Statistica Sinica **19** (2009), 1095 – 1117 (PREPRINT)

A Geometric Approach to Confidence Sets for Ratios: Fieller's Theorem, Generalizations, and Bootstrap

Ulrike von Luxburg and Volker H. Franz

Max Planck Institute for Biological Cybernetics, Tübingen, Germany
Justus–Liebig–Universität Giessen, Germany

Abstract: We present a geometric method to determine confidence sets for the ratio $E(Y)/E(X)$ of the means of random variables X and Y. This method reduces the problem of constructing confidence sets for the ratio of two random variables to the problem of constructing confidence sets for the means of one-dimensional random variables. It is valid in a large variety of circumstances. In the case of normally distributed random variables, the so-constructed confidence sets coincide with the standard Fieller confidence sets. Generalizations of our construction lead to definitions of exact and conservative confidence sets for very general classes of distributions, provided the joint expectation of (X, Y) exists and linear combinations of the form $aX + bY$ are well-behaved. Finally, our geometric method allows us to derive a very simple bootstrap approach for constructing conservative confidence sets for ratios that perform favorably in certain situations, in particular in the asymmetric heavy-tailed regime.

1. Introduction

In many practical applications we encounter the problem of estimating the ratio of two random variables X and Y. For example, we might wish to know how large one quantity is relative to another one, or to estimate the position at which a regression line intersects the abscissa (e.g., Miller (1986); Buonaccorsi (2001); see also Franz (submitted) for many references to practical studies involving ratios). While it is straightforward to construct an estimator for $E(Y)/E(X)$ by dividing the two sample means of X and Y, it is not obvious how confidence regions for this estimator can be determined. In the case where X and Y are jointly normally distributed, an exact solution to this problem has been derived by Fieller (1932, 1940, 1944, 1954); for more discussion see Kendall and Stuart (1961), Finney (1978), Miller (1986), and Buonaccorsi (2001).

In applications, practitioners often do not use Fieller's results and apply ad-hoc methods instead. Perhaps the main reason for this is that Fieller's confidence regions are perceived as counter-intuitive. In benign cases they form an interval which is not symmetric around the estimator, while in other cases the confidence region consists of two disjoint unbounded intervals, or even of the whole real line. Especially the latter case is highly unusual. The confidence region does not exclude any value at all — certainly not what one would expect from a "well-behaved" confidence region. However, different researchers (Gleser and Hwang (1987); Koschat (1987); Hwang (1995)) have shown that any method which cannot generate unbounded confidence limits for a ratio leads to arbitrary large deviations from the intended confidence level. For a discussion of the conditional confidence level, given that the Fieller confidence limits are bounded, see Buonaccorsi and Iyer (1984).

There have been several approaches to present Fieller's methods in a more intuitive way. Especially remarkable are the ones that rely on geometric arguments. Milliken (1982) attempted a geometric proof for Fieller's result when X and Y are independent normally distributed random variables. But his proof contained an error. Later, it was corrected and simplified by Guiard (1989). He considers the case $(X, Y) \sim N(\mu, \sigma^2 V)$, where the mean μ and the scale σ^2 of the covariance are unknown, but the covariance matrix V is known. Guiard presents a geometric construction of confidence regions, and then shows by an elegant comparison to a likelihood ratio test that the constructed regions are exact and coincide with Fieller's solution. The drawback of his proof is that it only works in the case where the covariance matrix V is known. Moreover, although the confidence sets are constructed by a geometric procedure, Guiard's proof relies on properties of the likelihood ratio test and does not give geometric insights into why the construction is correct. In this article we derive several simple geometric constructions for exact confidence sets for ratios. Our construction coincides with Guiard's if (X, Y) are normally distributed with known covariance matrix V, but it is also valid in the case where V is unknown. Our proof techniques are remarkably simple and purely geometric. The understanding gained then allows us to extend the geometric construction from normally distributed random variables to more general classes of distributions. While it is relatively straightforward to define confidence sets for elliptically symmetric distributions, another extension leads to a new construction of exact confidence sets for ratios for a very large class of distributions. The only assumptions we have to make is that the means of X and Y exist, and that it is possible to construct exact confidence sets for the means of linear combinations of the form $a_1X + a_2Y$. To our knowledge, these are the first *exact* confidence sets for ratios of very general classes of distributions. Finally, using the geometric insights also leads to a simple bootstrap procedure for confidence sets for ratios. This method is particularly well-suited for highly asymmetric and heavy-tailed distributions.

1.1 Definitions and notation

Consider a sample of n pairs $Z_i := (X_i, Y_i)_{i=1,\dots,n}$ drawn independently according to some underlying distribution. We first assume that this joint distribution is a two-dimensional normal distribution $N(\mu, C)$ with mean $\mu = (\mu_1, \mu_2)$ and covariance matrix $C = \left(\begin{smallmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{smallmatrix}\right)$, where both μ and C are unknown. Later we will study more general classes of distributions. Our goal is to estimate the ratio $\rho := \mu_2/\mu_1$ and to construct confidence sets for this ratio. To estimate the unknown means and the entries in the covariance matrix we use the standard estimators

$$\hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } \hat{\mu}_2 := \frac{1}{n}\sum_{i=1}^{n} Y_i, \tag{1.1}$$

$$\hat{c}_{11} := \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_1)^2 \text{ and } \hat{c}_{22} := \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mu}_2)^2, \tag{1.2}$$

$$\hat{c}_{12} := \hat{c}_{21} = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_1)(Y_i - \hat{\mu}_2). \tag{1.3}$$

Note that for convenience we rescaled the estimators $\hat{c_{ij}}$ by $1/n$ to reflect the variability of the estimators $\hat{\mu}_1$ and $\hat{\mu}_2$. As an estimator for the ratio $\rho = \mu_2/\mu_1$ we use $\hat{\rho} := \hat{\mu}_2/\hat{\mu}_1$. Note that our goal is to estimate $E(Y)/E(X)$ and not $E(Y/X)$. In fact, if X and Y are normally distributed, the latter quantity does not even exist. Moreover, when $\hat{\mu}_1$ and $\hat{\mu}_2$ are normally distributed, this implies that $E(\hat{\rho}) = E(\hat{\mu}_2/\hat{\mu}_1)$ does not exist as well. Hence, the estimator $\hat{\rho}$ is biased. For discussion on the bias of the estimator $\hat{\rho}$, see Beale (1962); Tin (1965); Durbin (1959); Rao (1981); Miller (1986); and Dalabehera and Sahoo (1995).

For $\alpha \in]0, 1[$, a *confidence set* (or *confidence region*) of level $1 - \alpha$ for ρ is a set R such that $P_\rho(\rho \in R) \geq 1 - \alpha$. If this statement holds with equality, then the confidence set R is called *exact*, otherwise it is called *conservative*. If the statement $P_\rho(\rho \in R) = 1 - \alpha$ only holds in the limit for the sample size $n \to \infty$, the confidence set R is called *asymptotically exact*. A confidence interval $[l, u]$ is called *equal-tailed* if $P_\rho(\rho < l) = P_\rho(\rho > u)$. It is called *symmetric around* $\hat{\rho}$ if it has the form $[\hat{\rho} - q, \hat{\rho} + q]$. For general background reading about confidence sets we refer to Chapter 20 of Kendall and Stuart (1961), Section 5.2 of Schervish (1995), and Chapter 4 of Shao and Tu (1995). For a real-valued random variable with distribution function F and $\alpha \in]0, 1[$, the α-quantile of F is the smallest x such that $F(x) = \alpha$. We denote this quantile by $q(F, \alpha)$. In the special case where F is induced by the Student-t distribution with f degrees of freedom, we denote the quantile by $q(t_f, \alpha)$.

