Biometrika (2015), **xx**, x, *pp*. 1–14

Printed in Great Britain

On Distributions of Ratios

BY SIMON A. BRODA
Department of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
s.a.broda@uva.nl

AND RAYMOND KAN
Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street Toronto, Ontario, Canada M5S 3E6
kan@chass.utoronto.ca

SUMMARY

Inversion formulae are derived that express the density and distribution function of a ratio of random variables in terms of the joint characteristic function of numerator and denominator. The resulting expressions are amenable to numerical evaluation and lead to simple asymptotic expansions. The expansions reduce to known results when the denominator is almost surely positive. Their accuracy is demonstrated with numerical examples.

Some key words: Characteristic Function; Fieller–Creasy Problem; Inversion Formula; Saddlepoint Approximation.

1. INTRODUCTION

The distribution functions of many random variables do not permit an analytic representation. By contrast, the characteristic and moment generating functions are often more tractable. Results which express the distribution or density function in terms of these are commonly referred to as inversion formulae.

In many cases, even the characteristic function is intractable, but the statistic may have a stochastic representation in terms of random variables whose joint characteristic function is readily available. The most common case, and the subject of this paper, is that of a ratio. Many test statistics and estimators are of this form. The denominator of such ratios is often related to some form of sample variance and therefore positive. This situation is quite fortunate, because there exist inversion formulae that express the density and distribution function of such a statistic in terms of the joint characteristic function of numerator and denominator. There are, however, important situations that require tail probabilities for ratios of two random variables that both take values on the entire real line. For example, such probabilities arise in certain inferential procedures for ratios of means or regression coefficients. Known as the Fieller–Creasy problem, this type of inference is required in a variety of contexts, such as the inverse regression problem, slope ratio assay, parallel line assay, bioequivalence, and the estimation of long-run multipliers in econometrics. The present manuscript derives inversion formulae for these tail probabilities and the associated densities. We first show that the standard results for ratios with positive denominator continue to apply in a certain special case, even if both numerator and denominator may take values on the entire real line. Our second result is fully general, but involves a double integral that must typically be evaluated numerically, potentially hampering use of the result in applications. We therefore expand the integral in a uniform asymptotic series. The first few terms in the series, commonly referred to as saddlepoint approximations, are given explicitly. Saddlepoint approximations were introduced to statistics by Daniels (1954) and have found numerous applications. We do not attempt to provide a full bibliography here, but refer to Butler (2007). Daniels (1954, 1983) had already considered ratios of random variables, but his result is limited to the case with positive denominator. Inferential procedures that violate this assumption and hence require the results of the present paper include Bayesian methods (Press, 1969; Kappenman et al., 1970; Hunter & Lamboy, 1981), the unstudentized percentile bootstrap (Wu, 1986, Sec. 10; Davison & Hinkley, 1997, Sec. 5·7), and generalized confidence intervals (Lee & Lin, 2004; Bebu et al., 2009). The latter two will be considered as numerical examples below.

2. INVERSION FORMULAE

Consider a random variable X and let $F_X(x)$ and $\varphi_X(s)$ denote the associated distribution and characteristic functions. Gurland (1948) and Gil-Pelaez (1951) show that at every point of continuity of F_X,

$$F_X(x) = H(c) - \frac{1}{2\pi \mathrm{i}} \oint_c \varphi_X(s) e^{-\mathrm{i}sx} \frac{\mathrm{d}s}{s} = H(c) - \frac{1}{\pi} \int_{-c\mathrm{i}}^{-c\mathrm{i}+\infty} \mathrm{Im}\left\{ \frac{\varphi_X(s)}{s} e^{-\mathrm{i}sx} \right\} \mathrm{d}s, \quad (1)$$

where $c = 0$, $\mathrm{i} \equiv \sqrt{-1}$, $\oint_c \equiv \lim_{\epsilon \downarrow 0} \left(\int_{-c\mathrm{i}-\infty}^{-c\mathrm{i}-\epsilon} + \int_{-c\mathrm{i}+\epsilon}^{-c\mathrm{i}+\infty} \right)$, and $H(x) \equiv \{1 + \mathrm{sgn}(x)\}/2$ is the Heaviside function. Care must be taken in interpreting the integral sign in (1). Wendel (1961) has shown that depending on φ_X, the integral may fail to converge absolutely at either limit of integration. The weakest known condition for absolute convergence is $E\{\log(1 + |X|)\} < \infty$ (Rosén, 1961). Consequently, Gil-Pelaez relied on Riemann integrals in his derivation, and Gurland employed principal value integrals. If φ_X is analytic in a horizontal strip $\mathcal{S}$ containing, possibly as its boundary, the real axis, then one may choose $c \in \mathbb{R}$ such that $s - c\mathrm{i} \in \mathcal{S}, s \in \mathbb{R}$. A judicious choice will improve the convergence of the integral at the lower limit of integration. Shephard (1991) provides a multivariate generalization of (1). In the bivariate case, if (X, Y) has a finite mean and $\varphi_{X,Y}(s, t)$ is absolutely integrable, Shephard shows that

$$F_{X,Y}(x, y) = G_{x,y}(c_1, c_2) - \frac{1}{(2\pi)^2} \oint_{c_2} \oint_{c_1} \chi(s, t) \mathrm{d}s\mathrm{d}t \quad (2)$$

$$= G_{x,y}(c_1, c_2) - \frac{1}{2\pi^2} \int_{-c_2\mathrm{i}}^{-c_2\mathrm{i}+\infty} \int_{-c_1\mathrm{i}}^{-c_1\mathrm{i}+\infty} \mathrm{Re}\left\{\chi(s, t) + \chi(s, -t^*)\right\} \mathrm{d}s\mathrm{d}t, \quad (3)$$

where $c_1 = c_2 = 0$, t^* is the complex conjugate of t, $G_{x,y}(c_1, c_2) \equiv H(c_1)F_Y(y) + H(c_2)F_X(x) - H(c_1)H(c_2)$, and $\chi(s, t) \equiv \varphi_{X,Y}(s, t) \exp(-\mathrm{i}sx - \mathrm{i}ty)/(st)$. The assumption of a finite mean ensures that the integral in (3) converges absolutely, thus removing the need for principal value integrals as in the similar result of Gurland (1948). For analytic characteristic functions, the properties of the integral can be improved by an appropriate choice of c_1 and c_2 as in the univariate case. We will not assume analyticity of the characteristic function until Section 4 below, so we set $c_1 = c_2 = 0$ but provide a derivation of (3) in Appendix B.

Our interest is in the density and and distribution function of $R \equiv X/Y$. If Y is almost surely positive and r is not an atom of R, then

$$F_R(r) = \mathrm{pr}\,(R < r) = \mathrm{pr}\,(X < rY) = \mathrm{pr}\,(X - rY < 0) = \mathrm{pr}\,(W_r < 0)\,,$$

where $W_r \equiv X - rY$ and the subscript r in W_r will be suppressed below. Provided that $E\{\log(1 + |X - rY|)\} < \infty$, an application of (1) then shows that

$$F_R(r) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \operatorname{Im}\{\varphi_{X,Y}(s, -rs)\}\frac{\mathrm{d}s}{s}.$$

If in addition, the mean of Y is finite, then $\varphi_{X,Y}$ is differentiable with respect to its second argument (e.g., Cramér, 1946, p. 101). Let $\varphi_2(s,t) \equiv \partial/\partial t\, \varphi_{X,Y}(s,t)$. If $\varphi_2(s, -rs)$ is absolutely integrable, then the dominated convergence theorem implies that one may differentiate under the integral sign (e.g., Cramér, 1946, pp. 67f.), so that the density of R is

$$f_R(r) = \frac{1}{\pi}\int_0^\infty \operatorname{Im}\{\varphi_2(s, -rs)\}\,\mathrm{d}s. \tag{4}$$

Equation (4) was first established in Geary (1944).

A more general expression is needed when both X and Y can take values on the entire real line. One such result is given in Theorem 2 of Gurland (1948), which shows that if 0 is not an atom of Y and F_R is continuous at r, then with 1_A denoting the indicator function of the set A,

$$F_R(r) = \frac{1}{2} - \frac{1}{2\pi}\int_{-\infty}^\infty \operatorname{Im}\left\{\varphi^+(s,-rs) + \varphi^-(s,-rs)\right\}\frac{\mathrm{d}s}{s}, \quad \varphi^\pm(s,t) \equiv E\left(e^{\mathrm{i}sX+\mathrm{i}tY}1_{\pm Y>0}\right),$$

and principal values are to be taken if the integral fails to converge absolutely. The problem with this result is that explicit expressions for φ^+ and φ^- are typically not available. It is therefore preferable to express the density and distribution function of R in terms of $\varphi_{X,Y}$ directly. Two such results will be derived in the next section.

3. Two Inversion Theorems for Ratios

Our first result shows that (4) remains valid if some linear combination of X and Y is almost surely positive or negative, that is, if X and Y form a definite pair, defined as follows.

DEFINITION 1 (DEFINITE PAIR). *We call two real-valued random variables a definite pair if there exists $\beta \in \{\mathbb{R} \cup \infty\}$ such that* $\operatorname{pr}(X - \beta Y < 0) = \delta$, *for* $\delta \in \{0, 1\}$.

