



# Minimax estimation of a bivariate cumulative distribution function

Rafał Połoczański<sup>1</sup> · Maciej Wilczyński<sup>1</sup> 

Received: 19 July 2018 / Published online: 24 September 2019  
© The Author(s) 2019

## Abstract

The problem of estimating a bivariate cumulative distribution function  $F$  under the weighted squared error loss and the weighted Cramer–von Mises loss is considered. No restrictions are imposed on the unknown function  $F$ . Estimators, which are minimax among procedures being affine transformation of the bivariate empirical distribution function, are found. Then it is proved that these procedures are minimax among all decision rules. The result for the weighted squared error loss is generalized to the case where  $F$  is assumed to be a continuous cumulative distribution function. Extensions to higher dimensions are briefly discussed.

**Keywords** Minimax estimation · Cumulative distribution function · Loss function

## 1 Introduction

Minimax estimation of a one dimensional cumulative distribution function (c.d.f.) was initiated in 1955 by Aggarwal (1955) and has been extensively studied since then (see the references given in the next paragraph). To the best of our knowledge, extensions of this approach to higher dimensions have not been investigated. In this paper we therefore consider estimating a bivariate c.d.f. and we generalize to this case some known results concerning minimax estimation of a univariate c.d.f. We also briefly discuss a multivariate generalization of these results.

Minimax estimation of a univariate c.d.f.  $F$  was considered by many authors. Using an invariance structure relative to the group of continuous and strictly increasing transformations, Aggarwal (1955) found the best invariant estimator of a continuous c.d.f.

---

✉ Maciej Wilczyński  
Maciej.Wilczynski@pwr.edu.pl  
Rafał Połoczański  
Rafal.Poloczanski@pwr.edu.pl

<sup>1</sup> Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

$F$  under the invariant loss  $L(F, \widehat{F}) = \int_{\mathbb{R}} |F - \widehat{F}|^r dF, r \geq 1$ . Here  $\widehat{F}$  stands for an estimate of  $F$ , based on a sample from  $F$ . Ferguson (1967, pages 191–197) generalized this result to the case that  $L(F, \widehat{F}) = \int_{\mathbb{R}} (F - \widehat{F})^2 h(F) dF$ , where  $h(\cdot)$  is a continuous and positive function. He also asked whether the best invariant estimates are minimax among the larger class of (not necessarily invariant) procedures. Yu (1992b) established the minimaxity of the best invariant procedure in Ferguson’s setup. In particular he found the minimax estimator of a continuous c.d.f.  $F$  under the loss of the form  $L(F, \widehat{F}) = \int_{\mathbb{R}} (F - \widehat{F})^2 F^{-\delta} (1 - F)^{-\gamma} dF$ , where  $\delta, \gamma \in \{0, 1\}$  are fixed numbers. Analog minimaxity findings were obtained by Mohammadi and van Zwet (2002) (entropy loss), Ning and Xie (2007) (Linex loss), and Stępień–Baran Stepien (2010) (strictly convex loss). Phadia and Yu (1992) proved minimaxity of the empirical distribution function under the Kolmogorov–Smirnov loss  $\sup_{t \in \mathbb{R}} |F(t) - \widehat{F}(t)|$ . Jafari Jozani et al. (2014) considered the problem of estimating a continuous distribution function  $F$ , as well as meaningful functions  $\tau(F)$ , under a large class of loss functions. They obtained best invariant estimators and established their minimaxity for Hölder continuous  $\tau$ ’s and strict bowl-shaped losses with a bounded derivative. Phadia (1973) considered a different model in which it is not assumed that  $F$  is continuous. He found the minimax estimator of  $F$  under the noninvariant loss function  $L(F, \widehat{F}) = \int_{\mathbb{R}} (F - \widehat{F})^2 F^{-\delta} (1 - F)^{-\gamma} dw$ , where  $\delta, \gamma \in \{0, 1\}$  are fixed numbers and  $w$  is a given non-null finite measure on  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , with  $\mathcal{B}_{\mathbb{R}}$  denoting the  $\sigma$ -algebra of Borel sets on  $\mathbb{R}$ . Yu (1992a) considered minimax estimation of  $F$  with a more general loss function  $L(F, \widehat{F}) = \int_{\mathbb{R}} G(F - \widehat{F})h(F) dw$ , where  $G(\cdot)$  is quadratic and  $h(\cdot)$  is continuous and positive. He proved that this problem is equivalent to that of finding a minimax estimator of a binomial proportion  $p$  under the loss  $L_B(p, d) = G(p - d)h(p)$ . His result was generalized by Jokiel-Rokita and Magiera (2007) to the case where  $G(\cdot)$  is convex.

In the first part of the paper we generalize the result of Phadia (1973) to the two-dimensional case. Let  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  be i.i.d. two-dimensional random vectors from an unknown bivariate (not necessarily continuous) c.d.f.  $F$  on  $\mathbb{R}^2$ . On the basis of this sample we find a minimax estimator  $\widehat{F}_1$  of  $F$  under the weighted squared error loss

$$L_1(\widehat{F}, F) = \iint_{\mathbb{R}^2} (\widehat{F}(s, t) - F(s, t))^2 (F(s, t))^{-\delta} (1 - F(s, t))^{-\gamma} dW(s, t). \tag{1}$$

Here  $W$  is a given non-null finite measure on  $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$  and  $\delta, \gamma \in \{0, 1\}$  are fixed numbers. In the case where  $\delta \neq 0$  or  $\gamma \neq 0$ , the loss functions  $L_1$  is more sensitive to departures from  $F$  in the tails of the distribution. We also show that the decision rule  $\widehat{F}_1$  remains minimax even if  $F$  is assumed to be absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^2$ . In the second part of the paper we find a minimax estimator  $\widehat{F}_2$  of an arbitrary bivariate c.d.f.  $F$  under the invariant weighted Cramer–von Mises loss

$$L_2(\widehat{F}, F) = \iint_{\mathbb{R}^2} (\widehat{F}(s, t) - F(s, t))^2 (F(s, t))^{-\delta} (1 - F(s, t))^{-\gamma} dF(s, t). \tag{2}$$

The latter result cannot be viewed as a natural generalization of the above mentioned result of Yu (1992b), because we derive  $\widehat{F}_2$  not assuming that  $F$  must be continuous. Since we use a much larger class of c.d.f.'s, the minimax inference for the loss  $L_2$  becomes easier, but is still nontrivial.

In the univariate case, the minimax estimators  $\phi$  and  $d_0$  found by Phadia (1973) and Yu (1992b), respectively, are linear functions of the empirical distribution function (e.d.f.). Therefore, it is not surprising that the minimax procedures  $\widehat{F}_1$  and  $\widehat{F}_2$  are linear functions of the bivariate e.d.f. Moreover, for each  $\delta, \gamma \in \{0, 1\}$ ,  $\widehat{F}_1 = \widehat{F}_2$  and  $\widehat{F}_1$  has the same form as  $\phi$  except that the univariate e.d.f. is replaced by the bivariate e.d.f. On the other hand,  $\phi \neq d_0$  and the relationship between  $\widehat{F}_2$  and  $d_0$  is slightly different than that between  $\widehat{F}_1$  and  $\phi$ . The reason for this difference is that the parameter space considered by Yu (1992b) contains only continuous c.d.f.'s.

### 2 Formulation of the problem

In the problem of estimating a univariate distribution function, the action space is restricted to the functions  $a(\cdot) : \mathbb{R} \rightarrow [0, 1]$ , which are nondecreasing. To define the appropriate action space to estimate a bivariate distribution function we note, that a two-dimensional analog of a nondecreasing function of one variable is a *2-increasing* function.

**Definition 1** A function  $a : \mathbb{R}^2 \rightarrow [0, 1]$  is 2-increasing if

$$a(s_2, t_2) - a(s_2, t_1) - a(s_1, t_2) + a(s_1, t_1) \geq 0$$

for any rectangle  $[s_1, t_1] \times [s_2, t_2] \in \mathbb{R}^2$ .

