Parameter estimation for a bivariate beta distribution with arbitrary beta marginals and positive correlation

Susanne Trick[1,2] · Constantin A. Rothkopf[1,2,3] · Frank Jäkel[1,2]

Abstract

We discuss a bivariate beta distribution that can model arbitrary beta-distributed marginals with a positive correlation. The distribution is constructed from six independent gamma-distributed random variates. While previous work used an approximate and sometimes inaccurate method to compute the distribution's covariance and estimate its parameters, here, we derive all product moments and the exact covariance, which can be computed numerically. Based on this analysis we present an algorithm for estimating the parameters of the distribution using moment matching. We evaluate this inference method in a simulation study and demonstrate its practical use on a data set consisting of predictions from two correlated forecasters. Furthermore, we generalize the bivariate beta distribution to a correlated Dirichlet distribution, for which the proposed parameter estimation method can be used analogously.

Keywords Bivariate beta distribution · Correlated beta distribution · Covariance · Moment matching

1 Introduction

Probabilistic forecasts are important in many domains, among them finance and economics, business and marketing, politics, public health, engineering, and meteorological, ecological,

---

Constantin A. Rothkopf and Frank Jäkel contributed equally to this work

---

✉ Susanne Trick
  susanne.trick@tu-darmstadt.de

  Constantin A. Rothkopf
  constantin.rothkopf@cogsci.tu-darmstadt.de

  Frank Jäkel
  jaekel@psychologie.tu-darmstadt.de

1   Centre for Cognitive Science, Technical University of Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany

2   Institute of Psychology, Technical University of Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany

3   Frankfurt Institute for Advanced Studies, Goethe University, Ruth-Moufang-Str. 1, 60438 Frankfurt, Germany

and environmental science. For binary events these forecasts take the form of probability estimates that are either provided by humans or machine learning algorithms. In order to model these probability estimates one can use an arbitrary distribution on the interval [0, 1], such as the beta distribution, beta-generated distributions, the Kumaraswamy distribution, or any distribution on the real numbers transformed through the logistic function. In many cases the probabilities will be correlated, e.g. if different experts forecast the outcome of an election or the probability of rain. In order to be able to model such correlated probabilities, therefore, a bivariate distribution is needed. For example, one can use bivariate generalizations of the Kumaraswamy distribution [1], bivariate beta-generated distributions [2], or a multivariate Gaussian with logistic transformations [3]. However, the most common choice for modeling such probabilities is the beta distribution, since it is the standard distribution for probabilities in Bayesian statistics and is simply more familiar to practitioners than the other distributions [4, 5]. Therefore, in this paper we focus on bivariate beta distributions.

While a multitude of constructions for bivariate beta distributions have been proposed, they have different constraints and properties, which limit their applicability. We will first review previous constructions of bivariate beta distributions together with their respective properties and then examine the most promising construction that can model arbitrary beta marginals with a positive correlation [5]. For this construction of modeling arbitrary beta marginals with positive correlation, so far, there has not been an exact method for parameter inference and it has thus rarely been used. Here, we therefore introduce a new estimation method for this bivariate beta distribution with arbitrary beta marginals and positive correlation.

Many bivariate beta distributions have been proposed in the literature [1, 2, 5–21]. Among them, some approaches have been derived from general families of bivariate distributions, such as the Farlie-Gumbel-Morgenstern family of distributions [e.g. 7] and the Sarmanov family of distributions [e.g. 8]. Others derived a bivariate beta distribution from a bivariate extension of the F distribution [9, 13], whereas Nadarajah and Kotz [11] proposed different bivariate beta distributions constructed from products of univariate beta-distributed random variables. In general, using different copulas one can construct different bivariate distributions with the same beta marginals [2, 20]. However, as beta-distributed random variables can easily be constructed from normalized gamma-distributed random variables, it is natural to try and generalize this construction to the bivariate case. In this vein, several authors have introduced correlations through shared gamma-distributed random variables [1, 5, 10, 14, 18]. The most straightforward case of this construction has been studied by Olkin and Liu [10], building on the work of Libby and Novick [6]. They use three independent gamma-distributed random variables U_1, U_2, U_3 with respective shape parameters v_1, v_2, v_3 and same scale parameter to construct

$$X' = \frac{U_1}{U_1 + U_3} \quad \text{and} \quad Y' = \frac{U_2}{U_2 + U_3}. \tag{1}$$

Using the standard construction of beta variates from gamma variates, the joint distribution of the random variables X' and Y' is a bivariate beta distribution with marginal distributions Beta(v_1, v_3) for X' and Beta(v_2, v_3) for Y'. The correlation between X' and Y', which is obtained through the shared latent variable U_3 and its parameter v_3, is in the range [0,1]. For high values of v_3, the correlation tends to 0 whereas for low values of v_3, it tends to 1. However, if v_3 is high, the values of X' and Y' also tend to 0 and if v_3 is low they tend to 1 accordingly. This behavior severely limits the usefulness of the distribution for most applications. A further limitation is the constraint that the marginal distributions share the same second parameter v_3. Thus, the bivariate beta distribution proposed by Olkin and Liu

does not allow for arbitrary beta marginals, which limits its flexibility in modeling probability forecasts.

Arnold and Ng [14] proposed a more flexible construction for a bivariate beta distribution. They use five independent gamma-distributed random variables $U_1, \ldots, U_5$ with shape parameters $v_1, \ldots, v_5$ and scale parameter 1 to define two correlated random variables

$$X = \frac{U_1 + U_3}{U_1 + U_3 + U_4 + U_5} \quad \text{and} \quad Y = \frac{U_2 + U_4}{U_2 + U_3 + U_4 + U_5} \tag{2}$$

with marginal distributions Beta($v_1 + v_3, v_4 + v_5$) for X and Beta($v_2 + v_4, v_3 + v_5$) for Y. Compared to Olkin and Liu [10], this construction of a bivariate beta distribution can generate all correlations in the range [-1,1] and marginal distributions with differing second parameters. Nevertheless, because of how the two marginals share parameters, not all combinations of parameters of the marginal beta distributions are possible. For example, the marginals Beta(10, 4) for X and Beta(1, 1) for Y cannot be obtained.

Olkin and Trikalinos [18] base their construction of a bivariate beta distribution on the Dirichlet distribution. $U = (U_{00}, U_{01}, U_{10}, U_{11})$ is drawn from a 4-dimensional Dirichlet distribution with parameters $v_1, \ldots, v_4$. By just using three of its components, U_{00}, U_{01}, U_{10}, new random variables

$$X = U_{00} + U_{10} \quad \text{and} \quad Y = U_{01} + U_{10} \tag{3}$$

are constructed, with marginal distributions Beta($v_1 + v_3, v_2 + v_4$) for X and Beta($v_2 + v_3, v_1 + v_4$) for Y. As Dirichlet-distributed random variables can also be constructed from gamma random variables, we can equivalently construct X and Y in Eq. (3) from four independent gamma-distributed random variables $U_1, \ldots, U_4$ with shape parameters $v_1, \ldots, v_4$ and equal scale parameter 1, with

$$X = \frac{U_1 + U_3}{U_1 + U_2 + U_3 + U_4} \quad \text{and} \quad Y = \frac{U_2 + U_3}{U_1 + U_2 + U_3 + U_4}. \tag{4}$$

As can easily be seen from this construction, all correlations in the range [-1,1] can be generated. In particular, the correlation tends to -1 if v_3 and v_4 tend to 0 and U_3 and U_4 will be negligible compared to U_1 and U_2. In this case $X \approx \frac{U_1}{U_1 + U_2} \approx 1 - Y$. Similarly, the higher the values of v_3 and v_4 relative to v_1 and v_2, the more negligible U_1 and U_2 will be and the correlation increases to 1 until $X \approx \frac{U_3}{U_3 + U_4} \approx Y$. Less obviously, a correlation of 0 is obtained in case $v_1 \cdot v_2 = v_3 \cdot v_4$ [18]. Still, this construction of a bivariate beta distribution does not allow arbitrary beta marginal distributions. Since all v_i are constrained to be positive, for some combinations of marginal distributions the resulting system of linear equations for the parameters v_i has no solution. For example, the two marginals Beta(2, 2) for X and Beta(1, 1) for Y cannot be generated, regardless of their correlation.