If X is positive with probability one, then $X - \beta Y$ is almost surely positive for $\beta = 0$, but it is less apparent that two random variables can form a definite pair even if both can take positive and negative values. As a simple example, let $X = 2Z_1^2 - Z_2^2$ and $Y = Z_1^2 - 2Z_2^2$, where Z_1 and Z_2 are independent standard Gaussian, so that $W = X - rY = (2 - r)Z_1^2 + (2r - 1)Z_2^2$. Then $\operatorname{pr}(W < 0) = 0$ for $1/2 \le r \le 2$. Our result is based on the following identity.

LEMMA 1. *If X and Y form a definite pair such that* $\operatorname{pr}(X - \beta Y < 0) = \delta$ *for* $\delta \in \{0, 1\}$, *and if* 0 *is not an atom of Y, then*

$$\operatorname{pr}(R < r) = 2\delta H(r - \beta) + (1 - 2\delta)\Big\{\operatorname{pr}(Y < 0) + \operatorname{sgn}(r - \beta)\operatorname{pr}(W < 0)\Big\},$$

where $R \equiv X/Y$ and $H(x) \equiv \{1 + \operatorname{sgn}(x)\}/2$ is the Heaviside function.

Proof. Appendix B. □

The following result follows at once.

THEOREM 1. *If X and Y form a definite pair such that* $\operatorname{pr}(X - \beta Y < 0) = \delta$ *for* $\delta \in \{0, 1\}$, 0 *is an atom of neither Y nor $W \equiv X - rY$,* $E\{\log(1 + |Y|)\} < \infty$, *and* $E\{\log(1 + |W|)\} <$

∞, then

$$F_R(r) = H(r-\beta) - \frac{(1-2\delta)}{\pi}\int_0^\infty \mathrm{Im}\left\{\varphi_{X,Y}(0,s) + \mathrm{sgn}(r-\beta)\varphi_{X,Y}(s,-rs)\right\}\frac{\mathrm{d}s}{s}. \quad (5)$$

If in addition, Y has a finite mean and $\varphi_2(s,-rs)$ is absolutely integrable, then R has density

$$f_R(r) = \frac{\mathrm{sgn}(r-\beta)}{\pi(2\delta-1)}\int_0^\infty \mathrm{Im}\left\{\varphi_2(s,-rs)\right\}\mathrm{d}s = \left|\frac{1}{\pi}\int_0^\infty \mathrm{Im}\left\{\varphi_2(s,-rs)\right\}\mathrm{d}s\right|. \quad (6)$$

Proof. Observe that r is an atom of R if and only if 0 is an atom of W. For the distribution function, use (1) in Lemma 1, together with $\varphi_{W,Y}(s,t) = \varphi_{X,Y}(s,t-rs)$. For the density, finiteness of the mean guarantees the existence of φ_2. The dominated convergence theorem implies that one may differentiate under the integral sign if $\varphi_2(s,-rs)$ is absolutely integrable. □

Our second result provides general inversion formulae for ratios that remain valid when (5) and (6) fail. We start from the following observation.

LEMMA 2. *If 0 is an atom of neither Y nor $W \equiv X - rY$, then*

$$F_R(r) = \mathrm{pr}\left(W<0\right) + \mathrm{pr}\left(Y<0\right) - 2\,\mathrm{pr}\left(W<0, Y<0\right). \quad (7)$$

Proof. Observe that r is an atom of R if and only if 0 is an atom of W. Hence

$$\begin{aligned} F_R(r) &= \mathrm{pr}\left(X>rY, Y<0\right) + \mathrm{pr}\left(X<rY, Y>0\right) \\ &= \mathrm{pr}\left(W<0\right) + \mathrm{pr}\left(Y<0\right) - 2\,\mathrm{pr}\left(W<0, Y<0\right). \end{aligned}$$ □

We then have the following result.

THEOREM 2. *If (X,Y) has a finite mean, $\varphi_{X,Y}$ is absolutely integrable, and 0 is not an atom of $W \equiv X - rY$, then for $|r| < \infty$,*

$$F_R(r) = \frac{1}{2} + \frac{1}{\pi^2}\int_0^\infty\int_0^\infty \mathrm{Re}\left\{\varphi_{X,Y}(s,t-rs) - \varphi_{X,Y}(s,-t-rs)\right\}\frac{\mathrm{d}s}{s}\frac{\mathrm{d}t}{t}$$

and

$$f_R(r) = \frac{1}{\pi^2}\int_0^\infty\int_{-\infty}^\infty \mathrm{Re}\left\{\varphi_2(s,-t-rs)\right\}\mathrm{d}s\frac{\mathrm{d}t}{t} \quad (8)$$

whenever this integral converges absolutely.

Proof. The assumptions imply that (W,Y) has a finite mean, and that its characteristic function $\varphi_{W,Y} = \varphi_{X,Y}(s,t-rs)$ is absolutely integrable. Hence (3) applies. Combining it with (7) completes the proof for the distribution function. Integrability of $\varphi_{W,Y}(s,t)$ ensures that R has a density. By dominated convergence,

$$f_R(r) = \frac{1}{\pi^2}\int_0^\infty\int_0^\infty \mathrm{Re}\left\{\varphi_2(s,-t-rs) - \varphi_2(s,t-rs)\right\}\mathrm{d}s\frac{\mathrm{d}t}{t}$$

whenever the integral is absolutely convergent. The result follows upon noting that $\varphi_{X,Y}(s,t)$ and $\varphi_{X,Y}(-s,-t)$ are complex conjugates, so that $-\mathrm{Re}\left\{\varphi_2(s,t-rs)\right\} = \mathrm{Re}\left\{\varphi_2(-s,-t+rs)\right\}$. □

4. Saddlepoint Approximation

4·1. *Density Approximation*

The first step in deriving a saddlepoint approximation to the density is to rewrite (8) in a form amenable to saddlepoint methods, as follows.

Lemma 3. *Suppose that X and Y have a joint density, and that their joint cumulant generating function $\mathbb{K}(s,t) \equiv \log E\{\exp(sX+tY)\}$ converges on the open set $\mathcal{T} \ni (0,0)$. Let $\bar{X}$ and $\bar{Y}$ denote the mean of n independent copies of X and Y, respectively. Then the density of $R \equiv \bar{X}/\bar{Y}$ is*

$$f_R^n(r) = 2I_2 - \mathrm{sgn}(c) I_1(0), \quad (0,c) \in \mathcal{T} \setminus (0,0), \tag{9}$$

where

$$I_1(t) \equiv \frac{n}{2\pi\mathrm{i}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} e^{n\mathbb{K}(s,t-rs)} \mathbb{K}_2(s,t-rs)\mathrm{d}s, \quad I_2 \equiv \frac{1}{2\pi\mathrm{i}} \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} I_1(t) \frac{\mathrm{d}t}{t}, \tag{10}$$

and $\mathbb{K}_i(\cdot,\cdot)$ denotes the derivative of $\mathbb{K}(s,t)$ with respect to its ith argument.

Proof. From (7), the distribution function of R is, in obvious notation,

$$F_R^n(r) = F_{\bar{W}}^n(0) + F_{\bar{Y}}^n(0) - 2F_{\bar{W},\bar{Y}}^n(0,0), \tag{11}$$

where $\bar{W} \equiv \bar{X} - r\bar{Y}$. The joint characteristic function of $(\bar{W},\bar{Y})$ is $\exp\{n\mathbb{K}(\mathrm{i}s, \mathrm{i}t - \mathrm{i}rs)\}$. Using it in (2) with $c_1, c_2 < 0$ such that $(c_1, c_2 - rc_1) \in \mathcal{T}$ and changing variables,

$$F_{\bar{W},\bar{Y}}^n(0,0) = \left(\frac{1}{2\pi\mathrm{i}}\right)^2 \int_{c_2-\mathrm{i}\infty}^{c_2+\mathrm{i}\infty} \int_{c_1-\mathrm{i}\infty}^{c_1+\mathrm{i}\infty} e^{n\mathbb{K}(s,t-rs)} \frac{\mathrm{d}s}{s}\frac{\mathrm{d}t}{t}.$$

Similarly, using the left member of (1) with $c_3 < 0$ such that $(c_3, -rc_3) \in \mathcal{T}$,

$$F_{\bar{W}}^n(0) = -\frac{1}{2\pi\mathrm{i}} \int_{c_3-\infty}^{c_3+\infty} e^{n\mathbb{K}(s,-rs)} \frac{\mathrm{d}s}{s},$$

and a similar expression for $F_{\bar{Y}}^n(0)$. Differentiating (11), the density of R is obtained as $f_R^n(r) = I_1(0) + 2I_2$ with $c = c_2$, and where we have set $c_1 = c_3 = 0$, which is permissible because differentiation has removed the pole at $s = 0$. If $c_2 > 0$, then the residue at the origin must be subtracted, which is precisely $I_1(0)$. □

The plan is to apply a standard Laplace approximation to $I_1(t)$, and then approximate I_2 by a saddlepoint approximation, modified as in Bleistein (1966) to accommodate the pole at the origin. Let $c_r \equiv (1,-r)^{\mathrm{T}}$. For any r such that a solution $\tilde{s}$, called the inner sadddlepoint, to

$$\mathbb{K}_1(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_2(\tilde{s}, t - r\tilde{s}) = 0 \tag{12}$$

exists, applying a standard Laplace approximation to $I_1(t)$ yields

$$I_1(t) = \left(\frac{n}{2\pi}\right)^{1/2} e^{nh(t)} g_0(t) \left\{1 + \mathcal{O}\left(n^{-1}\right)\right\}, \tag{13}$$

where $h(t) \equiv \mathbb{K}(\tilde{s}, t - r\tilde{s})$, $g_0(t) = \mathbb{K}_2(\tilde{s}, t - r\tilde{s})/\{c_r^{\mathrm{T}}\mathbb{K}''(\tilde{s}, t - r\tilde{s})c_r\}^{1/2}$, and $\mathbb{K}''(\cdot,\cdot) = \{\mathbb{K}_{ij}(\cdot,\cdot)\}$ denotes the Hessian of $\mathbb{K}$. To approximate I_2, we will need the following result.

