APPLICATIONES MATHEMATICAE
36,4 (2009), pp. 441–452

SARALEES NADARAJAH (Manchester)

ON THE USE OF DIFFERENCE OF TWO PROPORTIONS

Abstract. Differences of two *proportions* occur most frequently in biomedical research. However, as far as published work is concerned, only approximations have been used to study the distribution of such differences. In this note, we derive the exact probability distribution of the difference of two proportions for seven flexible beta type distributions. The expressions involve several well known special functions. The use of these results with respect to known approximations is illustrated.

1. Introduction. Differences of two *proportions* arise in many areas of the sciences, engineering and medicine. The need for the probability distribution of the difference of two proportions arises especially when one is interested in comparing the performances of two *entities*. Two example are:

- Suppose that there are two drugs, say A and B, and that one wishes to compare their efficiencies. The drugs are tested on a group of patients. Let X_1 denote the proportion of effectiveness for drug A. Let X_2 denote the proportion of effectiveness for drug B. Then the probability distribution of $D = X_1 - X_2$ can be used to study whether drug A is more effective than drug B or vice versa.
- Suppose that there are two pesticides, say A and B, and that one wishes to compare their efficiencies. The pesticides are tested on a field of crops. Let X_1 denote the proportion of effectiveness for pesticide A. Let X_2 denote the proportion of effectiveness for pesticide B. Then the probability distribution of $D = X_1 - X_2$ can be used to study whether pesticide A is more effective than pesticide B or vice versa.

2000 *Mathematics Subject Classification*: 33C90, 62E99.
Key words and phrases: beta distribution, biomedical research, proportions, special functions.

DOI: 10.4064/am36-4-4

The exact probability distribution of the difference of two proportions does not appear to have been known in the literature. Various approximations have been used for this distribution. We refer the readers to Radhakrishna et al. (1992) and Newcombe (1998) for reviews of known approximation methods. Two of the most commonly used approximations are the *normal approximation* and the *beta approximation*. The normal approximation suggests that $D = X_1 - X_2$ follows a truncated normal distribution when X_1 and X_2 are independent beta type random variables. The beta approximation suggests that $D = X_1 - X_2$ follows a beta distribution.

In this note, we derive the exact probability distribution of the difference of two proportions, i.e. the difference of two beta type random variables X_1 and X_2. We consider seven of the most known beta type distributions for X_1 and X_2. For each distribution, we derive an exact analytical expression for the probability density function (pdf) of the difference $D = X_1 - X_2$. These derivations are given in Section 2. The exact results are compared to the normal and beta approximations in Section 3. It turns out that these approximations perform poorly.

2. Exact distributions of $D = X_1 - X_2$. In this section, we provide seven theorems for the derivation of the pdf of $D = X_1 - X_2$. Theorem 1 considers the case that X_1 and X_2 have the standard beta distribution. The remaining theorems consider various generalized beta distributions. Theorems 2 and 3 consider the generalized beta distributions due to Libby and Novick (1982) and McDonald (1984), respectively. Theorems 4 and 5 consider the non-central and doubly non-central beta distributions. Theorems 6 and 7 consider the generalized beta distributions due to Gordy (1998) and Armero and Bayarri (1994), respectively. The details of the derivation of the pdf of $D = X_1 - X_2$ are given for Theorem 1, but omitted for the other theorems. The full details can be obtained from the author.

THEOREM 1. *Suppose X_1 and X_2 are independent beta random variables given by the pdfs*

$$f_1(x_1) = \frac{x_1^{a_1-1}(1-x_1)^{b_1-1}}{B(a_1, b_1)} \tag{1}$$

and

$$f_2(x_2) = \frac{x_2^{a_2-1}(1-x_2)^{b_2-1}}{B(a_2, b_2)}, \tag{2}$$

respectively, for $0 < x_1, x_2 < 1$ and $a_1, a_2, b_1, b_2 > 0$, where $B(\cdot, \cdot)$ denotes the beta function defined by

