The bivariate gamma distribution with Laplace transform $(1 + as + bt + cst)^{-q}$: history, characterisations, estimation.

Gérard Letac, Université Paul Sabatier,
Toulouse, France.

Conference at York, 11/30/07.

*

*with Jean-Yves Tourneret and Florent Chatelain, Irit-Enseeiht, Toulouse.

Vocabulary. While the names of distributions in $\mathbb{R}$ are generaly unambiguous, at the contrary in the jungle of distributions in $\mathbb{R}^n$ almost nothing is codified outside of Wishart and Gaussian cases. The scenario is usually as follows : choose a one dimensional *thingy* type (quite often an exponential dispersion model, namely a natural exponential family and all its real powers of convolution) as gamma or negative binomial, then any law in $\mathbb{R}^n$ whose margins are of thingy type are said to be multidimensional thingy. Although the study of *all* distributions with given marginals are rather in the non parametric domain of study, actually each author who isolates some parametric family will declare that he has THE multidimensional thingy family.

This is what we are also doing today by calling *bivariate gamma* any distribution $\nu(dx, dy)$ on $(0, \infty)^2$ such that there exists 4 parameters a, b, c, q with $q > 0$ satisfying

$$\int_0^\infty \int_0^\infty e^{-sx-ty}\nu(dx, dy) = \frac{1}{(1 + as + bt + cst)^q}$$

(To be fair it is called Kibble and Moran distribution by Johnson and Kotz). When $(X, Y) \sim \nu$ doing $t = 0$ shows that X is gamma distributed with shape parameter q et and scale parameter a :

$$\gamma_{q,a}(dx) = \frac{1}{\Gamma(q)} e^{-\frac{x}{a}} \left(\frac{x}{a}\right)^{q-1} \mathbf{1}_{(0,\infty)}(x) \frac{dx}{a}.$$

Same for Y. This shows that a et b are > 0. For if a or b are nonpositive the distribution ν is concentrated on an other quadrant, or on an axis. No generality is lost by postulating $a, b > 0$ and therefore $\nu(dx, dy)$ concentrated on $(0, \infty)^2$. The hypothesis $q > 0$ is more important and $q < 0$ would lead us to distributions concentrated on a finite number of points with binomial margins.

Constraints for parameters and density. We are going to see later on that if $a, b, q > 0$, the distribution ν does exist if and only if $0 \leq c \leq ab$. If $c > 2ab$ the function $(s, t) \mapsto -q \log(1 + as + bt + cst)$ is not convex inside the suitable branch of hyperbola, but the proof that $ab < c \leq 2ab$ is impossible is subtlier. Note that this implies that ν is infinitely divisible (as observed by Vere-Jones (1967)). Note also that is (X, Y) is bivariate gamma with parameters a, b, c, q then $(X/a, Y/b)$ is bivariate gamma with parameters $1, 1, c/ab, q$. If $c = ab$ obviously X et Y are independent.

The case $a = b = c$ (and thus $c \geq 1$ since $c \leq ab$) is called the *standard* case with parameter $r \in [0, 1)$ with $c = \frac{1}{1-r}$. It seems to be more interesting than the $a = b = 1$ case (although the $a = b = 1$ case is natural in the context of Lancaster probabilities that we shall consider later on).

If $c = 0$ then $(X, Y) \sim (aZ, bZ)$ with $Z \sim \gamma_{p,1}$. If $0 < c \leq ab$ the distribution ν has a non familiar density. To compute it, let us use natural exponential families (NEF).

The NEF $F(\mu_{\rho,q})$

Let $\rho > 0$, let $q > 0$ and consider the measure $\mu_{\rho,q}(dx, dy)$ with density

$$\frac{(xy)^{q-1}}{\Gamma(q)} f_q(\rho xy)$$

on $(0, \infty)^2$ where $f_q(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!\Gamma(q+n)}$. Its Laplace transform

$$L_{\mu_{\rho,q}}(\theta_1, \theta_2) = \int\int e^{\theta_1 x + \theta_2 y} \mu_{\rho,q}(dx, dy)$$

is defined on

$$\Theta(\mu_{\rho,q}) = \{(\theta_1, \theta_2)\ ;\ \theta_1 < 0, \theta_2 < 0,\ \theta_1\theta_2 - \rho > 0\}$$

and is equal to $L_{\mu_{\rho,q}}(\theta_1, \theta_2) = (\theta_1\theta_2 - \rho)^{-q}$ on this set. Consider the NEF $F(\mu_{\rho,q})$ generated by $\mu_{\rho,q}$.