Many of the geometric arguments in this paper are based on orthogonal projections of the two-dimensional plane to a one-dimensional subspace. In the two-dimensional plane, the line L_ρ through the origin with slope ρ and the line $L_{\rho\perp}$ orthogonal to L_ρ are

$$L_\rho := \{(x,y) \in \mathbb{R}^2 |\ y = \rho x\} \qquad \text{and} \qquad L_{\rho\perp} := \{(x,y) \in \mathbb{R}^2 |\ y = (-1/\rho)x\}.$$

For an arbitrary unit vector $a = (a_1, a_2)' \in \mathbb{R}^2$, let $\pi_a : \mathbb{R}^2 \to \mathbb{R},\ x \mapsto a'x = a_1x_1 + a_2x_2$ be the orthogonal projection of the two-dimensional plane on the one-dimensional subspace spanned by a, that is on the line L_r with slope $r = a_2/a_1$. We write π_r for the projection on L_r, and $\pi_{r\perp}$ for the projection on $L_{r\perp}$. Let $C \in \mathbb{R}^{2\times 2}$ be a covariance matrix with eigenvectors $v_1, v_2 \in \mathbb{R}^2$ and eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$. Consider the ellipse centered at some point $\mu \in \mathbb{R}^2$ such that its principal axes have the directions of v_1, v_2 and lengths $q\sqrt{\lambda_1}$ and $q\sqrt{\lambda_2}$ for some $q > 0$. We denote this ellipse by $E(C, \mu, q)$ and call it the covariance ellipse corresponding to C, centered at μ and scaled by q. This ellipse can also be described as the set of points $z \in \mathbb{R}^2$ which satisfy $(z - \mu)'C^{-1}(z - \mu) = q^2$.

2. Exact confidence regions for normally distributed random variables

Let us start with a few geometric observations. For given $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$, the ratio $\rho = \mu_2/\mu_1$ can be depicted as the slope of the line L_ρ in the two-dimensional plane that passes through the origin and the point (μ_1, μ_2). The estimated ratio $\hat\rho$ is the slope of the line through the origin and the point $\hat\mu = (\hat\mu_1, \hat\mu_2)$ (cf. Figure 2.1). Consider a confidence interval $R = [l, u] \subset \mathbb{R}$ that contains the estimator $\hat\rho$. The lower and upper limits of this interval correspond to the slopes of the two lines passing through the origin and the points $(1, l)$ and $(1, u)$, respectively. Let W denote the wedge enclosed by those two lines. The slopes of the lines inside the wedge exactly correspond to the ratios inside the interval R. The other way round, the interval $[l, u]$ can be reconstructed from the wedge as the intersection of the wedge with the line $x = 1$ (cf. Figure 2.1).

2.1 Geometric construction of exact confidence sets

In the following we want to construct an appropriate wedge containing $\hat\mu$ such that the region obtained by intersection with the line $x = 1$ yields an exact confidence region for ρ of level $1 - \alpha$. This wedge will be constructed as the smallest wedge containing a certain ellipse around the estimated mean $(\hat\mu_1, \hat\mu_2)$.

Construction 1 (Geometric construction of exact confidence regions R_{geo} for ρ in case of normal distributions)

1. *Estimate the means $\hat\mu_1$ and $\hat\mu_2$ and the covariance matrix $\hat C$ according to* (1.1) *to* (1.3)*.*

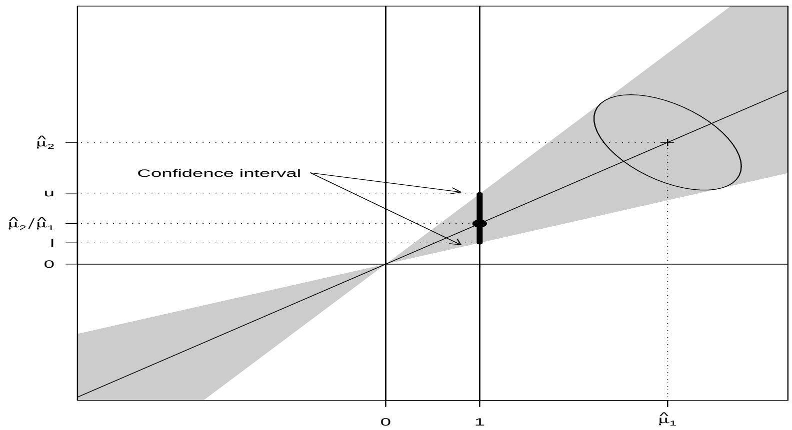

Figure 2.1: *The ratio $\hat{\mu}_2/\hat{\mu}_1$ can be depicted as the slope of the line through the points $(0,0)$ and $(\hat{\mu}_1, \hat{\mu}_2)$. The ratios inside $[l, u]$ correspond to the slopes of all lines in the wedge spanned by the lines with slopes l and u. For a given wedge, the corresponding interval $[l, u]$ can be obtained by intersecting the wedge with the line $x = 1$.*

2. *Let $q := q(t_{n-1}, 1-\alpha/2)$ be the $(1-\alpha/2)$-quantile of the Student-t distribution with $n-1$ degrees of freedom.*
3. *In the two-dimensional plane, plot the ellipse $E = E(\hat{C}, \hat{\mu}, q)$ centered at the estimated joint mean $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)$, with shape according to the estimated covariance matrix $\hat{C}$, and scaled by the number q computed in Step 2.*
4. *Depending on the position of the ellipse, distinguish between the following cases (see Figure 2.2).*
   (a) *If $(0,0)$ is not inside E, construct the two tangents to E through the origin $(0,0)$, and let W be the wedge enclosed by those tangents. Take R_{geo} to be the intersection of W with the line $x = 1$. Depending on whether the y-axis lies inside W or not, this results in an exclusive unbounded or a bounded confidence region.*
   (b) *If $(0,0)$ is inside E, choose $R_{geo} =]-\infty, \infty[$ (completely unbounded case).*

Let us give some intuitive reasons why the three cases make sense. In the first case, the denominator $\hat{\mu}_1$ is significantly different from 0. Observe that this is the case if and only if the ellipse E does not intersect the y-axis. Here we do not expect any difficulties from dividing by $\hat{\mu}_1$ as the denominator is "safely away from 0". Our uncertainty about the value of ρ is restricted to some interval around ρ, which corresponds to the bounded case. The situation is more complicated if the denominator is not significantly different from 0, that is the ellipse intersects the y-axis. As we divide by a number potentially close to 0, we cannot control the absolute value of the outcome, which may be arbitrarily large, nor can we be sure about its sign. Hence, regions of the form $]-\infty, c_1]$ and $[c_2, \infty[$ should be part of the confidence region. If, additionally, we are confident that the numerator is not too small, then we expect that ρ is not very close to 0. This is reflected by the "exclusive unbounded case". If, on the other hand, the numerator is not significantly different from 0, then we cannot guarantee anything: with dividing 0 by 0 any outcome is conceivable. Here the confidence set should coincide with the whole real line: the "completely unbounded" case.

Theorem 1 ($R_{\mathbf{geo}}$ is an exact confidence set for ρ) *Let $(X_i, Y_i)_{i=1,\dots,n}$ be an i.i.d. sample drawn from the distribution $N(\mu, C)$ with unknown μ and C, and let R_{geo} be the regions of Construction 1. Then R_{geo} is an exact confidence region of level $1-\alpha$ for ρ, that is for all μ and C we have $P(\rho \in R_{geo}) = 1-\alpha$.*

Proof. Let $a = (a_1, a_2)' \in \mathbb{R}^2$ be an arbitrary unit vector. We denote by $U := \pi_a(X, Y)$ the projection of the joint random variable (X, Y) on the subspace spanned by a. Then U is distributed according to $N(a'\mu, a'Ca)$. The independent sample points $(X_i, Y_i)_{i=1,\dots n}$ are mapped by π_a to independent sample points $(U_i)_{i=1,\dots,n}$. It is easy to see that the length of the interval $I := \pi_a(E)$ is $2q(a'\hat{C}a)^{1/2}$. Taking into account the choice of q in Construction 1 as the $(1-\alpha/2)$-quantile of the Student-t distribution, by the normality assumption on (X, Y) we can conclude that the projected ellipse $\pi_a(E)$ is a $(1-\alpha)$-confidence interval for the mean $\pi_a(\mu)$ of the projected random variables:

$$
\begin{aligned}
1-\alpha &= P\left(\pi_a(\mu) \in [\pi_a(\hat{\mu}) - q(a'\hat{C}a)^{1/2},\ \pi_a(\hat{\mu}) + q(a'\hat{C}a)^{1/2}]\right) \\
&= P\left(\pi_a(\mu) \in \pi_a(E)\right).
\end{aligned}
$$

