

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


An unexpected encounter with Cauchy and Lévy

Citation

Pillai, Natesh S., and Xiao-Li Meng. 2016. “An Unexpected Encounter with Cauchy and Lévy.” The Annals of Statistics 44 (5) (October): 2089–2097. doi:10.1214/15-aos1407.

Published version

https://doi.org/10.1214/15-AOS1407

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:28702414

Terms of use

This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

The Annals of Statistics
2016, Vol. 44, No. 5, 2089–2097
DOI: 10.1214/15-AOS1407
© Institute of Mathematical Statistics, 2016

AN UNEXPECTED ENCOUNTER WITH CAUCHY AND LÉVY

BY NATESH S. PILLAI AND XIAO-LI MENG

Harvard University

The Cauchy distribution is usually presented as a mathematical curiosity, an exception to the Law of Large Numbers, or even as an "Evil" distribution in some introductory courses. It therefore surprised us when Drton and Xiao [*Bernoulli* **22** (2016) 38–59] proved the following result for $m = 2$ and conjectured it for $m \geq 3$. Let $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_m)$ be i.i.d. $\mathrm{N}(0, \Sigma)$, where $\Sigma = \{\sigma_{ij}\} \geq 0$ is an $m \times m$ and *arbitrary* covariance matrix with $\sigma_{jj} > 0$ for all $1 \leq j \leq m$. Then

$$Z = \sum_{j=1}^{m} w_j \frac{X_j}{Y_j} \sim \mathrm{Cauchy}(0, 1),$$

as long as $\vec{w} = (w_1, \ldots, w_m)$ is independent of (X, Y), $w_j \geq 0$, $j = 1, \ldots, m$, and $\sum_{j=1}^{m} w_j = 1$. In this note, we present an elementary proof of this conjecture for any $m \geq 2$ by linking Z to a geometric characterization of Cauchy(0, 1) given in Willams [*Ann. Math. Stat.* **40** (1969) 1083–1085]. This general result is essential to the large sample behavior of Wald tests in many

Tribute: Like many statisticians, we first learned of the name Peter Hall through his many publications. But one of us (Meng) also had an opportunity to work with him directly, and below is a story from this experience.

"When I was a co-editor of *Statistica Sinica*, I had the great fortune to have Peter serving as an AE. It was a great fortune because one of (any) editor's main tasks is to ensure that each submission is handled promptly. Being a preeminent figure of our profession and with such a prolific research output, Peter had no fewer excuses than any of us to not let his AE ship be his highest priority. It therefore surprised me the first time I asked him to handle a submission. I sent him the request close to midnight. By the time I got on my computer the next morning, a full report from Peter was in, detailing his reasons why the paper did not merit a full review!

I never knew what time zone he was on, or whether he had ever slept, but it was a recurring theme that he was the minimal order statistics in terms of response time, yet his report was never rushed in its reasoning or judgement. Eventually I realized that this must be one of the key reasons that he could be the most prolific scholar of our day–always handling whatever came to his desk (or disk) right away, and always with his quick mind fully engaged. It was such an encouragement and inspiration to me for my own editorial work–what excuses did I have not to give it high priority, as I was only publishing a fraction of what Peter published?"

Peter was one of those among us whose human capacity cannot be modeled adequately by the ordinary normal model, but rather by Cauchy or even Levy. We therefore thank the editor Runze Li for giving us this opportunity to dedicate this article to Peter in his memory. Peter, you will be sorely missed, not the least by all the authors who wait anxiously for the review reports on their submissions.

Received May 2015; revised October 2015.

MSC2010 subject classifications. Primary 62E15, 62H10; secondary 62H15.

Key words and phrases. Cauchy distribution, Lévy distribution, Wald test, ratio of Gaussians.

applications such as factor models and contingency tables. It also leads to other unexpected results such as

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{w_i w_j \sigma_{ij}}{X_i X_j} \sim \text{Lévy}(0,1).$$

This generalizes the "super Cauchy phenomenon" that the average of m i.i.d. standard Lévy variables (i.e., inverse chi-squared variables with one degree of freedom) has the same distribution as that of a single standard Lévy variable multiplied by m (which is obtained by taking $w_j = 1/m$ and Σ to be the identity matrix).