An element $P(\theta_1,\theta_2)$ of this NEF has the form

$$P(\theta_1,\theta_2)(dx,dy) =$$

$$(\theta_1\theta_2-\rho)^q e^{\theta_1 x+\theta_2 y}\frac{(xy)^{q-1}}{\Gamma(q)} f_q(\rho xy)\mathbf{1}_{(0,\infty)^2}(x,y)dxdy.$$

Thus

$$\int\int e^{-sx-ty}P(\theta_1,\theta_2)(dx,dy) = \frac{L(\theta_1-s,\theta_2-t)}{L(\theta_1,\theta_2)}$$

$$= \frac{(\theta_1\theta_2-\rho)^q}{((\theta_1-s)(\theta_2-t)-\rho)^q}.$$

Introduce $r = \frac{\rho}{\theta_1\theta_2}$. We are going to see that r is the correlation coefficient of $P(\theta_1, \theta_2)$. Obviously the parameters

$$\begin{aligned} a &= \frac{-\theta_2}{\theta_1\theta_2 - \rho} = -\frac{1}{\theta_1(1-r)}, \\ b &= \frac{-\theta_1}{\theta_1\theta_2 - \rho} = -\frac{1}{\theta_2(1-r)}, \\ c &= \frac{1}{\theta_1\theta_2 - \rho} = \frac{1}{\theta_1\theta_2(1-r)} \end{aligned}$$

seem more attractive for $P(\theta_1, \theta_2)$ since

$$\int\int e^{-sx-ty} P(\theta_1, \theta_2)(dx, dy) = \frac{1}{(1 + as + bt + cst)^q}.$$

with the conversion table

$$\theta_1 = -\frac{b}{c}, \qquad \theta_2 = -\frac{a}{c}$$

$$r = \frac{ab - c}{ab}, \ \ \rho = \frac{ab - c}{c^2}.$$

In the standard case $a = b = c = \frac{1}{1-r}$ we have $\theta_1 = \theta_2 = -1$ and $\rho = r$. In the 'Lancaster' case $a = b = 1$ and $c = 1 - r$ we have $\theta_1 = \theta_2 = -\frac{1}{1-r}$ and $\rho = \frac{r}{(1-r)^2}$.

The classical objects of the NEF $F(\mu_{\rho,q})$. We have seen that $\Theta(\mu_{\rho,q})$ is the interior of the hyperbola branch

$$\{\theta_1 < 0, \theta_2 < 0; \theta_1\theta_2 - \rho > 0\}.$$

The cumulant function is

$$k(\theta_1, \theta_2) = -q \log(\theta_1\theta_2 - \rho)$$

Its differential is

$$k'(\theta_1, \theta_2) = \frac{-q}{\theta_1\theta_2 - \rho}(\theta_2, \theta_1) = (qa, qb) = (m_1, m_2).$$

The domain of the means is $M(F) = (0, \infty)^2$. This gives r as a function of m_1, m_2 and ρ :

$$1 - r = \frac{q^2}{2\rho m_1 m_2}\left(\sqrt{1 + \frac{4\rho m_1 m_2}{q^2}} - 1\right)$$

or in a simpler form :

$$\frac{1}{1-r} = \frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{4\rho m_1 m_2}{q^2}}.$$

Here is the variance function either expressed in canonical parameters, or in the mean parameters, or in the classical parameters :

$$\begin{aligned} k''(\theta_1,\theta_2) &= \frac{q}{(\theta_1\theta_2-\rho)^2}\left[\begin{array}{cc} \theta_2^2 & \rho \\ \rho & \theta_1^2 \end{array}\right] \\ &= \frac{1}{q}\left[\begin{array}{cc} m_1^2 & rm_1m_2 \\ rm_1m_2 & m_2^2 \end{array}\right] \\ &= q\left[\begin{array}{cc} a^2 & ab-c \\ ab-c & b^2 \end{array}\right] \end{aligned}$$

(do not forget that r is the complicated function of m_1 and m_2 given above). One characterisation of this bivariate gamma NEF is due to Barlev *et al.* (1994). It says that if a NEF concentrated on $(0,\infty)^2$ has a variance function of the form $\frac{1}{q}\left[\begin{array}{cc} m_1^2 & * \\ * & m_2^2 \end{array}\right]$ this implies that it is a bivariate gamma NEF.