Let  $\widehat{F}(\mathbf{Z}; s, t)$  be an estimator of a bivariate distribution function  $F(s, t)$ , based on the sample  $\mathbf{Z} = ((X_1, Y_1), \dots, (X_n, Y_n))$ . To simplify the notation we will also write  $\widehat{F}(\cdot, \cdot)$  and  $\widehat{F}(s, t)$  for  $\widehat{F}(\mathbf{Z}, \cdot, \cdot)$  and  $\widehat{F}(\mathbf{Z}, s, t)$ , respectively. We assume that for each realization  $\mathbf{z}$  of  $\mathbf{Z}$ , the decision rule  $\widehat{F}(\mathbf{z}; \cdot, \cdot)$  is an element of the action space

$$\mathcal{A} = \{a : a = a(s, t) \text{ is a 2-increasing function on } \mathbb{R}^2 \text{ with values in } [0, 1]\}.$$

It is obvious that any bivariate c.d.f.  $F$  belongs to the class  $\mathcal{A}$ . However,  $\mathcal{A}$  contains estimates  $a(\cdot, \cdot)$  which are not c.d.f.'s, because they do not satisfy the conditions  $a(-\infty, -\infty) = 0, a(-\infty, y) = 0, a(x, -\infty) = 0$  and  $a(\infty, \infty) = 1$ . Such estimates are often referred to as a defective distribution functions. We include these estimates in the action space to obtain satisfactory results. The necessity of using defective distribution functions to estimate a univariate distribution function was recognized by Aggarwal (1955).

The family of all estimators  $\widehat{F}$ , which satisfy the above condition we denote by  $\mathcal{D}$ , i.e. we put

$$\mathcal{D} = \{\widehat{F}(\mathbf{z}; \cdot, \cdot) : \mathbb{R}^{2n} \rightarrow \mathcal{A}\}.$$

We use the symbol  $\mathcal{D}_A$  to denote the subclass of  $\mathcal{D}$ , which consists of the following affine estimators

$$\mathcal{D}_A = \left\{ \widehat{F} \in \mathcal{D} : \widehat{F}(s, t) = a \sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t) + b : a, b \in \mathbb{R} \right\}.$$

The estimates from  $\mathcal{D}_A$  are computationally tractable. The important member of  $\mathcal{D}_A$  is the empirical bivariate c.d.f. defined by

$$\widehat{F}_{emp}(s, t) = \frac{\sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t)}{n}, \quad (s, t) \in \mathbb{R}^2.$$

Let  $\mathbb{E}_F$  denote the expectation with respect to the c.d.f.  $F$ . Then, the risk function of an estimate  $\widehat{F}$  under the loss function  $L_i$   $i = 1, 2$  [see (1) and (2)] is given by

$$R_i(F, \widehat{F}) = \mathbb{E}_F(L_i(F, \widehat{F})).$$

When the value of  $i$  is clear from the context, we write  $R$  instead of  $R_i$ . We are interested in finding the minimax estimator of  $F$ , i.e. the estimator  $\widehat{F}_N \in \mathcal{D}$ , for which the following equation holds

$$\sup_{F \in \mathcal{F}} R(\widehat{F}_N, F) = \inf_{\widehat{F} \in \mathcal{D}} \sup_{F \in \mathcal{F}} R(\widehat{F}, F) := \rho_N.$$

Here  $\mathcal{F}$  is the family of all bivariate distribution on  $\mathbb{R}^2$ , i.e.

$$\mathcal{F} = \{F : F \text{ is a cumulative distribution function on } \mathbb{R}^2\}.$$

Unfortunately, many estimators from the class  $\mathcal{D}$  are not computationally tractable and finding  $\widehat{F}_N$  may be a difficult task. Therefore, we first look for affine minimax estimators. We say that a decision rule  $\widehat{F}_A \in \mathcal{D}_A$  is minimax affine estimator of  $F$  if

$$\sup_{F \in \mathcal{F}} R(\widehat{F}_A, F) = \inf_{\widehat{F} \in \mathcal{D}_A} \sup_{F \in \mathcal{F}} R(\widehat{F}, F) := \rho_A.$$

The quantities  $\rho_N$  and  $\rho_A$  are called the minimax risk and minimax affine risk, respectively. Clearly,  $\rho_N \leq \rho_A$ . We prove that under both  $L_1$  and  $L_2$  these two risks are equal, i.e.  $\rho_A = \rho_N$ . Hence  $\widehat{F}_A$  is minimax among all estimators from the class  $\mathcal{D}$ .

We also discuss the minimax approach in the case where an unknown distribution  $F$  is assumed to be absolutely continuous with respect to the Lebesgue measure. Therefore, we denote

$$\mathcal{F}_{AC} = \{F : F \text{ is an absolutely continuous distribution function on } \mathbb{R}^2\}.$$

### 3 Auxiliary results

As we have mentioned above, Yu (1992a) proved that results on minimax estimation of a binomial proportion  $p$  can help to find minimax estimators of a univariate c.d.f.. It

is not surprising that these results also can be applied in the bivariate case. For the sake of completeness we recall here the following well known facts concerning inference on binomial proportion, that will be used in the next sections (see, e.g. Hodges and Lehmann 1950, Olkin and Sobel 1979, Lehmann and Casella 1998, pages 311–312 or Phadia 1973).

Suppose that on the basis of an observation  $X$  from the binomial distribution  $B(n, p)$  we wish to estimate the unknown success probability  $p$  under the weighted squared error loss  $L_B(d, p) = \frac{(d-p)^2}{p^\delta(1-p)^\gamma}$ , with  $\delta, \gamma \in \{0, 1\}$ . Let the numbers  $a_1 = a_1(\delta, \gamma)$ ,  $b_1 = b_1(\delta, \gamma)$  and  $r_1 = r_1(\delta, \gamma)$  satisfy the following equation

$$\begin{aligned} & \inf_{a, b \in \mathbb{R}} \sup_{p \in [0, 1]} \left[ \frac{a^2 np(1-p) + (b + (an - 1)p)^2}{p^\delta(1-p)^\gamma} \right] \\ &= \sup_{p \in [0, 1]} \left[ \frac{a_1^2 np(1-p) + (b_1 + (a_1 n - 1)p)^2}{p^\delta(1-p)^\gamma} \right] = r_1 \end{aligned} \tag{3}$$

(the term inside the first brackets is the risk function of an affine estimator  $d(X) = aX + b$ ). Then  $d_1(X) = a_1X + b_1$  is the minimax estimator of  $p$  and  $r_1$  is the minimax risk, i.e.

$$\inf_d \sup_{p \in [0, 1]} \mathbb{E}_p \left[ \frac{(d(X) - p)^2}{p^\delta(1-p)^\gamma} \right] = \sup_{p \in [0, 1]} \mathbb{E}_p \left[ \frac{(a_1X + b_1 - p)^2}{p^\delta(1-p)^\gamma} \right] = r_1, \tag{4}$$

where the infimum is over measurable functions  $d : \mathbb{R} \rightarrow \mathbb{R}$ . The constants  $a_1, b_1, r_1$  are given by

$$\begin{aligned} \delta = 0, \gamma = 0: & a_1 = \frac{1}{n + \sqrt{n}}, b_1 = \frac{1}{2} \frac{1}{\sqrt{n} + 1}, r_1 = \frac{1}{4(\sqrt{n} + 1)^2}, \\ \delta = 1, \gamma = 0: & a_1 = \frac{1}{n + \sqrt{n}}, b_1 = 0, r_1 = \frac{1}{(\sqrt{n} + 1)^2}, \\ \delta = 0, \gamma = 1: & a_1 = \frac{1}{n + \sqrt{n}}, b_1 = \frac{1}{\sqrt{n} + 1}, r_1 = \frac{1}{(\sqrt{n} + 1)^2}, \\ \delta = 1, \gamma = 1: & a_1 = \frac{1}{n}, b_1 = 0, r_1 = \frac{1}{n}. \end{aligned} \tag{5}$$

In each of the four cases,  $d_1(X)$  has constant risk and is the Bayes estimator of  $p$  when  $p$  has the beta prior  $B(\alpha, \beta)$  with  $(\alpha, \beta)$  equal to  $(\sqrt{n}/2, \sqrt{n}/2)$  in the first case,  $(1, \sqrt{n})$  in the second,  $(\sqrt{n}, 1)$  in the third and  $(1, 1)$  in the fourth. Therefore,  $d_1(X)$  is minimax (see Lehmann and Casella 1998, Corollary 1.5, page 311).