This equation is true for all unit vectors a. Now we want to consider the particular projection $\pi_{\rho_\perp}$ on the line $L_{\rho_\perp}$ (that is, we choose $a = (\rho/\sqrt{1+\rho^2}, -1/\sqrt{1+\rho^2})$). Showing that $\pi_{\rho_\perp}(\mu) \in \pi_{\rho_\perp}(E) \iff \rho \in R_{\text{geo}}$ will complete our proof. As in the construction of R_{geo} we distinguish two cases. If the origin is not inside the ellipse E we can construct the wedge W as described in the construction of R_{geo}. In this case we have the following geometric equivalences (see Figure 2.3):

$$\pi_{\rho_\perp}(\mu) \in \pi_{\rho_\perp}(E) \iff 0 \in \pi_{\rho_\perp}(E) \iff E \cap L_\rho \neq \emptyset \iff L_\rho \subset W \iff \rho \in R_{\text{geo}}.$$

In the second case, the origin is inside the ellipse E. In this case it is clear that $\pi_{\rho_\perp}(\mu) = 0$ is always inside $\pi_{\rho_\perp}(E)$. On the other hand, by definition the region R_{geo} coincides with $]-\infty, \infty[$ in this case, and thus $\rho \in R_{\text{geo}}$ is true. ☺

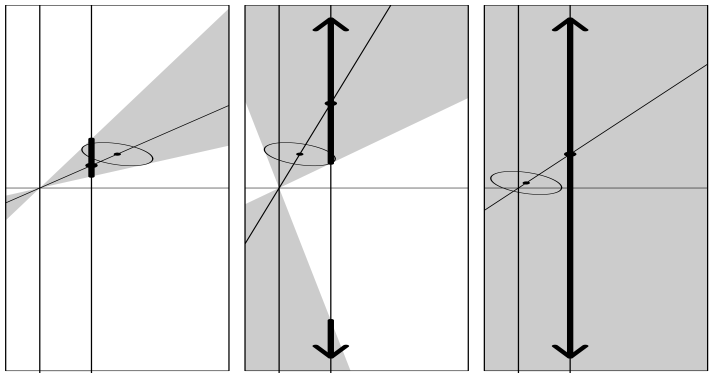

Figure 2.2: *The three cases in the construction of the confidence set R_{geo}: bounded, exclusive unbounded, and completely unbounded.*

2.2 Comparison to Fieller's confidence sets

Now we want to compare the confidence regions obtained by Construction 1 to the classic confidence sets constructed by Fieller (1932, 1940, 1944, 1954). To this end let us first state Fieller's result according to Subsection 4, p. 176-177 of Fieller (1954). We reformulate his definition in our notation.

Definition 2 (Fieller's confidence regions for $\boldsymbol{\rho}$ in case of normal distributions) *Compute the quantities*

$$q^2_{exclusive} := \frac{\hat{\mu}_1^2}{\hat{c}_{11}}, \quad q^2_{complete} := \frac{\hat{\mu}_2^2\hat{c}_{11} - 2\hat{\mu}_1\hat{\mu}_2\hat{c}_{12} + \hat{\mu}_1^2\hat{c}_{22}}{\hat{c}_{11}\hat{c}_{22} - \hat{c}_{12}^2}, \quad \textit{and}$$

$$l_{1,2} = \frac{1}{\hat{\mu}_1^2 - q^2\hat{c}_{11}}\left((\hat{\mu}_1\hat{\mu}_2 - q^2\hat{c}_{12}) \pm \sqrt{(\hat{\mu}_1\hat{\mu}_2 - q^2\hat{c}_{12})^2 - (\hat{\mu}_1^2 - q^2\hat{c}_{11})(\hat{\mu}_2^2 - q^2\hat{c}_{22})}\right),$$

with q as in the definition of the confidence regions R_{geo}. Then define the confidence set $R_{Fieller}$ as follows:

$$R_{Fieller} = \begin{cases}]-\infty,\infty[& \textit{if } q^2_{complete} \le q^2 \\ ]-\infty, \min\{l_1,l_2\}] \quad \cup \quad [\max\{l_1,l_2\},\infty[& \textit{if } q^2_{exclusive} < q^2 < q^2_{complete} \\ [\min\{l_1,l_2\}, \max\{l_1,l_2\}] & \textit{otherwise} \end{cases}$$

The three cases result in completely unbounded, exclusive unbounded, and bounded confidence sets, respectively.

Theorem 3 (Fieller) *Let $(X_i,Y_i)_{i=1,\dots,n}$ be an i.i.d. sample drawn from the distribution $N(\mu,C)$ with unknown μ and C. Then $R_{Fieller}$ as given in Definition 2 is an exact confidence region of level $1-\alpha$ for ρ.*

Proof of Fieller's theorem (sketch). Consider the function

$$T_{r,\hat{C}}(x) := \frac{x_2 - rx_1}{\sqrt{\hat{c_{22}} - 2r\hat{c_{12}} + r^2\hat{c_{11}}}}, \tag{2.1}$$

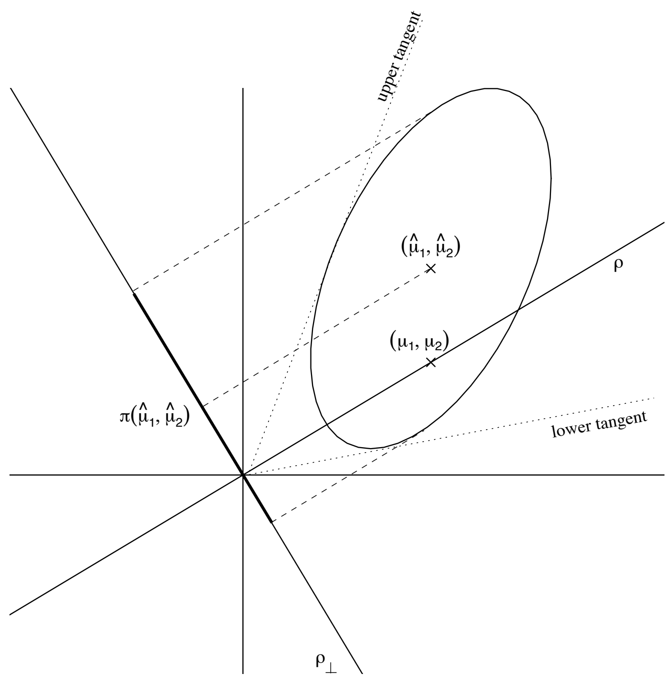

Figure 2.3: *Projection of the ellipse E on the subspace spanned by $\rho_\perp$.*

where $r \in \mathbb{R}$ is a parameter and $\hat{C}$ denotes the sample covariance matrix. When applied to $r = \rho$ and $x = \hat{\mu}$, the statistic $T_{\rho,\hat{C}}(\hat{\mu})$ has a Student-t distribution with $n-1$ degrees of freedom. The set $R_{\text{Fieller}} := \{r \in \mathbb{R} | \, T_{r,\hat{C}}(\hat{\mu}) \in [-q, q]\}$ now satisfies (by the definition of q as a Student-t quantile)

$$P(\rho \in R_{\text{Fieller}}) = P(T_{\rho,\hat{C}}(\hat{\mu}) \in [-q, q]) = 1 - \alpha.$$

Solving $-q \leq T_{r,\hat{C}}(\hat{\mu}) \leq q$ for r leads to a quadratic inequality whose solutions are given by Fieller's theorem. ☺

Let us make a few comments about this proof. The most important property of the statistic $T_{\rho,\hat{C}}(\hat{\mu})$ is the fact that its distribution does not depend on ρ. That is, it is a pivotal quantity. Otherwise, solving the inequalities $-q \leq T_{r,\hat{C}}(\hat{\mu}) \leq q$ for r would not lead to an expression independent of ρ. Moreover, note that the mapping $T_{\rho,\hat{C}}$ projects the points on the line $L_{\rho_\perp}$, and additionally scales them such that the projected sample mean has variance 1. In particular it is interesting to note that because $T_{\rho,\hat{C}}(\mu) = 0$, the set $J_\rho = [T_{\rho,\hat{C}}(\hat{\mu}) - q, T_{\rho,\hat{C}}(\hat{\mu}) + q]$ is a $(1-\alpha)$-confidence interval for the projected mean $T_{\rho,\hat{C}}(\mu)$:

$$P(T_{\rho,\hat{C}}(\mu) \in J_\rho) = P(0 \in [T_{\rho,\hat{C}}(\hat{\mu}) - q, T_{\rho,\hat{C}}(\hat{\mu}) + q]) \;\; = P(T_{\rho,\hat{C}}(\hat{\mu}) \in [-q, q]) = 1 - \alpha.$$

This property will be used later to generalize Fieller's confidence set to more general distributions. Also note that solving the inequalities $-q \leq T_{r,\hat{C}}(\hat{\mu}) \leq q$ coincides with the construction of the wedge in the geometric construction. The wedge can be seen as exactly the lines with slope r such that the projection of $\hat{\mu}$ on $L_{\hat{r_\perp}}$ is still within $[-q, q]$. Based on all these observations it is natural to expect a close relation between R_{Fieller} and R_{geo}, which indeed exists.