How to get quickly the density. Starting from $(1+as+bt+cst)^{-q} = \mathbb{E}(e^{-sX-tY})$ with

$$a, b, c > 0$$

1. We choose α and $\beta > 0$ such that $X_1 = \alpha X$ and $Y_1 = \beta Y$ satisfy

$$(1+c_1 s + c_1 t + c_1 st)^{-q} = \mathbb{E}(e^{-sX_1 - tY_1})$$

pour some $c_1 = \alpha a = \beta b = c\alpha\beta$. Thus (X_1, Y_1) is standard with parameters r and q. We get

$$\alpha = \frac{b}{c},\ \beta = \frac{a}{c},\ c_1 = \frac{ab}{c} = \frac{1}{1-r}$$

2. We now compute the density for the standard case

$$
\begin{aligned}
&\mathbb{E}(e^{-sX_1-tY_1}) \\
&= \frac{(1-r)^q}{(1-r+s+t+st)^q} \\
&= \frac{(1-r)^q}{((1+s)(1+t)-r)^q} \\
&= \frac{(1-r)^q}{\Big((1+s)(1+t)(1-\frac{r}{(1+s)(1+t)})\Big)^q} \\
&= \frac{(1-r)^q}{(1+s)^q(1+t)^q}\sum_{n=0}^{\infty}\frac{(q)_n}{n!}\frac{r^n}{(1+s)^n(1+t)^n} \\
&= \frac{(1-r)^q}{\Gamma(q)}\sum_{n=0}^{\infty}\frac{\Gamma(q+n)}{n!}\frac{r^n}{(1+s)^{q+n}(1+t)^{q+n}} \\
&= \frac{(1-r)^q}{\Gamma(q)}\sum_{n=0}^{\infty}\frac{r^n}{n!\Gamma(q+n)} \\
&\quad \times \int_0^\infty\int_0^\infty e^{-sx_1-ty_1}(x_1y_1)^{q+n-1}dx_1dy_1 \\
&= \frac{(1-r)^q}{\Gamma(q)}\int_0^\infty\int_0^\infty e^{-sx_1-ty_1}(x_1y_1)^{q-1}f_q(rx_1y_1)dx_1dy_1
\end{aligned}
$$

Mixing. In other terms if

$$N, X_0, \ldots, X_n \ldots, Y_0, \ldots, Y_n, \ldots$$

are independent and such that $X_0 \sim Y_0 \sim \gamma_{q,1}$, $X_i \sim Y_i \sim \gamma_{1,1}$ for $i \geq 1$ and if N follows a negative binomial distribution :

$$\Pr(N = n) = \frac{(q)_n}{n!}(1-r)^q r^n$$

denote

$$U_n = X_0 + \ldots + X_n, \quad V_n = Y_0 + \ldots + Y_n.$$

Under these circumstances (U_N, V_N) follows a standard gamma bivariate distribution with parameters r and q : thus this distribution is a mixing of gamma distributions.

Why $0 \leq c \leq ab$? (and not only $0 \leq c \leq 2ab$ for which $\log((1+as+bt+cst)$ is concave). This is equivalent to ask : why $r \in [0,1]$ and not only $r \in [-1,1]$? The preceeding calculation contains the answer ; if $r < 0$ the density is $\frac{(1-r)^q}{\Gamma(q)}(x_1y_1)^{q-1}f_q(rx_1y_1)$. But

$$\left(\frac{z}{2}\right)^{q-1} f_q\left(\frac{1}{4}z^2\right) = I_{q-1}(z)$$

Since the classical Bessel function $I_{q-1}(z)$ has its zeros on the imaginary axis and since they are all simple and in an infinite number, this implies that f_q is not positive on $(-\infty, 0)$.