### 4 Minimax estimator under the loss $L_1$

Since for each  $s, t \in \mathbb{R}$  the random variables  $\mathbb{1}(X_1 \leq s, Y_1 \leq t), \mathbb{1}(X_2 \leq s, Y_2 \leq t), \dots, \mathbb{1}(X_n \leq s, Y_n \leq t)$  are i.i.d. Bernoulli trials with probability of success  $F(s, t)$ , it

follows that the risk function of an affine estimator  $\widehat{F}_{ab}(s, t) = a \sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t) + b$  has the form

$$R_1(\widehat{F}_{ab}, F) = \mathbb{E}_F \left[ \iint_{\mathbb{R}^2} (\widehat{F}_{ab}(s, t) - F(s, t))^2 h(F(s, t)) dW(s, t) \right] \tag{6}$$

$$= \iint_{\mathbb{R}^2} \frac{[a^2 n F(s, t)(1 - F(s, t)) + (b + (an - 1)F(s, t))^2]}{[F(s, t)]^\delta [1 - F(s, t)]^\gamma} dW(s, t),$$

where  $h(p) = p^{-\delta}(1 - p)^{-\gamma}$ ,  $p \in (0, 1)$ . To find the minimax affine rule, we first derive the lower bound for the minimax affine risk  $\rho_A$ . The method used here is closely related to that of Phadia (1973). For any fixed integer  $k \geq 1$  and any  $p$  in  $(0, 1)$ , let  $F_{kp}$  be the bivariate c.d.f. defined by  $F_{kp} = pF_{k1} + (1 - p)F_{k2}$ , where  $F_{k1}$  and  $F_{k2}$  are the c.d.f.'s corresponding to uniform distributions over the squares  $A_1 = [-(k + 1), -k]^2$  and  $A_2 = [k, k + 1]^2$ , respectively. Since the integrand in (6) is nonnegative and since  $F_{kp}(s, t) = p$  on  $[-k, k]^2$ , it follows that the risk of  $\widehat{F}_{ab}$  at the point  $F_{kp}$  satisfies the following inequality

$$R_1(\widehat{F}_{ab}, F_{kp}) \geq \iint_{[-k, k]^2} [a^2 n F_{kp}(s, t)(1 - F_{kp}(s, t)) + (b + (an - 1)F_{kp}(s, t))^2] h(F_{k,p}(s, t)) dW(s, t)$$

$$= \iint_{[-k, k]^2} [a^2 np(1 - p) + (b + (an - 1)p)^2] h(p) dW(s, t)$$

$$= \left[ \frac{a^2 np(1 - p) + (b + (an - 1)p)^2}{p^\delta (1 - p)^\gamma} \right] \iint_{[-k, k]^2} dW(s, t). \tag{7}$$

We use this inequality to prove the following lemma.

**Lemma 1** *Let  $\delta, \gamma \in \{0, 1\}$  be fixed and let  $a_1, b_1, r_1$  be the corresponding numbers defined by (5). Then  $\widehat{F}_A = a_1 \sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t) + b_1$  is the minimax affine rule under the loss function  $L_1$  and the minimax affine risk is given by*

$$\rho_A = \inf_{a, b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_1(\widehat{F}_{ab}, F) = \sup_{F \in \mathcal{F}} R_1(\widehat{F}_A, F) = r_1 \iint_{\mathbb{R}^2} dW(s, t).$$

**Proof** Since (7) holds for any positive integer  $k$ , we conclude by (3) that

$$\rho_A = \inf_{a, b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_1(\widehat{F}_{ab}, F) \geq \inf_{a, b \in \mathbb{R}} \sup_{p \in [0, 1]} \sup_{k \geq 1} R_1(\widehat{F}_{ab}, F_{kp})$$

$$\begin{aligned} &\geq \inf_{a,b \in \mathbb{R}} \sup_{p \in [0,1]} \left[ \frac{a^2 np(1-p) + (b + (an-1)p)^2}{p^\delta(1-p)^\gamma} \right] \iint_{\mathbb{R}^2} dW(s,t) \\ &= r_1 \iint_{\mathbb{R}^2} dW(s,t). \end{aligned}$$

By straightforward calculations it is easy to verify that the integrand in the second line of (6) does not depend of  $F$  and equals  $r_1$  when  $a = a_1$  and  $b = b_1$  (see again Phadia 1973). Therefore, the risk function of the estimator  $\widehat{F}_A = \widehat{F}_{a_1 b_1}$  is constant and equal to  $r_1 \iint_{\mathbb{R}^2} dW(s,t)$ . This completes the proof, because

$$\begin{aligned} \inf_{a,b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_1(\widehat{F}_{ab}, F) &= \rho_A \geq r_1 \iint_{\mathbb{R}^2} dW(s,t) \\ &= \sup_{F \in \mathcal{F}} R_1(\widehat{F}_{a_1 b_1}, F) \geq \inf_{a,b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_1(\widehat{F}_{ab}, F). \end{aligned}$$

□

**Remark 1** Note that the constants  $a_1, b_1$  and  $r_1$ , which define both the form of a minimax affine rule and the corresponding minimax affine risk, depend on both  $\delta$  and  $\gamma$ . However, for simplicity of notation, this dependence is not shown throughout this paper.

The next theorem, which is the main result of this section, states that the estimator  $\widehat{F}_{a_1 b_1}$  is minimax in  $\mathcal{D}$ . The proof is based on a method of Yu (1992a) (cf. also Jokieli-Rokita and Magiera 2007).

**Theorem 1** Let  $\delta, \gamma \in \{0, 1\}$  be fixed and let  $a_1, b_1, r_1$  be the corresponding numbers defined by (5). Then the estimator  $\widehat{F}_A = a_1 \sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t) + b_1$  of an arbitrary bivariate c.d.f.  $F$  is minimax under the loss function  $L_1$  and the minimax risk is given by

$$\rho_N = \inf_{\widehat{F} \in \mathcal{D}} \sup_{F \in \mathcal{F}} R_1(\widehat{F}, F) = \sup_{F \in \mathcal{F}} R_1(\widehat{F}_A, F) = r_1 \iint_{\mathbb{R}^2} dW(s,t).$$

**Proof** To prove the theorem, it suffices to show that  $\sup_{F \in \mathcal{F}} R_1(\widehat{F}, F) \geq \rho_A$  for any  $\widehat{F} \in \mathcal{D}$ . Let  $k > 0$  be a fixed integer and let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be i.i.d. random vectors from the c.d.f.  $F_{kp}$  defined above. Since  $F_{kp}(s, t) = p$  on  $[-k, k]^2$ , it follows that for any  $\widehat{F} \in \mathcal{D}$ ,

$$\begin{aligned} R_1(\widehat{F}, F_{kp}) &= \mathbb{E}_{F_{kp}} [L_1(\widehat{F}, F_{kp})] \\ &\geq \mathbb{E}_{F_{kp}} \left[ \iint_{[-k,k]^2} (\widehat{F}(s,t) - F_{kp}(s,t))^2 h(F_{k,p}(s,t)) dW(s,t) \right] \end{aligned}$$

$$= \mathbb{E}_{F_{kp}} \left[ \iint_{[-k,k]^2} \frac{(\widehat{F}(s, t) - p)^2}{p^\delta(1 - p)^\gamma} dW(s, t) \right]. \tag{8}$$

Note first that the joint distribution of the vector  $\mathbf{Z} = ((X_1, Y_1), \dots, (X_n, Y_n))$  is given by

$$\begin{aligned} & \prod_{i=1}^n [p1_{A_1}(x_i, y_i) + (1 - p)1_{A_2}(x_i, y_i)] \\ &= \prod_{\{i:(x_i, y_i) \in A_1\}} p1_{A_1}(x_i, y_i) \times \prod_{\{i:(x_i, y_i) \in A_2\}} 1_{A_2}(x_i, y_i)(1 - p) = p^{n_1}(1 - p)^{n-n_1}, \end{aligned}$$

where  $n_1$  is the value of the random variable  $N_1$  which counts the number of observations  $(x_j, y_j)$  that fall into the square  $A_1$ . Since  $N_1$  is the sufficient statistic for  $p$ , we may assume that  $\widehat{F}$  depends on  $\mathbf{Z}$  only through  $N_1$ , i.e.  $\widehat{F}(\mathbf{Z}; s, t) = \widetilde{F}(N_1; s, t)$  for some Borel measurable function  $\widetilde{F} : \mathbb{R}^3 \rightarrow \mathbb{R}$ . Let the numbers  $\delta(i)$ ,  $i = 1, \dots, n$ , corresponding to  $\widehat{F}$ , be defined by  $\delta(i) \cdot \iint_{[-k,k]^2} dW(s, t) = \iint_{[-k,k]^2} \widetilde{F}(i; s, t) dW(s, t)$ .