Lemma 4 (Bleistein, 1966). *If $g_0(t)$ and $h(t)$ are real functions of t, analytic in a strip containing $c \neq 0$ and the imaginary axis, and $h(t)$ has a unique saddle point $\hat{t}_r \neq 0$ on the real* *axis in the interior of this strip, then*

$$\frac{1}{2\pi\mathrm{i}}\int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} g_0(t)e^{nh(t)}\frac{\mathrm{d}t}{t} = e^{nh(0)}g_0(0)\Big\{1_{c>0} - \Phi(\hat{w}\sqrt{n})\Big\} + \frac{e^{nh(\hat{t}_r)}}{(2\pi n)^{1/2}}\left\{\frac{g_0(\hat{t}_r)}{\hat{u}} - \frac{g_0(0)}{\hat{w}} + \mathcal{O}\left(n^{-1}\right)\right\},$$

where $\Phi(\cdot)$ *is the standard normal distribution function,* $\hat{w} \equiv \mathrm{sgn}(\hat{t}_r)\left[-2\big\{h(\hat{t}_r) - h(0)\big\}\right]^{1/2}$, $\hat{u} \equiv \hat{t}_r\{h''(\hat{t}_r)\}^{1/2}$*, and for each* r*, the saddlepoint* $\hat{t}_r$ *solves* $h'(\hat{t}_r) = 0$.

To apply the result, we require the first and second derivatives of $h(t)$. By virtue of (12),

$$h'(t) = \tilde{s}'(t)\left\{\mathbb{K}_1(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_2(\tilde{s}, t - r\tilde{s})\right\} + \mathbb{K}_2(\tilde{s}, t - r\tilde{s}) = \mathbb{K}_2(\tilde{s}, t - r\tilde{s}), \qquad (14)$$

where $\tilde{s}'(t)$ denotes the derivative of $\tilde{s}$ with respect to t. This is found by differentiating (12), which yields $\tilde{s}'(t) = -\left\{\mathbb{K}_{12}(\tilde{s}, t - r\tilde{s}) - r\mathbb{K}_{22}(\tilde{s}, t - r\tilde{s})\right\}/\{c_r^{\mathrm{T}}\mathbb{K}''(\tilde{s}, t - r\tilde{s})c_r\}$. With this, the second derivative evaluates to $h''(t) = |\mathbb{K}''(\tilde{s}, t - r\tilde{s})|/\{c_r^{\mathrm{T}}\mathbb{K}''(\tilde{s}, t - r\tilde{s})c_r\}$. The saddlepoint $\hat{t}_r$ is found by equating (14) to zero. Equivalently, $\hat{t}_r = \hat{t} + r\hat{s}$, where the outer saddlepoint $(\hat{s}, \hat{t})$ solves the system

$$\mathbb{K}'\left(\hat{s}, \hat{t}\right) \equiv \left\{\mathbb{K}_1(\hat{s}, \hat{t})\ \mathbb{K}_2(\hat{s}, \hat{t})\right\}^{\mathrm{T}} = 0. \qquad (15)$$

To apply the lemma, we assume that $\hat{t}_r \neq 0$ so that $\hat{t} \neq -r\hat{s}$ for the remainder of the proof; the other case will be dealt with separately. It is further observed that $(\hat{s}, \hat{t})$ is independent of r so that this system needs only be solved once for any given cumulant generating function, and that (15) implies that $g_0(\hat{t}_r) = \mathbb{K}_2(\hat{s}, \hat{t})/\{c_r^{\mathrm{T}}\mathbb{K}''(\hat{s}, \hat{t})c_r\}^{1/2} = 0$. Let $\tilde{s}_0 \equiv \tilde{s}(0)$, i.e., the inner saddlepoint corresponding to $t = 0$, and define $\tilde{w}_0 \equiv \mathrm{sgn}\left(\tilde{s}_0\right)\left\{-2\mathbb{K}\left(\tilde{s}_0, -r\tilde{s}_0\right)\right\}^{1/2}$ and $\tilde{g}_0 \equiv g_0(0) = \mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)/\{c_r^{\mathrm{T}}\mathbb{K}''(\tilde{s}_0, -r\tilde{s}_0)c_r\}^{1/2}$. Then

$$I_2 = \sqrt{n}\phi(\tilde{w}_0\sqrt{n})\tilde{g}_0\left\{1_{c>0} - \Phi(\hat{w}\sqrt{n}) - \phi(\hat{w}\sqrt{n})/(\hat{w}\sqrt{n}) + \mathcal{O}\left(n^{-1}\right)\right\},$$

where ϕ is the standard normal density. Combining the two approximations according to (9) produces the desired result.

THEOREM 3. *Suppose that* X *and* Y *have a joint density with respect to Lebesgue measure on* $\mathbb{R}^2$*, and that their joint cumulant generating function* $\mathbb{K}(s, t) \equiv \log E\left\{\exp(sX + tY)\right\}$ *converges on the open set* $\mathcal{T} \ni (0, 0)$*, with gradient* $\mathbb{K}'(s, t)$ *and Hessian* $\mathbb{K}''(s, t)$*. Let* $\bar{X}$ *and* $\bar{Y}$ *denote the mean of* n *independent copies of* X *and* Y*, respectively. If there exists a solution* $(\hat{s}, \hat{t})$ *to* $\mathbb{K}'\left(\hat{s}, \hat{t}\right) = 0$ *such that* $\hat{t} \neq -r\hat{s}$*, and a solution* $\tilde{s}_0$ *to*

$$c_r^{\mathrm{T}}\mathbb{K}'\left(\tilde{s}_0, -r\tilde{s}_0\right) = 0, \qquad c_r \equiv (1, -r)^{\mathrm{T}}, \qquad (16)$$

then the density of the ratio $R \equiv \bar{X}/\bar{Y}$ *is* $f_R^n(r) = \hat{f}_n^{(1)}(r)\left\{1 + \mathcal{O}\left(n^{-1}\right)\right\}$*, where*

$$\hat{f}_n^{(1)}(r) \equiv \sqrt{n}\phi(\tilde{w}_0\sqrt{n})\tilde{g}_0\left[1 - 2\left\{\Phi(\hat{w}\sqrt{n}) + \phi(\hat{w}\sqrt{n})/(\hat{w}\sqrt{n})\right\}\right], \qquad (17)$$

$$\tilde{g}_0 \equiv \mathbb{K}_2(\tilde{s}_0, -r\tilde{s}_0)/\left\{c_r^{\mathrm{T}}\mathbb{K}''(\tilde{s}_0, -r\tilde{s}_0)c_r\right\}^{1/2}, \qquad \tilde{w}_0 \equiv \mathrm{sgn}\left(\tilde{s}_0\right)\left\{-2\mathbb{K}\left(\tilde{s}_0, -r\tilde{s}_0\right)\right\}^{1/2},$$

$$\hat{w} \equiv \mathrm{sgn}\left(\hat{t} + r\hat{s}\right)\left[-2\left\{\mathbb{K}\left(\hat{s}, \hat{t}\right) - \mathbb{K}\left(\tilde{s}_0, -r\tilde{s}_0\right)\right\}\right]^{1/2}.$$

Higher order approximations are provided in the Supplementary Material.