Magnussen [5] introduced yet another construction based on six gamma variates. While all the constructions thus far constrain the parameters of the marginal beta distributions, this construction does allow for arbitrary beta marginals with positive correlation, thus providing the necessary flexibility to model probability forecasts. Magnussen's distribution is a special case of a more general 8-parameter bivariate beta distribution introduced by Arnold and Ng [14] and reviewed in Arnold and Ghosh [1], which even allows for positive and negative correlations. However, in many applications, it is enough to model positive correlations, for which the less complex 6-parameter distribution is sufficient. For example, if X and Y are probability estimates elicited from two skilled forecasters, we do not expect negative correlations. But we do want to allow for the possibility that their marginal forecasts have different distributions that should not be tied together by parameter constraints on the marginals.

Hence, the bivariate beta distribution proposed by Magnussen [5], which can model arbitrary beta-distributed marginals with a positive correlation, is an appropriate distribution for modeling correlated probability forecasts.

However, just like any other distribution, the bivariate beta distribution can only be used if its parameters can be estimated correctly. While Magnussen [5] proposes a moment matching approach for fitting the distribution's parameters, this approach relies on a rough and sometimes inaccurate approximation for the covariance. Also, Magnussen [5] did not discuss the fact that very similar data can be generated with different parameter values, which makes it hard to statistically infer the parameters of the bivariate beta distribution from data without constraining the distribution.

Therefore, in this work we introduce an alternative approach for estimating this bivariate beta distribution's parameters. First, we will derive the full joint distribution, which is missing in the work of Magnussen [5], probably because it is intractable. We will then clarify the relationship between Magnussen's distribution and the Olkin-Liu distribution [10]. Using this relationship with the Olkin-Liu distribution we derive all product moments and in particular the exact covariance function (and in passing we correct a small mistake in the product moments from Olkin and Liu [10]). For parameter inference, we propose to match moments numerically using the exact covariance we derived. While other estimation methods such as Bayesian inference could be used [22], here we focus on moment matching due to its simplicity and efficiency. In order to make parameter inference unambiguous, we additionally show how to reasonably constrain the distribution's parameters. We evaluate the proposed parameter estimation method in a simulation study and demonstrate its practical use on a real data set consisting of predictions from two correlated forecasters. We discuss the relationship between the distribution's parameters and the correlation and then extend the bivariate beta distribution to a correlated Dirichlet distribution, for which the proposed parameter estimation method can be used analogously.

The remainder of the paper is structured as follows. In Sect. 2 we discuss the bivariate beta distribution with arbitrary beta marginals, including its joint distribution in Sect. 2.1, its moments in Sect. 2.2, correlation and covariance in Sect. 2.3, and parameter inference in Sect. 2.4. Section 3 shows how to generalize the bivariate beta distribution to a correlated Dirichlet distribution.

2 A bivariate beta distribution with arbitrary beta marginals

Magnussen [5] uses six independent gamma-distributed random variables A_1, A_2, B_1, B_2, D_1, D_2 that are distributed according to

$$
\begin{aligned}
A_i &\sim \text{Gamma}(\alpha_i, 1) \quad i = 1, 2 \\
B_i &\sim \text{Gamma}(\beta_i, 1) \quad i = 1, 2 \\
D_i &\sim \text{Gamma}(\delta_i, 1) \quad i = 1, 2,
\end{aligned}
\tag{5}
$$

to construct two bivariate-beta-distributed random variables

$$
X = \frac{A_1 + D_1}{A_1 + A_2 + D_1 + D_2} \quad \text{and} \quad Y = \frac{B_1 + D_1}{B_1 + B_2 + D_1 + D_2}.
\tag{6}
$$

The resulting marginal distributions of X and Y are Beta(a_1, a_2) and Beta(b_1, b_2) with

$$
a_1 = \alpha_1 + \delta_1 \qquad a_2 = \alpha_2 + \delta_2 \qquad b_1 = \beta_1 + \delta_1 \qquad b_2 = \beta_2 + \delta_2.
\tag{7}
$$

The marginals follow immediately from the definition because the sum of gamma random variables of the same scale is gamma-distributed with the same scale but with the original shape parameters summed. In contrast to other constructions that were discussed above [10, 14, 18], this construction allows for arbitrary marginal distributions. In particular, when δ_1 and δ_2 tend to zero, we can model arbitrary independent marginal distributions $\text{Beta}(\alpha_1, \alpha_2)$ and $\text{Beta}(\beta_1, \beta_2)$.

Since all parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \delta_1, \delta_2$ need to be positive by definition, for fixed marginal distributions $\text{Beta}(a_1, a_2)$ for X and $\text{Beta}(b_1, b_2)$ for Y it must hold that $\delta_1 < \delta_1^{\max} = \min(a_1, b_1)$ and $\delta_2 < \delta_2^{\max} = \min(a_2, b_2)$. Therefore, for most marginal distributions the maximum correlation that can be generated is below 1. The higher the difference between two marginal distributions, the lower the possible maximum correlation. A perfect correlation approaching 1 can, of course, only be generated for equal marginal distributions, i.e. if $a_1 = b_1$ and $a_2 = b_2$ and $\alpha_1, \alpha_2, \beta_1,$ and β_2 tend to 0, as also noted by Magnussen [5]. Note that this limitation applies to other bivariate distributions that do not allow for arbitrary marginal beta distributions as well [e.g. 18].

The construction of this bivariate beta distribution can also be seen as a pairwise combination of three beta distributions. First transform the six independent gamma-distributed random variables (5) into three independent gamma- and three independent beta-distributed random variables,

$$
\begin{aligned}
U_1 &= A_1 + A_2, & U_1 &\sim \text{Gamma}(\upsilon_1, 1) \\
U_2 &= B_1 + B_2, & U_2 &\sim \text{Gamma}(\upsilon_2, 1) \\
U_3 &= D_1 + D_2, & U_3 &\sim \text{Gamma}(\upsilon_3, 1) \\
W_1 &= \frac{A_1}{A_1 + A_2}, & W_1 &\sim \text{Beta}(\alpha_1, \alpha_2) \\
W_2 &= \frac{B_1}{B_1 + B_2}, & W_2 &\sim \text{Beta}(\beta_1, \beta_2) \\
W_3 &= \frac{D_1}{D_1 + D_2}, & W_3 &\sim \text{Beta}(\delta_1, \delta_2)
\end{aligned}
\tag{8}
$$

with

$$
\upsilon_1 = \alpha_1 + \alpha_2 \qquad \upsilon_2 = \beta_1 + \beta_2 \qquad \upsilon_3 = \delta_1 + \delta_2.
\tag{9}
$$

With these definitions we can then rewrite construction (6) as

$$
\begin{aligned}
X &= \frac{U_1}{U_1 + U_3} \cdot W_1 + \frac{U_3}{U_1 + U_3} \cdot W_3 = X'W_1 + (1 - X')W_3 \\
Y &= \frac{U_2}{U_2 + U_3} \cdot W_2 + \frac{U_3}{U_2 + U_3} \cdot W_3 = Y'W_2 + (1 - Y')W_3,
\end{aligned}
\tag{10}
$$

where X' and Y' are defined as in (1) but with $\upsilon_1, \upsilon_2,$ and υ_3 as in (9). Furthermore, X' and Y' are independent of W_1, W_2, W_3. If parameters δ_1 and δ_2 and with them U_3 tend to 0, $X \approx W_1$ and $Y \approx W_2$ are independent with marginal distributions $\text{Beta}(\alpha_1, \alpha_2)$ for X and $\text{Beta}(\beta_1, \beta_2)$ for Y. Mixing in the shared component W_3 by increasing the values of parameters δ_1 and δ_2 increases the correlation between X and Y. If U_1 and U_2 are negligible compared to U_3 because δ_1 and δ_2 dominate the parameters, the correlation will be close to 1 with $X \approx W_3 \approx Y$ and hence X and Y have the same marginal distribution $\text{Beta}(\delta_1, \delta_2)$, as mentioned before.