$$B(a, b) = \int_0^1 w^{a-1}(1-w)^{b-1} dw.$$

Then the pdf of the difference $D = X_1 - X_2$ can be expressed as

$$f_D(d) = \Gamma(a_1+b_1)\Gamma(a_2+b_2)$$

$$\times \begin{cases} \dfrac{\Gamma(a_1+b_1)\Gamma(a_2+b_2)(-d)^{a_2-1}(1+d)^{a_1+b_2-1}}{\Gamma(b_1)\Gamma(a_2)\Gamma(a_1+b_2)} \\ \qquad \times F_1\left(a_1, 1-b_1, 1-a_1, a_1+b_2; 1+d, \dfrac{1+d}{d}\right) \\ \qquad\qquad \text{if } -1 \le d \le 0, \\ \dfrac{\Gamma(a_1+b_1)\Gamma(a_2+b_2)d^{a_1-1}(1-d)^{a_2+b_1-1}}{\Gamma(b_2)\Gamma(a_1)\Gamma(a_2+b_1)} \\ \qquad \times F_1\left(a_2, 1-b_2, 1-a_1, a_2+b_1; 1-d, \dfrac{d-1}{d}\right) \\ \qquad\qquad \text{if } 0 \le d \le 1 \end{cases}$$

for $-1 \le d \le 1$, where F_1 denotes the Appell hypergeometric function of the first kind defined by

$$F_1(a, b, b', c; z, \xi) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{(a)_{k+l}(b)_k(b')_l z^k \xi^l}{(c)_{k+l} k! l!},$$

where $(c)_k = c(c+1)\cdots(c+k-1)$ denotes the ascending factorial.

Proof. Since

$$f_D(d) = \begin{cases} \int_0^{1+d} f_1(t) f_2(t-d)\, dt & \text{if } -1 \le d \le 0, \\ \int_0^{1-d} f_1(d+t) f_2(t)\, dt & \text{if } 0 \le d \le 1, \end{cases}$$

one can write

$$f_D(d) = \frac{\Gamma(a_1+b_1)\Gamma(a_2+b_2)}{\Gamma(a_1)\Gamma(a_2)\Gamma(b_1)\Gamma(b_2)} J(d) \tag{3}$$

if $-1 \le d \le 0$, where

$$J(d) = \int_0^{1+d} t^{a_1-1}(1-t)^{b_1-1}(t-d)^{a_2-1}(1+d-t)^{b_2-1}\, dt. \tag{4}$$

Setting $w = t/(1+d)$, (4) can be rewritten as

$$J(d) = (-d)^{a_2-1}(1+d)^{a_1+b_2-1} K(d), \tag{5}$$

where

$$K(d) = \int_0^1 w^{a_1-1}(1-w)^{b_2-1}\{1-(1+d)w\}^{b_1-1}\left\{1 - \frac{(1+d)w}{d}\right\}^{a_2-1} dw. \tag{6}$$

Using equation (3.211) in Gradshteyn and Ryzhik (2000), one can express (6) as

$$K(d) = B(a_1, b_2) F_1\left(a_1, 1-b_1, 1-a_1, a_1+b_2; 1+d, \frac{1+d}{d}\right). \tag{7}$$

Combining (3), (5) and (7) yields (3) for $-1 \le d \le 0$. The result for $0 \le d \le 1$ can be established similarly. ■

Theorem 2. *If X_1 and X_2 are independent random variables given by the pdfs*

$$f_1(x_1) = \frac{p_1 x_1^{a_1 p_1 - 1}\{1 - x_1^{p_1}\}^{b_1 - 1}}{B(a_1, b_1)}$$

and

$$f_2(x_2) = \frac{p_2 x_2^{a_2 p_2 - 1}\{1 - x_2^{p_2}\}^{b_2 - 1}}{B(a_2, b_2)},$$

respectively, for $0 < x_1, x_2 < 1$ and $a_1, a_2, b_1, b_2, p_1, p_2 > 0$, then the pdf of $D = X_1 - X_2$ can be expressed as

$$f_D(d) = C d^{a_1 p_1 + a_2 p_2 - 1} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} A(k,l) d^{p_1 k + p_2 l} I_{k,l}(d)$$