Theorem 4 (R_{geo} and R_{Fieller} coincide) *The confidence region R_{geo} defined in Construction 1 coincides with $R_{Fieller}$ as given in Definition 2.*

Proof. *(Sketch)* First one has to show that the three cases in Fieller's theorem coincide with the three cases in the geometric approach. Second, one then has to verify that the numbers l_1 and l_2 in Fieller's theorem coincide with the slopes of the tangents to the ellipse. Both steps can be solved by straightforward but lengthy calculations. Details can be found in von Luxburg and Franz (2004). ☺

Note that in the proof of Fieller's theorem, we did not directly use the fact that we have *paired* samples $(X_i, Y_i)_{i=1,...,n}$. Indeed, Fieller's theorem and its proof are also valid in the more general setting where we are given two independent samples $X_1, ..., X_n$ and $Y_1, ..., Y_m$ with a different number of sample points, and use unbiased estimators for the means μ_1, μ_2 and independent unbiased estimators for the (co)variances $\hat{c_{ij}}$. In this case one has to take care to choose the degrees of freedom in the Student-t-distribution appropriately, see Buonaccorsi (2001) and Section 3.3.3 of Rencher (1998).

3. Exact confidence sets for general random variables

In this section we extend our geometric approach to non-normally distributed random variables.

3.1 Elliptically symmetric distributions

In the normally distributed case, the main reason why Construction 1 leads to exact confidence sets is that the projected and studentized mean is Student-t distributed, no matter in which direction we project. More generally, such a property holds for all elliptically symmetric random variables. Elliptically symmetric random variables can be written in the form $\mu + AY$ where μ is a shift parameter, A is a matrix with $AA' = C$, and Y is any spherically symmetric random variable generated by a distribution H on $\mathbb{R}_+$. For a brief overview of spherical and elliptical distributions see Eaton (1981); for an extensive treatment see Fang, Kotz, and Ng (1990). In particular, if X is an elliptically symmetric random variable with shift μ, covariance C, and generator H, then the statistic $T_{r,\hat{C}}(\hat{\mu})$ introduced in (2.1) is a pivotal quantity which has the same distribution for all $r \in \mathbb{R}$. Denote the distribution function of this statistic by G. To extend Construction 1 to the case of elliptically symmetric distributions, all we have to do is to choose the parameter q as the quantile $q(G, 1-\alpha/2)$ of the distribution G. With similar arguments as in the last sections, one can see that the resulting confidence set is exact.

3.2 Confidence sets for a general class of distributions

Once we leave the class of elliptically symmetric random variables, the distributions of the projected means are no longer independent of the direction of the projection. Consequently, the techniques presented above do not apply. However, there is a surprisingly simple way to circumvent this problem. To see this, let us re-interpret Construction 1 as depicted in Figure 3.1. Previously, to determine whether $r \in \mathbb{R}$ should be an element of R_{geo}, we checked whether the line with slope r is inside the wedge enclosing the ellipse E. But note that the same result can be achieved if we project the sample on the line $L_{r_\perp}$, construct a one-dimensional confidence set J_r for the mean on $L_{r_\perp}$, and check whether $0 \in J_r$ or not. This observation is key to the following construction.

Construction 2 (Exact confidence sets $\boldsymbol{R}_{\mathbf{gen}}$ for $\boldsymbol{\rho}$ in case of general distributions)

1. *For each $r \in \mathbb{R}$, project the sample points on $L_{r_\perp}$ to get the new points $U_{r,i} = \pi_{r_\perp}(X_i, Y_i)$, $i = 1, \ldots, n$.*
2. *For each $r \in \mathbb{R}$, construct a confidence set J_r for the mean of $U_{r,i}$, that is, a set such that $P(\pi_{r_\perp}(\mu) \in J_r) = 1 - \alpha$.*
3. *Define the confidence set R_{gen} for ρ as $R_{gen} = \{r \in \mathbb{R} \mid 0 \in J_r\}$.*

The big advantage of this construction is that the projection in the direction of the true value ρ is not singled out as a "special" projection, we simply look at *all projections*. Hence, Construction 2 does not require any knowledge about ρ.

Theorem 5 ($R_{\mathbf{gen}}$ is an exact confidence set for ρ) *Let $(X_i, Y_i)_{i=1,\ldots,n} \in \mathbb{R}^2$ be i.i.d. pairs of random variables with an arbitrary distribution such that the joint mean of (X, Y) exists. If the confidence sets J_r used in Construction 2 exist and are exact (resp. conservative resp. liberal) confidence sets of level $(1-\alpha)$ for the means of $\pi_{r_\perp}((X_i, Y_i))_{i=1,\ldots,n}$, then R_{gen} is an exact (resp. conservative resp. liberal) confidence set for ρ.*

Proof. In the exact case, we have to prove that the true ratio ρ satisfies $P(\rho \in R_{\text{gen}}) = (1-\alpha)$. By definition of R_{gen}, for each $r \in \mathbb{R}$ we have that $r \in R_{\text{gen}} \iff 0 \in J_r$. In particular, this also holds for $r = \rho$. Moreover, the projection corresponding to the true ratio ρ projects the true mean μ on the origin of the coordinate system. By linearity, the projection of the true mean $\pi_{\rho_\perp}(\mu)$ equals the mean of the projected random variables. By construction of J_r we know that the latter is inside J_r with probability exactly $(1-\alpha)$. So we

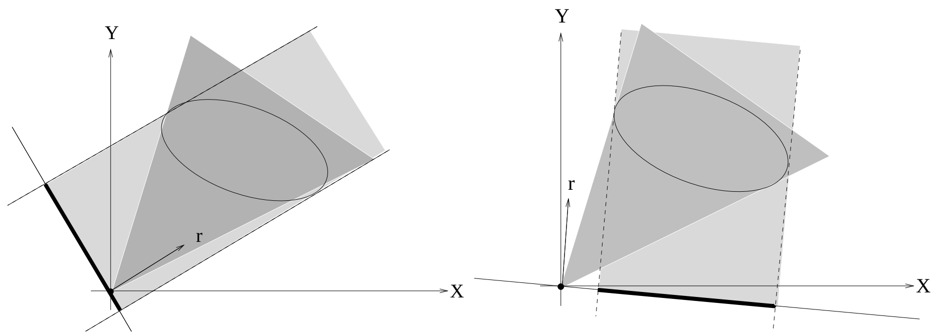

Figure 3.1: *(Second geometric interpretation) By definition, ratio r is an element of Fieller's confidence set R_{geo} if the line L_r (depicted by the little arrow) is inside the wedge enclosing the covariance ellipse. This is the case if and only if the origin is inside the projection $J_r := \pi_{r_\perp}(E)$ of the ellipse on the line $L_{r_\perp}$. The left panel shows a case where $r \in R_{geo}$, the right panel a case where $r \notin R_{geo}$.*

can conclude that $P(\rho \in R_{\text{gen}}) = P(0 \in J_\rho) = P(\pi_{\rho_\perp}(\mu) \in J_\rho) = 1 - \alpha$. ☺

To our knowledge, Construction 2 is the first construction of exact confidence sets for general distributions. It reduces the problem of estimating confidence sets for the ratio of two random variables to the problem of estimating confidence sets for the means of one-dimensional random variables. At first glance this looks very promising. However, the crux of the matter for applying this construction in practice is that one has to know the analytic form of the distribution of the projected means. For this one has to be able to derive an analytic expression for general linear combinations of X and Y. While there might be some special cases in which this is tractable, for the vast majority of distributions such an analytic form is not easy to obtain. As a consequence, while being of theoretic interest, Construction 2 is of limited relevance for practical applications.

4. Conservative confidence sets for more general random variables

Our geometric principles can also be used to derive very simple conservative confidence sets for general distributions. The main idea is to replace the ellipse used in Construction 1 by a more general convex set $M \subset \mathbb{R}^2$. A straightforward idea is to choose M as a $(1-\alpha)$-confidence set for the bivariate joint mean $\mu \in \mathbb{R}^2$, that is a set such that $P(\mu \in M) = 1 - \alpha$. Then, as above, we can construct the wedge W around M which is given by the two enclosing tangents and choose a confidence region R_{cons} by intersecting the wedge with the line $x = 1$, distinguishing between the same three cases as above.