Then,

$$\begin{aligned} & \iint_{[-k,k]^2} (\widetilde{F}(i; s, t) - p)^2 dW(s, t) = \iint_{[-k,k]^2} [\widetilde{F}(i; s, t) - \delta(i) + \delta(i) - p]^2 dW(s, t) \\ &= \iint_{[-k,k]^2} [\widetilde{F}(i; s, t) - \delta(i)]^2 dW(s, t) + 2[\delta(i) - p] \iint_{[-k,k]^2} [\widetilde{F}(i; s, t) - \delta(i)] dW(s, t) \\ & \quad + \iint_{[-k,k]^2} [\delta(i) - p]^2 dW(s, t) \\ &= \iint_{[-k,k]^2} [\widetilde{F}(i; s, t) - \delta(i)]^2 dW(s, t) + 2[\delta(i) - p] \cdot 0 + \iint_{[-k,k]^2} [\delta(i) - p]^2 dW(s, t) \\ &\geq \iint_{[-k,k]^2} [\delta(i) - p]^2 dW(s, t). \end{aligned}$$

This immediately shows that

$$\mathbb{E}_{F_{kp}} \left[ \iint_{[-k,k]^2} (\widetilde{F}(N_1; s, t) - p)^2 dW(s, t) \right] \geq \mathbb{E}_{F_{kp}} \left[ \iint_{[-k,k]^2} (\delta(N_1) - p)^2 dW(s, t) \right].$$



Since  $N_1$  has the binomial distribution  $B(n, p)$ , the last inequality implies by (8) and (4) that

$$\begin{aligned} \inf_{\widehat{F} \in \mathcal{D}} \sup_{p \in [0,1]} R_1(\widehat{F}, F_{kp}) &\geq \inf_{\delta} \sup_{p \in [0,1]} \mathbb{E}_p \left[ \frac{(\delta(N_1) - p)^2}{p^\delta(1-p)^\gamma} \right] \iint_{[-k,k]^2} dW(s, t) \\ &= r_1 \iint_{[-k,k]^2} dW(s, t). \end{aligned}$$

Here we use the fact that  $\widehat{F}(\mathbf{Z}; \cdot, \cdot) = \widetilde{F}(N_1; \cdot, \cdot)$ . Letting  $k \rightarrow \infty$  we obtain the lower bound

$$\begin{aligned} \sup_{F \in \mathcal{F}} R_1(\widehat{F}, F) &\geq \lim_{k \rightarrow \infty} \inf_{\widehat{F} \in \mathcal{D}} \sup_{p \in [0,1]} R_1(\widehat{F}, F_{kp}) \\ &\geq \lim_{k \rightarrow \infty} r_1 \iint_{[-k,k]^2} dW(s, t) = r_1 \iint_{\mathbb{R}^2} dW(s, t) \end{aligned}$$

which proves that under the loss  $L_1$ , the decision rule  $\widehat{F}_A := \widehat{F}_{a_1 b_1}$  is minimax among all estimators. □

**Remark 2** Since for any integer  $k \geq 1$  and any  $p \in (0, 1)$ , the c.d.f.  $F_{kp}$  is absolutely continuous (with respect to Lebesgue measure), it follows that we have proved a stronger result than the statement of the last theorem. In fact, we have shown that under the loss  $L_1$ ,  $\widehat{F}_A$  is the minimax estimator of an unknown absolutely continuous bivariate c.d.f.  $F$ . Moreover, Theorem 1 can be easily generalized to dimensions  $d > 2$ . In a slight modification of the proof,  $F_{kp}$  is a  $d$ -variate c.d.f., which equals  $p$  over the hypercube  $[-k, k]^d$ .

### 5 Minimax estimation under the loss function $L_2$

Under the loss  $L_2$  the risk of an affine decision rule  $\widehat{F}_{ab}$  is given by [cf. (6)]

$$R_2(\widehat{F}_{ab}, F) = \iint_{\mathbb{R}^2} \frac{[a^2 n F(s, t)(1 - F(s, t)) + (b + (an - 1)F(s, t))^2]}{[F(s, t)]^\delta [1 - F(s, t)]^\gamma} dF(s, t). \tag{9}$$

To find the minimax affine estimator, we first derive the lower bound for the minimax affine risk  $\rho_A$ . For this purpose, we choose a suitable family of the bivariate c.d.f.'s and consider estimation of  $F$  in the resulting submodel. Let  $(\tilde{x}_n)_{n \geq 1}$  be a given increasing sequence of points from  $(0, 1)$  and let  $m \geq 1$  be a fixed integer. We put  $\tilde{x}_0 = \tilde{y}_0 = 0$  and  $\tilde{y}_i = 1 - \tilde{x}_i$  for  $i \geq 1$ . Let the set  $\mathcal{S}_m$  be defined by

$$\mathcal{S}_m = \{s = (s_0, \dots, s_m) \in [0, 1]^{m+1} : s_0 + \dots + s_m = 1\}.$$

For any probability vector  $\mathbf{p} = (p_0, \dots, p_m) \in \mathcal{S}_m$ , we denote by  $F_{m,p}$  the bivariate c.d.f. which corresponds to a discrete random vector  $(X, Y)$  with the support  $\{(\tilde{x}_0, \tilde{y}_0), \dots, (\tilde{x}_m, \tilde{y}_m)\}$  and with the joint probability mass function given by

$$f_{m,p}(\tilde{x}_i, \tilde{y}_i) = \Pr(X = \tilde{x}_i, Y = \tilde{y}_i) = p_i, \quad i = 0, \dots, m.$$

The c.d.f.  $F_{m,p}$  satisfies

$$F_{m,p}(\tilde{x}_i, \tilde{y}_i) = \Pr(X \leq \tilde{x}_i, Y \leq \tilde{y}_i) = \begin{cases} p_0, & \text{when } i = 0, \\ p_0 + p_i, & \text{when } i = 1, \dots, m, \end{cases}$$

which implies that for any integer  $k \geq 0$ ,

$$\begin{aligned} & \iint_{\mathbb{R}^2} \frac{[F_{m,p}(s, t)]^k}{[F_{m,p}(s, t)]^\delta [1 - F_{m,p}(s, t)]^\gamma} dF_{m,p}(s, t) \\ &= \frac{(p_0)^k}{(p_0)^\delta (1 - p_0)^\gamma} p_0 + \sum_{i=1}^m \frac{(p_0 + p_i)^k}{(p_0 + p_i)^\delta (1 - p_0 - p_i)^\gamma} p_i. \end{aligned} \tag{10}$$

We use the last equality to find the lower bound for the minimax affine risk  $\rho_A$ . For each  $p_0 \in [0, 1]$  and each integer  $m \geq 1$ , let  $\mathbf{p}_0^{(m)}$  denote the corresponding vector from  $\mathcal{S}_m$  given by

$$\mathbf{p}_0^{(m)} = \left( p_0, \frac{1 - p_0}{m}, \dots, \frac{1 - p_0}{m} \right).$$

Then, for any integer  $k \geq 0$ ,

$$\begin{aligned} & \lim_{m \rightarrow \infty} \iint_{\mathbb{R}^2} \frac{[F_{m,p_0^{(m)}}(s, t)]^k}{[F_{m,p_0^{(m)}}(s, t)]^\delta [1 - F_{m,p_0^{(m)}}(s, t)]^\gamma} dF_{m,p_0^{(m)}}(s, t) \\ &= \lim_{m \rightarrow \infty} \left[ \frac{(p_0)^k p_0}{(p_0)^\delta (1 - p_0)^\gamma} + \frac{[p_0 + (1 - p_0)/m]^k (1 - p_0)}{[p_0 + (1 - p_0)/m]^\delta [1 - p_0 - (1 - p_0)/m]^\gamma} \right] \\ &= \frac{(p_0)^k}{(p_0)^\delta (1 - p_0)^\gamma}. \end{aligned}$$

Using (9) and applying the last equality with  $k = 0, 1, 2$ , we therefore find that

$$\lim_{m \rightarrow \infty} R_2(\widehat{F}_{ab}, F_{m,p_0^{(m)}}) = \left[ \frac{a^2 n p_0 (1 - p_0) + [b + (an - 1)p_0]^2}{p_0^\delta (1 - p_0)^\gamma} \right].$$