for $-1 \le d \le 1$, where

$$I_{k,l}(d) = \begin{cases} B_{1/2}(1 - a_1 p_1 - p_1 k - a_2 p_2 - p_2 l, a_2 p_2 + p_2 l) \\ \qquad - B_{d/(1+3d)}(1 - a_1 p_1 - p_1 k - a_2 p_2 - p_2 l, a_2 p_2 + p_2 l) \\ \qquad\qquad \text{if } -1 \le d \le 0, \\ B_{1-d}(1 - a_1 p_1 - p_1 k - a_2 p_2 - p_2 l, a_2 p_2 + p_2 l) \quad \text{if } 0 \le d \le 1, \end{cases}$$

$$A(k,l) = \frac{(1-b_1)_k (1-b_2)_l}{k!\,l!}, \qquad C = \frac{p_1 p_2}{B(a_1, b_1) B(a_2, b_2)},$$

where $B_x(\cdot,\cdot)$ denotes the incomplete beta function defined by

$$B_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\, dt.$$

Theorem 3. *If X_1 and X_2 are independent random variables given by the pdfs*

$$f_1(x_1) = \frac{\lambda_1^{a_1} x_1^{a_1 - 1}(1 - x_1)^{b_1 - 1}}{B(a_1, b_1)\{1 - (1 - \lambda_1) x_1\}^{a_1 + b_1}}$$

and

$$f_2(x_2) = \frac{\lambda_2^{a_2} x_2^{a_2-1}(1-x_2)^{b_2-1}}{B(a_2,b_2)\{1-(1-\lambda_2)x_2\}^{a_2+b_2}},$$

respectively, for $0 < x_1, x_2 < 1$ *and* $a_1, a_2, b_1, b_2, \lambda_1, \lambda_2 > 0$, *then the pdf of* $D = X_1 - X_2$ *can be expressed as*

$$f_D(d) = C \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} A(k,l) I_{k,l}(d)$$

for $-1 \le d \le 1$, *where*

$$I_{k,l}(d) = \begin{cases} B(b_2, a_1+k)(-d)^{a_2+l-1}(1+d)^{a_1+b_2+k-1} \\ \quad \times F_1\left(a_1+k, 1-b_1, 1-a_2-l, a_1+b_2+k; 1+d, \frac{1+d}{d}\right) \\ \qquad\qquad\qquad \text{if } -1 \le d \le 0, \\ B(b_1, a_2+l) d^{a_1+k-1}(1-d)^{a_2+b_1+l-1} \\ \quad \times F_1\left(a_2+l, 1-b_2, 1-a_1-k, a_2+b_1+l; 1-d, \frac{d-1}{d}\right) \\ \qquad\qquad\qquad \text{if } 0 \le d \le 1, \end{cases}$$

$$A(k,l) = \frac{(a_1+b_1)_k (a_2+b_2)_l (1-\lambda_1)^k (1-\lambda_2)^l}{k!l!},$$

$$C = \frac{\lambda_1^{a_1}\lambda_2^{a_2}}{B(a_1,b_1)B(a_2,b_2)}.$$

THEOREM 4. *If* X_1 *and* X_2 *are independent random variables given by the pdfs*

$$f_1(x_1) = \frac{x_1^{a_1-1}(1-x_1)^{b_1-1}}{\Gamma(b_1)} \exp\left(-\frac{\lambda_1}{2}\right) \sum_{k=0}^{\infty} \frac{\Gamma(a_1+b_1+k)(\lambda_1 x_1)^k}{\Gamma(a_1+k)2^k k!}$$

and

$$f_2(x_2) = \frac{x_2^{a_2-1}(1-x_2)^{b_2-1}}{\Gamma(b_2)} \exp\left(-\frac{\lambda_2}{2}\right) \sum_{k=0}^{\infty} \frac{\Gamma(a_2+b_2+k)(\lambda_2 x_2)^k}{\Gamma(a_2+k)2^k k!},$$

respectively, for $0 < x_1, x_2 < 1$ *and* $a_1, a_2, b_1, b_2, \lambda_1, \lambda_2 > 0$, *then the pdf of* $D = X_1 - X_2$ *can be expressed as*

