

Addendum: Bayesian estimation in Kibble's bivariate gamma distribution

George Iliopoulos, Dimitris Karlis and Ioannis Ntzoufras

Key words and phrases: Downton's bivariate exponential distribution; Kibble's bivariate gamma distribution; Markov chain Monte Carlo; regenerative simulation; reversible jump.

Abstract:

The paper describes Bayesian estimation for the parameters of Kibble's (1941) bivariate gamma distribution. The density of this distribution can be written as a mixture, allowing for a simple data augmentation scheme. An MCMC algorithm is constructed to facilitate Bayesian estimation. We show that the resulting chain is geometrically ergodic and thus a regenerative sampling procedure is applicable allowing for estimation of ergodic means' standard errors. Bayesian hypothesis testing procedures are developed to test both the dependence hypothesis of the two variables as well as the hypothesis that their means are equal. A reversible jump MCMC algorithm is proposed to carry out this model selection problem. Real and simulated datasets are used to illustrate the proposed methodology.

7. ADDENDUM

In this supplement we additional Tables and Figures for the examples of our paper entitled 'Bayesian estimation in Kibble's bivariate gamma distribution'.

List of Tables

1	Model Diagnostics for all datasets	2
2	Quantile Diagnostics for all datasets	3
3	Summary of Badly Fitted Quantiles for all datasets	4
4	Percentage of extreme observations	5

List of Figures

1	Ergodic evolution of log-Bayes factors of model m_4 vs. m_i ($i = 1, 2, 3$) for simulated examples 1-6.	6
---	---	---

Sim.	Data	p_1^*	p_2^*	p_3^*	p_4^*	
1	1	0.295	0.195	0.165	0.415	✓
	2	0.275	0.195	0.400	0.407	✓
	3	0.390	0.282	0.166	0.419	✓
	4	0.388	0.284	0.392	0.406	✓
2	1	0.259	0.338	0.472	0.160	✓
	2	0.250	0.328	0.238	0.168	✓
	3	0.456	0.465	0.473	0.164	✓
	4	0.454	0.472	0.222	0.163	✓
3	1	0.057	0.003	0.004	0.248	X
	2	0.054	0.002	0.110	0.245	X
	3	0.396	0.134	0.003	0.234	X
	4	0.373	0.115	0.217	0.266	✓
4	1	0.472	0.267	0.000	0.256	X
	2	0.438	0.271	0.352	0.435	✓
	3	0.296	0.461	0.000	0.251	X
	4	0.293	0.456	0.391	0.457	✓
5	1	0.175	0.109	0.000	0.404	X
	2	0.080	0.092	0.237	0.319	X
	3	0.439	0.318	0.000	0.402	X
	4	0.343	0.344	0.286	0.317	✓
6	1	0.159	0.119	0.000	0.229	X
	2	0.108	0.062	0.046	0.238	X
	3	0.405	0.415	0.000	0.220	X
	4	0.345	0.399	0.242	0.262	✓
Ex2	1	0.373	0.016	0.000	0.041	X
	2	0.077	0.008	0.002	0.259	X
	3	0.300	0.164	0.000	0.040	X
	4	0.453	0.186	0.259	0.244	✓

Table 1: Model Diagnostics for all datasets; (10000 iterations and additional 1000 iterations as burn-in); $p_j^* = \min(p_i, 1 - p_i)$ for $j = 1, 2, 3, 4$, last column again indicates badly fitted model: ✓ if all p-values > 0.10, X otherwise.