Hence, by (3) and (5), we obtain the following lower bound for the risk of affine estimators

$$\begin{aligned} \rho_A &= \inf_{a,b} \sup_{F \in \mathcal{F}} R_2(\widehat{F}_{a,b}, F) \geq \inf_{a,b} \sup_{p_0 \in [0,1]} \sup_{m \geq 1} R_2(\widehat{F}_{ab}, F_{m,p_0^{(m)}}) \\ &\geq \inf_{a,b} \sup_{p_0 \in [0,1]} \lim_{m \rightarrow \infty} R_2(\widehat{F}_{ab}, F_{m,p_0^{(m)}}) \\ &= \inf_{a,b \in \mathbb{R}} \sup_{p_0 \in [0,1]} \left[ \frac{a^2 n p_0 (1 - p_0) + (b + (an - 1)p_0)^2}{p_0^\delta (1 - p_0)^\gamma} \right] = r_1. \end{aligned}$$

**Lemma 2** Let  $\delta, \gamma \in \{0, 1\}$  be fixed and let  $a_1, b_1, r_1$  be the corresponding constants defined (5). Then  $\widehat{F}_A = a_1 \sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t) + b_1$  is the minimax affine rule under the loss  $L_2$  and the minimax affine risk is given by

$$\rho_A = \inf_{a,b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_2(\widehat{F}_{ab}, F) = \sup_{F \in \mathcal{F}} R_2(\widehat{F}_A, F) = r_1.$$

**Proof** The risk function of  $\widehat{F}_A = \widehat{F}_{a_1 b_1}$  is constant and equal to  $r_1$ , because the integrand in (9) does not depend of  $F$  and equals  $r_1$  when  $a = a_1$  and  $b = b_1$  (cf. the proof of Lemma 1). This completes the proof, because  $r_1$  is the lower bound for  $\rho_A$  and hence

$$\inf_{a,b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_2(\widehat{F}_{ab}, F) = \rho_A \geq r_1 = \sup_{F \in \mathcal{F}} R_2(\widehat{F}_A, F) \geq \inf_{a,b \in \mathbb{R}} \sup_{F \in \mathcal{F}} R_2(\widehat{F}_{ab}, F). \quad \square$$

To prove that  $\widehat{F}_A$  is minimax among all estimators we use the Bayes approach. More precisely, we take a specific sequence of priors on  $\mathcal{F}$  such that the corresponding sequence of Bayes risks converges to the supremum of the risk for  $\widehat{F}_A$ . Since the corresponding limit of Bayes risks is the lower bound for the minimax risk, we conclude that  $\widehat{F}_A$  is minimax.

Suppose that we know a priori that  $(X_1, Y_1), \dots, (X_n, Y_n)$  are i.i.d. random vectors from  $F_{m,p}$ , where  $m$  is a fixed positive integer and  $\mathbf{p} = (p_0, \dots, p_m) \in \mathcal{S}_m$  is an unknown probability vector. Then, the joint distribution of  $\mathbf{Z} = ((X_1, Y_1), \dots, (X_n, Y_n))$  is given by

$$\prod_{i=1}^n f_{m,p}(x_i, y_i) = \prod_{j=1}^m p_j^{N_j},$$

where  $N_k = \#\{i : (x_i, y_i) = (\tilde{x}_k, \tilde{y}_k)\}$  for  $k = 0, \dots, m$ . It is clear that  $\mathbf{N} = (N_0, \dots, N_m)$  has the multinomial distribution on  $m + 1$  categories with  $n$  draws and probability vector  $\mathbf{p} = (p_0, \dots, p_m)$ . Note that for any estimator  $\widehat{F} \in \mathcal{D}$ , we obtain

$$L_2(\widehat{F}, F_{m,p}) = \frac{(\widehat{F}(\tilde{x}_0, \tilde{y}_0) - p_0)^2}{(p_0)^\delta (1 - p_0)^\gamma} p_0 + \sum_{i=1}^m \frac{(\widehat{F}(\tilde{x}_i, \tilde{y}_i) - p_0 - p_i)^2}{(p_0 + p_i)^\delta (1 - p_0 - p_i)^\gamma} p_i$$

Since  $N$  is the sufficient statistics for  $\mathbf{p}$ , we may assume that  $\widehat{F}(\tilde{x}_i, \tilde{y}_i)$ ,  $i = 0, \dots, m$ , depends on  $\mathbf{Z}$  only through  $N$  i.e.  $\widehat{F}(\mathbf{Z}; \tilde{x}_i, \tilde{y}_i) = d_i(N)$  for some real-valued Borel measurable function  $d_i$ . Therefore, the problem of estimating  $F_{m,p}$  with the loss  $L_2$  and the sample  $\mathbf{Z}$  from  $F_{m,p}$  is equivalent to estimation of multinomial probabilities  $\mathbf{p} = (p_0, \dots, p_m)$  under the loss

$$L(\mathbf{d}, \mathbf{p}) = \frac{(d_0 - p_0)^2}{(p_0)^\delta(1 - p_0)^\gamma} p_0 + \sum_{i=1}^m \frac{(d_i - p_0 - p_i)^2}{(p_0 + p_i)^\delta(1 - p_0 - p_i)^\gamma} p_i \tag{11}$$

and the sample  $N$ . This means that the corresponding two minimax risks are equal, i.e.

$$\inf_{\widehat{F} \in \mathcal{F}} \sup_{\mathbf{p} \in \mathcal{S}_m} R_2(\widehat{F}, F_{m,p}) = \inf_{\mathbf{d} \in \mathcal{D}_m} \sup_{\mathbf{p} \in \mathcal{S}_m} R(\mathbf{d}, \mathbf{p}),$$

where  $R(\mathbf{d}, \mathbf{p}) = \mathbb{E}_p [L(\mathbf{d}(N), \mathbf{p})]$  and  $\mathcal{D}_m$  is the set of all estimators of  $\mathbf{p} = (p_0, \dots, p_m) \in \mathcal{S}_m$ .

To find the minimax risk in the latter problem we use the Bayes approach. Let  $\alpha_0, \dots, \alpha_m$  be any given positive numbers and let  $\alpha = \sum_{i=0}^m \alpha_i$ . Suppose that the unknown vector  $\mathbf{p} = (p_0, \dots, p_m)$  has the Dirichlet distribution  $\pi = \mathcal{D}(\alpha_0, \dots, \alpha_m)$  with the parameter vector  $(\alpha_0, \dots, \alpha_m)$ . Then, the random vector  $(p_0, \dots, p_{m-1})$  has the Lebesgue p.d.f.

$$f(p_0, \dots, p_{m-1}) = \frac{\Gamma(\alpha)}{\prod_{i=0}^m \Gamma(\alpha_i)} \prod_{i=0}^m p_i^{\alpha_i-1} \times I_{\mathcal{S}_m}(\mathbf{p}),$$

where  $p_m = 1 - (p_0 + \dots + p_{m-1})$  and  $I_{\mathcal{S}_m}(\cdot)$  is the indicator function of the set  $\mathcal{S}_m$ . Since the Dirichlet prior is conjugate to the multinomial distribution it follows that the posterior of  $\mathbf{p}$  given  $N = \mathbf{n}$  is  $\mathcal{D}(\alpha_0 + n_0, \dots, \alpha_m + n_m)$ , i.e.

$$f(p_0, \dots, p_{m-1} | \mathbf{n}) = \frac{\Gamma(\alpha + n)}{\prod_{i=0}^m \Gamma(n_i + \alpha_i)} \prod_{i=0}^m p_i^{n_i + \alpha_i - 1} \times I_{\mathcal{S}_m}(\mathbf{p}).$$