$$f_D(d) = \exp\left(-\frac{\lambda_1+\lambda_2}{2}\right) \times \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \Gamma(a_1+b_1+k)\Gamma(a_2+b_2+l)\, \frac{\lambda_1^k \lambda_2^l}{2^{k+l}k!l!}\, I_{k,l}(d)$$

for $-1 \le d \le 1$, *where*

$$I_{k,l}(d) = \begin{cases} \dfrac{(-d)^{a_2+l-1}(1+d)^{a_1+b_2+k-1}}{\Gamma(b_1)\Gamma(a_2+l)\Gamma(a_1+b_2+k)} \\ \quad \times F_1\left(a_1+k, 1-b_1, 1-a_1-l, a_1+b_2+k; 1+d, \dfrac{1+d}{d}\right) \\ \hfill \text{if } -1 \le d \le 0, \\ \dfrac{d^{a_1+k-1}(1-d)^{a_2+b_1+l-1}}{\Gamma(b_2)\Gamma(a_1+k)\Gamma(a_2+b_1+l)} \\ \quad \times F_1\left(a_2+l, 1-b_2, 1-a_1-k, a_2+b_1+l; 1-d, \dfrac{d-1}{d}\right) \\ \hfill \text{if } 0 \le d \le 1. \end{cases}$$

THEOREM 5. *If X_1 and X_2 are independent random variables given by the pdfs*

$$f_1(x_1) = x_1^{a_1-1}(1-x_1)^{b_1-1}\exp\left(-\frac{\lambda_1+\mu_1}{2}\right) \times \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\Gamma(a_1+b_1+k+l)(\lambda_1 x_1)^k\{\mu_1(1-x_1)\}^l}{\Gamma(a_1+k)\Gamma(b_1+l)2^{k+l}k!l!}$$

and

$$f_2(x_2) = x_2^{a_2-1}(1-x_2)^{b_2-1}\exp\left(-\frac{\lambda_2+\mu_2}{2}\right) \times \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\Gamma(a_2+b_2+k+l)(\lambda_2 x_2)^k\{\mu_2(1-x_2)\}^l}{\Gamma(a_2+k)\Gamma(b_2+l)2^{k+l}k!l!},$$

respectively, for $0 < x_1, x_2 < 1$ and $a_1, a_2, b_1, b_2, \lambda_1, \lambda_2, \mu_1 > 0, \mu_2 > 0$, then the pdf of $D = X_1 - X_2$ can be expressed as

$$f_D(d) = C\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} A(k,l,m,n) J_{k,l,m,n}(d)$$

for $-1 \le d \le 1$, where

$$J_{k,l,m,n}(d) = \begin{cases} \dfrac{(-d)^{a_2+m-1}(1+d)^{a_2+b_2+k+n-1}}{\Gamma(a_1+b_2+k+n)\Gamma(a_2+m)\Gamma(b_1+l)} \\ \quad \times F_1\left(a_1+k, 1-b_1-l, 1-a_2-m, a_1+b_2+k+n; 1+d, \dfrac{1+d}{d}\right) \\ \hfill \text{if } -1 \le d \le 0, \\ \dfrac{d^{a_1+k-1}(1-d)^{a_2+b_1+l+m-1}}{\Gamma(a_2+b_1+l+m)\Gamma(a_1+k)\Gamma(b_2+n)} \\ \quad \times F_1\left(a_2+m, 1-b_2-n, 1-a_1-k, a_2+b_1+l+m; 1-d, \dfrac{d-1}{d}\right), \\ \hfill \text{if } 0 \le d \le 1, \end{cases}$$

$$A(k,l,m,n) = \frac{\lambda_1^k \mu_1^l \lambda_2^m \mu_2^n \Gamma(a_1+b_1+k+l)\Gamma(a_2+b_2+m+n)}{2^{k+l+m+n} k!l!m!n!},$$

$$C = \exp\left(-\frac{\lambda_1+\mu_1+\lambda_2+\mu_2}{2}\right).$$

THEOREM 6. *If X_1 and X_2 are independent random variables given by the pdfs*

$$f_1(x_1) = \frac{x_1^{a_1-1}(1-x_1)^{b_1-1}\exp(-\gamma_1 x_1)}{B(a_1,b_1)\ {}_1F_1(a_1;a_1+b_1;-\gamma_1)}$$