2.1 Joint distribution

X and Y in (10) are linear transformations of X' and Y'. Given W_1, W_2, W_3 it is easy to recover X' and Y' from observed X and Y,

$$X' = \frac{X - W_3}{W_1 - W_3} = \frac{|X - W_3|}{|W_1 - W_3|} = f_1(X)$$

$$Y' = \frac{Y - W_3}{W_2 - W_3} = \frac{|Y - W_3|}{|W_2 - W_3|} = f_2(Y). \qquad (11)$$

Note that we can ignore the sign because according to (10) X is always between W_1 and W_3 and Y between W_2 and W_3, so that the numerator and denominator always have the same sign.

As X' and Y' jointly follow the Olkin-Liu distribution [10],

$$p'(x', y') = \frac{(x')^{\upsilon_1 - 1} (1 - x')^{\upsilon_2 + \upsilon_3 - 1} (y')^{\upsilon_2 - 1} (1 - y')^{\upsilon_1 + \upsilon_3 - 1}}{B(\upsilon_1, \upsilon_2, \upsilon_3) (1 - x'y')^{\upsilon_1 + \upsilon_2 + \upsilon_3}}, \qquad (12)$$

where $B(\upsilon_1, \upsilon_2, \upsilon_3) = \frac{\Gamma(\upsilon_1)\Gamma(\upsilon_2)\Gamma(\upsilon_3)}{\Gamma(\upsilon_1 + \upsilon_2 + \upsilon_3)}$, the joint distribution of X and Y given W_1, W_2, W_3 is

$$p(x, y \mid w_1, w_2, w_3) = \left| \frac{df_1(x)}{dx} \frac{df_2(y)}{dy} \right| p'(f_1(x), f_2(y))$$

$$= \frac{1}{|w_1 - w_3||w_2 - w_3|}$$

$$\frac{\left(\frac{|x - w_3|}{|w_1 - w_3|}\right)^{\upsilon_1 - 1} \left(1 - \frac{|x - w_3|}{|w_1 - w_3|}\right)^{\upsilon_2 + \upsilon_3 - 1} \left(\frac{|y - w_3|}{|w_2 - w_3|}\right)^{\upsilon_2 - 1} \left(1 - \frac{|y - w_3|}{|w_2 - w_3|}\right)^{\upsilon_1 + \upsilon_3 - 1}}{B(\upsilon_1, \upsilon_2, \upsilon_3) \left(1 - \frac{|x - w_3|}{|w_1 - w_3|} \frac{|y - w_3|}{|w_2 - w_3|}\right)^{\upsilon_1 + \upsilon_2 + \upsilon_3}} \qquad (13)$$

$$= \frac{|w_1 - w_3||w_2 - w_3|}{B(\upsilon_1, \upsilon_2, \upsilon_3)}$$

$$\cdot \frac{|x - w_3|^{\upsilon_1 - 1}|x - w_1|^{\upsilon_2 + \upsilon_3 - 1}|y - w_3|^{\upsilon_2 - 1}|y - w_2|^{\upsilon_1 + \upsilon_3 - 1}}{(|w_1 - w_3||w_2 - w_3| - |x - w_3||y - w_3|)^{\upsilon_1 + \upsilon_2 + \upsilon_3}}$$

with x between w_1 and w_3 and y between w_2 and w_3 according to (10). We have not been able to integrate out w_1, w_2, w_3 from their joint distribution with x and y. However, we suspect that even if the joint density for X and Y could be expressed in terms of special functions, computing those might not be efficient enough for parameter inference for which we will resort to moment matching. Example plots with smoothed samples for the joint density are shown in Fig. 1 for several parameter settings showing different marginal distributions for X and Y and different correlations between X and Y. Sampling from the bivariate beta distribution is realized with JAGS [23].

2.2 Moments

As the marginal distributions for X and Y are beta-distributed, their moments are readily available, even in closed form. Computation of the product moments $E(X^k Y^l)$ is more challenging but can be realized with help of the work of Olkin and Liu [10]. Looking at construction (10), X and Y are a linear combination of independent beta-distributed random variables W_1, W_2, W_3 with weights X' and Y'. Thus, we can express the product moments
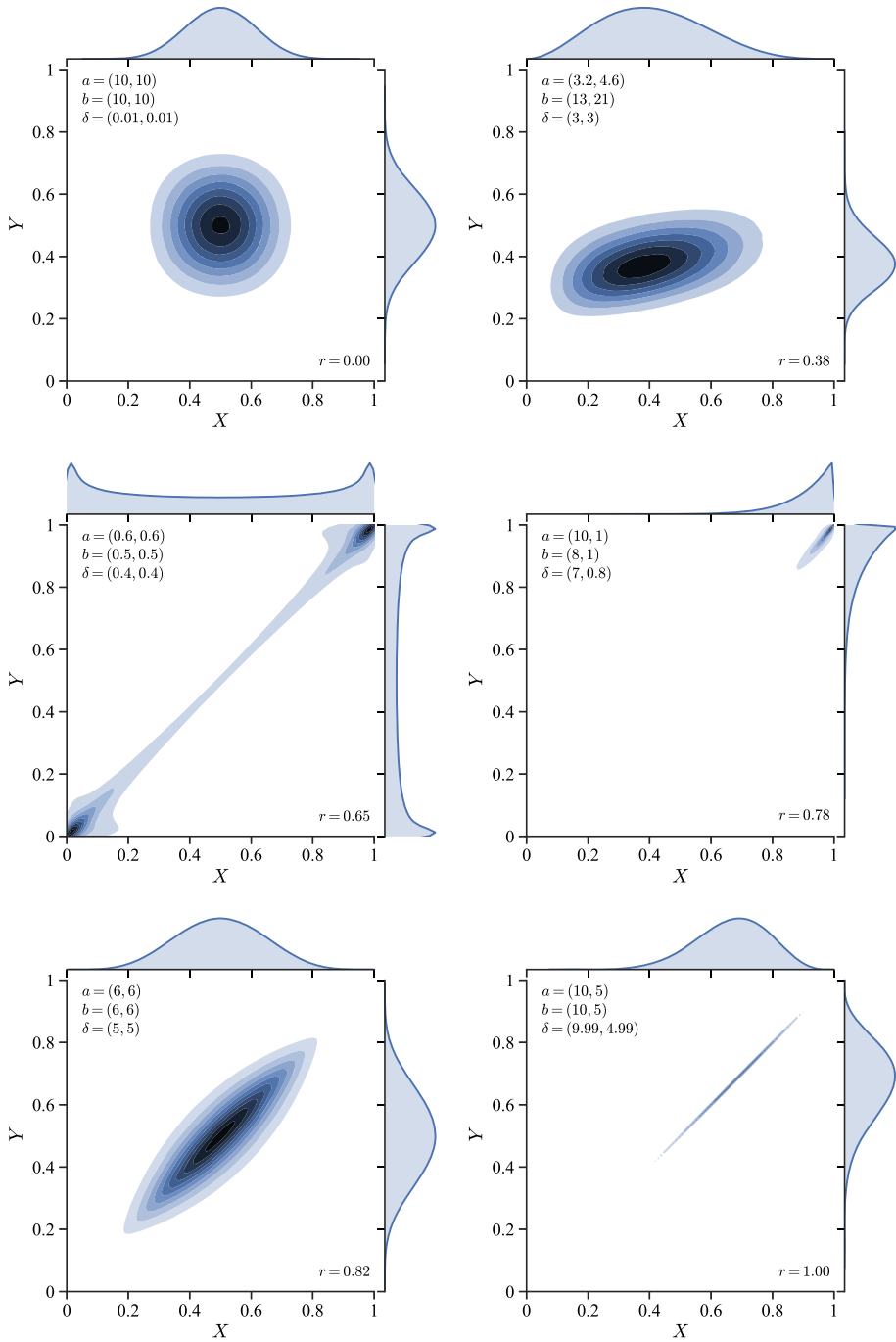
Fig. 1 Joint densities of bivariate beta distributions with selected parameter values. The plots were created with kernel density estimation based on 10 million samples of the respective distributions. Note that the smoothing is inaccurate at the borders and produces artifacts close to zero and one as a consequence of smoothing with a symmetric kernel

as

$$\mathrm{E}\left(X^k Y^l\right) = \mathrm{E}\left(\left(X' W_1 + (1 - X') W_3\right)^k \left(Y' W_2 + (1 - Y') W_3\right)^l\right). \tag{14}$$