Let  $\mathbf{d}^\pi = (d_1^\pi, \dots, d_m^\pi)$  be the Bayes estimator of  $\mathbf{p} = (p_0, \dots, p_m)$  under the loss  $L(\mathbf{d}, \mathbf{p})$ , given by (11). Then

$$d_i^\pi(N) = \begin{cases} \frac{\mathbb{E}[(p_0)^{1-\delta}(1 - p_0)^{-\gamma} p_0 | N]}{\mathbb{E}[(p_0)^{-\delta}(1 - p_0)^{-\gamma} p_0 | N]}, & i = 0, \\ \frac{\mathbb{E}[(p_0 + p_i)^{1-\delta}(1 - p_0 - p_i)^{-\gamma} p_i | N]}{\mathbb{E}[(p_0 + p_i)^{-\delta}(1 - p_0 - p_i)^{-\gamma} p_i | N]}, & i = 1, \dots, m, \end{cases}$$

provided that these posterior moments are finite. If  $\delta = \gamma = 0$ , then both integral and the resulting Bayes risk can be easily calculated, because for each  $k, l \in \{0, 1, 2, 3\}$ ,

$$\begin{aligned} \mathbb{E}_\pi \left( p_i^k p_0^l \right) &= \frac{\alpha_i (\alpha_i + 1) \cdots (\alpha_i + k - 1) \alpha_0 (\alpha_0 + 1) \cdots (\alpha_0 + l - 1)}{\alpha (\alpha + 1) \cdots (\alpha + k + l - 1)}. \end{aligned} \tag{12}$$

To find these moments in the case where  $\delta = 1$  or  $\gamma = 1$ , we use the following Liouville formula (cf. Fichtenholz 1992). Let a function  $\phi : [0, 1] \rightarrow \mathbb{R}$  be continuous and let  $p$  and  $q$  be any positive real numbers. If the integral  $\int_0^1 |\phi(u)| u^{p+q-1} du$  is finite, then the following identity holds

$$\iint_{\substack{x \geq 0, y \geq 0 \\ x+y \leq 1}} \phi(x+y) x^{p-1} y^{q-1} dx dy = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \int_0^1 \phi(u) u^{p+q-1} du. \tag{13}$$

Let  $a, b$  be any real numbers such that  $\alpha_0 + \alpha_i + a + 1 > 0$  and  $\alpha - \alpha_0 - \alpha_i + b > 0$ . Since the vector  $(p_0, p_i, 1 - p_0 - p_i)$  is distributed according to  $\mathcal{D}(\alpha_0, \alpha_i, \alpha - \alpha_0 - \alpha_i)$ , it follows from (13) that

$$\begin{aligned} C_i(a, b) &:= \mathbb{E}_\pi \left[ (p_0 + p_i)^a (1 - p_0 - p_i)^b p_i \right] = \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)\Gamma(\alpha_i)\Gamma(\alpha - \alpha_0 - \alpha_i)} \\ &\times \iint_{\substack{p_0 \geq 0, p_i \geq 0 \\ p_0 + p_i \leq 1}} (p_0 + p_i)^a (1 - p_0 - p_i)^b p_i p_0^{\alpha_0-1} p_i^{\alpha_i-1} (1 - p_0 - p_i)^{\alpha - \alpha_0 - \alpha_i - 1} dp_0 dp_i \\ &= \frac{\Gamma(\alpha)\alpha_i}{\Gamma(\alpha_0 + \alpha_i + 1)\Gamma(\alpha - \alpha_0 - \alpha_i)} \frac{\Gamma(\alpha_0 + \alpha_i + a + 1)\Gamma(\alpha - \alpha_0 - \alpha_i + b)}{\Gamma(\alpha + a + b + 1)} \end{aligned} \tag{14}$$

Moreover, if  $\alpha_0 + a + 1 > 0$  and  $\alpha - \alpha_0 + b > 0$ , then we also get

$$\begin{aligned} C_0(a, b) &:= \mathbb{E}_\pi \left[ (p_0)^a (1 - p_0)^b p_0 \right] \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)\Gamma(\alpha - \alpha_0)} \frac{\Gamma(\alpha_0 + a + 1)\Gamma(\alpha - \alpha_0 + b)}{\Gamma(\alpha + a + b + 1)}, \end{aligned} \tag{15}$$

because  $(p_0, 1 - p_0)$  has the distribution  $\mathcal{D}(\alpha_0, \alpha - \alpha_0)$ . In particular, since  $\delta, \gamma \in \{0, 1\}$ ,  $C_i(-\delta, -\gamma)$  and  $C_i(1 - \delta, -\gamma)$  are finite for all  $i = 0, \dots, m$  if  $(\alpha_0, \dots, \alpha_m)$  belongs to the set  $\mathcal{A}_m^{\delta, \gamma}$  defined by

$$\mathcal{A}_m^{\delta, \gamma} = \{(\alpha_0, \dots, \alpha_m) \in \mathbb{R}_+^{m+1} : \alpha - \alpha_0 - \alpha_i - \gamma > 0, \text{ for all } i = 1, \dots, m\}. \tag{16}$$

Let  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$ . Then, since the posterior of  $\mathbf{p}$  given  $N = \mathbf{n}$  is  $\mathcal{D}(\alpha_0 + n_0, \dots, \alpha_m + n_m)$ , it follows by (14) and (15), that the Bayes estimator  $\mathbf{d}^\pi = (d_0^\pi, \dots, d_m^\pi)$  of  $\mathbf{p} = (p_0, \dots, p_m)$  is

$$d_i^\pi(N) = \begin{cases} \frac{\mathbb{E}[(p_0)^{1-\delta}(1-p_0)^{-\gamma} p_0 | N]}{\mathbb{E}[(p_0)^{-\delta}(1-p_0)^{-\gamma} p_0 | N]} = \frac{N_0 + \alpha_0 + 1 - \delta}{n + \alpha + 1 - \delta - \gamma}, & i = 0, \\ \frac{\mathbb{E}[(p_0 + p_i)^{1-\delta}(1-p_0 - p_i)^{-\gamma} p_i | N]}{\mathbb{E}[(p_0 + p_i)^{-\delta}(1-p_0 - p_i)^{-\gamma} p_i | N]} = \frac{N_0 + N_i + \alpha_0 + \alpha_i + 1 - \delta}{n + \alpha + 1 - \delta - \gamma}, & i = 1, \dots, m, \end{cases}$$

Clearly, the finiteness of the above posterior means is implied by the fact that if  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$  then  $(\alpha_0 + n_0, \dots, \alpha_m + n_m) \in \mathcal{A}_m^{\delta, \gamma}$ , because  $\alpha_i + n_i \geq \alpha_i > 0$  and  $\alpha + n - (\alpha_0 + n_0) - (\alpha_i + n_i) - \gamma = \alpha - \alpha_0 - \alpha_i - \gamma + (n - n_0 - n_i) \geq \alpha - \alpha_0 - \alpha_i - \gamma > 0$  for each  $i = 1, \dots, m$ .

Since the random variables  $N_0$  and  $N_0 + N_i$ ,  $i = 1, \dots, m$  have the binomial distributions  $B(n, p_0)$  and  $B(n, p_0 + p_i)$ , respectively, the risk of  $\mathbf{d}^\pi$  is given by

$$R(\mathbf{d}^\pi, \mathbf{p}) = \mathbb{E}_p [L(\mathbf{d}^\pi(N), \mathbf{p})] = \frac{np_0(1-p_0) + (u_0 - up_0)^2}{(n+u)^2} \frac{p_0}{(p_0)^\delta(1-p_0)^\gamma} + \sum_{i=1}^m \frac{n(p_0 + p_i)(1-p_0 - p_i) + [u_i - u(p_0 + p_i)]^2}{(n+u)^2} \frac{p_i}{(p_0 + p_i)^\delta(1-p_0 - p_i)^\gamma},$$

where for simplicity of notation we write

$$u_i = \begin{cases} \alpha_0 + 1 - \delta & \text{when } i = 0, \\ \alpha_0 + \alpha_i + 1 - \delta, & \text{when } i = 1, \dots, m, \end{cases} \quad \text{and } u = \alpha + 1 - \delta - \gamma. \quad (17)$$

Then, by (14) and (15), the Bayes risk  $r(\pi) = \mathbb{E}_\pi [R(\mathbf{d}^\pi, \mathbf{p})]$  can be written as

$$r(\pi) = \sum_{i=0}^m \left[ \frac{nC_i(1-\delta, 1-\gamma) + u_i^2 C_i(-\delta, -\gamma) - 2u_i u C_i(1-\delta, -\gamma) + u^2 C_i(2-\delta, -\gamma)}{(n+u)^2} \right].$$

To simplify the last formula, we note that by (14) and (15)

$$C_i(1-\delta, -\gamma) = \frac{u_i}{u} C_i(-\delta, -\gamma), \quad C_i(2-\delta, -\gamma) = \frac{u_i(u_i + 1)}{u(u + 1)} C_i(-\delta, -\gamma).$$