and

$$f_2(x_2) = \frac{x_2^{a_2-1}(1-x_2)^{b_2-1}\exp(-\gamma_2 x_2)}{B(a_2,b_2)\ {}_1F_1(a_2;a_2+b_2;-\gamma_2)},$$

respectively, for $0 < x_1, x_2 < 1$ and $a_1, a_2, b_1, b_2, \gamma_1, \gamma_2 > 0$, where ${}_1F_1$ denotes the confluent hypergeometric function defined by

$${}_1F_1(\alpha;\beta;x) = \sum_{k=0}^{\infty} \frac{(\alpha)_k}{(\beta)_k}\frac{x^k}{k!},$$

then the pdf of $D = X_1 - X_2$ can be expressed as

$$f_D(d) = \begin{cases} C(-d)^{a_2-1}(1+d)^{a_1+b_2-1}\exp(\gamma_2 d)\displaystyle\sum_{k=0}^{\infty}\frac{(1-b_1)_k}{k!}(1+d)^k \\ \quad \times B(a_1+k,b_2)\Phi_1\left(a_1+k, 1-a_2, a_1+b_2+k; \dfrac{1+d}{d}, -(\gamma_1+\gamma_2)(1+d)\right) \\ \hfill \text{if } -1 \le d \le 0, \\ Cd^{a_1-1}(1-d)^{b_1+a_2-1}\exp(-\gamma_1 d)\displaystyle\sum_{k=0}^{\infty}\frac{(1-b_2)_k}{k!}(1-d)^k \\ \quad \times B(b_1,a_2+k)\Phi_1\left(a_2+k, 1-a_1, b_1+a_2+k; \dfrac{d-1}{d}, -(\gamma_1+\gamma_2)(1-d)\right) \\ \hfill \text{if } 0 \le d \le 1 \end{cases}$$

for $-1 \le d \le 1$, where

$$1/C = B(a_1,b_1)\ {}_1F_1(a_1;a_1+b_1;-\gamma_1)B(a_2,b_2)\ {}_1F_1(a_2;a_2+b_2;-\gamma_2)$$

and Φ_1 denotes the Appell hypergeometric series of the first kind defined by

$$\Phi_1(\alpha,\beta,\gamma;x,y) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(\alpha)_{k+l}(\beta)_k}{(\gamma)_{k+l}}\frac{x^k y^l}{k!l!}.$$

THEOREM 7. *If X_1 and X_2 are independent random variables given by the pdfs*

$$f_1(x_1) = \frac{x_1^{a_1-1}(1-x_1)^{b_1-1}/(1+z_1x_1)^{\gamma_1}}{B(a_1,b_1)\ {}_2F_1(\gamma_1,a_1;a_1+b_1;-z_1)}$$

and

$$f_2(x_2) = \frac{x_2^{a_2-1}(1-x_2)^{b_2-1}/(1+z_2x_2)^{\gamma_2}}{B(a_2,b_2)\ {}_2F_1(\gamma_2,a_2;a_2+b_2;-z_2)},$$

respectively, for $0 < x_1, x_2 < 1$ *and* $a_1, a_2, b_1, b_2, \gamma_1, \gamma_2, z_1, z_2 > 0$, *where* ${}_2F_1$ *denotes the Gauss hypergeometric function defined by*

$${}_2F_1(\alpha,\beta;\gamma;x) = \sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{(\gamma)_k} \frac{x^k}{k!},$$

then the pdf of $D = X_1 - X_2$ *can be expressed as*

$$f_D(d) = C \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{(-1)^{k+l}(\gamma_1)_k(\gamma_2)_l z_1^k z_2^l}{k!l!} I_{k,l}(d) \tag{8}$$