A few remarks are in order. First, it can be verified that the approximation is exact when X and Y are jointly Gaussian. Second, the approximate density is always non-negative. These two statements are proven in the Supplementary Material. Third, the term in front of the square brackets, and thus the approximation for I_1, is the standard saddlepoint approximation derived in Daniels (1954) for the case with $\text{pr}\,(Y < 0) = 0$. One may therefore interpret the term in curly braces as a correction for cases in which this requirement fails. Indeed, if X and Y form a definite pair, then $\hat{w}$ diverges to $\pm\infty$, and the two approximations coincide in absolute value. Since the term in square brackets is always greater than unity, the correction is, in general, upwards; hence using the absolute value of Daniels's approximation when it is not applicable will tend to underestimate the density. Fourth, it is seen that (17) is undefined when $\hat{w} = 0$, which happens whenever $\hat{t} = -r\hat{s}$. This singularity is, however, removable. Two cases can be distinguished: (i) $\mu_X \equiv E\,(X) \neq 0$ or $\mu_Y \equiv E\,(Y) \neq 0$ or both, so that $(\hat{s}, \hat{t}) \neq (0, 0)$. Then $f_n(r)$ has a removable singularity at $r_0 \equiv -\hat{t}/\hat{s}$, and the limiting value is

$$\hat{f}_n^{(1)}(r_0) = (2/\pi)^{1/2}\phi(\tilde{w}_0\sqrt{n})|\mathbb{K}''(\hat{s},\hat{t})|^{1/2}/\{c_{r_0}^{\mathrm{T}}\mathbb{K}''(\hat{s},\hat{t})c_{r_0}\}. \tag{18}$$

If, on the other hand, (ii) $\mu_X = \mu_Y = 0$, then $\hat{s} = \hat{t} = 0$. Consequently (17) is undefined for all r and should be replaced by the limit

$$\hat{f}_n^{(1)}(r) = \pi^{-1}|\Sigma|^{1/2}/(c_r^{\mathrm{T}}\Sigma c_r), \tag{19}$$

where $\Sigma \equiv \mathbb{K}''(0, 0)$ is the covariance matrix of (X, Y). The accuracy of the approximation is reduced to $f_R^n(r) = \hat{f}_n^{(1)}(r)\{1 + \mathcal{O}(n^{-1/2})\}$ in this case, because in (17), the $\mathcal{O}(1)$ term $1 - 2\Phi(\hat{w}\sqrt{n})$ vanishes between the square brackets. Furthermore, the right hand side of (19) is the density of a ratio of two correlated Gaussians, so that the density of $\bar{X}/\bar{Y}$ is approximated by that of a ratio of Gaussians with matching first and second moments, which is correct to the order stated. The asymptotic distribution in the non-zero mean case is quite different: suppose that $\mu_Y \neq 0$ and let $\lambda \equiv \mu_X/\mu_Y$. It is a standard result (see, e.g., Fuller, 1990, Theorem 1·3·7) that $\sqrt{n}\mu_Y(R - \lambda) \to \mathrm{N}(0, c_\lambda^{\mathrm{T}}\Sigma c_\lambda)$ in distribution. In approximation (17), the term in curly braces tends to unity as $n \to \infty$ for fixed r, so that the approximation will converge to that derived in Daniels (1954), and hence ultimately to a Gaussian density. The case with $\mu_Y = 0, \mu_X \neq 0$ can be treated by considering R^{-1}. It can be verified longhand that the saddlepoint approximation to the density of R^{-1}, $\hat{g}_n^{(1)}(r)$, say, obeys the symmetry relation $\hat{g}_n^{(1)}(r) = \hat{f}_n^{(1)}\left(r^{-1}\right)/r^2$.

4·2. *Tail Probability Approximation*

As discussed in the Supplementary Material, the method of proof used for the density approximation does not extend to tail probabilities. Instead, we rely on a result from Kolassa (2003) to approximate each probability in (11) separately. Kolassa's result is a multivariate version of the approximation of Hauschildt (1969) and Robinson (1982). After correcting a typographical error, it reads as follows.

THEOREM 4 (KOLASSA, 2003). *Suppose the d-dimensional random vector X has a density and a joint cumulant generating function $\mathbb{K}(\tau) \equiv \log E\,\{\exp(\tau^{\mathrm{T}}X)\}$ with gradient $\mathbb{K}'(\tau)$, Hessian $\mathbb{K}''(\tau)$, and third order derivatives $\mathbb{K}_{ijk}(\tau) \equiv \partial^3/(\partial\tau_i\partial\tau_j\partial\tau_k)\mathbb{K}(\tau),\ i, j, k \in \{1, \ldots, d\}$. Choose a compact subset $\mathcal{C}$ of the range of $\mathbb{K}'(\tau)$. Denote by $\bar{X}$ the mean of n independent copies of X, and for fixed $\bar{x} \in \mathcal{C}$, let the saddlepoint $\hat{\tau}$ solve $\mathbb{K}'(\hat{\tau}) = \bar{x}$. Then, provided that $\hat{\tau} > 0$,*

$$\mathrm{pr}\left(\bar{X} > \bar{x}\right) = e^{n(\widehat{\mathbb{K}} - \hat{\tau}^{\mathrm{T}}\bar{x})}\Bigg[e^{n\hat{\tau}^{\mathrm{T}}\widehat{\mathbb{K}}''\hat{\tau}/2}\Bigg\{I(0, n\widehat{\mathbb{K}}'', \hat{\tau}) + \frac{n}{6}\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d}\mathbb{K}_{ijk}(\hat{\tau})I(e_i + e_j + e_k, n\widehat{\mathbb{K}}'', \hat{\tau})\Bigg\} + \mathcal{O}(n^{-1})\Bigg].$$

Here, $\widehat{\mathbb{K}}$ and $\widehat{\mathbb{K}}''$ denote the cumulant generating function and its Hessian evaluated at $\hat{\tau}$, e_j is a $1 \times d$ vector with all components zero except for a 1 at position j, and, for Σ a positive definite matrix and $m = (m_1, \ldots, m_d)$,

$$I(m, \Sigma, \hat{\tau}) \equiv \frac{1}{(2\pi \mathrm{i})^d}\int_{\hat{\tau}-\mathrm{i}\infty}^{\hat{\tau}+\mathrm{i}\infty} e^{\tau^{\mathrm{T}}\Sigma\tau/2 - \tau^{\mathrm{T}}\Sigma\hat{\tau}}\prod_{j=1}^{d}\frac{(\tau_j - \hat{\tau}_j)^{m_j}}{\tau_j}\mathrm{d}\tau. \tag{20}$$

In Theorem 4 and below, vector inequalities are to be interpreted as holding elementwise. Applying the theorem requires a means of evaluating the function I. Kolassa provides a recursive algorithm for this purpose, which expresses I in terms of the multivariate normal distribution and its derivatives. Appendix A presents explicit expressions for the relevant cases when $d = 1$ and $d = 2$. Kolassa defines the function I only for $\hat{\tau} > 0$; when some elements of $\hat{\tau}$ are zero, it can be defined as the appropriate limit. For our purposes, it will prove convenient to also allow $\hat{\tau}_j < 0$ for some or all j. Let D denote a diagonal matrix, with elements $d_{jj} = 1$ if $\hat{\tau}_j \geq 0$ and $d_{jj} = -1$ if $\hat{\tau}_j < 0$. Then the following relationship holds.

$$I(m, \Sigma, \hat{\tau}) = I(m, D\Sigma D, D\hat{\tau})\prod_{j=1}^{d} d_{jj}^{m_j+1}. \tag{21}$$

All elements of $D\hat{\tau}$ are nonnegative, so the expressions in Appendix A apply. When applied to the present problem, the following result is obtained.

THEOREM 5. *Under the conditions of Theorem 3, and if there exists $\check{t}_0$ such that $\mathbb{K}_2(0, \check{t}_0) = 0$, then $F_R^n(r) = \hat{F}_n^{(1)}(r) + \mathcal{O}\left(n^{-1}\right)$, where*

$$\hat{F}_n^{(1)}(r) \equiv \mathrm{sgn}_0(\hat{t}_r)\left\{P_1 - H_0(\tilde{s}_0)\right\} + \mathrm{sgn}_0(\hat{s})\left\{P_2 - H_0(\check{t}_0)\right\} - 2\left\{P_3 - H_0(\hat{t}_r)H_0(\hat{s})\right\},$$

$$P_1 \equiv e^{n\left(\tilde{\kappa}_0^{(0)} + \tilde{s}_0^2\tilde{\kappa}_0^{(2)}/2\right)}\left\{I(0, n\tilde{\kappa}_0^{(2)}, \tilde{s}_0) + n\tilde{\kappa}_0^{(3)}I(3, n\tilde{\kappa}_0^{(2)}, \tilde{s}_0)/6\right\},$$

$$P_2 \equiv e^{n\left(\check{\kappa}_0^{(0)} + \check{t}_0^2\check{\kappa}_0^{(2)}/2\right)}\left\{I(0, n\check{\kappa}_0^{(2)}, \check{t}_0) + n\check{\kappa}_0^{(3)}I(3, n\check{\kappa}_0^{(2)}, \check{t}_0)/6\right\},$$

$$P_3 \equiv e^{n\left(\hat{\kappa}^{(0,0)} + \hat{\tau}^{\mathrm{T}}\widehat{\mathbb{K}}''\hat{\tau}/2\right)}\left\{\hat{I}_{0,0} + \frac{n}{6}\sum_{j=0}^{3}\binom{3}{j}\hat{\kappa}^{(3-j,j)}\hat{I}_{3-j,j}\right\},$$

$H_0(s) \equiv 1_{s\geq 0}$, $\mathrm{sgn}_0(s) \equiv 2H_0(s) - 1$, $\tilde{s}_0$ is as in (16)*, $(\hat{s}, \hat{t})$ is as in* (15)*, $\hat{t}_r \equiv \hat{t} + r\hat{s}$, $\hat{\tau}_r \equiv (\hat{s}, \hat{t}_r)$, $\hat{\tau} = (\hat{s}, \hat{t})$, $\hat{\mathcal{K}} \equiv [(\hat{\kappa}^{(2,0)}, \hat{\kappa}^{(1,1)})^{\mathrm{T}}, (\hat{\kappa}^{(1,1)}, \hat{\kappa}^{(0,2)})^{\mathrm{T}}]$, $\widehat{\mathbb{K}}'' \equiv \mathbb{K}''(\hat{s}, \hat{t})$, $\check{\kappa}_0^{(j)} \equiv \mathbb{K}_{2^j}(0, \check{t}_0)$,*

$$\tilde{\kappa}_0^{(i)} \equiv \sum_{k=0}^{i}\binom{i}{k}(-r)^k\mathbb{K}_{1^{i-k}2^k}(\tilde{s}_0, -r\tilde{s}_0), \qquad \hat{\kappa}^{(i,j)} \equiv \sum_{k=0}^{i}\binom{i}{k}(-r)^k\mathbb{K}_{1^{i-k}2^{j+k}}(\hat{s}, \hat{t}),$$

$\mathbb{K}_{1^i2^j}(s, t) \equiv \partial^{i+j}\mathbb{K}(s, t)/\partial s^i\partial t^j$, $\hat{I}_{i,j} \equiv I\{(i, j), n\hat{\mathcal{K}}, \hat{\tau}_r\}$, and explicit expressions for evaluating the function I are given in Appendix A.