Since X' and Y' are independent of W_1, W_2, W_3, it is possible to compute the expectation if the moments of W_1, W_2, W_3, X', Y' and the product moments of X' and Y' are known. W_1, W_2, W_3 as well as the marginals of X' and Y' are beta-distributed, so their moments can be computed straightforwardly in closed form. Furthermore, X' and Y' are jointly Olkin-Liu distributed according to (12), and Olkin and Liu [10] have shown how to compute their product moments. However, note that the derivation of $\mathrm{E}((X')^k (Y')^l)$ in equation (2.2) in the work of Olkin and Liu [10] is incorrect and should read

$$\mathrm{E}\left((X')^k (Y')^l\right) = \sum_{j=0}^{\infty} d A(j) \frac{B(v_1 + k + j, v_2 + v_3)}{B(v_1 + j, v_2 + v_3)} \frac{B(v_2 + l + j, v_1 + v_3)}{B(v_2 + j, v_1 + v_3)} \tag{15}$$

$$= \frac{\Gamma(v_1 + v_3)\Gamma(v_2 + v_3)\Gamma(v_1 + k)\Gamma(v_2 + l)\Gamma(\Upsilon)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)\Gamma(\Upsilon + k)\Gamma(\Upsilon + l)}$$

$$\sum_{j=0}^{\infty} \frac{(v_1 + k)_j (v_2 + l)_j (\Upsilon)_j}{(\Upsilon + k)_j (\Upsilon + l)_j} \frac{1}{j!} \tag{16}$$

$$= h \cdot {}_3F_2(v_1 + k, v_2 + l, \Upsilon; \Upsilon + k, \Upsilon + l; 1), \tag{17}$$

with

$$d = \frac{\Gamma(v_1 + v_3)\Gamma(v_2 + v_3)}{\Gamma(v_3)\Gamma(\Upsilon)} \tag{18}$$

$$A(j) = \frac{\Gamma(v_1 + j)}{\Gamma(v_1)} \frac{\Gamma(v_2 + j)}{\Gamma(v_2)} \frac{\Gamma(\Upsilon)}{\Gamma(\Upsilon + j)} \frac{1}{j!} \tag{19}$$

$$h = \frac{\Gamma(v_1 + v_3)\Gamma(v_2 + v_3)\Gamma(v_1 + k)\Gamma(v_2 + l)\Gamma(\Upsilon)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)\Gamma(\Upsilon + k)\Gamma(\Upsilon + l)}, \tag{20}$$

where $\Upsilon = v_1 + v_2 + v_3$ and ${}_pF_q$ is the generalized hypergeometric function. Equations (15), (18), and (19) are taken directly from Olkin and Liu [10] with $a = v_1, b = v_2, c = v_3$. Equations (16), (17), and (20) are our corrections of their equations.

2.3 Correlation and covariance

The correlation r between X and Y is

$$r = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{21}$$

with the known variances of the beta marginals

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{a_1 a_2}{(a_1 + a_2)^2 (a_1 + a_2 + 1)} \\
&= \frac{(\alpha_1 + \delta_1)(\alpha_2 + \delta_2)}{(\alpha_1 + \delta_1 + \alpha_2 + \delta_2)^2 (\alpha_1 + \delta_1 + \alpha_2 + \delta_2 + 1)} \\
\mathrm{Var}(Y) &= \frac{b_1 b_2}{(b_1 + b_2)^2 (b_1 + b_2 + 1)} \\
&= \frac{(\beta_1 + \delta_1)(\beta_2 + \delta_2)}{(\beta_1 + \delta_1 + \beta_2 + \delta_2)^2 (\beta_1 + \delta_1 + \beta_2 + \delta_2 + 1)}.
\end{aligned} \tag{22}$$

For the covariance of X and Y Magnussen [5] gives an approximate solution, namely

$$\text{Cov}(X, Y) \approx \frac{a_1 a_2 \delta_2 + (1 + b_1)(1 + b_2)\delta_1}{(a_1 + b_1)(a_2 + b_2)(1 + a_1 + b_1)(1 + a_2 + b_2)}, \quad (23)$$

where a_1, a_2, b_1, and b_2 are defined based on $\alpha_1, \alpha_2, \beta_1, \beta_2, \delta_1, \delta_2$ as in (7). This approximation is inaccurate for small values of these parameters, e.g. for $a_1 = a_2 = b_1 = b_2 = 4, \delta_1 = \delta_2 = 3$, the approximated covariance is $\text{Cov}(X, Y) = 0.024$, while the true covariance computed from 10^6 samples of the bivariate beta distribution is $\text{Cov}(X, Y) = 0.020$. This might seem like a small difference but it results in an overestimated correlation of $r = 0.854$ as opposed to the true correlation of $r = 0.730$. Even more worryingly, for $a_1 = a_2 = b_1 = b_2 = 1, \delta_1 = \delta_2 = \frac{4}{5}$, the approximated covariance is $\text{Cov}(X, Y) = \frac{1}{9}$, which results in an estimated correlation of $r = \frac{4}{3}$, which is greater than 1 and therefore wrong by definition.

Given the connection to the Olkin–Liu distribution [10], which we derived in Sect. 2.2, we therefore proceed to compute the exact covariance between X and Y:

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) \quad (24)$$

where

$$\begin{aligned}
\text{E}(X) &= \frac{a_1}{a_1 + a_2} = \frac{\alpha_1 + \delta_1}{\alpha_1 + \delta_1 + \alpha_2 + \delta_2} \\
\text{E}(Y) &= \frac{b_1}{b_1 + b_2} = \frac{\beta_1 + \delta_1}{\beta_1 + \delta_1 + \beta_2 + \delta_2}
\end{aligned} \quad (25)$$

are readily available as the means of the beta marginals. We can compute $\text{E}(XY)$ from (14) with $k = l = 1$, which results in

$$\begin{aligned}
\text{E}(XY) &= \text{E}\left((X'W_1 + (1 - X')W_3)(Y'W_2 + (1 - Y')W_3)\right) \\
&= \text{E}(X'Y')\text{E}(W_1)\text{E}(W_2) + \left(\text{E}(X') - \text{E}(X'Y')\right)\text{E}(W_1)\text{E}(W_3) \\
&\quad + \left(\text{E}(Y') - \text{E}(X'Y')\right)\text{E}(W_2)\text{E}(W_3) \\
&\quad + \left(1 - \text{E}(Y') - \text{E}(X') + \text{E}(X'Y')\right)\text{E}(W_3^2)
\end{aligned} \quad (26)$$

with the moments of the beta marginals from (8)

$$\begin{aligned}
\text{E}(W_1) &= \frac{\alpha_1}{\alpha_1 + \alpha_2} \\
\text{E}(W_2) &= \frac{\beta_1}{\beta_1 + \beta_2} \\
\text{E}(W_3) &= \frac{\delta_1}{\delta_1 + \delta_2} \\
\text{E}(W_3^2) &= \frac{\delta_1(\delta_1 + 1)}{(\delta_1 + \delta_2 + 1)(\delta_1 + \delta_2)}
\end{aligned} \quad (27)$$

and (10)

$$\begin{aligned}
\text{E}(X') &= \frac{\upsilon_1}{\upsilon_1 + \upsilon_3} \\
\text{E}(Y') &= \frac{\upsilon_2}{\upsilon_2 + \upsilon_3}
\end{aligned} \quad (28)$$

with υ_1, υ_2, and υ_3 as defined in (9). $E(X'Y')$ can be specialized from the moments (17) above with $k = l = 1$:

$$E(X'Y') = h \; {}_3F_2(\upsilon_1 + 1, \upsilon_2 + 1, \Upsilon; \Upsilon + 1, \Upsilon + 1; 1) \quad \text{with}$$
$$h = \frac{\Gamma(\upsilon_1 + \upsilon_3)\Gamma(\upsilon_2 + \upsilon_3)}{\Gamma(\upsilon_3)\Gamma(\Upsilon + 1)} \cdot \frac{\upsilon_1 \upsilon_2}{\Upsilon} \tag{29}$$

and $\Upsilon = \upsilon_1 + \upsilon_2 + \upsilon_3$ as before. According to Olkin and Liu [10] there is no closed-form solution for (17) and (29). However, using the generalized hypergeometric function the product moments and thus the covariance between X and Y can be computed numerically (e.g. by using the hyper function in sympy or the HypergeometricPFQ function in Mathematica). Note that with Raabe's test one can show that the generalized hypergeometric function ${}_3F_2$ in (29) converges. According to Raabe's test the series converges if

$$\lim_{j \to \infty} \rho_j > 1 \text{ with } \rho_j = j\left(\frac{c_j}{c_{j+1}} - 1\right), \tag{30}$$

where c_j is the j-th element of the series described by the generalized hypergeometric function ${}_3F_2$ in (29), which is

$$c_j = \frac{(\upsilon_1 + 1)_j(\upsilon_2 + 1)_j(\Upsilon)_j}{(\Upsilon + 1)_j(\Upsilon + 1)_j} \frac{1}{j!}. \tag{31}$$

Since $\lim_{j \to \infty} \rho_j = 1 + \upsilon_3 > 1$, the generalized hypergeometric function ${}_3F_2$ in (29) converges. However, this analysis also suggests that convergence of the series might be very slow for small $\upsilon_3 = \delta_1 + \delta_2$.

2.4 Parameter inference

Magnussen [5] used the method of moments to infer the parameters a_1, a_2, b_1, b_2 for the marginal distributions. Given these parameters he then matched the empirical correlation to the correlation for the parameters δ_1, δ_2 (given marginal parameters a_1, a_2, b_1, b_2) using the approximate solution for the covariance given in (23). However, as shown in Sect. 2.3 this approximation can lead to very inaccurate correlation estimates. An additional problem with the distribution is that very similar data can be generated with different parameter values, as one can see in Fig. 2(a) and (b). Increasing δ_1 and simultaneously decreasing δ_2 or vice versa while keeping the marginal parameters a_1, a_2 and b_1, b_2 fixed, can result in very similar correlations, which is shown in Fig. 2(c). Since two distributions with very different parameters can lead to extremely similar data, it is hard to statistically infer the parameters δ_1 and δ_2 from data: The empirical correlation alone does not provide enough constraints and differences in higher moments can be subtle.

Therefore, in order to make parameter inference unambiguous, we decided to constrain the 6-parameter bivariate beta distribution to five parameters: two for each marginal and one parameter to control the correlation. A reasonable way to constrain δ_1 and δ_2 is to set

$$\delta_2 = \frac{\delta_2^{\max}}{\delta_1^{\max}}\delta_1 \tag{32}$$

with $\delta_1^{\max} = \min(a_1, b_1)$, $\delta_2^{\max} = \min(a_2, b_2)$, because this enables the maximum possible correlation between X and Y when the maximum values for δ_1 and δ_2 are attained and the shared component between X and Y is as big as it can be without violating the marginal constraints.
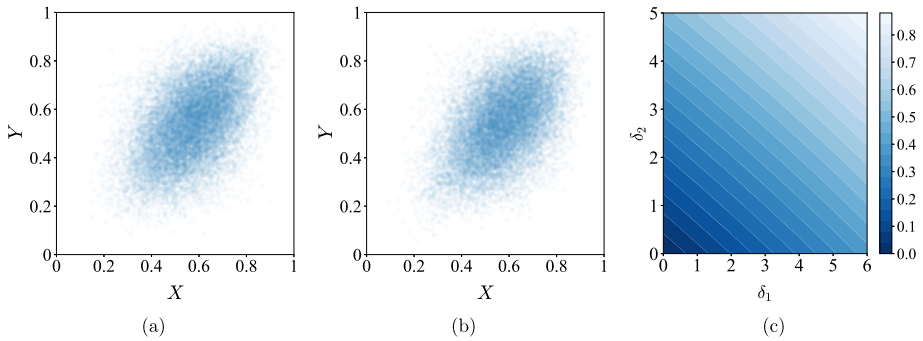
Fig. 2 Different parameters can generate similar data. (**a**) and (**b**) show samples generated from two bivariate beta distributions with the same marginal parameters $a = (8, 6)$, $b = (6, 5)$ and correlation parameters $\delta = (3.28, 2.73)$ for (**a**) and $\delta = (5, 1.6)$ for (**b**). Both, the data in (**a**) and (**b**) show a correlation of $r = 0.47$. Correspondingly, (**c**) shows the correlations generated by different combinations of δ_1 and δ_2 given marginal parameters $a = (8, 6)$ and $b = (6, 5)$. Different combinations of δ_1 and δ_2 can lead to very similar correlations

2.4.1 Moment matching

Using this constraint (32), the model-inherent constraints $\delta_1 < \delta_1^{\max}$, $\delta_2 < \delta_2^{\max}$, and the formula for the correlation derived in Sect. 2.3, we can now optimize the parameters numerically to match the empirical moments. First, the marginal parameters a_1, a_2, b_1, b_2 are obtained using the standard procedure of moment matching for the beta distribution. Given the estimated marginal parameters, an estimate of δ_1 (and with it δ_2) can be obtained numerically by minimizing the quadratic deviation between the theoretical correlation and the empirical correlation. To avoid the undefined cases $\delta_1 \leq 0$ and $\delta_1 \geq \delta_1^{\max}$ we bound the optimization between ε and $\delta_1^{\max} - \varepsilon$ with $\varepsilon = 0.001$. Unless the empirical correlation is bigger than the maximum correlation that can be attained with the matched marginals or smaller than 0, it can be matched exactly for some δ_1. Otherwise δ_1 will take on its maximum value $\delta_1^{\max} - \varepsilon$ or its minimal value ε. We implement inference in Python using the package mpmath [24] for the numerical computation of the generalized hypergeometric function and the package scipy [25] for optimization.

As an example we used this numerical moment matching approach on 5000 data points generated with parameters $a_1 = 8$, $a_2 = 6$, $b_1 = 6$, $b_2 = 5$, $\delta_1 = 3.28$, $\delta_2 = 2.73$, equivalent to the data shown in Fig. 2(a). We inferred the parameter values $\hat{a}_1 = 8.143$, $\hat{a}_2 = 6.193$, $\hat{b}_1 = 5.882$, $\hat{b}_2 = 4.931$, $\hat{\delta}_1 = 3.286$, $\hat{\delta}_2 = 2.754$. Figure 3(a) shows the correlations implied by different values for parameter δ_1 compared to the desired correlation of $\hat{r} = 0.47$. As one can see, the difference between r and $\hat{r}$ is zero for the inferred $\hat{\delta}_1 = 3.286$. Due to our constraint (32), $\hat{\delta}_2 = 2.754$ can be computed from δ_1. In this case there is an almost linear relationship between δ_1 and r but this is not true in general, especially for smaller parameter values.

This can be seen in a second example where we applied our moment matching approach on 5000 data points generated with parameters $a_1 = 0.2$, $a_2 = 0.9$, $b_1 = 0.4$, $b_2 = 1$, $\delta_1 = 0.1$, $\delta_2 = 0.45$ and received the inferred parameter values $\hat{a}_1 = 0.197$, $\hat{a}_2 = 0.903$, $\hat{b}_1 = 0.403$, $\hat{b}_2 = 1.017$, $\hat{\delta}_1 = 0.101$, $\hat{\delta}_2 = 0.462$. As seen in Fig. 3(b), for this second example the relationship between δ_1 and the correlation is not well approximated by a linear function, in contrast to the first example in Fig. 3(a). Still, inference works as for the first example shown above.