Moments of the standard bivariate gamma. Recall that the distribution of (X,Y) such that

$$\mathbb{E}(e^{-sX-tY})=(1-r)^q(1-r+s+t+st)^{-q}$$

has been called standard bivariate gamma distribution with covariance r and shape parameter q, and therefore with density

$$(1-r)^q\frac{(xy)^{q-1}}{\Gamma(q)}f_q(rxy)e^{-x-y}.$$

We get easily from this the moments : $\mathbb{E}(X^sY^t)=$

$$\frac{(1-r)^q\Gamma(q+s)\Gamma(q+t)}{\Gamma(q)^2}\ {}_2F_1(q+s,q+t;q;r)=$$

$$\frac{(1-r)^{-s-t}\Gamma(q+s)\Gamma(q+t)}{\Gamma(q)^2}\ {}_2F_1(-s,-t;q;r)$$

and therefore $\mathbb{E}((XY)^s)$

$$\leq\frac{(1-r)^{-2s}\Gamma(q+s)^2}{\Gamma(q)^2}\ {}_2F_1(-s,-t;q;1)$$

$$=(1-r)^{-s-t}\frac{\Gamma(q+2s)}{\Gamma(q)}$$

For instance

$$
\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(Y) = \frac{q}{1-r} \\
\mathbb{E}(X^2) &= \mathbb{E}(Y^2) = \frac{q(q+1)}{(1-r)^2} \\
\sigma^2(X) &= \sigma^2(Y) = \frac{q}{(1-r)^2} \\
\mathbb{E}(XY) &= \frac{1}{(1-r)^2} q^2 (1 + \frac{r}{q}) \\
\mathbb{E}(X^2Y^2) &= \frac{1}{(1-r)^4} q^2 (q+1)^2 (1 + \frac{4r}{q} + \frac{2r^2}{q(q+1)})
\end{aligned}
$$

Joint distribution of $(P,S)=(XY,X+Y)$, standard case. The distribution is

$$(1-r)^q\frac{p^{q-1}}{\Gamma(q)}f_q(rp)e^{-s}\frac{1}{\sqrt{s^2-4p}}\mathbf{1}_{s>2\sqrt{p}}(p,s)dpds.$$

Writing $h_q(z)=\sum_{n=0}^{\infty}\frac{(q)_n}{n!\Gamma(2q+2n)}z^n$ and imitating the way that the density of (X,Y) has been found one gets the law of S

$$(1-r)^q s^{2q-1}h_q(rs^2)\mathbf{1}_{s>0}(s)ds$$

but for the density of $P=XY$ nothing can be obtained from the joint distribution of (P,S) and from the Mellin transform

$$\mathbb{E}(P^s)=(1-r)^{-2s}\frac{\Gamma(q+s)^2}{\Gamma(q)^2}\ {}_2F_1(-s,-s;q;r).$$

History. These distributions have been introduced by Wicksell (1933) and independently by Kibble (1941) and mentioned by Moran (1967), by Barndorff Nielsen(1980), Wang (1982), Seshadri (1988), Barlev and 5 coauthors(1994) and Johnson and Kotz (1972). Angelo Koudou (1995) and Philippe Bernardorff (2003) study them in their thesis, defended at the Université Paul Sabatier at Toulouse.

Bivariate gamma and Wishart distributions

In one dimension, gamma distributions with when the shape parameter is a half integer are known as chi squares. If the shape parameter is not a half integer, probabilistic interpretations are rather rare. Here is one which is due to Yor (1990) :

$$\left(\int_0^\infty e^{-\frac{t}{p}+B(t)}dt\right)^{-1} \sim \gamma_{p,1}$$

when B is a standard Brownian motion. The most natural generalization of the chi square distribution is the Wishart distribution : if $Z_1, \ldots, Z_N$ are iid $N(0, \Sigma)$ in $\mathbb{R}^r$ (written by columns) the Wishart distribution is the law of the random matrix

$$W = \frac{1}{2}(Z_1 Z_1^T + \ldots + Z_N Z_N^T).$$

It satisfies

$$\mathbb{E}(e^{-\operatorname{tr}(\theta W)}) = \det(I + \Sigma^{-1}\theta)^{-N/2}.$$

(θ is a symmetric positive definite matrix).