Another simple but effective way to choose the convex set M it to take the axis-parallel rectangle $A := I_1 \times I_2$, where the intervals $I_1 := [l_1, u_1]$ and $I_2 := [l_2, u_2]$ are confidence intervals for the one-dimensional means μ_1 of X and μ_2 of Y. Formally, this leads to the following construction.

Construction 3 (Geometric construction of conservative confidence regions $R_{\mathbf{cons}}$ for ρ for general distributions)

1. *Construct exact (or conservative) confidence intervals I_1 and I_2 of level $(1-\alpha/2)$ for the means of X and Y, respectively. In the two-dimensional plane, define the rectangle $A = I_1 \times I_2$.*
2. (a) *If $(0,0)$ is not inside A, construct the two tangents to A through the origin $(0,0)$, and let W be the wedge enclosed by those tangents. Define the confidence region R_{cons} as the intersection of W with the line $x = 1$. Depending on whether the y-axis lies inside W or not, this results in an exclusive unbounded or a bounded confidence region*
   (b) *If $(0,0)$ is inside A, choose the confidence region as $R_{cons} =]-\infty, \infty[$.*

Theorem 6 ($R_{\mathbf{cons}}$ is a conservative confidence set for ρ) *Let $(X_i, Y_i)_{i=1,\dots,n} \in \mathbb{R}^2$ be i.i.d. pairs of random variables with arbitrary distribution such that the joint mean of (X, Y) exists. If the confidence sets I_1 and I_2 used in Construction 3 exist and are exact or conservative confidence sets of level $(1-\alpha)$ for the means of X and Y, then R_{cons} is a conservative confidence set for ρ of level $(1-2\alpha)$.*

Proof. The proof of this theorem is immediate:

$$\begin{aligned} P(\rho \in R_{\text{cons}}) &= P(\mu \in W) \geq P(\mu \in A) = P(\mu_1 \in I_1 \text{ and } \mu_2 \in I_2) \\ &= 1 - P(\mu_1 \notin I_1 \text{ or } \mu_2 \notin I_2) \geq 1 - (P(\mu_1 \notin I_1) + P(\mu_2 \notin I_2)) = 1 - 2\alpha. \end{aligned}$$ ☺

Interestingly, it can be seen easily that the set R_{cons} constructed using the rectangle coincides with the set obtained by "dividing" the one-dimensional confidence intervals I_2 by I_1, namely $R_{\text{cons}} = I_2/I_1 := \{y/x \mid y \in I_2, x \in I_1\}$. The latter is a heuristic for confidence sets for ratios which can sometimes be found in the literature, usually without any theoretical justification. Our geometric method now reveals effortlessly that it is statistically safe to use this heuristic, but that it will lead to conservative confidence sets of level $1 - 2\alpha$.

One could think of even more general ways to construct a convex set $M \subset \mathbb{R}^2$ as base for the conservative geometric construction. For example, instead of using axis-parallel projections as in Construction 3, one could base the convex set M on projections in arbitrary directions (for example, using the two projections in the directions ρ and $\rho_\perp$, or even using more than two projections). However, we would like to stress one big advantage of using the axis-parallel rectangle. While the exact generalizations presented in Section 3 requires construction of confidence sets for the means of arbitrary linear combinations of the form $aX + bY$, for the rectangle construction we only need to be able to construct exact confidence sets for the marginal distributions of X and Y, respectively. One can envisage many situations where distributional assumptions on X and Y are reasonable, but where the distributions of projections of the form $aX + bY$ cannot be computed in closed form. In such a situation, the rectangle construction can serve as an easy loophole. The price we pay is the one of obtaining conservative confidence sets for the ratio instead of exact ones. But, in many cases, obtaining confidence sets which are provably conservative might be preferred over using heuristics, with unknown guarantees, to approximate exact confidence sets.

5. Bootstrap confidence sets

In the last sections we have seen how exact and conservative confidence sets for ratios of general classes of distributions can be constructed. In practice, the application of those methods is limited since we still need to know the exact distributions of the projections of (X, Y). In this section we want to investigate how approximate confidence sets can be constructed in cases where the underlying distributions are unknown. It is natural to consider bootstrap procedures for this purpose (e.g., Efron (1979); Efron and Tibshirani (1993); Shao and Tu (1995); Davison and Hinkley (1997)). However, if the variance of the statistics of interest does not exist, as is usually the case for $\hat{\rho}$, bootstrap confidence regions can be erroneous (Athreya (1987); Knight (1989)). Moreover, standard bootstrap methods that attempt to bootstrap the statistic $\hat{\rho}$ directly cannot result in unbounded confidence regions. This is problematic, as it has been shown that any method which is not able to generate unbounded confidence limits for a ratio can lead to arbitrary large deviations from the intended confidence level (Gleser and Hwang (1987); Koschat (1987); Hwang (1995)). Hence, bootstrapping $\hat{\rho}$ directly is not an option. Instead, in the literature there are several approaches to use bootstrap methods based on the studentized statistic $T_{\rho,\hat{C}}(\hat{\mu})$ introduced in (2.1). A simple approach along those lines is taken in Choquet, L'Ecuyer, and Léger (1999). The authors use standard bootstrap methods to construct a confidence interval $[q_1, q_2]$ for the mean of the statistic $T_{\hat{\rho},\hat{C}}(\hat{\mu})$. As confidence set for the ratio, they then use the interval $[\hat{\rho} - q_2 S_{\hat{\rho}}, \hat{\rho} - q_1 S_{\hat{\rho}}]$ where $S_{\hat{\rho}}$ is the estimated standard deviation of $\hat{\rho}$. However, this approach is problematic as well: the confidence sets do not have the qualitative behavior of those of Fieller, and as they are always finite, the coverage probability can be arbitrarily small.

5.1 Bootstrap approach by Hwang and its geometric interpretation

A more promising bootstrap approach for ratios has been presented by Hwang

(1995). He suggests using standard bootstrap methods to construct confidence sets for the mean of $T_{\hat{\rho},\hat{C}}(\hat{\mu})$. To determine the confidence set for the ratio, he then proceeds, as Fieller, to solve a quadratic equation to determine the confidence set for the ratio. Hwang (1995) argues that his confidence sets are advantageous when dealing with asymmetric distributions such as exponential distributions. However, we need to be careful here. Hwang (1995) only treats the case of one-sided confidence sets, where he constructs a confidence set of the form $]-\infty, q]$ for $T_{\hat{\rho},\hat{C}}(\hat{\mu})$ and then solves the quadratic equation $T_{\hat{\rho},\hat{C}}(\hat{\mu})^2 \leq q^2$. This leads to the well-known bounded, exclusively unbounded, and completely unbounded cases. However, the two-sided case is more involved and is not discussed in his paper. If one uses symmetric bootstrap confidence sets of the form $[-q, q]$ for $T_{\hat{\rho},\hat{C}}(\hat{\mu})$, then one can proceed by solving one quadratic inequality similar to that above. However, if one wants to exploit the fact that the distribution might not be symmetric, one would have to use asymmetric (for example equal-tailed) confidence sets of the form $[q_1, q_2]$ for $T_{\hat{\rho},\hat{C}}(\hat{\mu})$, in which case solving $q_1 \leq T_{\hat{\rho},\hat{C}}(\hat{\mu}) \leq q_2$ can lead to unpleasant effects. To satisfy both inequalities simultaneously, one has to solve two different quadratic inequalities. The joint solution can not only attain the three Fieller types, but all possible intersections of two Fieller-type sets. For example, one can obtain confidence sets for the ratio which are only unbounded on one side, such as $]-\infty, l] \cup [l', u]$. These are quite implausible. As discussed earlier: in cases where the denominator is not significantly different from 0, the confidence set should be unbounded on both ends; otherwise, the confidence set of the ratio would reflect a certainty about the sign of the denominator that is not present in the confidence set of the denominator itself. Consequently, we believe that Hwang's approach should only be used with symmetric (and not with equal-tailed) confidence sets for $T_{\hat{\rho},\hat{C}}(\hat{\mu})$. In this case, Hwang's bootstrap approach can easily be interpreted in our geometric approach and is in fact very similar to Fieller's approach: as in Construction 1, one forms the covariance ellipse centered at $\hat{\mu}$ using the estimated covariance matrix $\hat{C}$. But instead of using quantiles of the Student-t distribution to determine the width q of the ellipse, one now uses bootstrap quantiles for this purpose. Then one proceeds exactly as in the Fieller case. This geometric interpretation reveals that Hwang's approach relies on a crucial assumption on the distribution of the sample means: their covariance structure has to be elliptical. Consequently, his bootstrap approach cannot be considered distribution-free.