(a)   (b)
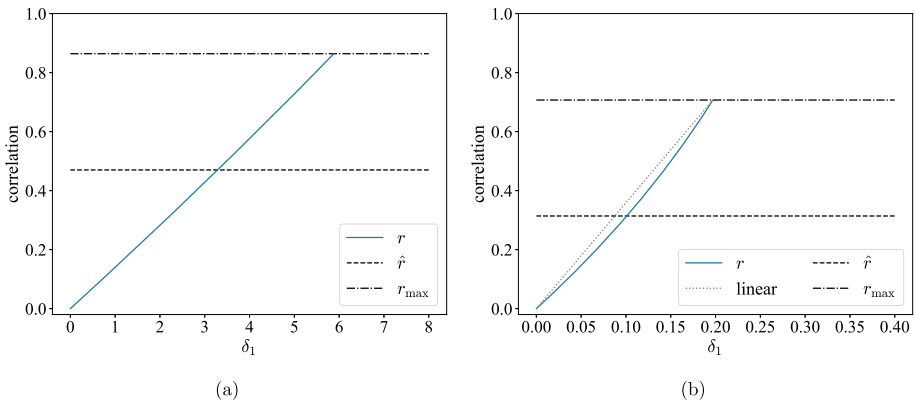
Fig. 3 The correlations r implied by different values of correlation parameter δ_1 compared to the desired correlation $\hat{r}$ given estimates of the marginal parameters a_1, a_2, b_1, b_2. δ_2 is not displayed since it can be computed from δ_1 using constraint (32). r_{max} is the maximum correlation that can be reached for the given marginal parameters. (**a**) shows the first inference example with $a_1 = 8.143, a_2 = 6.193, b_1 = 5.882, b_2 = 4.931, \hat{r} = 0.47$, and $r_{max} = 0.864$. The difference between r and $\hat{r}$ is zero for $\delta_1 = 3.286$. In (**b**) we show the second inference example with $a_1 = 0.197, a_2 = 0.903, b_1 = 0.403, b_2 = 1.017, \hat{r} = 0.314$, and $r_{max} = 0.707$. The difference between r and $\hat{r}$ is zero for $\delta_1 = 0.101$. The dotted linear reference line additionally shows that the relationship between the correlation and δ_1 is non-linear

2.4.2 Simulation study

The performance of the proposed approach for parameter inference was evaluated in a simulation study. We generated data from a bivariate beta distribution using different marginal parameters, correlation parameters, and different numbers of generated samples N and inferred δ_1 from these data using the proposed moment matching approach. For half of all considered simulations, X and Y were chosen to have the same marginal parameters to be able to generate the full range of correlations from 0 to 1, hence $a_1 = b_1, a_2 = b_2$. All possible combinations of the values in $[0.5, 1, 2, 3, 4, 5]$ for a_1 and a_2 were tested while omitting symmetric cases with $a_1 \leq a_2$, resulting in 21 different marginal distributions. For the remaining simulations we considered differing marginal distributions. We chose a subset of 7 marginal distributions from the set of marginals above as $[[0.5, 0.5], [1, 1], [1, 4], [2, 2], [2, 4], [3, 3], [4, 5]]$ and tested inference for all $\binom{7}{2} = 21$ combinations of different marginals in this set. Thus, in total we inferred parameters for 42 combinations of marginals. The correlation parameters δ_1 were chosen as $p \cdot \delta_1^{max}$ with $p = 0.01, 0.05, 0.25, 0.5, 0.75, 0.95, 0.99$ respectively, in order to show how inference works for data with different correlations between 0 and the maximum possible correlation for the respective marginals. δ_2 was obtained using the constraint in (32). To test the effect of the number of samples N on the accuracy of parameter inference we evaluated with $N = 100, 500, 1000$. For each of the $42 \cdot 7 \cdot 3 = 882$ parameter settings, we repeated the data simulation and inference process 50 times, resulting in 44100 inference results. These results are displayed in Fig. 4, for $N = 100, 500, 1000$ in (a), (b), (c) respectively. We can see that the inferred $\hat{\delta}_1$ match the true value of δ_1, the better the higher the number of available data samples N. The average standard deviations of all inferred $\hat{\delta}_1$ are 0.171 for $N = 100$, 0.078 for $N = 500$, and 0.055 for $N = 1000$. Note that although the generalized hypergeometric function $_3F_2$ in (29) is guaranteed to converge, as we showed in Sect. 2.3, it can happen that

(a) $N = 100$ (b) $N = 500$ (c) $N = 1000$
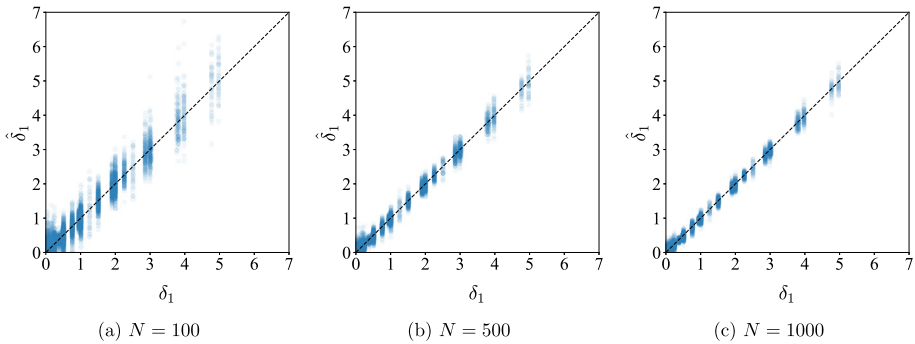
Fig. 4 Results of a simulation study for evaluating the proposed moment matching approach for parameter inference. We generated data from a bivariate beta distribution using different marginal distributions with different correlations and different numbers of generated samples N and inferred δ_1 from these data. We show inferred $\hat{\delta}_1$ against true δ_1 for $N = 100$ samples (**a**), $N = 500$ samples (**b**), and $N = 1000$ samples (**c**). $\hat{\delta}_1$ matches δ_1, the better the higher the number of available data samples N

its numerical computation fails due to very slow convergence. In our simulation study, this error occurred for 2.5% of all inference computations, only for low values of δ_1.

2.4.3 Application on a real data set of correlated forecasts

The bivariate beta distribution has broad applicability in many fields as diverse as Bayesian analysis, where it can model the correlation among priors for Binomial distributions [14], the modeling of proportions of hardwood forests over time, where it serves to estimate decadal changes in the relative land use of a region [5], the modeling of proportions of electorate voting in a two candidate election, proportions of substances in mixtures, or brand shares [16], and utility assessment [6]. Furthermore, the bivariate beta distribution can be used for modeling probabilities produced by two correlated forecasters. Correlations between forecasters are quite common, e.g. two bookmakers who base their odds on common information will produce correlated odds. For the same reason experts in risk assessment will often produce correlated forecasts. Similarly, different machine classifiers produce correlated predictions when trained on the same data [26]. These correlations should be taken into account when their predictions are combined, e.g. in different techniques for classifier fusion [27], since combining correlated classifiers can otherwise lead to overconfidence and high generalization error [28]. Here, we use such a data set consisting of the predictions of two classifiers as an illustrative example for the application of the proposed inference method. Two classifiers, a Bayes Net and a Random Forest, were trained on Alen Shapiro's chess (King-Rook vs. King-Pawn) data set [29, 30].[1] The task is to predict if King+Pawn will win a chess match against King+Rook based on 36 categorical features of a chess position. For training both classifiers, we used 10-fold cross-validation. The two classifiers' predictions on the respective 10 test sets form the data set we evaluate on. We only considered the predicted probabilities of winning King+Pawn for all 1527 match instances actually won by King+Pawn. The predicted probabilities of winning King+Rook for the matches actually won by King+Rook might follow a different distribution and are therefore excluded in our example. With the bivariate beta distribution we now model the winning probabilities the two classifiers predicted for King+Pawn.