Data	Model	Var.	d							
			0.01	0.05	0.25	0.50	0.75	0.95	0.99	
1	1	x	0.116	0.093	0.448	0.486	0.483	0.255	0.266	
		y	0.439	0.347	0.455	0.124	0.468	0.148	0.324	
	2	x	0.116	0.091	0.425	0.493	0.483	0.254	0.262	
		y	0.434	0.340	0.461	0.118	0.480	0.149	0.325	
	3	x	0.107	0.081	0.370	0.424	0.413	0.204	0.334	
		y	0.445	0.379	0.385	0.094	0.374	0.233	0.383	
	4	x	0.102	0.076	0.369	0.414	0.407	0.198	0.333	
		y	0.446	0.377	0.385	0.093	0.373	0.238	0.401	
	2	1	x	0.251	0.165	0.123	0.488	0.161	0.068	0.269
			y	0.146	0.104	0.290	0.424	0.220	0.123	0.222
		2	x	0.247	0.159	0.122	0.486	0.151	0.059	0.266
			y	0.141	0.099	0.276	0.434	0.214	0.114	0.217
3		x	0.226	0.119	0.226	0.306	0.354	0.179	0.403	
		y	0.131	0.085	0.190	0.390	0.428	0.269	0.134	
4		x	0.216	0.112	0.223	0.302	0.345	0.171	0.395	
		y	0.129	0.083	0.185	0.378	0.422	0.260	0.136	
3		1	x	0.340	0.240	0.126	0.119	0.166	0.007	0.226
			y	0.060	0.085	0.248	0.200	0.115	0.003	0.008
		2	x	0.360	0.223	0.120	0.107	0.146	0.005	0.228
			y	0.058	0.081	0.231	0.178	0.090	0.001	0.006
	3	x	0.294	0.139	0.428	0.473	0.363	0.115	0.481	
		y	0.076	0.169	0.351	0.293	0.350	0.117	0.118	
	4	x	0.292	0.140	0.442	0.466	0.341	0.094	0.498	
		y	0.075	0.158	0.354	0.281	0.341	0.100	0.111	
	4	1	x	0.370	0.355	0.329	0.327	0.104	0.452	0.132
			y	0.047	0.461	0.309	0.356	0.157	0.139	0.355
		2	x	0.369	0.367	0.341	0.234	0.054	0.396	0.140
			y	0.044	0.432	0.316	0.259	0.057	0.082	0.458
3		x	0.354	0.418	0.464	0.206	0.262	0.381	0.077	
		y	0.052	0.475	0.194	0.447	0.346	0.295	0.483	
4		x	0.349	0.454	0.484	0.100	0.224	0.354	0.069	
		y	0.048	0.496	0.181	0.498	0.226	0.244	0.378	
5		1	x	0.109	0.227	0.069	0.357	0.413	0.136	0.395
			y	0.082	0.258	0.421	0.481	0.473	0.133	0.273
		2	x	0.112	0.209	0.060	0.367	0.427	0.071	0.301
			y	0.082	0.242	0.390	0.482	0.437	0.122	0.327
	3	x	0.125	0.146	0.191	0.359	0.283	0.371	0.411	
		y	0.078	0.353	0.379	0.253	0.263	0.341	0.438	
	4	x	0.121	0.132	0.174	0.289	0.197	0.287	0.480	
		y	0.070	0.333	0.392	0.229	0.218	0.363	0.469	
	6	1	x	0.009	0.054	0.009	0.235	0.131	0.044	0.410
			y	0.042	0.010	0.045	0.048	0.078	0.113	0.133
		2	x	0.008	0.045	0.001	0.175	0.080	0.025	0.407
			y	0.038	0.009	0.033	0.025	0.031	0.050	0.099
3		x	0.022	0.135	0.099	0.326	0.420	0.308	0.274	
		y	0.049	0.022	0.270	0.416	0.423	0.400	0.477	
4		x	0.021	0.130	0.056	0.279	0.363	0.303	0.249	
		y	0.053	0.022	0.232	0.390	0.410	0.411	0.494	
Ex2		1	x	0.425	0.488	0.123	0.011	0.047	0.079	0.072
			y	0.119	0.128	0.496	0.243	0.017	0.086	0.036
		2	x	0.491	0.454	0.066	0.000	0.004	0.009	0.380
			y	0.128	0.091	0.401	0.071	0.001	0.069	0.230
	3	x	0.268	0.213	0.403	0.282	0.497	0.473	0.022	
		y	0.237	0.386	0.113	0.224	0.319	0.423	0.112	
	4	x	0.397	0.234	0.275	0.224	0.448	0.323	0.124	
		y	0.354	0.459	0.016	0.144	0.144	0.484	0.351	

