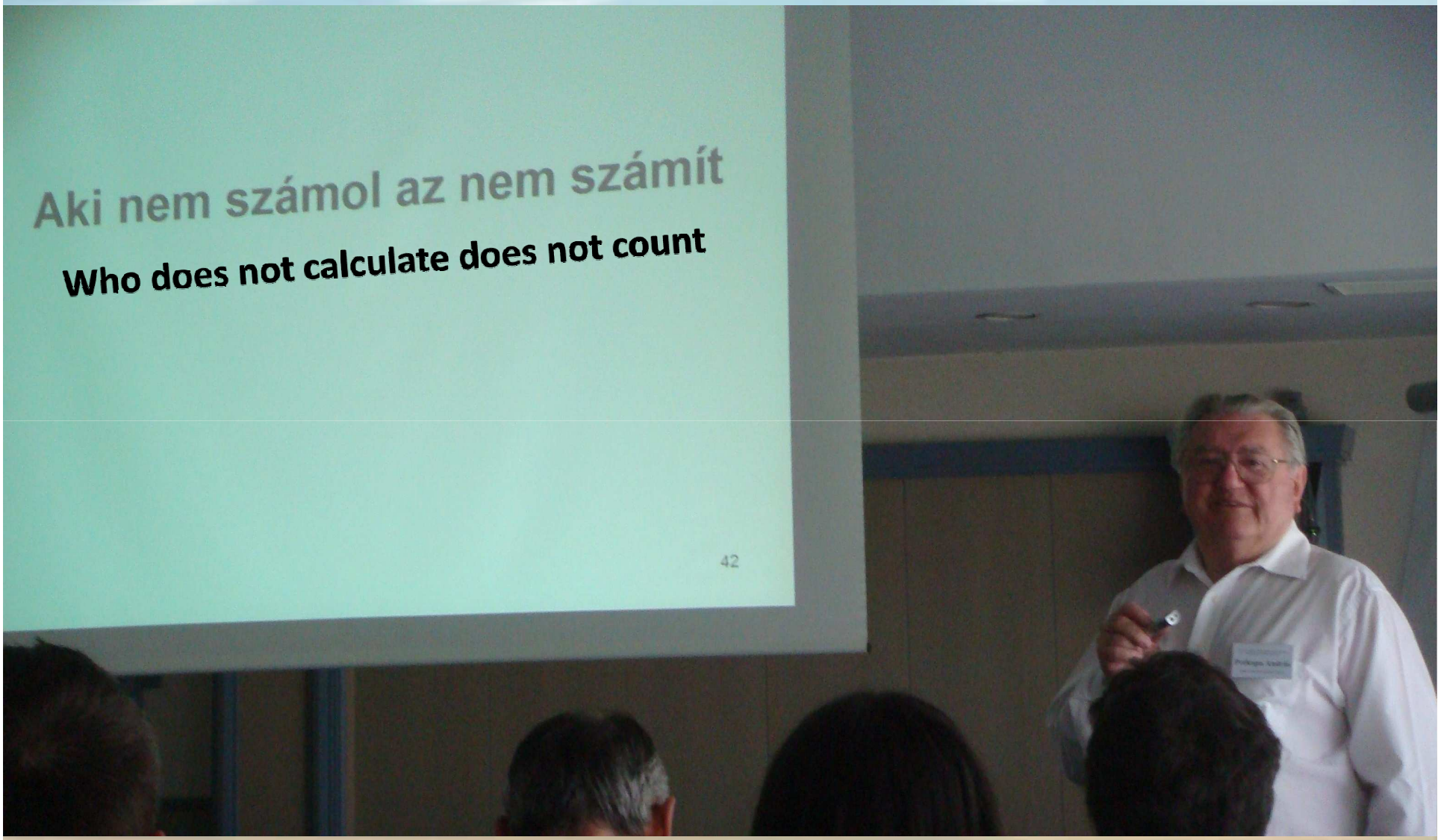
H. Teicher (1954) proved the following recursion for the calculation of the bivariate Poisson probabilities:

$$(y+1) f_P^{\lambda_1+\lambda_2,\lambda_2+\lambda_3}(x+1, y+1) =$$
$$= \lambda_2 f_P^{\lambda_1+\lambda_2,\lambda_2+\lambda_3}(x, y) + \lambda_3 f_P^{\lambda_1+\lambda_2,\lambda_2+\lambda_3}(x+1, y),\ x, y = 0,1,2,\ldots$$

Then having a tableau of the bivariate Poisson probabilities a simple summation results the values of the Poisson cdf:

$$F_P^{\lambda_1+\lambda_2,\lambda_2+\lambda_3}(h,k) = \sum_{x=0}^{h} \sum_{y=0}^{k} f_P^{\lambda_1+\lambda_2,\lambda_2+\lambda_3}(x, y)$$