Moreover, we also have  $C_i(1-\delta, 1-\gamma) = \frac{u_i(u - u_i)}{u(u + 1)} C_i(-\delta, -\gamma)$ , which implies that

$$\begin{aligned}
 r(\pi) &= \sum_{i=0}^m \left[ \frac{nC_i(1-\delta, 1-\gamma) + C_i(-\delta, -\gamma)(u_i^2 - 2u_i^2 + uu_i(u_i + 1)/(u + 1))}{(n + u)^2} \right] \\
 &= \sum_{i=0}^m \frac{nC_i(1-\delta, 1-\gamma) + uC_i(1-\delta, 1-\gamma)}{(n + u)^2} = \sum_{i=0}^m \frac{C_i(1-\delta, 1-\gamma)}{n + u}.
 \end{aligned}$$

Finally, by (14), (15) and (17), we obtain the following formula for the Bayes risk of  $d^\pi$

$$\begin{aligned}
 r(\pi) &= \frac{1}{n + \alpha + 1 - \delta - \gamma} \left[ \frac{\Gamma(\alpha)}{\Gamma(\alpha_0)\Gamma(\alpha - \alpha_0)} \frac{\Gamma(\alpha_0 + 2 - \delta)\Gamma(\alpha - \alpha_0 + 1 - \gamma)}{\Gamma(\alpha + 3 - \delta - \gamma)} \right. \\
 &\quad \left. + \sum_{i=1}^m \frac{\Gamma(\alpha)\alpha_i}{\Gamma(\alpha_0 + \alpha_i + 1)\Gamma(\alpha - \alpha_0 - \alpha_i)} \frac{\Gamma(\alpha_0 + \alpha_i + 2 - \delta)\Gamma(\alpha - \alpha_0 - \alpha_i + 1 - \gamma)}{\Gamma(\alpha + 3 - \delta - \gamma)} \right].
 \end{aligned} \tag{18}$$

Before stating the main result of this section, we introduce some notation. If  $\delta = \gamma = 0$  we put  $n_0 = 4$  and when  $\delta = 1$  or  $\gamma = 1$  we set  $n_0 = 1$ .

**Theorem 2** *Let  $\delta, \gamma \in \{0, 1\}$  be fixed and let  $a_1, b_1, r_1$  be the corresponding constants defined by (5). If  $n \geq n_0$  then the estimator  $\widehat{F}_A = a_1 \sum_{i=1}^n \mathbb{1}(X_i \leq s, Y_i \leq t) + b_1$  of  $F$  is minimax under the weighted Cramer–von Mises loss function  $L_2(\widehat{F}, F)$  and the minimax risk is given by*

$$\rho_N = \inf_{\widehat{F} \in \mathcal{D}} \sup_{F \in \mathcal{F}} R_2(\widehat{F}, F) = \sup_{F \in \mathcal{F}} R_2(\widehat{F}_A, F) = r_1.$$

**Proof** Lemma 2 implies that for any integer  $m > 1$  and any  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$ , we have

$$\begin{aligned}
 r_1 &= \sup_{F \in \mathcal{F}} R_2(\widehat{F}_A, F) \geq \inf_{d \in \mathcal{D}_m} \sup_{p \in \mathcal{S}_m} R(d, p) \geq \inf_{d \in \mathcal{D}_m} \mathbb{E}_\pi [R(d, p)] \\
 &= \mathbb{E}_\pi [R(d^\pi, p)] = r(\alpha_0, \dots, \alpha_m),
 \end{aligned}$$

where  $r(\alpha_0, \dots, \alpha_m)$  stands for the Bayes risk given by the right-hand side of (18). It is clear that to prove minimaxity of  $\widehat{F}_A$  it suffices to find a sequence of points  $((\alpha_0^{(m)}, \dots, \alpha_m^{(m)}))$  such that

$$\begin{aligned}
 (\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) &\in \mathcal{A}_m^{\delta, \gamma} \text{ for each } m > 1 \text{ and } \lim_{m \rightarrow \infty} r(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) \\
 &= r_1(\delta, \gamma)
 \end{aligned} \tag{19}$$

(see, e.g. Lehmann and Casella 1998, Theorem 1.12, page 316).

Suppose first that  $\delta = 1$  and  $\gamma = 0$ . Then, for any integer  $m > 1$  and any  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$ ,

$$\begin{aligned} r(\alpha_0, \dots, \alpha_m) &= \frac{\alpha_0(\alpha - \alpha_0) + \sum_{i=1}^m \alpha_i(\alpha - \alpha_0 - \alpha_i)}{\alpha(\alpha + 1)(n + \alpha)} \\ &= \frac{\alpha(\alpha - \alpha_0) - \sum_{i=1}^m \alpha_i^2}{\alpha(\alpha + 1)(n + \alpha)} \end{aligned}$$

[cf. (18)]. Now, let  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) = \frac{1}{m+1} (\sqrt{n}, \dots, \sqrt{n})$ . Then, by (16),  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) \in \mathcal{A}_m^{\delta, \gamma}$ , because  $\alpha_0^{(m)}, \dots, \alpha_m^{(m)}$  are positive numbers that satisfy

$$\alpha^{(m)} - \alpha_0^{(m)} - \alpha_i^{(m)} - \gamma = \alpha^{(m)} - \alpha_0^{(m)} - \alpha_i^{(m)} = \sqrt{n} - \frac{2\sqrt{n}}{m+1} > 0, \quad i = 1, \dots, m.$$

To complete the proof of the theorem for the case  $\delta = 1$  and  $\gamma = 0$ , we conclude that

$$\begin{aligned} r(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) &= \frac{\sqrt{n} \left( \sqrt{n} - \frac{\sqrt{n}}{m+1} \right) - m \left( \frac{\sqrt{n}}{m+1} \right)^2}{\sqrt{n}(\sqrt{n} + 1)(n + \sqrt{n})} \xrightarrow{m \rightarrow \infty} \frac{1}{(\sqrt{n} + 1)^2} \\ &= r_1(1, 0). \end{aligned}$$

Now we consider the case  $\delta = 0$  and  $\gamma = 1$ . Then, for any integer  $m > 1$  and any  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$ ,

$$\begin{aligned} r(\alpha_0, \dots, \alpha_m) &= \frac{\alpha_0(\alpha_0 + 1) + \sum_{i=1}^m \alpha_i(\alpha_0 + \alpha_i + 1)}{\alpha(\alpha + 1)(n + \alpha)} \\ &= \frac{\alpha(\alpha_0 + 1) + \sum_{i=1}^m \alpha_i^2}{\alpha(\alpha + 1)(n + \alpha)}, \end{aligned}$$

[again cf. (18)]. Let  $(\varepsilon_m)$  be any sequence of real numbers converging to zero such that  $\varepsilon_m > 2/(m - 1)$  when  $m > 2$ . Let  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) = (\sqrt{n} - 1, (1 + \varepsilon_m)/m, \dots, (1 + \varepsilon_m)/m)$ . If  $n > 1$  then, by (16),  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) \in \mathcal{A}_m^{\delta, \gamma}$ , because  $\alpha_0^{(m)}, \dots, \alpha_m^{(m)}$  are positive numbers that satisfy

$$\begin{aligned} \alpha^{(m)} - \alpha_0^{(m)} - \alpha_i^{(m)} - \gamma &= (\sqrt{n} + \varepsilon_m) - (\sqrt{n} - 1) - (1 + \varepsilon_m)/m - 1 \\ &= \varepsilon_m - (1 + \varepsilon_m)/m > 0 \end{aligned}$$

for each  $i = 1, \dots, m$ . Simple calculation yields

$$\begin{aligned} r(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) &= \frac{(\sqrt{n} + \varepsilon_m) \sqrt{n} + m [(1 + \varepsilon_m)/m]^2}{(\sqrt{n} + \varepsilon_m) (\sqrt{n} + \varepsilon_m + 1) (n + \sqrt{n} + \varepsilon_m)} \xrightarrow{m \rightarrow \infty} \frac{1}{(\sqrt{n} + 1)^2} = r_1(0, 1), \end{aligned}$$



which proves minimaxity of  $\widehat{F}_A$  when  $\delta = 0$  and  $\gamma = 1$ .