5.2 A geometric bootstrap approach

We now want to suggest a bootstrap approach which potentially is more suited to deal with highly asymmetric distributions. To this end, we adapt the geometric Construction 3 in a straightforward manner: we use bootstrap methods to construct the one-dimensional confidence intervals I_1 and I_2 used in Construction 3, and then proceed exactly as in Construction 3. The advantage of this approach is obvious: we do not need to make any assumptions on the distribution, can easily use asymmetric confidence intervals I_1 and I_2, and still obtain Fieller-type behavior (as opposed to Hwang's method, which does not have this behavior when using asymmetric bootstrap sets). Moreover, our construction does not assume elliptical covariance structure, and can, for example, be used for heavy-tailed distributions that are not in the domain of attraction of the normal law. In this sense, the geometric bootstrap approach can be applied in situations where both Fieller's and Hwang's confidence sets fail.

Note that one can easily come up with other, more involved bootstrap methods based on the geometric method. For example, one can use more than two projections, can use projections which are not parallel to the coordinate axes, or can even base the wedge on more general two-dimensional convex sets in the plane. A completely different approach can be based on bootstrapping polar representations of the data (along the lines of Koschat (1987)). We tried all those alternative approaches in our simulations. However, given that none of them outperformed the existing methods, we refrain from discussing more details due to space constraints.

5.3 Simulation study

In this section we present numerical simulations to compare the bootstrap approach of Hwang, our geometric bootstrap approach, and Fieller's standard confidence set.

Setup. For X and Y we use three different types of distributions: Normal distributions with means fixed at 1 and variance between 0.1 and 10. Exponential distributions with means between 0.1 and 10. Exponential distributions are highly asymmetric, but still in the domain of attraction of the normal law. Pareto distributions with density $p(x) = ak^a/x^{a+1}$, cf. Chapter 20 of Johnson, Kotz, and Balakrishnan (1994). The parameter a in the Pareto density is called the tail index and is denoted by $\mathrm{Tail}(X)$. For a $\mathrm{Pareto}(k, a)$ distributed random variable, all moments of order larger than a exist, the smaller moments do not exist. In particular, for $a \in]1, 2[$, the expectation exists, but the variance does not exist. In this case the distribution is heavy-tailed and not in the domain of attraction of the normal law. In our experiments, we varied a between 1.1 and 2.5 and chose k such that the expectation of $\mathrm{Pareto}(k, a)$ is 1, that is, $k = (a-1)/a$. For some simulations we also used an inverted Pareto distribution (a Pareto distribution which has been flipped around its mean, so that its tail goes in the negative direction).

For each fixed distribution of X and Y, we independently sampled $n = 20$ ($n = 100$, $n = 1000$, respectively) data points X_i and Y_i. Then we computed the

Fieller confidence set according to Definition 2, our geometric bootstrap confidence sets based on Construction 3 as introduced in Section 5.2, and Hwang's bootstrap confidence sets. Each simulation was repeated $R = 1000$ times to obtain the empirical coverage. As nominal coverage probability we always chose 90% (for investigating coverage, this is more meaningful than the level 95% as it leaves more room for deviations in both directions). To construct the bootstrap confidence sets for the one-dimensional means of X and Y (in the geometric method) and the projection $T_{\hat{\rho},\hat{C}}(\hat{\mu})$ (in Hwang's method) we used different bootstrap methods. As the default bootstrap method we used bootstrap-t (cf. Efron and Tibshirani (1993)). We also tried several other standard methods such as the percentile or the bias corrected and accelerated (BCA) method (cf. Efron and Tibshirani (1993)), but did not observe qualitatively different behavior. To deal with heavy-tailed distributions, we applied methods based on subsampling self-normalizing sums, as introduced by Hall and LePage (1996), see also Romano and Wolf (1999). Here one has to choose a single parameter, the size m of the subsamples. We did not use any automatic method to optimize over m, but, based on values reported in Romano and Wolf (1999), fixed it to $m = 10$ (resp. $40, 400$) for $n = 20$ (resp. $100, 1000$). For all bootstrap methods, we tried both equal-tailed and symmetric confidence sets, in all cases with $B = 2000$ bootstrap samples. We report the bootstrap results using notations such as Hwang(symmetric, bootstrap-t) or Geometric(equal-tailed, subsampling). The terms in parentheses always refer to the construction of the confidence sets for the respective one-dimensional projections: the first term is either "symmetric" or "equal-tailed", the second one "bootstrap-t" or "subsampling".

Evaluation. In all settings we evaluated the empirical coverage (see Table 5.1) and the number of bounded confidence sets (see Table 5.2). Due to space constraints we cannot show the results for all parameter settings, nor any graphical evaluation. For instructive color plots of all evaluations, see the supplementary material to this paper (von Luxburg and Franz (2007)).

Coverage properties in case of finite variance. We start with the case where both X and Y are normally distributed (Table 5.1.A). Here Fieller's confidence set is exact, and indeed we can see that it achieves very good coverage values. In terms of absolute deviation from the nominal confidence level, Hwang performs comparably to Fieller. The difference is that Fieller tends to be slightly conservative, while Hwang tends to be slightly liberal. As predicted, the geometric method is conservative and achieves higher than nominal coverage. For all three methods, the results based on different sample sizes and different bootstrap constructions are qualitatively similar (see the Supplement).

A **Empirical coverage: X ∼ normal, Y ∼ normal, n=100, nominal level 0.90**
Geometric(symmetric,bootstrap-t) / Hwang(symmetric,bootstrap-t) / Fieller

	Var(X)				
Var(Y)	0.1	1	3	5	10
0.1	1.0 / .90 / .93	.97 / .90 / .94	.96 / .89 / .93	.95 / .89 / .93	.94 / .89 / .93
1	.97 / .91 / .94	.99 / .87 / .92	.99 / .91 / .94	.97 / .88 / .92	.96 / .90 / .93
3	.94 / .89 / .92	.99 / .90 / .94	1.0 / .89 / .91	1.0 / .89 / .93	.99 / .89 / .93
5	.97 / .90 / .94	.98 / .91 / .94	.99 / .88 / .93	1.0 / .91 / .94	.99 / .89 / .93
10	.95 / .90 / .93	.96 / .90 / .93	.97 / .88 / .92	.99 / .88 / .93	.99 / .90 / .93

B **Empirical coverage: X ∼ exponential, Y ∼ normal, n=20, nominal level 0.90**
Geometric(equal-tailed,bootstrap-t) / Hwang(equal-tailed,bootstrap-t) / Fieller

	mean(X)			
Var(Y)	1	3	5	10
0.1	.93 / .84 / .89	.90 / .84 / .87	.92 / .85 / .89	.92 / .87 / .90
1	.98 / .88 / .91	.98 / .87 / .91	.98 / .87 / .92	.98 / .88 / .92
3	.98 / .89 / .92	.98 / .89 / .92	.98 / .89 / .93	.97 / .87 / .90
5	.97 / .87 / .93	.96 / .87 / .92	.95 / .86 / .90	.95 / .87 / .92
10	.95 / .87 / .92	.95 / .89 / .92	.95 / .87 / .92	.95 / .86 / .91

C **Empirical coverage: X ∼ pareto, Y ∼ paretoinverted, n=100, nominal level 0.90**
Geometric(equal-tailed,subsampling) / Hwang(equal-tailed,subsampling) / Fieller

	Tail(X)				
Tail(Y)	1.1	1.5	1.9	2.1	2.5
1.1	.73 / .17 / .25	.85 / .26 / .35	.79 / .23 / .32	.78 / .23 / .30	.71 / .18 / .25
1.5	.85 / .28 / .38	.96 / .59 / .70	.98 / .65 / .76	.98 / .64 / .75	.97 / .63 / .73
1.9	.82 / .24 / .33	.98 / .63 / .73	.99 / .72 / .82	.99 / .76 / .85	.99 / .75 / .82
2.1	.78 / .20 / .28	.98 / .65 / .75	.99 / .76 / .85	.99 / .77 / .85	.99 / .78 / .87
2.5	.74 / .19 / .26	.97 / .62 / .73	.99 / .75 / .84	.99 / .79 / .86	.99 / .83 / .89