1. Cauchy distribution: Evil or angel? Many of us may recall the surprise or even a mild shock we experienced the first time we encountered the Cauchy distribution. *What does it mean that it does not have a mean?* Surely one can always take a sample average, and surely it should converge to something by the Law of Large Numbers (LLN). But then we learned that the LLN does not apply to the Cauchy distribution. Obviously, there cannot be an upper bound on how things may vary in general, and hence it is not difficult to imagine a distribution with infinite variance. But the nonexistence of the mean, which is not the same as the mean is infinite, is harder to envision intuitively. Therefore, some introductory courses (e.g., at our institution) have given the Cauchy distribution the nickname "Evil," because it has created a few excruciating moments (no pun intended), even for some of our best young minds when they tried hard to understand the meaning of not having a mean.

Of course, gain often comes with pain, because soon we would learn something deeper. The nonexistence of the mean for Cauchy is a reflection of the fact that the sample average of an i.i.d. Cauchy sample actually does converge, except it does not converge to a conventional mean, that is, a deterministic number. Rather, it converges trivially to a Cauchy random variable, and more surprisingly, the limiting distribution is the same as that for each term in the entire sequence, as indexed by the sample size. In this sense, the Cauchy distribution is as nice as an angel, because probabilistically its sample average sequence never deviates from its starting point, a dream case for anyone who studies probabilistic behavior of a random sequence.

Through settling a conjecture set forth in [4], we prove in this article that this nice property *can* hold even when the i.i.d. assumption is violated (and the terms are not trivially identical). Specifically, let $\Sigma = \{\sigma_{ij}\} \geq 0$ and $\sigma_{jj} > 0$ for all $j = 1, \ldots, m$, and let X, Y be independent variables distributed as $\mathrm{N}(0, \Sigma)$. We denote the row vectors as $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_m)$. Let $\vec{w} = (w_1, \ldots, w_m)$ be a random vector such that

$$\vec{w} \perp \{X, Y\}, \qquad \sum_{j=1}^{m} w_j = 1 \quad \text{and} \quad w_j \geq 0, \qquad j = 1, \ldots, m. \tag{1.1}$$

Define the random variable

$$Z = \sum_{j=1}^{m} w_j \frac{X_j}{Y_j}. \tag{1.2}$$

In [4], the w_j's were assumed to be fixed constants, but by a conditioning argument, it is trivial to generalize from deterministic $\vec{w}$ to a random $\vec{w}$ as long as it is independent of (X, Y). Therefore, throughout this article we will present the more general random (but independent) $\vec{w}$ version of the results presented in [4] and in the literature with fixed $\vec{w}$, whenever appropriate.

When $\Sigma = \sigma^2 \mathrm{I}_{m\times m}$, it is well known that Z has the standard Cauchy distribution on $\mathbb{R}$ with p.d.f. $\pi^{-1}(1+z^2)^{-1}$, denoted by Cauchy$(0, 1)$; Cauchy(μ, σ) then denotes the distribution of $\mu + \sigma Z$. The fascination with this result is evident from the number of different approaches proposed in the literature to prove it, such as by characteristic functions, convolutions, multivariate change of variables, and most recently by a trigonometric approach [2].

Nevertheless, for arbitrary Σ (so that the terms X_j/Y_j, $j = 1, \ldots, m$ are no longer independent in general), few would expect that Z might remain to be Cauchy$(0, 1)$. However, through simulations Drton and Xiao [4] suspected that this was indeed the case, and via a rather complex and indirect argument involving the Residue theorem, they were able to prove it when $m = 2$ and some cases of perfect correlation when $m > 2$. They conjectured that the result should hold for $m > 2$ for an arbitrary Σ, but their argument does not seem to be easily generalizable to the $m > 2$ case, nor was it feasible to invoke induction because of the dependence induced by Σ.

Seriously intrigued by the findings and the conjecture in [4], we worked for a while trying to extend their complex analytic approach. By using copulas of Cauchy distributions and also the Residue theorem, we ultimately succeeded in finding a proof for all $m \geq 2$. However, we were not satisfied by our lengthy proof because it did not provide any geometric interpretation or statistical insight. We therefore continued to search for a simpler and more inspiring approach. Thanks to an elegant but less well-known geometric characterization of Cauchy$(0, 1)$ given in [15] and in [2, 16], we are able to provide an elementary and geometrically appealing proof of the following result, conjectured by Drton and Xiao in [4].