for $-1 \le d \le 1$, *where*

$$I_{k,l}(d) = \begin{cases} B(a_1+k,b_2)(-d)^{a_2+l-1}(1+d)^{a_1+b_2+k-1} \\ \quad \times F_1\left(a_1+k, 1-b_1, 1-a_2-l, a_1+b_2+k; 1+d, \dfrac{1+d}{d}\right) \\ \hfill \text{if } -1 \le d \le 0, \\ B(b_1,a_2+l)d^{a_1+k-1}(1-d)^{a_2+b_1+l-1} \\ \quad \times F_1\left(a_2+l, 1-b_2, 1-a_1-k, a_2+b_1+l; 1-d, \dfrac{d-1}{d}\right) \\ \hfill \text{if } 0 \le d \le 1, \end{cases}$$

$$1/C = B(a_1,b_1)\, {}_2F_1(\gamma_1,a_1;a_1+b_1;-z_1) B(a_2,b_2)\, {}_2F_1(\gamma_2,a_2;a_2+b_2;-z_2).$$

3. Discussion. We have derived analytical expressions for the pdf of the difference of two proportios by considering seven flexible beta type distributions. The expressions involve the Appell hypergeometric function of the first kind, the Appell hypergeometric series of the first kind, the confluent hypergeometric function, the Gauss hypergeometric function, and the incomplete beta function. These functions are well known and well established in the mathematics literature; see Lebedev (1972), Erdélyi et al. (1981), Prudnikov et al. (1986) and Gradshteyn and Ryzhik (2000) for their detailed properties. In-built numerical routines for computing them are available in most mathematical packages, e.g. Maple, Mathematica and Matlab.

We feel that the results in Section 2 can be of help in deriving an exact approach for the various modelling problems that involve differences of two proportions. Their use can be illustrated, for example, by comparing the derived pdfs with the approximations mentioned in Section 1. Suppose X_1 and X_2 are independent beta random variables with the pdfs specified by (1) and (2), respectively. The beta approximation for $D = X_1 - X_2$ assumes that $D \approx 2X - 1$ with X taken to be a beta random variable with

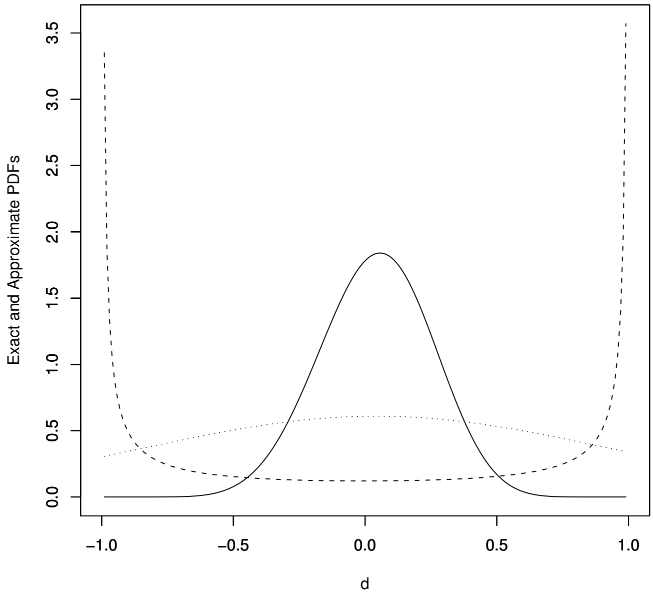

Fig. 1. The exact (solid curve), beta approximated (dashed curve) and normal approximated (dotted curve) pdfs of $D = X_1 - X_2$ for $a_1 = 0.9$, $b_1 = 0.8$, $a_2 = 0.7$ and $b_2 = 0.5$.