Suppose now that  $\delta = 1$  and  $\gamma = 1$ . Then, for any integer  $m > 1$  any  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$ , we have

$$r(\alpha_0, \dots, \alpha_m) = \frac{\alpha_0 + \sum_{i=1}^m \alpha_i}{\alpha(n + \alpha - 1)} = \frac{\alpha}{\alpha(n + \alpha - 1)} = \frac{1}{n + \alpha - 1}.$$

Define  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) = \frac{1}{m + 1} (1 + \varepsilon_m, \dots, 1 + \varepsilon_m)$ , where  $(\varepsilon_m)$  is the sequence given above. Then, by (16),  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) \in \mathcal{A}_m^{\delta, \gamma}$ , because  $\alpha_0^{(m)}, \dots, \alpha_m^{(m)}$  are positive numbers and

$$\alpha^{(m)} - \alpha_0^{(m)} - \alpha_i^{(m)} - \gamma = (1 + \varepsilon_m) - \frac{2(1 + \varepsilon_m)}{(m + 1)} - 1 > 0 \quad i = 1, \dots, m.$$

To prove minimaxity of  $\widehat{F}_A$  in the case, where  $\delta = 1, \gamma = 1$ , we note that

$$r(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) = \frac{1}{n + 1 + \varepsilon_m - 1} \xrightarrow{m \rightarrow \infty} \frac{1}{n} = r_1(1, 1).$$

Consider now the last case where  $\delta = 0$  and  $\gamma = 0$  and assume that  $n \geq 4$ . Then, for any integer  $m > 1$  and any point  $(\alpha_0, \dots, \alpha_m) \in \mathcal{A}_m^{\delta, \gamma}$ , the Bayes risk (18) can be rewritten as

$$\begin{aligned} r(\alpha_0, \dots, \alpha_m) &= \frac{\alpha_0(\alpha_0 + 1)(\alpha - \alpha_0) + \sum_{i=1}^m \alpha_i(\alpha_0 + \alpha_i + 1)(\alpha - \alpha_0 - \alpha_i)}{(n + \alpha + 1)\alpha(\alpha + 1)(\alpha + 2)} \\ &= \frac{\alpha(\alpha_0 + 1)(\alpha - \alpha_0) + (\alpha - 1 - 2\alpha_0) \sum_{i=1}^m \alpha_i^2 - \sum_{i=1}^m \alpha_i^3}{(n + \alpha + 1)\alpha(\alpha + 1)(\alpha + 2)} \end{aligned}$$

Define  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) = \left( \frac{\sqrt{n} + \varepsilon_m}{2} - 1, \frac{\sqrt{n} + \varepsilon_m}{2m}, \dots, \frac{\sqrt{n} + \varepsilon_m}{2m} \right)$ , where  $(\varepsilon_m)$  is the sequence given above. Then, by (16),  $(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}) \in \mathcal{A}_m^{\delta, \gamma}$ , because, for  $n \geq n_0 = 4, \alpha_0^{(m)}, \dots, \alpha_m^{(m)}$  are positive numbers and for each  $i = 1, \dots, m$  the following condition holds

$$\begin{aligned} \alpha^{(m)} - \alpha_0^{(m)} - \alpha_i^{(m)} - \gamma &= (\sqrt{n} + \varepsilon_m - 1) - \left( \frac{\sqrt{n} + \varepsilon_m}{2} - 1 \right) - \frac{\sqrt{n} + \varepsilon_m}{2m} - 0 \\ &= \frac{\sqrt{n} + \varepsilon_m}{2} - \frac{\sqrt{n} + \varepsilon_m}{2m} > 0. \end{aligned}$$

Since  $\alpha^{(m)} = \sqrt{n} + \varepsilon_m - 1$  and hence  $\alpha^{(m)} - 1 - 2\alpha_0^{(m)} = 0, \alpha_0^{(m)} + 1 = \alpha^{(m)} - \alpha_0^{(m)} = (\sqrt{n} + \varepsilon_m)/2$ , it follows that

$$\begin{aligned}
 & r\left(\alpha_0^{(m)}, \dots, \alpha_m^{(m)}\right) \\
 &= \frac{(\sqrt{n} + \varepsilon_m - 1) [(\sqrt{n} + \varepsilon_m)/2]^2 - m [(\sqrt{n} + \varepsilon_m)/(2m)]^3}{(n + \sqrt{n} + \varepsilon_m)(\sqrt{n} + \varepsilon_m - 1)(\sqrt{n} + \varepsilon_m)(\sqrt{n} + \varepsilon_m + 1)} \xrightarrow{m \rightarrow \infty} \frac{1}{4(\sqrt{n} + 1)^2} \\
 &= r_1(0, 0).
 \end{aligned}$$

This proves minimaxity of  $\widehat{F}_A$  in the case, where  $\delta = 0$ ,  $\gamma = 0$ .  $\square$

**Remark 3** Theorem 2 can be generalized to higher dimensions. In order to do this, we must modify the c.d.f.'s  $F_{m,p}$  defined at the beginning of this section. For simplicity of notation, we consider only the three-dimensional case. Let  $(\tilde{x}_n)_{n \geq 1}$  be an increasing sequence of points from  $(0, 1)$ , and let  $\tilde{x}_0 = \tilde{y}_0 = \tilde{z}_0 = 0$  and  $\tilde{y}_i = \tilde{z}_i = 1 - \tilde{x}_i$  for  $i \geq 1$ . For any  $m \geq 1$  and any probability vector  $\mathbf{p} = (p_0, \dots, p_m) \in \mathcal{S}_m$ , let  $F_{m,p}$  be the c.d.f. of a discrete random vector  $(X, Y, Z)$  with the support  $\{(\tilde{x}_0, \tilde{y}_0, \tilde{z}_0), \dots, (\tilde{x}_m, \tilde{y}_m, \tilde{z}_m)\}$  and with the joint probability mass function given by

$$f_{m,p}(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i) = \Pr(X = \tilde{x}_i, Y = \tilde{y}_i, Z = \tilde{z}_i) = p_i, \quad i = 0, \dots, m.$$

Then analogously to the bivariate case, the c.d.f.  $F_{m,p}$  satisfies

$$F_{m,p}(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i) = \Pr(X \leq \tilde{x}_i, Y \leq \tilde{y}_i, Z \leq \tilde{z}_i) = \begin{cases} p_0, & \text{when } i = 0, \\ p_0 + p_i, & \text{when } i = 1, \dots, m, \end{cases}$$

which implies that the 3-variate analog of the equality (10) holds. The rest of the generalization is straightforward.

**Acknowledgements** We are grateful to the referees for their many valuable comments and suggestions, which substantially improved this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Aggarwal O (1955) Some minimax invariant procedures for estimating a cumulative distribution function. *Ann Math Stat* 26:450–463
- Ferguson TS (1967) *Mathematical statistics*. Academic Press, New York
- Fichtenholz GM (1992) *Differential- und Integralrechnung*. University Books for Mathematics. Johann Ambrosius Barth Verlag GmbH, Leipzig
- Hodges JJJ, Lehmann EL (1950) Some problems in minimax point estimation. *Ann Math Stat* 21:182–197
- Jafari Jozani M, Leblanc A, Marchand E (2014) On continuous distribution functions minimax and best invariant estimators and integrated balanced loss functions. *Can J Stat* 42(3):470–486
- Jokiel-Rokita A, Magiera R (2007) Minimax estimation of a cumulative distribution function by converting to a parametric problem. *Metrika* 66(1):61–73
- Lehmann E, Casella G (1998) *Theory of point estimation*. Springer, New York
- Mohammadi L, van Zwet WR (2002) Minimax invariant estimator of a continuous distribution function under entropy loss. *Metrika* 56:31–42

- Ning J, Xie M (2007) Minimax invariant estimator of continuous distribution function under linex loss. *J Syst Sci Complex* 20(1):119–126
- Olkin I, Sobel M (1979) Admissible and minimax estimation for the multinomial distribution and for independent binomial distributions. *Ann Stat* 7:284–290
- Phadia E (1973) Minimax estimation of a cumulative distribution function. *Ann Stat* 1:209–230
- Phadia E, Yu Q (1992) Minimality of the best invariant estimator of a distribution function under the Kolmogorov–Smirnov loss. *Ann Stat* 20(4):2192–2195
- Stepien A (2010) Minimax invariant estimator of a continuous distribution function under a general loss function. *Metrika* 72:37–49
- Yu Q (1992) A general method of finding a minimax estimator of a distribution function when no equalizer rule is available. *Can J Stat* 20(3):281–290
- Yu Q (1992) Minimax invariant estimator of a continuous distribution function. *Ann Inst Stat Math* 44:729–735

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