4. Bivariate Poisson distribution (5)



When $\lambda_1 = \lambda_3 = 0.1$ **and** $\lambda_2 = 0.4$ **by the recursion of H. Teicher (1954) we get for the bivariate Poisson probabilities the following tableau:**

$x \backslash y$	0	1	2	3	4
0	0.548812	0.054881	0.002744	0.000092	0.000002
1	0.054881	0.225013	0.022226	0.001107	0.000037
2	0.002744	0.022226	0.046114	0.004501	0.000223
3	0.000092	0.001107	0.004501	0.006299	0.000618
4	0.000002	0.000037	0.000223	0.000618	0.000645

From the tableau we get: $f_P^{0.5,0.5}(1,1) = 0.225013.$

The series expansion by the Charlier polynomials gives:

$$f_P^{0.5,0.5}(1,1) = \left(0.5\mathrm{e}^{-0.5}\right)^2 \left\{1 + 0.4 + 9\frac{0.4^2}{2!} + 25\frac{0.4^3}{3!} + 49\frac{0.4^4}{4!} + 81\frac{0.4^5}{5!} + 121\frac{0.4^6}{6!} + 169\frac{0.4^7}{7!} + 225\frac{0.4^8}{8!} + \cdots\right\} \approx 0.225013.$$

From the tableau we get: $F_P^{0.5,0.5}(1,1) = 0.548812 + 2 \cdot 0.054881 + 0.225013 = 0.883587.$

The series expansion by the Charlier polynomials gives:

$$F_P^{0.5,0.5}(1,1) = \left(\mathrm{e}^{-0.5} + 0.5\mathrm{e}^{-0.5}\right)^2 + \left(0.5\mathrm{e}^{-0.5}\right)^2 \left\{0.4 + \frac{0.4^2}{2!} + 9\frac{0.4^3}{3!} + 25\frac{0.4^4}{4!} + 49\frac{0.4^5}{5!} + 81\frac{0.4^6}{6!} + 121\frac{0.4^7}{7!} + 169\frac{0.4^8}{8!} + \cdots\right\} \approx 0.883587.$$

The Statistician (2003)
52, *Part* 3, *pp.* 381–393

Analysis of sports data by using bivariate Poisson models

Dimitris Karlis

Athens University of Economics and Business, Greece

and Ioannis Ntzoufras

University of the Aegean, Chios, Greece

[Received November 2001. Final revision April 2003]

Summary. Models based on the bivariate Poisson distribution are used for modelling sports data. Independent Poisson distributions are usually adopted to model the number of goals of two competing teams. We replace the independence assumption by considering a bivariate Poisson model and its extensions. The models proposed allow for correlation between the two scores, which is a plausible assumption in sports with two opposing teams competing against each other. The effect of introducing even slight correlation is discussed. Using just a bivariate Poisson distribution can improve model fit and prediction of the number of draws in football games. The model is extended by considering an inflation factor for diagonal terms in the bivariate joint distribution. This inflation improves in precision the estimation of draws and, at the same time, allows for overdispersed, relative to the simple Poisson distribution, marginal distributions. The properties of the models proposed as well as interpretation and estimation procedures are provided. An illustration of the models is presented by using data sets from football and water-polo.

Keywords: Bivariate Poisson regression; Difference of Poisson variates; Inflated distributions; Soccer

The bivariate binomial distribution is defined as:

$$\begin{aligned} \xi_1 &= \eta_1 + \eta_2 \\ \xi_2 &= \eta_2 + \eta_3 \end{aligned}$$

where η_1, η_2, η_3 are independent, binomially distributed with parameters n_1, n_2, n_3 and p.

Denote the one and two variate binomial probability functions by:

$$f_b^{(n_i,p)}(x) = P(\eta_i = x), \quad i = 1,2,3$$ and $$f_b^{(n_1+n_2,p),(n_2+n_3,p)}(x,y) = P(\xi_1 = x, \xi_2 = y)$$

Then the probability function of the bivariate binomial distribution can be written as:

$$f_b^{(n_1+n_2,p),(n_2+n_3,p)}(x,y) = f_b^{(n_1+n_2,p)}(x) f_b^{(n_2+n_3,p)}(y) \cdot$$
$$\cdot \left\{ 1 + \sum_{r=1}^{\min(n_1+n_2-x, n_2+n_3-y, n_2)} \binom{n_2}{r} (pq)^r K_r^{(n_1+n_2,p)}(x) K_r^{(n_2+n_3,p)}(y) \right\}$$