Proof. See Appendix B. □

Similar to the density approximation, it can be verified that the tail probability approximation in Theorem 5 is exact if X and Y are jointly Gaussian.

5. Applications

5·1. *Generalized Confidence Intervals for Ratios of Normal Means and Regression Coefficients*

Let $\{u_i\}_{i=1}^{n_U}$ and $\{v_i\}_{i=1}^{n_V}$ denote unpaired samples from two normal populations with unknown means μ_U and μ_V and unknown variances. Inference is desired for $\theta \equiv \mu_U/\mu_V$. This situation arises in direct assays. Lee & Lin (2004) derive a generalized confidence interval, in the sense of Weerahandi (1993), for θ. Unlike Fieller's (1940) procedure, their approach allows for heteroskedasticity. A $100(1-\alpha)\%$ equi-tailed interval is $(r_{\alpha/2}, r_{1-\alpha/2})$, with r_α the α quantile of $R \equiv (a+bT_1)/(c+dT_2)$, where $a = \bar{u} \equiv n_U^{-1}\sum_{i=1}^{n_U} u_i$, $b = -s_U/\sqrt{\nu_1}$, $c = \bar{v} \equiv n_V^{-1}\sum_{i=1}^{n_V} v_i$, $d = -s_V/\sqrt{\nu_2}$, $s_U^2 \equiv n_U^{-1}\sum_{i=1}^{n_U}(u_i-\bar{u})^2$, $s_V^2 \equiv n_V^{-1}\sum_{i=1}^{n_V}(v_i-\bar{v})^2$, $T_1 \equiv Z_1(X_1/\nu_1)^{-1/2}$, $T_2 \equiv Z_2(X_2/\nu_2)^{-1/2}$, $\nu_1 \equiv n_U - 1$, $\nu_2 \equiv n_V - 1$, Z_1, Z_2 are independent standard normal variates, and X_1, X_2 are distributed as $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ respectively, independently from each other and the Z_i. Thus T_1 and T_2 are independently distributed as Student's t with respective degrees of freedom ν_1 and ν_2. The Student's t distribution does not admit a moment generating function, but we may write $R = X/Y$, where $X \equiv [a\{X_1X_2/(\nu_1\nu_2)\}^{1/2} + bZ_1(X_2/\nu_2)^{1/2}]$ and $Y \equiv [c\{X_1X_2/(\nu_1\nu_2)\}^{1/2} + dZ_2(X_1/\nu_1)^{1/2}]$. As shown in the Supplementary Material, the joint moment generating function of X and Y is $\mathbb{M}(s,t) \equiv \exp\{\mathbb{K}(s,t)\} = (\nu_1/\omega_1)^{\nu_1/2}(\nu_2/\omega_2)^{\nu_2/2}\Gamma\{(\nu_1+1)/2\}\Gamma\{(\nu_2+1)/2\}{}_2\mathrm{F}_1(\nu_1,\nu_2;\{\nu_1+\nu_2+1\}/2;z)/[\Gamma(1/2)\Gamma\{(\nu_1+\nu_2+1)/2\}]$, where $\omega_1 \equiv \nu_1 - d^2t^2$, $\omega_2 \equiv \nu_2 - b^2s^2$, and $z \equiv (as+ct)/(4\omega_1\omega_2)^{1/2} + 1/2$, defined for all (s,t) such that $\omega_1, \omega_2 > 0$ and $z < 1$. Here ${}_2\mathrm{F}_1(\cdot,\cdot;\cdot;\cdot)$ denotes the Gauss hypergeometric function. This can be evaluated efficiently based on the algorithm given in the Supplementary Material. It is not clear how (X,Y) could be written as a mean of independent copies of identically distributed random variables, so we apply the approximations formally with $n=1$. Nevertheless, and as demonstrated in Figure 1, the accuracy of the approximation improves as $\nu_1, \nu_2 \longrightarrow \infty$ and the distribution of (X,Y) tends to a Gaussian, for which the saddlepoint approximation is exact. The second order approximation offers little improvement here and is hence not shown.

For illustration, we consider the data given in Finney (1978, Table 2.3.1), for which $\bar{u} = 19{\cdot}9$, $s_U^2 = 68{\cdot}15$, $\bar{v} = 16{\cdot}8$, $s_V^2 = 75{\cdot}87$, and $n_U = n_V = 7$, so that $\mathrm{pr}(Y<0) = 0{\cdot}162\%$. Note that $\mathrm{pr}(Y<0)$ is the p-value of a t-test for $H_0: \mu_V = 0$ against $H_1: \mu_V > 0$. In this direct assay, the data represent fatal doses, in microlitres per kilogram of cats' body weight, of two tinctures of strophanthus. The generalized confidence intervals for μ_U/μ_V based on exact calculations and the saddlepoint approximation are, respectively, $(0{\cdot}597, 2{\cdot}574)$ and $(0{\cdot}614, 2{\cdot}490)$.

Bebu et al. (2009) provide an extension of the method to ratios of regression coefficients, as required in slope ratio and parallel line assays. Let $\hat{\beta}_i$ and $\hat{\sigma}_i$, $i \in \{1,\ldots,p\}$, denote the ordinary least squares estimates and associated standard errors from a linear regression with errors $(u_1,\ldots,u_n)^{\mathrm{T}}$ distributed as $\mathrm{N}(0,\sigma^2 I_n)$. Bebu et al. show that a $100(1-\alpha)\%$ equi-tailed generalized confidence interval for β_i/β_j can be constructed as above, but with r_α now representing the α quantile of $R \equiv (\hat{\beta}_i - \hat{\sigma}_i\tilde{T}_1)/(\hat{\beta}_j - \hat{\sigma}_j\tilde{T}_2)$, where $(\tilde{T}_1, \tilde{T}_2)$ is distributed as bivariate Student's t with $\nu \equiv n-p$ degrees of freedom and correlation ρ equal to the usual least squares estimate of $\mathrm{corr}(\hat{\beta}_i, \hat{\beta}_j)$. The assumption of elliptical, rather than independent, random variables simplifies the analysis considerably. The exact density of R has been obtained in Press (1969) in the context of Bayesian analysis of the same problem. As shown there, it suffices to consider the case with $\rho = 0$, as the distribution of R is the same as that of $\tilde{R} + q$, where $\tilde{R} \equiv \{\tilde{a} + \tilde{b}Z_1(X/\nu)^{-1/2}\}/\{\tilde{c} + \tilde{d}Z_2(X/\nu)^{-1/2}\}$, $q \equiv \rho\hat{\sigma}_i/\hat{\sigma}_j$, $\tilde{a} \equiv \hat{\beta}_i - q\hat{\beta}_j$, $\tilde{b} \equiv -\hat{\sigma}_i(1-\rho^2)^{1/2}$, $\tilde{c} \equiv \hat{\beta}_j$, $\tilde{d} \equiv -\hat{\sigma}_j$, and X is distributed as χ^2_ν, independently of Z_1, Z_2. Next, using the same argument as above, $\tilde{R} = \tilde{X}/\tilde{Y}$, where $\tilde{X} \equiv \tilde{a}X/\nu + \tilde{b}Z_1(X/\nu)^{1/2}$ and $\tilde{Y} \equiv \tilde{c}X/\nu + \tilde{d}Z_2(X/\nu)^{1/2}$. The joint cumulant generating function of $(\tilde{X}, \tilde{Y})$ is $-\nu \log[1 - \{\tilde{b}^2 s^2 + \tilde{d}^2 t^2 + 2(\tilde{a}s + \tilde{c}t)\}/\nu]/2$, defined for all (s,t) such that the argument to the logarithm is positive. The saddlepoints for approximating $\mathrm{pr}(\tilde{R} < \tilde{r})$ can be stated explicitly as $\hat{s} = -\tilde{a}/\tilde{b}^2$, $\hat{t} = \check{t}_0 = -\tilde{c}/\tilde{d}^2$, and $\tilde{s}_0 = (\tilde{c}\tilde{r} - \tilde{a})/(\tilde{b}^2 + \tilde{d}^2\tilde{r}^2)$, so that the entire approximation can be computed analytically. Bebu et al. (2009) illustrate their procedure for a parallel line essay of estrogen. Inference is desired for $\gamma \equiv \exp(\beta_3/\beta_2)$, where β_2 and β_3 are respectively the coefficients on the log dose and the test substance dummy. In this example, $\hat{\beta}_2 = 21{\cdot}86$, $\hat{\beta}_3 = -34{\cdot}84$, $\hat{\sigma}_2 = 5{\cdot}93$, $\hat{\sigma}_3 = 11{\cdot}40$, $\rho = -0{\cdot}93$, $n = 33$, and $p = 3$, so that $\mathrm{pr}(Y < 0) = 0{\cdot}045\%$. Bebu et al. (2009) obtain the generalized confidence interval $(0{\cdot}132, 0{\cdot}367)$ by simulation. Exact calculations lead to the interval $(0{\cdot}133, 0{\cdot}364)$. The saddlepoint approximation yields $(0{\cdot}133, 0{\cdot}363)$.