For $r = 2$ write $W = \begin{bmatrix} X & Z \\ Z & Y \end{bmatrix}$. Suppose that $\theta = \mathrm{diag}(s,t)$ and denote $\Sigma^{-1} = \begin{bmatrix} a & \gamma \\ \gamma & b \end{bmatrix}$ and denote $c = ab - \gamma^2$. One gets exactly

$$\mathbb{E}(e^{-sX-tY}) = \frac{1}{(1+as+bt+cst)^{N/2}}:$$

As a result : the diagonal of a Wishart matrix on the (2,2) real symmetric matrices is a bivariate gamma.

This result can be extended in several directions : by replacing $N/2$ by $q \geq 1/2$; by replacing the (2,2) positive definite matrices by the Lorentz cone

$$C_d = \{(x, y, z_1, \ldots, z_d)\,;x, y > 0,\ xy > z_1^2 + \cdots + z_d^2\}$$

and writing symbolically its elements as $w = \begin{bmatrix} x & z \\ z & y \end{bmatrix}$. The Wishart distributions on C_d of shape parameter q are the elements of the natural exponential family generated by the measure on C_d with density

$$(xy - z_1^2 - \cdots - z_d^2)^{q-1-\frac{d}{2}}$$

if $q > 1/2$ (and some singular distribution on ∂C_d if $q = 1/2$). If (X, Y, Z) follows such a Wishart law on C_d then (X, Y) is bivariate gamma. The cases of ordinary Wishart distributions, or complex or quaternionic are the cases $d = 1, 2, 4$.

Note that for $0 < q < 1/2$ the bivariate gamma does exist but *cannot* be the marginal distribution of (X, Y) for a Wishart distribution $(X, Y, Z_1, \ldots, Z_d)$.

The bivariate gamma are Lancaster probabilities Suppose that $\alpha(dx)$ and $\beta(dy)$ are distributions on $\mathbb{R}$ such that $\int e^{-a|x|}\alpha(dx)$ and $\int e^{-a|y|}\beta(dy)$ are finite for some $a > 0$. Let (P_n) and (Q_n) be the sequences of corresponding orthonormal polynomials. We say that the distribution $\mu(dx, dy)$ on $\mathbb{R}^2$ with margins α and β is of Lancaster type if either their exists a sequence (ρ_n) of real numbers such that μ is absolutely continuous with respect to $\alpha(dx)\beta(dy)$ and is

$$\mu(dx, dy) = \left[\sum_{n=0}^{\infty} \rho_n P_n(x) Q_n(y)\right] \alpha(dx)\beta(dy)$$

or μ is the weak limit of such distributions. In both cases we have

$$\mathbb{E}_\mu(P_n(X)|Y) = \rho_n Q_n(Y),\ \ \mathbb{E}_\mu(Q_n(Y)|X) = \rho_n P_n(X).$$

D'jachenko (1962) has observed that the bivariate gamma is a Lancaster probabilility with $\rho_n = r^n$ where $r \in [0,1]$ is the correlation coefficient. This is a not quite obvious fact and relies on a classical formula on Laguerre polynomials. Note that this is not the standard bivariate gamma but the one with $a = b = 1$ and $c = 1 - r$. To this, add a Tyan and Thomas theorem which says that in general the $(\rho_n)'_{n\geq 0}$s appearing in the definition of Lancaster probabilities are sequences of moments, and you get an elegant integral representation :

Proposition. $\mu(dx, dy)$ is a Lancaster probability for the margins $\alpha = \beta = \gamma_{q,1}$ if and only if there exists a probability $\nu(dr)$ on $[0,1]$ such that

$$\int_0^\infty \int_0^\infty e^{-sx-ty}\mu(dx,dy) =$$

$$\int_0^1 \frac{\nu(dr)}{(1+s+t+(1-r)st)^q}.$$

Dans ce cas $\int_0^1 r^n \nu(dr) = \rho_n$.

Other details on Lancaster probabilities can be found in Koudou (1996) and the paper on Gibbs sampling by Diaconis and two coauthors to appear in *Statistical Science* in 2008.