Where $K_r^{(n_1+n_2,p)}(x)$ **and** $K_r^{(n_2+n_3,p)}(y)$ **are the so called Krawtchouk polynomials**

and $q = 1 - p$ **.**

The Krawtchouk polynomials are defined as:

$$f_b^{(n,p)}(x)\, K_r^{(n,p)}(x) = (-\nabla)^r\, f_b^{(n-r,p)}(x), \quad r = 0,1,\ldots,n,$$

where

$$\nabla f(x) = f(x) - f(x-1).$$

They have the explicit expression:

$$K_r^{(n,p)}(x) = \sum_{\nu=0}^{r} (-1)^{r-\nu} \binom{n-x}{r-\nu} \binom{x}{\nu} p^{r-\nu} (1-p)^{\nu}, \quad r = 0,1,\ldots,n.$$

and satisfy the orthogonal relation:

$$\sum_{x=0}^{n} K_r^{(n,p)}(x) K_s^{(n,p)}(x) \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{r} p^r (1-p)^r \delta_{r,s}, \quad r,s = 0,1,\ldots,n.$$

Applying the summation formula:

$$\sum_{x=0}^{h} f_B^{(n,p)}(x) K_r^{(n,p)}(x) = -f_B^{(n,p)}(h) K_{r-1}^{(n,p)}(h)$$

for the bivariate binomial cumulative probability distribution function we get:

$$F_B^{(n_1+n_2,p),(n_2+n_3,p)}(h,k) = \sum_{x=0}^{h} \sum_{y=0}^{k} f_B^{(n_1+n_2,p),(n_2+n_3,p)}(x,y) = F_B^{(n_1+n_2,p)}(h) F_B^{(n_2+n_3,p)}(k) +$$

$$\sum_{r=1}^{\min(n_1+n_2-h,n_2+n_3-k,n_2)} \binom{n_2}{r} (pq)^r f_B^{(n_1+n_2,p)}(h) K_{r-1}^{(n_1+n_2,p)}(h) f_B^{(n_2+n_3,p)}(k) K_{r-1}^{(n_2+n_3,p)}(k),$$

where: $F_B^{(n,p)}(h) = \sum_{x=0}^{h} f_B^{(n,p)}(x) = \sum_{x=0}^{h} \binom{n}{x} p^x q^{n-x}.$

By the three term recurrence formula we can easily calculate the values of the Krawtchouk polynomials:

$$(r+1) K_{r+1}^{(n,p)}(x) + pq(n-r+1) K_{r-1}^{(n,p)}(x) = [x - r - (n-2)p] K_r^{(n,p)}(x)$$

$$K_0^{(n,p)}(x) \equiv 1, \quad K_1^{(n,p)}(x) = -np + x.$$

Let us regard again the random variables:

$$\begin{aligned} \xi_1 &= \eta_1 + \eta_2 \\ \xi_2 &= \eta_2 + \eta_3 \end{aligned}$$

where η_1, η_2, η_3 are independent, binomially distributed with parameters n_1, n_2, n_3 and p.

The probability function of the bivariate binomial distribution can be calculated as:

$$P(\xi_1 = x, \xi_2 = y) = \sum_{s=\max(x-n_1, y-n_3)}^{\min(x,y)} P(\eta_1 = x-s, \eta_3 = y-s \mid \eta_2 = s) P(\eta_2 = s) =$$

$$= \sum_{s=\max(x-n_1, y-n_3)}^{\min(x,y)} \binom{n_1}{x-s} p^{x-s} q^{n_1-x+s} \binom{n_3}{y-s} p^{y-s} q^{n_3-y+s} \binom{n_2}{s} p^s q^{n_2-s} =$$

$$= q^{n_1+n_2+n_3} \sum_{s=\max(x-n_1, y-n_3)}^{\min(x,y)} \binom{n_1}{x-s} \binom{n_3}{y-s} \binom{n_2}{s} \left(\frac{p}{q}\right)^{x+y-s}$$

When $n_1 = 2, n_2 = 1, n_3 = 3, p = 0.2, q = 0.8$ **we get for the bivariate binomial probabilities the following tableau:**

	0	1	2	3	4	sum
0	0.262144	0.196608	0.049152	0.004096	0.000000	0.512000
1	0.131072	0.163840	0.073728	0.014336	0.000256	0.383232
2	0.016384	0.045056	0.027648	0.006400	0.000512	0.09600
3	0.000000	0.004096	0.003072	0.000768	0.000064	0.008000
sum	0.409600	0.409600	0.153600	0.025600	0.000832	0.999233

Thank You for the Attention!
Happy Birthday !