D **Empirical coverage: X ∼ pareto, Y ∼ paretoinverted, n=100, nominal level 0.90**
Geometric(symmetric,subsampling) / Hwang(symmetric,subsampling) / Fieller

	Tail(X)				
Tail(Y)	1.1	1.5	1.9	2.1	2.5
1.1	.72 / .38 / .26	.81 / .43 / .35	.76 / .37 / .30	.76 / .36 / .29	.70 / .31 / .27
1.5	.79 / .43 / .33	.95 / .71 / .71	.97 / .73 / .74	.96 / .75 / .76	.96 / .73 / .74
1.9	.77 / .41 / .31	.96 / .75 / .75	.98 / .80 / .82	.99 / .81 / .83	.98 / .83 / .85
2.1	.75 / .39 / .30	.97 / .74 / .73	.99 / .83 / .84	.99 / .83 / .85	.99 / .86 / .87
2.5	.71 / .39 / .28	.96 / .74 / .74	.98 / .82 / .84	.99 / .86 / .88	.99 / .87 / .89

E **Empirical coverage: X ∼ pareto, Y ∼ paretoinverted, n=1000, nominal level 0.90**
Geometric(equal-tailed,bootstrap-t) / Hwang(equal-tailed,bootstrap-t) / Fieller

	Tail(X)				
Tail(Y)	1.1	1.5	1.9	2.1	2.5
1.1	.24 / .24 / .24	.35 / .33 / .33	.27 / .27 / .27	.27 / .26 / .26	.22 / .24 / .24
1.5	.33 / .30 / .30	.84 / .72 / .75	.86 / .74 / .77	.88 / .75 / .78	.87 / .75 / .77
1.9	.29 / .29 / .29	.89 / .78 / .80	.96 / .86 / .90	.95 / .85 / .88	.96 / .87 / .89
2.1	.25 / .25 / .25	.87 / .74 / .76	.97 / .84 / .88	.97 / .85 / .89	.98 / .87 / .91
2.5	.26 / .27 / .27	.86 / .75 / .77	.95 / .86 / .88	.98 / .87 / .91	.98 / .88 / .92

Table 5.1: *Empirical coverages of different methods. See text for details.*

A Fraction bounded: X ∼ normal, Y ∼ normal, n=100, nominal level 0.90
Geometric(symmetric,bootstrap-t) / Hwang(symmetric,bootstrap-t) / Fieller

		Var(X)				
		0.1	1	3	5	10
Var(Y)	0.1	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	.90 / .95 / .93	.56 / .67 / .60	.17 / .27 / .20
	1	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	.93 / .96 / .94	.52 / .63 / .57	.15 / .24 / .18
	3	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	.91 / .95 / .92	.54 / .64 / .58	.18 / .26 / .21
	5	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	.91 / .96 / .93	.52 / .64 / .57	.19 / .28 / .21
	10	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	.92 / .95 / .93	.52 / .63 / .57	.19 / .28 / .22

B Fraction bounded: X ∼ exponential, Y ∼ normal, n=20, nominal level 0.90
Geometric(equal-tailed,bootstrap-t) / Hwang(equal-tailed,bootstrap-t) / Fieller

		mean(X)			
		1	3	5	10
Var(Y)	0.1	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	1	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	3	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	5	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	10	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0

C Fraction bounded: X ∼ pareto, Y ∼ paretoinverted, n=100, nominal level 0.90
Geometric(equal-tailed,subsampling) / Hwang(equal-tailed,subsampling) / Fieller

		Tail(X)				
		1.1	1.5	1.9	2.1	2.5
Tail(Y)	1.1	.99 / .41 / .91	1.0 / .69 / .98	1.0 / .84 / .99	1.0 / .88 / 1.0	1.0 / .93 / 1.0
	1.5	.99 / .48 / .89	1.0 / .76 / .98	1.0 / .88 / 1.0	1.0 / .94 / 1.0	1.0 / .96 / 1.0
	1.9	.98 / .48 / .90	1.0 / .80 / .98	1.0 / .92 / 1.0	1.0 / .96 / 1.0	1.0 / .98 / 1.0
	2.1	.99 / .50 / .91	1.0 / .80 / .98	1.0 / .94 / 1.0	1.0 / .96 / 1.0	1.0 / .99 / 1.0
	2.5	.99 / .47 / .89	1.0 / .82 / .99	1.0 / .95 / 1.0	1.0 / .97 / 1.0	1.0 / .99 / 1.0

D Fraction bounded: X ∼ pareto, Y ∼ paretoinverted, n=100, nominal level 0.90
Geometric(symmetric,subsampling) / Hwang(symmetric,subsampling) / Fieller

		Tail(X)				
		1.1	1.5	1.9	2.1	2.5
Tail(Y)	1.1	.45 / .43 / .90	.80 / .73 / .98	.94 / .86 / 1.0	.96 / .88 / 1.0	.99 / .91 / 1.0
	1.5	.48 / .53 / .91	.81 / .79 / .98	.94 / .91 / 1.0	.97 / .92 / 1.0	.99 / .96 / 1.0
	1.9	.47 / .55 / .91	.78 / .79 / .97	.92 / .93 / 1.0	.97 / .96 / 1.0	.99 / .98 / 1.0
	2.1	.46 / .52 / .90	.81 / .84 / .99	.93 / .93 / 1.0	.96 / .96 / 1.0	.99 / .98 / 1.0
	2.5	.45 / .52 / .90	.79 / .81 / .99	.94 / .94 / 1.0	.96 / .97 / 1.0	.98 / .98 / 1.0

E Fraction bounded: X ∼ pareto, Y ∼ paretoinverted, n=1000, nominal level 0.90
Geometric(equal-tailed,bootstrap-t) / Hwang(equal-tailed,bootstrap-t) / Fieller

		Tail(X)				
		1.1	1.5	1.9	2.1	2.5
Tail(Y)	1.1	.91 / .94 / .95	.99 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	1.5	.91 / .94 / .95	.99 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	1.9	.91 / .94 / .95	.98 / .99 / .99	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	2.1	.91 / .93 / .95	.99 / .99 / .99	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0
	2.5	.91 / .93 / .94	.99 / .99 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0	1.0 / 1.0 / 1.0

Table 5.2: *Fraction of bounded confidence sets of different methods. See text for details.*

To investigate the effect of symmetry, we considered the case where one of the random variables is exponentially distributed and thus highly asymmetric (Table 5.1.B). We can see that qualitatively, the three procedures behave as described above (Fieller slightly conservative, Hwang slightly liberal, geometric conservative), even for a small sample size $n = 20$ (results for larger n are similar, see supplement). The fact that the original distribution was asymmetric seems not to have much impact on the results.

Coverage properties in heavy-tailed regime. The general picture changes dramatically if we investigate the case of heavy-tailed distributions. Here we considered simulations with $X \sim$ Pareto, $Y \sim$ Paretoinverted. The reason for using the inverted Pareto distribution for Y (instead of the "standard" one) was that we wanted to study a general asymmetric case — the distribution of the projections on $L_{\rho_\perp}$ would be perfectly symmetric in cases where both X and Y are generated according to the same distribution. Results for $X, Y \sim$ Pareto can be found in the Supplement. In Table 5.1.C we can see that for the heavy-tailed parameters $a < 2$, both Fieller's and Hwang's confidence sets fail completely and lead to empirical coverage probabilities as low as 0.20 instead of 0.90. For Hwang, this happened no matter what bootstrap method we used (symmetric or equal-tailed, bootstrap-t or subsampling), see Tables 5.1.C to E and the Supplement. The method Geometric(equal-tailed, subsampling), on the other hand, performs much better than both Fieller's and Hwang's methods in the heavy-tailed regime $a < 2$. The overall coverage of the geometric method never drops below 0.70, a distinct improvement over the other two methods. It is interesting to observe that the good performance of the geometric method in the heavy-tailed regime decreases sharply if we use bootstrap-t instead of the subsampling bootstrap intervals (Table 5.1.E). The reason is that in the heavy-tailed case, bootstrap-t does not achieve good coverage for the one-dimensional projections, and then the coverage of the final confidence intervals suffers as well. Finally, when the Pareto tail parameter lies in the region $a > 2$, we are again in the domain of attraction of the normal law. Here all results resemble the ones already reported for the finite variance case.