The Wang characterisation (1982) Recall the characterisation by Lukacs (1956) of the ordinary gamma distributions : If X and Y are positive, independent, not constant and such that $Z = X/(X+Y)$ is independent of $X+Y$ then there exist $p, q, c > 0$ such that $X \sim \gamma_{p,c}$ and $Y \sim \gamma_{q,c}$. Proof : let a and $b \in (0,1)$ such that $\mathbb{E}(X|X+Y) = a(X+Y)$ and $\mathbb{E}(X^2|X+Y) = b(X+Y)^2$. Then

$$\mathbb{E}(\mathbb{E}(e^{-s(X+Y)}X|X+Y)) = a\mathbb{E}((X+Y)e^{-s(X+Y)})$$

$$\mathbb{E}(\mathbb{E}(e^{-s(X+Y)}X^2|X+Y)) = b\mathbb{E}((X+Y)^2e^{-s(X+Y)}).$$

Denoting $L_X(s) = \mathbb{E}(e^{-sX})$ we get the differential system in L_X and L_Y : $L'_X L_Y = a(L_X L_Y)'$ and $L''_X L_Y = b(L_X L_Y)''$ easily solved by intoducing $k_X = \log L_X$.

Wang does the same thing in two dimensions : if $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ in $(0, \infty)^2$ are independent, not constant and such that

$$Z = \left(\frac{X_1}{X_1 + Y_1}, \frac{X_2}{X_2 + Y_2}\right)$$

is independent of $X+Y$ then there exist $p, q, a, b, c > 0$ such that X and Y are bivariate gamma with

$$\mathbb{E}(e^{-sX_1 - tX_2}) = (1 + as + bt + cst)^{-p}$$

$$\mathbb{E}(e^{-sY_1 - tY_2}) = (1 + as + bt + cst)^{-q}.$$

The method of proof is the same as in one dimension. But see also a splendid paper by Letac and Wesołowski in TAMS (2008) which generalizes...

A parenthesis on the negative multinomial distribution in dimension 2. The distribution $NM(p,a,b,c)$ of (X,Y) in $\mathbb{N}^2$ with parameters p,a,b,c and constraints $0 < p$, $0 < a,b < 1$ and $0 < c < (1-a)(1-b)$ has generating function :

$$\mathbb{E}(x^X y^Y) = \left[\frac{(1-a)(1-b)-c}{1-ax-by+(ab-c)xy}\right]^p.$$

The same little trick as for the gamma gives

$$\Pr(X=m,\ Y=n) = ((1-a)(1-b)-c)^p a^m b^n \times$$

$$\sum_{k=0}^{\min(m,n)} \frac{(p)_k}{k!}\frac{(p+k)_{m-k}}{(m-k)!}\frac{(p+k)_{n-k}}{(n-k)!}\left(\frac{c}{ab}\right)^k.$$

by the following calculation

$$
\begin{aligned}
& \left[\frac{(1-a)(1-b)-c}{1-ax-by+(ab-c)xy}\right]^p \\
= \ & \left[\frac{(1-a)(1-b)-c}{(1-ax)(1-by)(1-\frac{cxy}{(1-ax)(1-by)}}\right]^p \\
= \ & \left[\frac{(1-a)(1-b)-c}{(1-ax)(1-by)}\right]^p \times \\
& \sum_{k=0}^{\infty} \frac{(p)_k}{k!} \frac{c^k x^k y^k}{(1-ax)^k (1-by)^k} \\
= \ & ((1-a)(1-b)-c)^p \times \\
& \sum_{k=0}^{\infty} \frac{(p)_k}{k!} \frac{c^k x^k y^k}{(1-ax)^{p+k} (1-by)^{p+k}} \\
= \ & ((1-a)(1-b)-c)^p \times \\
& \sum_{k,r,s=0}^{\infty} \frac{(p)_k}{k!} \frac{(p+k)_r}{r!} \frac{(p+k)_s}{s!} a^r b^s c^k x^{r+k} y^{s+k}
\end{aligned}
$$