THEOREM 1.1. *For any $\Sigma = \{\sigma_{ij}\} \geq 0$ such that $\sigma_{jj} > 0$ for all $j = 1, \ldots, m$ and $\vec{w}$ satisfying* (1.1)*, the random variable Z defined in* (1.2) *is distributed as Cauchy*$(0, 1)$.

A theoretical speculation from this unexpected result is that for a set of random variables $\{\xi_1, \ldots, \xi_m\}$, the dependence among them can be overwhelmed by the heaviness of their marginal tails (e.g., take $\xi_j = X_j/Y_j$) in determining the stochastic behavior of their linear combinations. We invite the reader to ponder with us whether this is a pathological phenomenon or something profound.

2. Applications and prior work. As discussed in [4], the Z in (1.2) naturally appears in many important applications. Following [4], let $q \in \mathbb{R}[x_1, \ldots, x_m]$ be a homogeneous m variate polynomial with gradient ∇q. Then by the δ-method, the variance of $q(X) \equiv q(X_1, \ldots, X_m)$ can be approximated by $\nabla q(X)\Sigma\nabla^\top q(X)$, resulting in the Wald statistics

$$W_{q,\Sigma}(X) = \frac{q^2(X)}{\nabla q(X)\Sigma\nabla^\top q(X)} \equiv \{\nabla[\log q(X)]\Sigma\nabla^\top[\log q(X)]\}^{-1}. \tag{2.1}$$

That is, quoting [4], "the random variable $W_{q,\Sigma}$ appears in the large sample behavior of Wald tests with Σ as the asymptotic covariance matrix of an estimator and the polynomial q appearing in a Taylor approximation to the function that defines the constraint to be tested." Thus, there are many applications in which a distribution theory for $W_{q,\Sigma}$ is needed. These include contingency tables [6–8], graphical models [3, Chapter 4], and the testing of so-called "tetrad constraints" in factor analysis [4, 9]. See [4] for more applications and an extensive list of references.

When $X \sim \mathrm{N}(0, \Sigma)$, by the arguments presented in Section 6 of [4], Theorem 1.1 implies the following result, also conjectured in [4], on the quadratic forms for Gaussian random variables.

THEOREM 2.1. *Let $\Sigma = \{\sigma_{ij}\} \geq 0$ and $\sigma_{jj}^2 > 0$ for all $j = 1, \ldots, m$, and $X = (X_1, \ldots, X_m) \sim \mathrm{N}(0, \Sigma)$. If $q(x_1, \ldots, x_m) = x_1^{a_1} \cdots x_m^{a_m}$ with nonnegative real exponents $a_1, \ldots, a_k$ such that $\sum_j a_j > 0$, then*

$$W_{q,\Sigma}(X) \sim \frac{1}{(\sum_{j=1}^m a_j)^2}\chi_1^2,$$

where χ_1^2 denotes a standard chi-squared variable with 1 degree of freedom.

An obvious surprising aspect of Theorem 2.1 is that the exact distribution of $W_{q,\Sigma}$ is free of Σ. A consequential but somewhat hidden surprise is revealed by expressing Theorem 2.1 in the following equivalent form.

THEOREM 2.2. *Let Σ be the same as in Theorem 2.1. For any $(w_1, \ldots, w_m)$ satisfying (1.1), and $X \sim \mathrm{N}(0, \Sigma)$, we have*

$$\left(\frac{w_1}{X_1}, \ldots, \frac{w_m}{X_m}\right)\Sigma\left(\frac{w_1}{X_1}, \ldots, \frac{w_m}{X_m}\right)^\top \sim \chi_1^{-2}, \tag{2.2}$$

where χ_1^{-2} denotes an inverse chi-squared variable with 1 degree of freedom.

These two results are equivalent when we observe that for $q(x_1, \ldots, x_m) = x_1^{a_1} \cdots x_m^{a_m}$, $\nabla \log q = (\frac{a_1}{x_1}, \ldots, \frac{a_m}{x_m})$. Theorem 2.2 therefore is merely a re-expression of Theorem 2.1 using the rightmost expression of (2.1), and by letting

$w_j = a_j / \sum_k a_k$, $j = 1, \ldots, m$. (The generalization to random but independent w_j's follows the conditioning argument discussed previously.)