---

[1] https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King-Pawn%29.
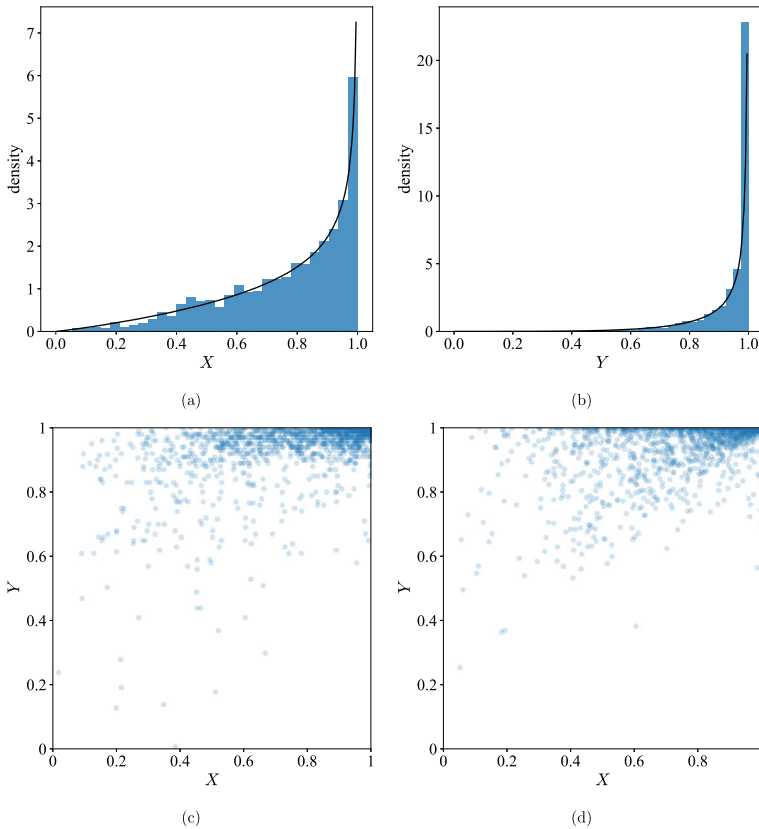
Fig. 5 The predicted winning probabilities from the chess classifier data set. (**a**) and (**b**) show the marginal distributions of X representing classifier 1 (Bayes Net), and Y representing classifier 2 (Random Forest). For both marginal distributions we show the relative frequencies of the predictions as well as the beta densities inferred with moment matching. (**c**) jointly shows X and Y, which are correlated with approximately $\hat{r} = 0.483$. (**d**) shows a simulated data set consisting of 1527 samples of a bivariate beta distribution with the parameters inferred for the chess classifier data set

Figure 5 shows the data and the inferred distribution. X is the predicted probability for King+Pawn winning of classifier 1, a Bayes Net, and Y the predicted probability of classifier 2, a Random Forest. In Fig. 5(a) and (b) we show histograms of the predictions of X and Y together with marginal beta densities that were inferred with moment matching: the parameters are $\hat{a}_1 = 2.094$, $\hat{a}_2 = 0.64$ for (a) and $\hat{b}_1 = 4.44$, $\hat{b}_2 = 0.288$ for (b). Figure 5(c) jointly shows X and Y with a correlation of approximately $\hat{r} = 0.483$. Matching this correlation, too, as described in Sect. 2.4.1, we obtain $\hat{\delta}_1 = 1.723$ and $\hat{\delta}_2 = 0.237$. The corresponding correlation is $r = 0.483$, which matches the data's empirical correlation up to numerical precision. Figure 5(d) shows a simulated data set consisting of 1527 samples drawn from a bivariate beta distribution with the inferred parameters, $\hat{a}_1 = 2.094$, $\hat{a}_2 = 0.64$, $\hat{b}_1 = 4.44$, $\hat{b}_2 = 0.288$, $\hat{\delta}_1 = 1.723$ and $\hat{\delta}_2 = 0.237$. As can be seen, the generated data set is similar to the real data set in Fig. 5(c) and the classifiers' predictions can thus be modeled reasonably well with this bivariate beta distribution.

Fig. 6 The exact correlation computed as in Sect. 2.3 compared to the approximate correlation computed with Eq. (33) for all parameter configurations used in the simulation study in Sect. 2.4.2
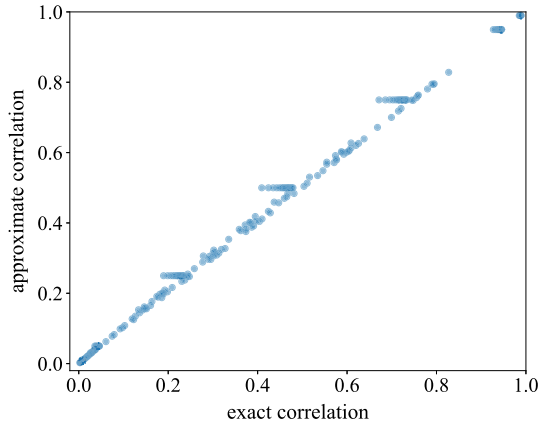
2.4.4 Relationship between δ_1 and the correlation

We found empirically that over a large range of parameters the following approximate relationship holds:

$$r \approx \frac{\delta_1}{\delta_1^{\max}} r_{\max}, \qquad (33)$$

while $\delta_1^{\max} = \min(a_1, b_1)$ and $r_{\max}$ is the maximum possible correlation for the given marginals. While this relationship could be used for approximate inference of δ_1, we still recommend the moment matching approach proposed in Sect. 2.4.1 that gives exact results. However, the shown relationship allows interpreting the correlation parameter δ_1. Particularly for equal marginal distribution with $a_1 = b_1$ and $a_2 = b_2$, for which $r_{\max} = 1$, this interpretation of δ_1 is very simple: The fraction of $\frac{\delta_1}{\delta_1^{\max}}$ approximately matches the generated correlation. For example, if $a_1 = b_1 = 2$ and $a_2 = b_2 = 4$, for $\delta_1 = 1$ we generate a correlation of $r = 0.468 \approx \frac{1}{2} r_{\max} = 0.5$ with $r_{\max} = 1$. For differing marginal distributions, interpreting δ_1 is more difficult because $r_{\max}$ must be computed numerically using the formulas given above. If, e.g., $a_1 = a_2 = 2$ and $b_1 = 1$, $b_2 = 4$, with $\delta_1 = 0.5$, we generate a correlation of $0.3 \approx \frac{0.5}{1} r_{\max} = 0.313$ with $r_{\max} = 0.627$. In Fig. 6, we plot the exact correlation and the approximated correlation computed with (33) for all parameter configurations used in the simulation study in Sect. 2.4.2. As can be seen, the relationship in (33) holds for all parameter values. The plateaus seen for approximate correlations of 0.25, 0.5, and 0.75 are a consequence of choosing $\delta_1 = p \cdot \delta_1^{\max}$ with $p = 0.25, 0.5, 0.75$ in the simulation study (Sect. 2.4.2) leading to approximate correlations of 0.25, 0.5, 0.75 for all simulations with equal marginals. We leave it as an open problem to show when approximation (33) holds to what accuracy.

3 Generalization to the correlated Dirichlet distribution

The bivariate beta distribution can be generalized to a correlated Dirichlet distribution [27] in order to model two positively correlated random vectors $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$ with the two marginal vectors being Dirichlet-distributed. A k-dimensional correlated Dirichlet distribution can be constructed from $3k$ gamma-distributed random

variables $A_1, \ldots, A_k,\ B_1, \ldots, B_k,\ D_1, \ldots, D_k$ with $3k$ parameters $\alpha_1, \ldots, \alpha_k,\ \beta_1, \ldots \beta_k,$ $\delta_1, \ldots \delta_k$ distributed according to

$$
\begin{aligned}
A_i &\sim \text{Gamma}(\alpha_i, 1) \quad i = 1, \ldots, k \\
B_i &\sim \text{Gamma}(\beta_i, 1) \quad i = 1, \ldots, k \\
D_i &\sim \text{Gamma}(\delta_i, 1) \quad i = 1, \ldots, k.
\end{aligned}
\tag{34}
$$

These random variables are used to construct the correlated Dirichlet-distributed random variables $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$ with

$$
X_i = \frac{A_i + D_i}{\sum_{i=1}^{k} A_i + \sum_{i=1}^{k} D_i} \qquad \text{and} \qquad Y_i = \frac{B_i + D_i}{\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} D_i}.
\tag{35}
$$

The two resulting marginal distributions are Dirichlet$(X; \alpha_1 + \delta_1, \ldots, \alpha_k + \delta_k)$ and Dirichlet$(Y; \beta_1 + \delta_1, \ldots, \beta_k + \delta_k)$.