Table 2: Quantile Diagnostics for all datasets; Cell values stand for $p_{X_d}^* = \min(p_{X_d}, 1 - p_{X_d})$ and $p_{Y_d}^* = \min(p_{Y_d}, 1 - p_{Y_d})$; where $p_{X_d} = P(X_d^{rep} > X_d)$ and $p_{Y_d} = P(Y_d^{rep} > Y_d)$, X_d and Y_d are the d -quantiles of X and Y respectively while X_d^{rep} and Y_d^{rep} are the corresponding quantiles of the predictive distribution (10000 iterations and additional 1000 iterations as burn-in).

Data	Model	Number of quantiles with $p_{X_d}^*$ or $p_{Y_d}^*$		
		< 0.025	< 0.05	< 0.10
1	1	0	0	1
	2	0	0	1
	3	0	0	2
	4	0	0	1
2	1	0	0	1
	2	0	0	2
	3	0	0	1
	4	0	0	1
3	1	3	3	5
	2	3	3	6
	3	0	0	1
	4	0	0	2
4	1	0	0	1
	2	0	1	4
	3	0	0	2
	4	0	1	2
5	1	0	0	2
	2	0	0	3
	3	0	0	1
	4	0	0	1
6	1	3	7	9
	2	5	10	12
	3	2	3	4
	4	2	2	4
ex2	1	2	3	7
	2	5	5	7
	3	1	1	1
	4	1	1	1

Table 3: Summary of Badly Fitted Quantiles for all datasets; (10000 iterations and additional 1000 iterations as burn-in).

Data	Model (m)	Number of values with $PMEO_{X_i}^*$ or $PMEO_{Y_i}^*$		
		< 0.01	< 0.025	< 0.05
1	1	0.015	0.045	0.090
	2	0.005	0.025	0.075
	3	0.010	0.055	0.095
	4	0.010	0.025	0.080
2	1	0.010	0.075	0.110
	2	0.010	0.040	0.095
	3	0.010	0.060	0.110
	4	0.010	0.035	0.085
3	1	0.050	0.070	0.095
	2	0.025	0.050	0.085
	3	0.015	0.065	0.090
	4	0.015	0.045	0.065
4	1	0.010	0.040	0.105
	2	0.005	0.010	0.020
	3	0.015	0.035	0.085
	4	0.005	0.015	0.020
5	1	0.030	0.055	0.090
	2	0.020	0.025	0.035
	3	0.035	0.060	0.090
	4	0.015	0.025	0.030
6	1	0.020	0.075	0.115
	2	0.010	0.025	0.045
	3	0.020	0.055	0.125
	4	0.005	0.010	0.045
EX2	1	0.030	0.045	0.085
	2	0.005	0.010	0.015
	3	0.020	0.040	0.080
	4	0.000	0.010	0.020

Table 4: Percentage of extreme observations; $PMEO_{X_i}$, $PMEO_{Y_i}$: probabilities of more extreme observation for X_i and Y_i respectively ($i = 1, \dots, n$); $PMEO_{X_i}^* = \min(PMEO_{X_i}, 1 - PME O_{X_i})$, $PMEO_{Y_i}^* = \min(PMEO_{Y_i}, 1 - PME O_{Y_i})$; $PMEO_{X_i} = P(X_i^{rep} < X_i)$ and $PMEO_{Y_i} = P(Y_i^{rep} < Y_i)$, (10000 iterations and additional 1000 iterations as burn-in).