Cox (1985) gives an example in which inference is desired for the ratio of the average percentages of fine gravel in two kinds of surface soil. If one is willing to assume independence between the samples, then the problem fits Bebu et al.'s framework. The quantity of interest is the ratio β_1/β_2 of two dummy regressors. For Cox' data, $\hat{\beta}_1 = 10{\cdot}91, \hat{\beta}_2 = 3{\cdot}94, \hat{\sigma}_1 = \hat{\sigma}_2 = 1{\cdot}83, n = 7, p = 2$, and $\rho = 0$. In this case $\mathrm{pr}(Y < 0) = 2{\cdot}63\%$, so that the denominator is not significantly different from zero at the 5% level, implying that Fieller's (1940) solution does not produce a proper interval when $1 - \alpha = 0{\cdot}95$. A 95% generalized confidence interval is $(-3{\cdot}310, 15{\cdot}680)$. The saddlepoint approximation yields $(-2{\cdot}240, 15{\cdot}370)$. A researcher might choose to truncate the interval from below at zero, as the regressand in this example is strictly positive.

5·2. *Bootstrap Inference for a Ratio of Sample Means*

As in Davison & Hinkley (1988, Sec. 4; 1997, Sec. 9·5·2), we consider the problem of approximating the bootstrap distribution of the ratio of sample means $\lambda \equiv \mu_X/\mu_Y$ in a paired sample $\{(x_i, y_i)\}_{i=1}^n$ drawn from an absolutely continuous distribution; an extension to structural inference in simultaneous equation models is discussed in the Supplementary Material. The bootstrap distribution is discrete, whereas our saddlepoint approximation is continuous. The validity of using the saddlepoint approximation in this fashion follows from arguments in the appendix of Jing et al. (1994); however, a continuity-corrected version of the saddlepoint approximation might be warranted. We do not pursue this here because the improvements afforded by such a modification are typically small in bootstrap applications (Davison & Hinkley, 1988, Sec. 8; 1997, p. 468). The data considered by Davison & Hinkley are strictly positive, allowing them to rely on the saddlepoint approximation of Daniels (1954, 1983), and to which our approximation (17) reduces in that case. Here, we consider the Cushny–Peebles data (Cushny & Peebles, 1905), see Table 1. The data describe the additional hours of sleep gained by 10 patients using two different sleep-inducing drugs.

Table 1. *Additional hours of sleep gained by 10 patients using sleep-inducing drugs A and B*

Drug										
A (y_i)	1·9	0·8	1·1	0·1	−0·1	4·4	5·5	1·6	4·6	3·4
B (x_i)	0·7	−1·6	−0·2	−1·2	−0·1	3·4	3·7	0·8	0·0	2·0

Source: Cushny & Peebles (1905, Table I.)

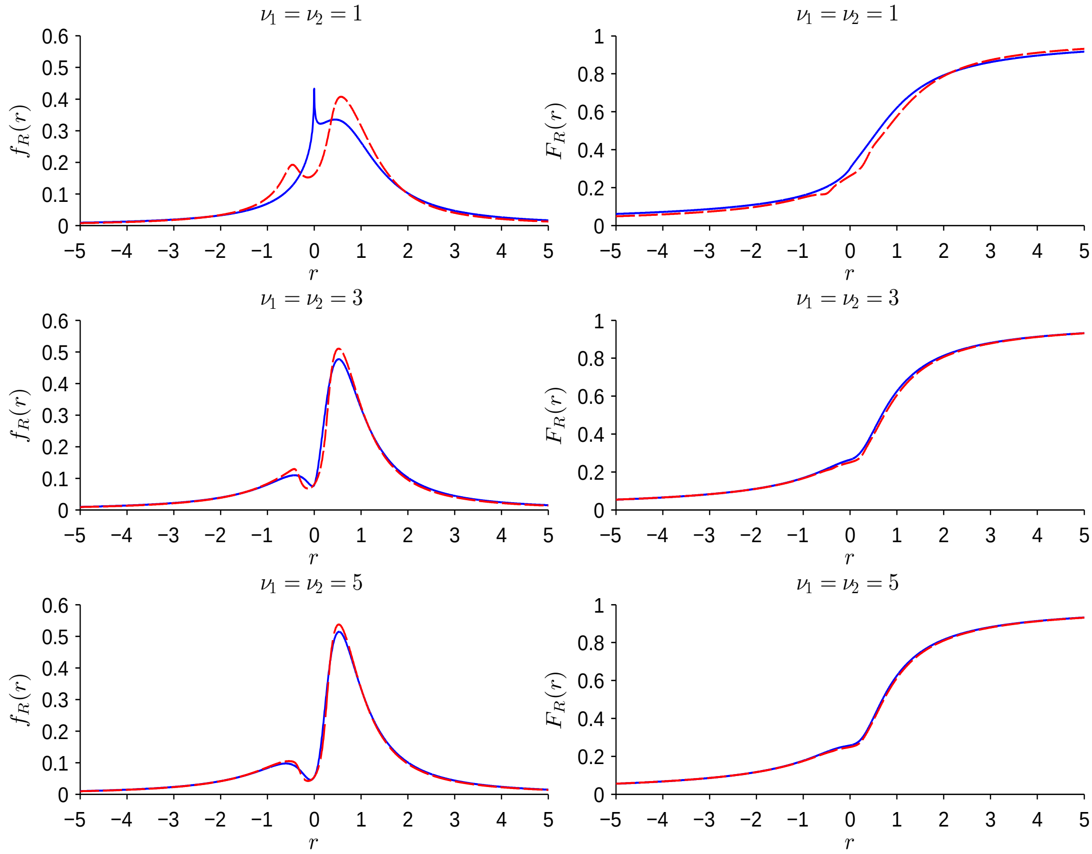

Fig. 1. Density and distribution of a ratio of independent Student's t variates, $a = 3$, $c = 2$, $b = 1$, and d chosen such that $\mathrm{pr}(Y < 0) = 25\%$. From top to bottom, $d = 2.0$, $d = 2.61$, and $d = 2.75$. Solid, exact density and distribution functions; broken lines, saddlepoint approximation.

The discreteness of the bootstrap distribution implies that we cannot rely on Theorem 2 to evaluate it, but the small sample size allows us to obtain it by enumerating all $(2n-1)!/\{n!(n-1)!\}$ possible realizations. Table 2 shows the exact distribution function, the saddlepoint approximation from Theorem 5, and the approximation obtained from a ratio of correlated Gaussians with fitted moments. The saddlepoint approximation tracks the exact bootstrap distribution accurately, including in the extreme tails. The percentile bootstrap confidence intervals for λ based on exact calculations and the saddlepoint approximation are, respectively, $(-0{\cdot}1812, 0{\cdot}5866)$ and $(-0{\cdot}1815, 0{\cdot}5867)$. For completeness, the exact and saddlepoint basic bootstrap confidence intervals are respectively given by $(0{\cdot}0572, 0{\cdot}8250)$ and $(0{\cdot}0571, 0{\cdot}8253)$. The marginal probability of the denominator becoming negative is quite small for the data at hand $(0{\cdot}0^5 648\%)$, but this is irrelevant because the numerator and denominator of R form a definite pair for this data set.

Acknowledgement

The authors thank Ron Butler, John Kolassa, an associate editor, and two anonymous referees for their helpful comments and suggestions. The second author gratefully acknowledges finan-

Table 2. *Approximations to the bootstrap distribution of $R = \bar{X}/\bar{Y}$ for the Cushny–Peebles data*

	Lower tail probability			Relative error	
r	Exact	SPA	RN	SPA	RN
$-0{\cdot}6$	0·28	0·28	0·83	0·36	193·36
$-0{\cdot}4$	0·73	0·73	1·57	0·40	114·39
$-0{\cdot}2$	2·22	2·24	3·48	0·73	56·80
0·0	7·87	7·82	9·24	$-0{\cdot}66$	17·36
0·2	26·89	27·00	27·33	0·42	1·64
0·4	68·40	68·36	67·72	$-0{\cdot}11$	$-2{\cdot}13$
0·6	98·24	98·22	96·78	$-0{\cdot}85$	$-82{\cdot}84$

SPA, saddlepoint approximation; RN, ratio of correlated normals with fitted moments. Relative error is normalized by $\min\{F_R^n(r), 1 - F_R^n(r)\}$. Entries are in percent.

cial support from National Bank Financial of Canada and the Social Sciences and Humanities Research Council of Canada, and from Tinbergen Institute for his visit to Amsterdam.