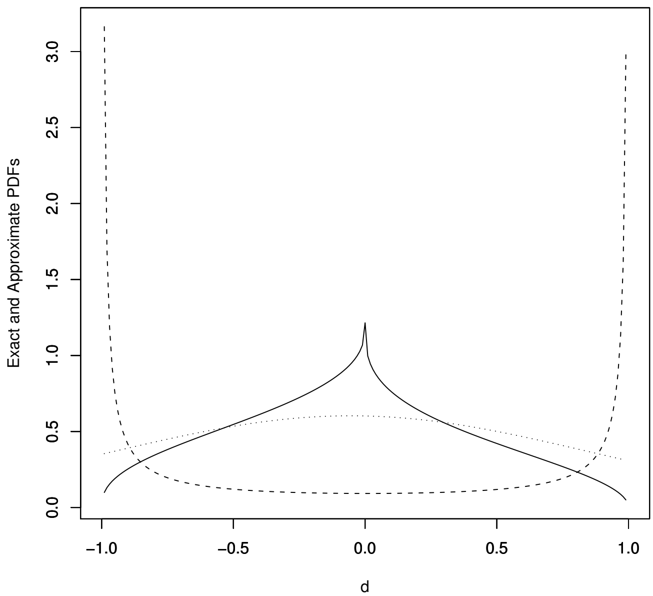

Fig. 2. The exact (solid curve), beta approximated (dashed curve) and normal approximated (dotted curve) pdfs of $D = X_1 - X_2$ for $a_1 = 0.9$, $b_1 = 0.8$, $a_2 = 0.7$ and $b_2 = 0.5$.

parameters, say, a and b. These parameters can be determined by setting $E(X_1) - E(X_2) = 2E(X) - 1$ and $\mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 4\,\mathrm{Var}(X)$, yielding

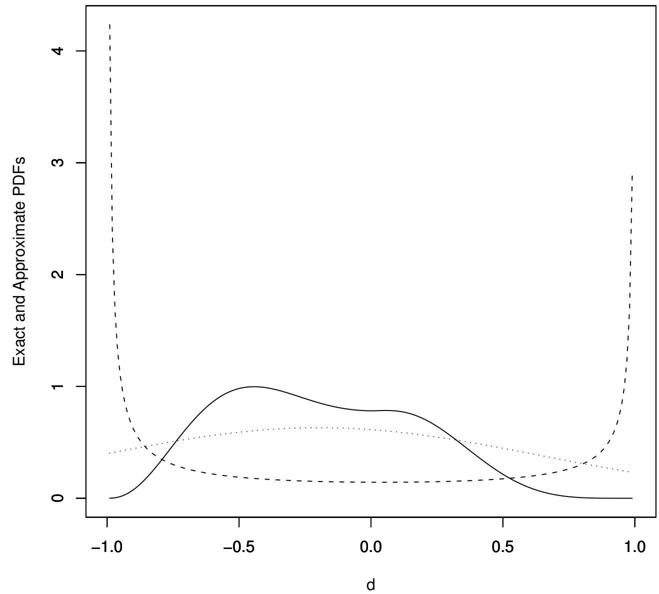

Fig. 3. The exact (solid curve), beta approximated (dashed curve) and normal approximated (dotted curve) pdfs of $D = X_1 - X_2$ for $a_1 = 0.6$, $b_1 = 0.8$, $a_2 = 5$ and $b_2 = 3$.

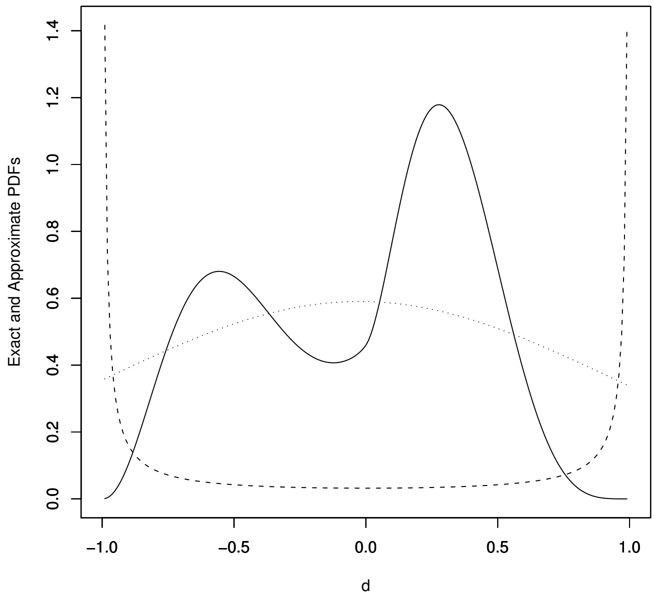

Fig. 4. The exact (solid curve), beta approximated (dashed curve) and normal approximated (dotted curve) pdfs of $D = X_1 - X_2$ for $a_1 = 0.3$, $b_1 = 0.2$, $a_2 = 5$ and $b_2 = 3$.