Although in one dimension we have

$$\lim_{n\to\infty} H_{1/n} NB(q, \frac{nc}{1+nc}) = \gamma_{q,c},$$

(here $H_{1/n}$ is the action on distribution of the dilation $x \mapsto H_{1/n}(x) = x/n$), in two dimensions it is impossible to find an affine transformation $A_n(x,y) = (\alpha_n x, \beta_n y)$ and a sequence (a_n, b_n, c_n) of parameters such that $\lim_{n\to\infty} A_n NM(q, a_n, b_n, c_n)$ converges towards a non singular bivariate gamma. The reason is : replacing x and y respectively by $x_n = e^{\alpha_n s}$ and $y_n = e^{\beta_n t}$ the limit of

$$\frac{(1-a_n)(1-b_n)-c_n}{1-a_n x_n - b_n y_n + (a_n b_n - c_n) x_n y_n}$$

has necessarily the form $\frac{1}{1+As+Bt}$ and never the form $\frac{1}{1+As+Bt+Cst}$.

Maximum likelihood estimation, standard case.

Recall that the density of the standard bivariate gamma has the form

$$(1-r)^q f_q(rxy)$$

with respect to the reference measure

$$\frac{1}{\Gamma(q)}(xy)^{q-1}e^{-x-y}\mathbf{1}_{(0,\infty)^2}(x,y)dxdy.$$

If we observe the sample $(X_1, Y_1), \ldots, (X_N, Y_N)$ we get

Proposition. In the standard case (q known, r unknown) the maximum likelihood estimator $\widehat{r}$ for $r \in [0,1)$ always exists. We have $\widehat{r} = 0$ if and only if $\frac{1}{N}\sum_{i=1}^N X_iY_i \leq q^2$. If not, $\widehat{r}$ is the unique solution of

$$\frac{q}{1-r} = \frac{1}{N}\sum_{i=1}^{N} X_iY_i\frac{f_{q+1}}{f_q}(rX_iY_i)$$

which needs a numerical treatment.

The trick is to show that $x \mapsto \log f_q(x)$ is concave on $(0, \infty)$. First $f_q'(x) = f_{q+1}(x)$ and $f_q''(x) = f_{q+2}(x)$. We change q into $q-1$. Therefore we have to show that

$$f_{q+1}(x) f_{q-1}(x) - f_q^2(x) < 0$$

for all $q > 1$ and $x > 0$. Denote $u_n(q) = \frac{1}{n!\Gamma(q+n)}$. Thus the coefficient of z^n of the entire function $f_{q+1}(z) f_{q-1}(z) - f_q^2(z)$ is

$$v_n(q) = \sum_{k=0}^{n} [u_k(q+1) u_{n-k}(q-1) - u_k(q) u_{n-k}(q)].$$

Let us show that $v_n(q) < 0$ for $q > 1$. We write :

$$
\begin{aligned}
v_n(q) &= \sum_{k=0}^{n} \frac{1}{k!(n-k)!}[\frac{1}{\Gamma(q+k+1)\Gamma(q+n-k-1)} \\
&\quad -\frac{1}{\Gamma(q+k)\Gamma(q+n-k)}] \\
&= -\frac{n+1}{\Gamma(q+n+1)\Gamma(q)} \\
&\quad +\sum_{k=0}^{n-1} \frac{n-2k-1}{k!(n-k)!} \; \frac{1}{\Gamma(q+k+1)\Gamma(q+n-k)} \\
&= -\frac{n+1}{\Gamma(q+n+1)\Gamma(q)} \\
&\quad +\frac{1}{2}\sum_{k=0}^{n-1} \; \frac{1}{\Gamma(q+k+1)\Gamma(q+n-k)} \\
&\quad \times[\frac{n-2k-1}{k!(n-k)!} + \frac{n-2(n-1-k)-1}{(n-1-k)!(k+1)!}] \\
&= -\frac{n+1}{\Gamma(q+n+1)\Gamma(q)} \\
&\quad -\frac{1}{2}\sum_{k=0}^{n-1} \; \frac{1}{\Gamma(q+k+1)\Gamma(q+n-k)} \\
&\quad \times\frac{(n-2k-1)^2}{(k+1)!(n-k)!} < 0.
\end{aligned}
$$

Maximum likelihood estimation, general case.