Supplementary material

Supplementary material available at *Biometrika* online includes mathematical proofs, the higher order terms for Theorem 3, an algorithm for evaluating the joint moment generating function in Section 5·1, and an extension of the application in Section 5·2. Matlab code for evaluating all expressions presented in the paper is available from the authors.

Appendices

A. *Explicit Expressions for I in* (20) *for $d = 1$ and $d = 2$*

The expressions below are valid if all elements of $\hat{\tau}$ are nonnegative, which can be achieved by an application of (21). For $d = 1$, $I(0, \Sigma, \hat{\tau}) = \Phi(-\hat{u})$ and $I(3, \Sigma, \hat{\tau}) = \left\{(\hat{u}^2 - 1)\phi(\hat{u}) - \hat{u}^3\Phi(-\hat{u})\right\}\Sigma^{-3/2}$, where $\hat{u} \equiv \hat{\tau}\Sigma^{1/2}$. For $d = 2$, let $\rho \equiv \Sigma_{12}/(\Sigma_{11}\Sigma_{22})^{1/2}$, $\tilde{\tau}_1 \equiv {\Sigma_{11}}^{1/2}\hat{\tau}_1 + \rho{\Sigma_{22}}^{1/2}\hat{\tau}_2$, $\tilde{\tau}_2 \equiv {\Sigma_{22}}^{1/2}\hat{\tau}_2 + \rho{\Sigma_{11}}^{1/2}\hat{\tau}_1$, and define $J_0 \equiv \Phi_2(-\tilde{\tau}_1, -\tilde{\tau}_2; \rho)$,

$$J_1 \equiv \frac{\phi(\tilde{\tau}_2)}{\Sigma_{22}^{1/2}}\Phi\left\{-\frac{\tilde{\tau}_1 - \rho\tilde{\tau}_2}{(1-\rho^2)^{1/2}}\right\}, \quad J_2 \equiv \frac{\phi(\tilde{\tau}_1)}{\Sigma_{11}^{1/2}}\Phi\left\{-\frac{\tilde{\tau}_2 - \rho\tilde{\tau}_1}{(1-\rho^2)^{1/2}}\right\}, \quad J_3 \equiv \frac{\phi_2(\tilde{\tau}_1, \tilde{\tau}_2; \rho)}{(\Sigma_{11}\Sigma_{22})^{1/2}},$$

where $\phi_2(\cdot, \cdot; \rho)$ and $\Phi_2(\cdot, \cdot; \rho)$ denote, respectively, the density and distribution function of a standard bivariate Gaussian with correlation ρ. The relevant functions can now be expressed as

$$I\{(0,\ 0), \Sigma, \hat{\tau}\} = J_0, \qquad I\{(3,\ 0), \Sigma, \hat{\tau}\} = \hat{\tau}_1^2(J_2 - \hat{\tau}_1 J_0) - (J_2 + \Sigma_{12}k_1)/\Sigma_{11},$$
$$I\{(2,\ 1), \Sigma, \hat{\tau}\} = k_1 + \hat{\tau}_1^2(J_1 - \hat{\tau}_2 J_0), \qquad I\{(1,\ 2), \Sigma, \hat{\tau}\} = k_2 + \hat{\tau}_2^2(J_2 - \hat{\tau}_1 J_0),$$
$$I\{(0,\ 3), \Sigma, \hat{\tau}\} = \hat{\tau}_2^2(J_1 - \hat{\tau}_2 J_0) - (J_1 + \Sigma_{12}k_2)/\Sigma_{22},$$

where $k_1 \equiv (\hat{\tau}_1 - \hat{\tau}_2\Sigma_{12}/\Sigma_{11})(\hat{\tau}_2 J_2 - J_3)$ and $k_2 \equiv (\hat{\tau}_2 - \hat{\tau}_1\Sigma_{12}/\Sigma_{22})(\hat{\tau}_1 J_1 - J_3)$.

B. *Proofs*

Proof of (3). Denote by $\mathcal{T}$ the convergence region of the joint moment generating function of X and Y, so that $\varphi_{X,Y}(-\mathrm{i}s, -\mathrm{i}t)$ is analytic for $(s,t) \in \mathcal{T}$. The inversion formula for the joint density is

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^2}\int_{-c_2\mathrm{i}-\infty}^{-c_2\mathrm{i}+\infty}\int_{-c_1\mathrm{i}-\infty}^{-c_1\mathrm{i}+\infty}\varphi_{X,Y}(s,t)e^{-\mathrm{i}sx-\mathrm{i}ty}\mathrm{d}s\mathrm{d}t.$$

If $c_1, c_2 < 0$, integrate between $-\infty$ and x and between $-\infty$ and y, to yield

$$F_{X,Y}(x, y) = \frac{1}{(2\pi \mathrm{i})^2} \int_{-c_2\mathrm{i}-\infty}^{-c_2\mathrm{i}+\infty} \int_{-c_1\mathrm{i}-\infty}^{-c_1\mathrm{i}+\infty} \varphi_{X,Y}(s, t) e^{-\mathrm{i}sx-\mathrm{i}ty} \frac{\mathrm{d}s}{s}\frac{\mathrm{d}t}{t}.$$

If $c_1 > 0$, then integration is from x to ∞, yielding $\mathrm{pr}(X > x, Y < y)$, and so on. This proves the result for $c_1c_2 \neq 0$. To prove it for $c_1 \to 0$, assume for simplicity that we may pick $c_1, c_2 > 0$; the other cases are treated similarly. By Cauchy's integral theorem, the path of integration in s can be deformed into one consisting of the ray $(-\infty; -\epsilon]$, a small semicircle in the upper half of the complex plane centered at the origin with radius $\epsilon < c_1$ traversed clockwise, and the ray $[\epsilon; \infty)$. As $\epsilon \downarrow 0$, the integral along the semicircle converges to $-\pi\mathrm{i}\varphi_{X,Y}(0, t)e^{-\mathrm{i}ty}/t$ by the residue theorem. The integral in t can be dealt with in the same manner. Combining the expressions and simplifying yields (2), from which (3) follows because $\chi(s, t)$ and $\chi(-s^*, -t^*)$ are complex conjugates. □

Proof of Lemma 1. Consider the case with $\mathrm{pr}\,(X - \beta Y < 0) = 0$ first. The result is trivial if $\beta = \infty$, i.e., if $\mathrm{pr}(Y < 0) = 1$. For the remainder of the proof, assume that β is finite. Writing $R = X/Y = R_1 + \beta$, where $R_1 \equiv (X - \beta Y)/Y$, it is seen that $R < \beta$, or equivalently $R_1 < 0$, if and only if $Y < 0$, as $X - \beta Y > 0$. If $r < \beta$, then $R < r$, or equivalently $R_1 < r - \beta$, if and only if $(X - \beta Y)/(r - \beta) < Y < 0$. Thus $\mathrm{pr}\,(R < r) = \mathrm{pr}\,(Y < 0) - \mathrm{pr}\,\{Y < (X - \beta Y)/(r - \beta)\} = \mathrm{pr}\,(Y < 0) - \mathrm{pr}\,(X - rY < 0)$. Similarly, if $r > \beta$, then $R < r$, or equivalently $R_1 < r - \beta$, if and only if $Y < 0$ or $Y > (X - \beta Y)/(r - \beta)$. Hence $\mathrm{pr}\,(R < r) = \mathrm{pr}\,(Y < 0) + \mathrm{pr}\,\{Y > (X - \beta Y)/(r - \beta)\} = \mathrm{pr}\,(Y < 0) + \mathrm{pr}\,(X - rY < 0)$. If $\mathrm{pr}\,(X - \beta Y < 0) = 1$, define $R = (-X)/(-Y)$ and proceed as above. □

Proof of Theorem 5. We begin by approximating $F_{\bar{W}}^{n}(0)$. The cumulant generating function of W is $\mathbb{K}(s, -rs)$. Define $\tilde{s}_0$ as in (16) and let $\tilde{\kappa}_0^{(i)} \equiv \partial^i/\partial s^i \mathbb{K}(s, -rs)\big|_{s=\tilde{s}_0}$ as given in Theorem 5. By Theorem 4, if $\tilde{s}_0 > 0$, then $\mathrm{pr}(\bar{W} > 0) = P_1 + \mathcal{O}(n^{-1})$. If $\tilde{s}_0 < 0$, the approximation is applied to $-\bar{W}$. The cumulant generating function of $-\bar{W}$ is $\mathbb{K}(-s, rs)$, so that the signs on $\tilde{s}_0$ and $\tilde{\kappa}_0^{(3)}$ are reversed, whereas $\tilde{\kappa}_0^{(0)}$ and $\tilde{\kappa}_0^{(2)}$ remain unaltered. Thus $\mathrm{pr}(-\bar{W} > 0) = \mathrm{pr}(\bar{W} < 0) = -P_1 + \mathcal{O}(n^{-1})$. Combining the two approximations yields

$$\mathrm{pr}\left(\bar{W} < 0\right) = H_0(\tilde{s}_0) - P_1 + \mathcal{O}\left(n^{-1}\right), \tag{B1}$$

which can be verified to remain valid if $\tilde{s}_0 = 0$. A similar derivation shows that

$$\mathrm{pr}\left(\bar{Y} < 0\right) = H_0(\check{t}_0) - P_2 + \mathcal{O}\left(n^{-1}\right), \tag{B2}$$