Analogous to the example for the bivariate beta distribution in Sect. 2.4.3, this correlated Dirichlet distribution can be used for modeling non-binary probabilistic predictions of experts, sensors, or classifiers. This is particularly useful for Bayesian approaches to classifier or expert fusion, which are the reason why we started working on this distribution in the first place. For example, in a companion paper we apply it to classifier fusion, but instead of using moment matching—as developed here—we use rather inefficient Markov-chain methods to sample from the posterior distribution over the parameters [27]. Being able to explicitly model the correlation between probabilistic classifiers or probability estimates given by human experts with the correlated Dirichlet distribution allows Bayes optimal fusion of classifiers or experts, avoids overconfidence of the ensemble and thereby improves its performance. Applications of classifier fusion are widespread. Popular examples are intrusion detection, fake news detection, detection of diseases in medicine, or recognition of human states such as emotions [31]. Another application of the correlated Dirichlet distribution is the generation of stochastic matrices with individual rows or columns being Dirichlet-distributed and correlated, which can be beneficial for Markov processes, in optimal control, or reinforcement learning.

The derivations of the product moments and the exact covariance of the correlated Dirichlet distribution are analogous to the derivations for the bivariate beta distribution shown in this work. Thus, the parameters of the correlated Dirichlet distribution can also be estimated using the proposed moment matching approach, extended to the higher dimensionality of the Dirichlet distribution.

References

1. Arnold, B.C., Ghosh, I.: Bivariate beta and Kumaraswamy models developed using the Arnold-Ng bivariate beta distribution. REVSTAT–Stat. J. 15(2), 223–250 (2017). https://doi.org/10.57805/revstat.v15i2.211

2. Samanthi, R.G., Sepanski, J.: A bivariate extension of the beta generated distribution derived from copulas. Commun. Stat.-Theory Methods 48(5), 1043–1059 (2019). https://doi.org/10.1080/03610926.2018.1429626

3. Pirs, G., Strumbelj, E.: Bayesian combination of probabilistic classifiers using multivariate normal mixtures. J. Mach. Learn. Res. 20, 51–1 (2019)

4. Johnson, N.L., Kotz, S., Balakrishnan, N.: Continuous Univariate Distributions, Volume 2 vol. 289. John Wiley & Sons, New York, USA (1995)

5. Magnussen, S.: An algorithm for generating positively correlated beta-distributed random variables with known marginal distributions and a specified correlation. Comput. Stat. Data Anal. 46(2), 397–406 (2004). https://doi.org/10.1016/S0167-9473(03)00169-5

6. Libby, D.L., Novick, M.R.: Multivariate generalized beta distributions with applications to utility assessment. J. Educ. Stat. 7(4), 271–294 (1982). https://doi.org/10.3102/10769986007004271

7. Gupta, A.K., Wong, C.: On three and five parameter bivariate beta distributions. Metrika 32(1), 85–91 (1985). https://doi.org/10.1007/BF01897803

8. Ting Lee, M.-L.: Properties and applications of the Sarmanov family of bivariate distributions. Commun. Stat. Theory Methods 25(6), 1207–1222 (1996). https://doi.org/10.1080/03610929608831759

9. Jones, M.: Multivariate t and beta distributions associated with the multivariate F distribution. Metrika 54(3), 215–231 (2002). https://doi.org/10.1007/s184-002-8365-4

10. Olkin, I., Liu, R.: A bivariate beta distribution. Stat. Probability Lett. 62(4), 407–412 (2003). https://doi.org/10.1016/S0167-7152(03)00048-8

11. Nadarajah, S., Kotz, S.: Some bivariate beta distributions. Statistics 39(5), 457–466 (2005). https://doi.org/10.1080/02331880500286902

12. Sarabia, J.M., Castillo, E.: Bivariate distributions based on the generalized three-parameter beta distribution. In: Advances in Distribution Theory, Order Statistics, and Inference, pp. 85–110. Birkhäuser Boston, Boston (2006). https://doi.org/10.1007/0-8176-4487-3_6

13. El-Bassiouny, A., Jones, M.: A bivariate F distribution with marginals on arbitrary numerator and denominator degrees of freedom, and related bivariate beta and t distributions. Stat. Methods Appl. 18(4), 465 (2009). https://doi.org/10.1007/s10260-008-0103-y

14. Arnold, B.C., Ng, H.K.T.: Flexible bivariate beta distributions. J. Multivariate Anal. 102(8), 1194–1202 (2011). https://doi.org/10.1016/j.jmva.2011.04.001

15. Bran-Cardona, P.A., Orozco-Castañeda, J., Nagar, D.K.: Bivariate generalization of the Kummer-beta distribution. Revista Colombiana de Estadística 34(3), 497–512 (2011)

16. Gupta, A.K., Orozco-Castañeda, J.M., Nagar, D.K.: Non-central bivariate beta distribution. Stat. Papers 52(1), 139–152 (2011). https://doi.org/10.1007/s00362-009-0215-y

17. Orozco-Castañeda, J.M., Nagar, D.K., Gupta, A.K.: Generalized bivariate beta distributions involving Appell's hypergeometric function of the second kind. Comput. Math. Appl. 64(8), 2507–2519 (2012). https://doi.org/10.1016/j.camwa.2012.06.006

18. Olkin, I., Trikalinos, T.A.: Constructions for a bivariate beta distribution. Stat. Probability Lett. 96, 54–60 (2015). https://doi.org/10.1016/j.spl.2014.09.013

19. Nadarajah, S., Shih, S.H., Nagar, D.K.: A new bivariate beta distribution. Statistics 51(2), 455–474 (2017). https://doi.org/10.1080/02331888.2016.1240681

20. Koutoumanou, E., Wade, A., Cortina-Borja, M.: Local dependence in bivariate copulae with beta marginals. Revista Colombiana de Estadística 40(2), 281–296 (2017). https://doi.org/10.15446/rce.v40n2.59404

21. David Sam Jayakumar, G., Sulthan, A., Samuel, W.: A new bivariate beta distribution of Kind-1 of Type-A. J. Stat. Manag. Syst. 22(1), 141–158 (2019). https://doi.org/10.1080/09720510.2018.1537593

22. Crackel, R., Flegal, J.: Bayesian inference for a flexible class of bivariate beta distributions. J. Stat. Comput. Simul. 87(2), 295–312 (2017). https://doi.org/10.1080/00949655.2016.1208202

23. Plummer, M.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing, vol. 124, pp. 1–10. Vienna, Austria (2003)

24. Johansson, F., et al.: Mpmath: a Python Library for Arbitrary-precision Floating-point Arithmetic (version 0.18). (2013). http://mpmath.org/

25. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov,

N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

26. Jacobs, R.A.: Methods for combining experts' probability assessments. Neural Comput. **7**(5), 867–888 (1995). https://doi.org/10.1162/neco.1995.7.5.867

27. Trick, S., Rothkopf, C.: Bayesian classifier fusion with an explicit model of correlation. In: International Conference on Artificial Intelligence and Statistics. PMLR, pp 2282–2310 (2022)

28. Ueda, N., Nakano, R.: Generalization error of ensemble estimators. In: Proceedings of International Conference on Neural Networks (ICNN'96), vol. 1, pp. 90–95 . https://doi.org/10.1109/ICNN.1996.548872. IEEE (1996)

29. Shapiro, A.D.: Structured Induction in Expert Systems. Addison-Wesley, Boston (1987)

30. Dua, D., Graff, C.: UCI Machine Learning Repository (2017). http://archive.ics.uci.edu/ml

31. Sesmero, M.P., Iglesias, J.A., Magán, E., Ledezma, A., Sanchis, A.: Impact of the learners diversity and combination method on the generation of heterogeneous classifier ensembles. Appl. Soft Comput. **111**, 107689 (2021). https://doi.org/10.1016/j.asoc.2021.107689

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.