Even less easy. It has no meaning to ask for concavity of the log of the density parameterized by $(\theta_1, \theta_2, \rho)$ since the domain of definition

$$\{(\theta_1, \theta_2, \rho)\ ; \theta_i < 0,\ \theta_1\theta_2 > \rho > 0\}$$

is not convex (watch the boundary $\theta_1 = \theta_2$ and $\theta_1\theta_2 = \rho$ to be convinced). Taking $\rho = r\theta_1\theta_2$, the domain $(-\infty, 0)^2 \times [0, 1)$ for (θ_1, θ_2, r) is more manageable, but for fixed r the log of the density in not concave with respect to (θ_1, θ_2) since, up to a linear function, it a a function of the product $\theta_1\theta_2$. This excludes concavity in the product space $(-\infty, 0)^2$ (watch the determinant of the Hessian matrix of $(\theta_1, \theta_2) \mapsto h(\theta_1\theta_2)$ when $\theta_1\theta_2$ becomes close to zero).

(θ_1, θ_2, r) can be estimated by writing the 3 equations of cancelation of the gradient of the log likelihood (the proof of the fact that the log likelihood goes to $-\infty$ at the boundary of the domain is quite delicate). One computes θ_1 and θ_2 with respect to r which becomes the solution of the following equation where $(\overline{X}, \overline{Y})$ is the empirical mean : $0 = g_N(r)$

$$= r - 1 + \frac{q}{\overline{XY}} \frac{1}{N} \sum_{i=1}^{N} X_i Y_i \frac{f_{q+1}}{f_q} [\frac{r}{(1-r)^2} \frac{q^2 X_i Y_i}{\overline{XY}}].$$

We content ourselves in studying it for large N. Using large numbers law the equation becomes $0 = g(r) =$

$$r - 1 + \frac{q}{\mathbb{E}(X)\mathbb{E}(Y)} \mathbb{E}\left[XY \frac{f_{q+1}}{f_q} (\frac{r}{(1-r)^2} \frac{q^2 XY}{\mathbb{E}(X)\mathbb{E}(Y)}) \right].$$

An important remark is that $g(r)$ depends only on the distribution of $(\frac{X}{\mathbb{E}(X)}, \frac{Y}{\mathbb{E}(Y)})$. Therefore, without loss of generality we may assume that the distribution of (X, Y) is standard with parameter r_0.

The crucial function g_{r_0}. As explained the above function g depends only on parameter r_0 and is $g_{r_0}(r) =$

$$r-1+\frac{(1-r_0)^2}{q}\mathbb{E}\left[XY\frac{f_{q+1}}{f_q}(\frac{r}{(1-r)^2}(1-r_0)^2XY)\right].$$

It is easily seen that $g_{r_0}(0) = \frac{r_0}{q} > 0$.

Knowing that $f_q(u) \sim \frac{1}{2\sqrt{\pi}}u^{-\frac{q}{2}+\frac{1}{4}}e^{2\sqrt{u}}$ if $u \to \infty$, we get that $\frac{f_{q+1}(u)}{f_q(u)} \sim \frac{1}{\sqrt{u}}$ and therefore $g(r) \sim -c\sqrt{1-r}$ with $c = 1 - \frac{1-r_0}{q}\mathbb{E}(\sqrt{XY})$. For seeing that $c > 0$ observe that

$$\begin{aligned}
\mathbb{E}(\sqrt{XY}) &= \frac{1}{1-r_0}\frac{\Gamma(q+\frac{1}{2})^2}{\Gamma(q)^2}\ {}_2F_1(-\frac{1}{2},-\frac{1}{2};q;r_0)\\
&\leq \frac{1}{1-r_0}\frac{\Gamma(q+\frac{1}{2})^2}{\Gamma(q)^2}\ {}_2F_1(-\frac{1}{2},-\frac{1}{2};q;1)\\
&= \frac{q}{1-r_0}
\end{aligned}$$

Finaly $g_{r_0}(r_0) = 0$. Possibly g_{r_0} is convex on $[0, 1]$, but studying g''_{r_0} is quite difficult. It may be not true when r_0 is close to 1. However all simulations show that $g_{r_0}(r) > 0$ if $0 < r < r_0$ and $g_{r_0}(r) < 0$ if $r_0 < r < 1$. In this case r_0 is the only solution of $g_{r_0}(r) = 0$.