where $\check{t}_0$ solves $\mathbb{K}_2(0, \check{t}_0) = 0$. Finally, define $(\hat{s}, \hat{t})$ as in (15) and let $\hat{t}_r \equiv \hat{t} + r\hat{s}$ as before. Assume for the moment that $\hat{s} > 0$ and $\hat{t}_r > 0$. The joint cumulant generating function of (W, Y) is $\mathbb{K}(s, t - rs)$. The saddlepoint is $\hat{\tau}_r \equiv (\hat{s}, \hat{t}_r)$. Let $\hat{\tau} \equiv (\hat{s}, \hat{t})$, $\widehat{\mathbb{K}}'' \equiv \mathbb{K}''(\hat{s}, \hat{t})$, and $\hat{\mathcal{K}} \equiv \mathbb{K}''_{W,Y}(\hat{s}, \hat{t}_r) = [(\hat{\kappa}^{(2,0)}, \hat{\kappa}^{(1,1)})^{\mathrm{T}}, (\hat{\kappa}^{(1,1)}, \hat{\kappa}^{(0,2)})^{\mathrm{T}}]$, where $\hat{\kappa}^{(i,j)} \equiv \partial^{i+j}/\partial s^i \partial t^j \mathbb{K}(s, t - rs)\big|_{s=\hat{s},\, t=\hat{t}_r}$ as given in Theorem 5. Simplification shows that $\hat{\tau}_r^{\mathrm{T}}\hat{\mathcal{K}}\hat{\tau}_r = \hat{\tau}^{\mathrm{T}}\widehat{\mathbb{K}}''\hat{\tau}$. By Theorem 4, $\mathrm{pr}\left(\bar{W} > 0, \bar{Y} > 0\right) = P_3 + \mathcal{O}\left(n^{-1}\right)$, so

$$\mathrm{pr}\left(\bar{W} < 0, \bar{Y} < 0\right) = \mathrm{pr}\left(\bar{W} < 0\right) + \mathrm{pr}\left(\bar{Y} < 0\right) + P_3 - 1 + \mathcal{O}\left(n^{-1}\right), \quad \hat{s}, \hat{t}_r > 0. \tag{B3}$$

Next, assume that $\hat{s} > 0$ and $\hat{t}_r < 0$. Applying Theorem 4 to $(W, -Y)$ switches the sign on $\hat{t}_r$, the off-diagonal elements of $\hat{\mathcal{K}}$, $\hat{\kappa}^{(3-j,j)}$ for odd j, and, by equation (21), on $\hat{I}_{i,j}$ for even j, but leaves $\hat{\kappa}^{(0,0)}$ and $\hat{\tau}^{\mathrm{T}}\widehat{\mathbb{K}}''\hat{\tau}$ unaltered. Consequently, $\mathrm{pr}\left(\bar{W} > 0, -\bar{Y} > 0\right) = \mathrm{pr}\left(\bar{W} > 0, \bar{Y} < 0\right) = -P_3 + \mathcal{O}\left(n^{-1}\right)$, so

$$\mathrm{pr}\left(\bar{W} < 0, \bar{Y} < 0\right) = \mathrm{pr}\left(\bar{Y} < 0\right) + P_3, \quad \hat{s} > 0, \quad \hat{t}_r < 0. \tag{B4}$$

Now assume that $\hat{s} < 0$ and $\hat{t}_r > 0$. Applying Theorem 4 to $(-W, Y)$ switches the sign on $\hat{s}$, the off-diagonal elements of $\hat{\mathcal{K}}$, $\hat{\kappa}^{(3-j,j)}$ for even j, and $\hat{I}_{i,j}$ for even i, but leaves $\hat{\kappa}^{(0,0)}$ and $\hat{\tau}^{\mathrm{T}}\widehat{\mathbb{K}}''\hat{\tau}$ unaltered. Consequently, $\mathrm{pr}\left(\bar{W} > 0, \bar{Y} > 0\right) = \mathrm{pr}\left(\bar{W} < 0, \bar{Y} > 0\right) = -P_3 + \mathcal{O}\left(n^{-1}\right)$, so that

$$\mathrm{pr}\left(\bar{W} < 0, \bar{Y} < 0\right) = \mathrm{pr}\left(\bar{W} < 0\right) + P_3, \quad \hat{s} < 0, \quad \hat{t}_r > 0. \tag{B5}$$

Finally, if both $\hat{s} < 0$ and $\hat{t}_r < 0$, applying Theorem 4 to $(-W, -Y)$ switches the sign on both $\hat{s}$ and $\hat{t}_r$, $\hat{\kappa}^{(3-j,j)}$ for all j, and $\hat{I}_{i,j}$ if $i+j$ is odd, but leaves $\hat{\kappa}^{(0,0)}$, $\hat{\mathcal{K}}$, and $\hat{\tau}^{\mathrm{T}}\widehat{\mathbb{K}}''\hat{\tau}$ unaltered. Thus

$$\mathrm{pr}\left(-\bar{W} > 0, -\bar{Y} > 0\right) = \mathrm{pr}\left(\bar{W} < 0, \bar{Y} < 0\right) = P_3 + \mathcal{O}\left(n^{-1}\right), \quad \hat{s}, \hat{t}_r < 0. \tag{B6}$$

Combining (B3)–(B6) yields

$$\mathrm{pr}\left(\bar{W} < 0, \bar{Y} < 0\right) = H_0(\hat{t}_r)\,\mathrm{pr}\left(\bar{W} < 0\right) + H_0(\hat{s})\,\mathrm{pr}\left(\bar{Y} < 0\right) + P_3 - H_0(\hat{t}_r)H_0(\hat{s}) + \mathcal{O}\left(n^{-1}\right),$$

which, together with (11) and the univariate approximations (B1) and (B2), gives the result. □

REFERENCES

BEBU, I., SEILLIER-MOISEIWITSCH, F. & MATHEW, T. (2009). Generalized confidence intervals for ratios of regression coefficients with applications to bioassays. *Biometrical J.* **51**, 1047–1058.

BLEISTEIN, N. (1966). Uniform expansions of integrals with stationary points near algebraic singularity. *Comm. Pure Appl. Math.* **19**, 353–370.

BUTLER, R. W. (2007). *Saddlepoint Approximations with Applications*. Cambridge: Cambridge University Press.

COX, C. P. (1985). Interval estimates for the ratio of the means of two normal populations with variances related to the means. *Biometrics* **41**, 261–265.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton, New Jersey: Princeton University Press.

CUSHNY, A. A. & PEEBLES, A. R. (1905). The action of optical isomers. II. Hyoscines. *J. Physiol.* **32**, 510–510.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–650.

DANIELS, H. E. (1983). Saddlepoint approximations for estimating equations. *Biometrika* **70**, 89–96.

DAVISON, A. C. & HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417–431.

DAVISON, A. C. & HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

FIELLER, E. C. (1940). The biological standardization of insulin. *J. Roy. Statist. Soc.* Suppl. **7**, 1–64.

FINNEY, D. J. (1978). *Statistical Method in Biological Assay*. High Wycombe: Charles Griffin and Company Ltd., 3rd ed.

FULLER, W. A. (1990). *Sampling Statistics*. Hoboken: Wiley.

GEARY, R. C. (1944). Extension of a theorem by Harald Cramér on the frequency distribution of the quotient of two variables. *J. Roy. Statist. Soc. Ser. B* **107**, 56–57.

GIL-PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika* **38**, 481–82.

GURLAND, J. (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19**, 228–237.

HAUSCHILDT, R. (1969). Saddlepoint approximations to distribution functions. M.Sc. Thesis, University of British Columbia.

HUNTER, W. G. & LAMBOY, W. F. (1981). A Bayesian analysis of the linear calibration problem. *Technometrics* **23**, 323–328.

JING, B.-Y., FEUERVERGER, A. & ROBINSON, J. (1994). On the bootstrap saddlepoint approximations. *Biometrika* **81**, 211–215.

KAPPENMAN, R. F., GEISSER, S. & ANTLE, C. E. (1970). Bayesian and fiducial solutions for the Fieller–Creasy problem. *Sankhyā Ser. B* **32**, 331–340.

KOLASSA, J. E. (2003). Multivariate saddlepoint tail probability approximations. *Ann. Statist.* **31**, 274–286.

LEE, J. C. & LIN, S.-H. (2004). Generalized confidence intervals for the ratio of means of two normal populations. *J. Statist. Plann. Inference* **123**, 49–60.

PRESS, S. J. (1969). The t-ratio distribution. *J. Amer. Statist. Assoc.* **64**, 242–252.

ROBINSON, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. Roy. Statist. Soc. Ser. B* **44**, 91–101.

ROSÉN, B. (1961). On the asymptotic distribution of sums of independent identically distributed random variables. *Ark. Mat.* **4**, 323–332.

SHEPHARD, N. G. (1991). From characteristic function to distribution function: A simple framework for the theory. *Econ. Theory* **7**, 519–529.

WEERAHANDI, S. (1993). Generalized confidence intervals. *J. Amer. Statist. Assoc.* **88**, 899–905.

WENDEL, J. G. (1961). The nonabsolute convergence of Gil-Pelaez' inversion integral. *Ann. Math. Statist.* **32**, 338–339.

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261–1295.