the solutions

$$a = D_1 \left\{ \frac{D_1(1 - D_1)}{D_2} - 1 \right\}, \qquad b = (1 - D_1) \left\{ \frac{D_1(1 - D_1)}{D_2} - 1 \right\},$$

where

$$D_1 = \frac{1}{2}\left\{\frac{a_1}{a_1+b_1} - \frac{a_2}{a_2+b_2} + 1\right\},$$
$$D_2 = \frac{1}{4}\left\{\frac{a_1 b_1}{(a_1+b_1)^2(a_1+b_1+1)} + \frac{a_2 b_2}{(a_2+b_2)^2(a_2+b_2+1)}\right\}.$$

The normal approximation for $D = X_1 - X_2$ assumes that $D \approx N(\mu, \sigma^2)$ truncated over the interval $[-1, 1]$. The parameters μ and σ can be obtained as the simultaneous solutions of the equations

$$\frac{1}{\Phi((1-\mu)/\sigma) - \Phi(-(1+\mu)/\sigma)} \int_{-1}^{1} x\phi\left(\frac{x-\mu}{\sigma}\right)dx = D_3,$$
$$\frac{1}{\Phi((1-\mu)/\sigma) - \Phi(-(1+\mu)/\sigma)} \int_{-1}^{1} x^2\phi\left(\frac{x-\mu}{\sigma}\right)dx = D_4,$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$
$$\Phi(x) = \int_{-\infty}^{x} \phi(y)dy,$$
$$D_3 = \frac{a_1}{a_1+b_1} - \frac{a_2}{a_2+b_2},$$
$$D_4 = \frac{a_1(a_1+1)}{a_1+b_1} + \frac{a_2(a_2+1)}{a_2+b_2} - \frac{2a_1a_2}{(a_1+b_1)(a_2+b_2)}.$$

The exact pdf of $D = X_1 - X_2$ is given by (3) of Theorem 1. Figures 1, 2, 3 and 4 show how the exact and approximate pdfs compare for some chosen values of a_1, b_1, a_2 and b_2. It is clear that both the approximations are very poor. The normal approximation appears to perform better than the beta approximation.

Acknowledgments. The author would like to thank the Editor and the referee for carefully reading the paper and for their comments which greatly improved the paper.

References

C. Armero and M. J. Bayarri (1994), *Prior assessments for prediction in queues*, J. Roy. Statist. Soc. Ser. D 43, 139–153.

A. Erdélyi, W. Magnus, F. Oberhettinger and F. G. Tricomi (1981), *Higher Transcendental Functions*, Volumes I–III, Krieger, Melbourne, FL.

M. B. Gordy (1988), *Computationally convenient distributional assumptions for common-value auctions*, Comput. Economics 12, 61–78.

I. S. Gradshteyn and I. M. Ryzhik (2000), *Tables of Integrals, Series, and Products*, 6th ed., Academic Press, San Diego.

N. N. Lebedev (1972), *Special Functions and Their Applications*. Dover Publ., New York.

D. L. Libby and M. R. Novick (1982), *Multivariate generalized beta distributions with applications to utility assessment*, J. Educat. Statist. 7, 271–294.

J. B. McDonald (1984), *Some generalized functions for the size distribution of income*, Econometrica 52, 647–664.

R. G. Newcombe (1998), *Interval estimation for the difference between independent proportions: Comparison of eleven methods*, Statist. Medicine 17, 873–890.

A. P. Prudnikov, Y. A. Brychkov and O. I. Marichev (1986), *Integrals and Series*, Vol. 1–3, Gordon and Breach, Amsterdam.

S. Radhakrishna, B. N. Murthy, N. G. K. Nair, P. Jayabal and R. Jayasri (1992), *Confidence-intervals in medical research*, Indian J. Medical Res. Sect. B 96, 199–205.

School of Mathematics
University of Manchester
Manchester M13 9PL, UK
E-mail: saralees.nadarajah@manchester.ac.uk

Received on 31.1.2008;
revised version on 3.11.2008 (1917)