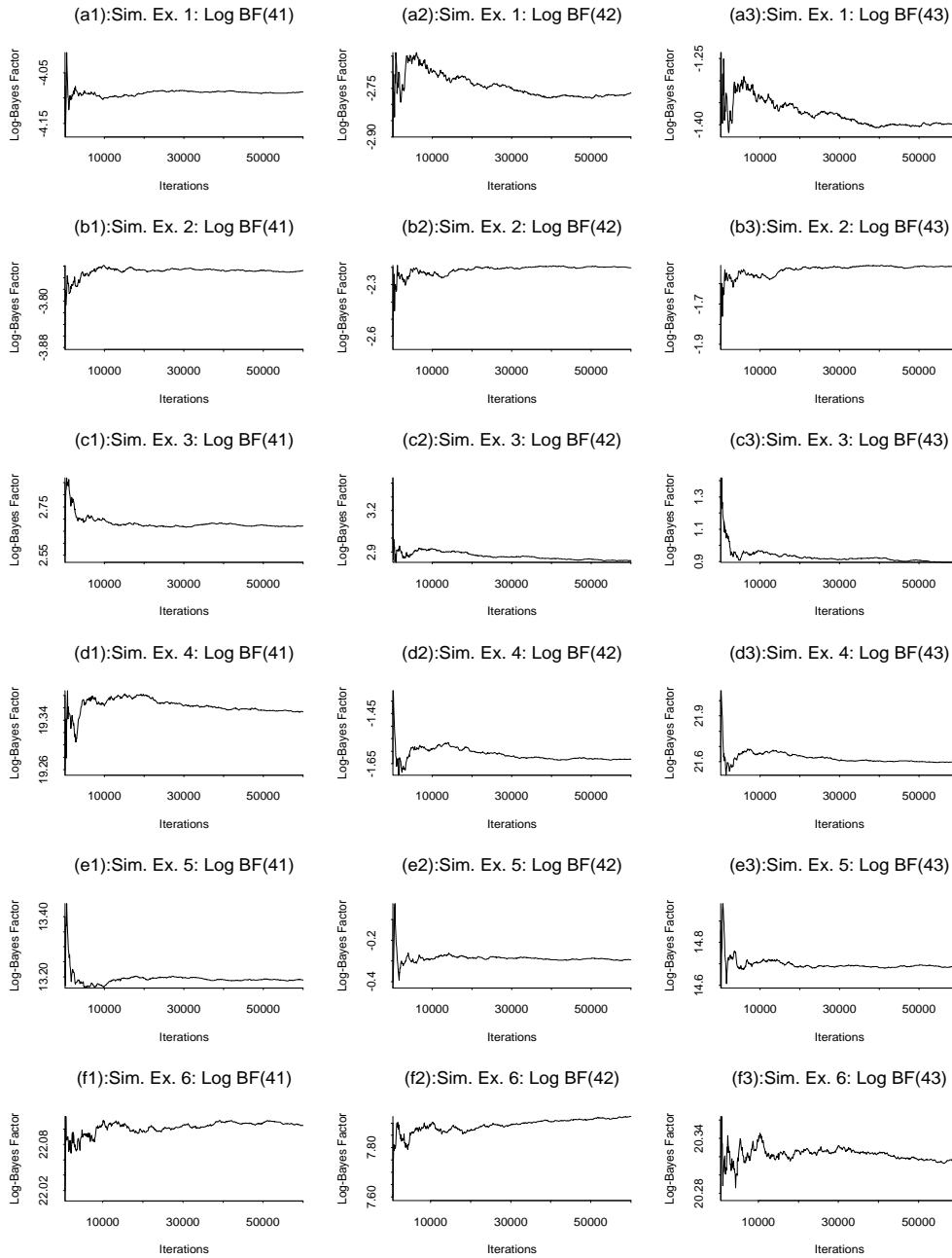


Figure 1: Ergodic evolution of log-Bayes factors of model m_4 vs. m_i ($i = 1, 2, 3$) for simulated examples 1–6. The i -th column corresponds to model m_i , $i = 1, 2, 3$

Received March 2004
Accepted September 2004

George Iliopoulos: geh@unipi.gr
Department of Statistics and Insurance Science
University of Pireaus
80 Karaoli and Dimitriou Str.
18534 Pireaus
GREECE

Dimitris Karlis: karlis@aueb.gr
Department of Statistics
Athens University of Economics
76 Patission Str, 104343 Athens
GREECE

Ioannis Ntzoufras: ntzoufras@aueb.gr
Department of Statistics
Athens University of Economics
76 Patission Str, 104343 Athens
GREECE