Interpretation of the results in terms of projections. The quality of all three methods depends crucially on the quality of the one-dimensional confidence sets under consideration. For distributions in the domain of attraction of the normal law, Fieller's confidence sets perform very well, even for highly asymmetric distributions. The likely reason is that even for small sample sizes, the distribution of the sample means is already close enough to normal that using bootstrap does not lead to any advantage over using a normal distribution assumption. In the heavy-tailed regime, both Hwang and Fieller fail. This is the case because neither of them achieves good coverage probabilities for the projected one-dimensional random variables $T_{\hat{\rho},\hat{C}}(\hat{\mu})$. Here the geometric method has a big advantage, because, instead of considering projections in arbitrary directions, it only has to deal with projections on the coordinate axes. That the coverage of the one-dimensional confidence sets on the projections is an important indicator of the quality of the confidence set for the ratio can also observed from the fact that the coverage of 0.70 achieved by Geometric(subsampling) in the case $a = 1.1$ (Tables 5.1.C to E) accords with values reported by Romano and Wolf (1999) for the coverage of confidence sets for the mean of Pareto distributions.

Number of bounded confidence sets. In Table 5.2, we compare the number of bounded confidence sets for the three methods. Often, these numbers do not differ much across the different methods. In some cases, Geometric(equal-tailed) performs favorably in that it has more bounded confidence sets than the other methods (see the Supplement). In the asymmetric heavy-tailed case it can be seen that when using symmetric rather than equal-tailed confidence sets in the geometric method, the number of bounded confidence sets decreases heavily (compare Tables 5.2.C and D). This is due to the fact that the one-dimensional confidence sets then become large in both directions (whereas the equal-tailed ones are only large in one direction). Hence, the origin is contained in the resulting rectangle much more often, which leads to unbounded confidence sets. This strongly speaks in favor of using equal-tailed bootstrap confidence sets rather than symmetric ones in the geometric method. Note that for Hwang's method, using equal-tailed confidence sets can lead to implausible confidence sets which are unbounded on one side, but bounded on the other (as explained above). In our experiments, such confidence sets indeed did occur, but not very often (about 20 times in 1000 repetitions).

6. Summary

The geometric approach shows that confidence sets for ratios can be derived from one-dimensional confidence sets for the mean of projections of (X, Y). Of course, the quality of the ratio confidence sets crucially depends on the quality of those one-dimensional confidence sets. For distributions that are in the domain of attraction of the normal law, we recommend using Fieller's confidence set instead of any bootstrap method. Here, Fieller's set works fine even for small sample sizes and asymmetric distributions. Hwang's set achieves comparable results in terms of absolute deviation, but as opposed to Fieller's sets its deviations tend to be to the liberal side, which should be avoided in our opinion. For asymmetric heavy-tailed distributions we recommend using our Geometric(equal-tailed, subsampling) method. This method can be seen as a natural generalization of the geometric interpretation of the Fieller method to a bootstrap scenario. Even though it does not work perfectly, its coverage outperforms Fieller's and Hwang's methods by a large margin, and the number of bounded confidence sets is often higher than for Fieller or Hwang. The performance of the geometric method of course depends on the performance of the bootstrap method used for the one-dimensional distributions. If one is able to improve the bootstrap intervals for the means of those distributions, one is very likely to further improve the coverage of the geometric confidence sets for the ratio.

Acknowledgments
We would like to thank Volker Guiard for helpful comments on an earlier version of this manuscript.

References

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics*, *15*(2), 724 – 731.

Beale, E. M. L. (1962). Some uses of computers in operational research. *Industrielle Organisation*, *31*, 51 – 52.

Buonaccorsi, J. P. (2001). Fieller's theorem. In A. H. El-Shaarawi and W. W. Piegorsch (Eds.), *Encyclopedia of Environmetrics* (pp. 773 – 775). Wiley.

Buonaccorsi, J. P. and Iyer, H. K. (1984). A comparison of confidence-regions and designs in estimation of a ratio. *Communications in statistics — Simulation and computation*, *13*, 723 – 741.

Choquet, D., L'Ecuyer, P., and Léger, C. (1999). Bootstrap confidence intervals for ratios of expectations. *ACM Trans. Model. Comput. Simul.*, *9*(4), 326 – 348.

Dalabehera, M. and Sahoo, L. N. (1995). Efficiencies of 6 almost unbiased ratio estimators under a particular model. *Statistical Papers*, *36*, 61-67.

Davison, A. and Hinkley, D. (1997). *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, *46*, 477 – 480.

Eaton, M. (1981). On the projections of istotropic distributions. *Ann. of Stat.*, *9*(4), 391 – 400.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. of Stat.*, *7*, 1 – 26.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Chapman & Hall, Boca Raton.

Fang, K., Kotz, S., and Ng, K. (1990). *Symmetric multivariate and related distributions.* New York: Chapman and Hall.

Fieller, E. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, *24*(3/4), 428 – 440.

Fieller, E. (1940). The biological standardization of insulin. *Supplement to the Journal of the Royal Statistical Society*, *7*(1), 1 – 64.

Fieller, E. (1944). A fundamental formula in the statistics of biological assays and some applications. *Quarterly Journal of Pharmacy and Pharmacology*, *17*, 117 – 123.

Fieller, E. (1954). Some problems in interval estimation. *J. Roy. Statist. Soc. Ser. B.*, *16*, 175 – 185.

Finney, D. J. (1978). *Statistical methods in biological assay* (3rd ed.). Griffin, London.

Franz, V. H. (submitted). *Ratios: A short guide to confidence limits and proper use.* (Preprint available at `http://arxiv.org/abs/0710.2024`)

Gleser, L. J. and Hwang, J. T. (1987). The nonexistence of 100(1–alpha)–percent confidence sets of finite expected diameter in errors-in-variables and related models. *Annals of Statistics*, *15*, 1351 – 1362.

Guiard, V. (1989). Some remarks on the estimation of the ratio of the expectation values of a two-dimensional normal random variable (correction of the theorem of Milliken). *Biometrical J.*, *31*(6), 681 – 697.

Hall, P. and LePage, R. (1996). On bootstrap estimation of the distribution of the studentized mean. *Ann. Inst. Statist. Math.*, *48*(3), 403 – 421.

Hwang, J. T. G. (1995). Fieller's problems and resampling techniques. *Statistica Sinica*, *5*, 161 – 171.

Johnson, N., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions (volume 1)* (Second ed.). New York: Wiley.

Kendall, M. and Stuart, A. (1961). *The Advanced Theory of Statistics* (Vol. 2). Charles Griffin & Company, London.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *The Annals of Statistics*, *17*(3), 1168 – 1175.

Koschat, M. A. (1987). A characterization of the Fieller solution. *The Annals of Statistics*, *15*(1), 462 – 468.

Miller, R. J. J. (1986). *Beyond ANOVA, Basics of applied statistics.* Wiley, New York.

Milliken, G. (1982). On a confidence interval about a parameter estimated by a ratio of normal random variables. *Comm. Statist. A—Theory Methods*, *11*(17), 1985 – 1995.

Rao, P. S. R. S. (1981). Efficiencies of 9 2-phase ratio estimators for the mean. *Journal of the American Statistical Association*, *76*, 434-442.

Rencher, A. (1998). *Multivariate Statistical Inference and Applications.* Wiley, New York.

Romano, J. and Wolf, M. (1999). Subsampling inference for the mean in the heavy-tailed case. *Metrika*, *50*, 55 – 69.

Schervish, M. (1995). *Theory of Statistics.* Springer, New York.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, New York.

Tin, M. (1965). Comparison of some ratio estimators. *Journal of the American Statistical Association*, *60*(309), 294 – 307.

von Luxburg, U. and Franz, V. (2004). *Confidence sets for ratios: A purely geometric approach to Fieller's theorem* (Technical Report No. 133). Max Planck Institute for Biological Cybernetics.

von Luxburg, U. and Franz, V. (2007). *Supplementary material to "A geometric approach to confidence sets for ratios: Fieller's theorem, generalizations, and bootstrap".*

Ulrike von Luxburg, Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany; Phone: +49 7071 601 540; Fax: +49 7071 601 552

E-mail: ulrike.luxburg@tuebingen.mpg.de

Volker H. Franz, Justus-Liebig-Univerität Giessen, FB06, Abt. Allgemeine Psychologie, Otto-Behaghel-Str. 10F, 35394 Giessen, Germany

E-mail: volker.franz@psychol.uni-giessen.de