When $\Sigma = \mathrm{I}_{m\times m}$, Theorem 2.2 recovers the not so well-known "super Cauchy phenomenon," that is, the average of m i.i.d. χ_1^{-2} is distributed exactly as m times a single χ_1^{-2}, by taking $w_j = m^{-1}$ for all j's in Theorem 2.2. The χ_1^{-2} distribution is also known as the standard Lévy distribution (with location parameter equal to 0 and scale parameter 1; see [1], page 33). This result can easily be verified via the characteristic function of the χ_1^{-2} distribution,

$$\phi(t) = e^{-\sqrt{-2\mathbf{i}t}},$$

because $\phi^m(t/m) = \phi(mt)$. We call this "super Cauchy phenomenon" because it says that for an i.i.d. Lévy sample of size m, their average is m times more variable than any one of them, exceeding the case of Cauchy where the average has the same variability as a single variable. That is, if we denote the α-percentile of the average of m i.i.d. samples by $p_\alpha^{(m)}$, then for the Cauchy sample we have $p_\alpha^{(m)} = p_\alpha^{(1)}$, but for the Lévy sample, we have $p_\alpha^{(m)} = m p_\alpha^{(1)}$.

Clearly, the characteristic function approach does not apply when Σ is not diagonal, but nevertheless Theorem 2.2 says that the above distributional result generalizes when Σ goes beyond the diagonal case. At the first glance, result (2.2) might seem to be a wishful thinking by a novice to probability or algebra, who (mistakenly) treats $X^{-1}\Sigma X^{-\top}$, where $X^{-1} \equiv \{X_1^{-1}, \ldots, X_m^{-1}\}$, as $(X\Sigma^{-1}X^\top)^{-1}$, which would then permit him to use the usual standardization trick by letting $Z = X\Sigma^{-1/2} \sim \mathrm{N}(0, I)$. However, this would have led him to guess that the left-hand side of (2.2), when $w_j = m^{-1}$, $j = 1, \ldots, m$, distributes as $m^{-2}(\sum_j Z_j^2)^{-1}$, which would then be $m^{-2}\chi_m^{-2}$, not χ_1^{-2}.

It is instructive to express the left-hand side of (2.2) as the average of m^2 terms when we take $w_j = m^{-1}$:

$$\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\sigma_{ij}}{X_i X_j} \sim \chi_1^{-2}. \tag{2.3}$$

This is a rather remarkable result because the left-hand side can only be made invariant (algebraically) to the variances σ_{jj} by expressing $X_j = \sqrt{\sigma_{jj}}X'_j$ with variance of X'_j equal to 1, but not to the correlations $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. Yet, (2.2) says that it is actually a pivotal quantity for $\{\rho_{ij}\}$.

There are also some works on multivariate Cauchy densities that are relevant to our problem. For instance, in [13] and [10], the authors studied the distribution theory for the ratio of two Gaussian random variables. Ferguson [5] derived a general result for the characteristic function of a multivariate Cauchy distribution. In [11], the authors studied a generalization of the bivariate Cauchy distribution. McCullagh [14] showed that it is natural to parametrize the family of Cauchy (μ, σ) distributions in the complex plane, as this location-scale family is closed

under Möbius transformations. Finally, we remark that Drton and Xiao [4] did not prove the $m = 2$ case of Theorem 1.1 directly; instead, they first proved of a special case of Theorem 2.1 [with $q(x_1, x_2) = x_1^{a_1} x_2^{a_2}$], generalizing a similar result of Glonek [8]. Drton and Xiao then obtained the conclusion of Theorem 1.1 for $m = 2$ as a corollary using change of variables.

3. Proof. The key idea of our proof relies on the following result from [15], as reformulated in [16]. The proof of this lemma as given in [15] is short, and it also relies on the Residue theorem for contour integration. Very recently, the author of [2] gave a geometric proof for the case of $m = 2$. See [12] for further generalizations of this lemma.

LEMMA 3.1. *Let* $\Theta_1 \sim \text{Unif}(-\pi, \pi]$, *and* $\{w_1, \ldots, w_m\}$ *be independent of* Θ_1, *where* $w_j \geq 0$ *and* $\sum_j w_j = 1$. *Then for any* $\{u_1, \ldots, u_m\}$, *where* $u_j \in \mathbb{R}$,

$$\sum_{j=1}^{m} w_j \tan(\Theta_1 + u_j) \sim \text{Cauchy}(0, 1).$$

Intuitively speaking, if $\Theta_1 \sim \text{Unif}(-\pi, \pi]$, then for any constant u_j, $(\Theta_1 + u_j) \bmod (2\pi) \sim \text{Unif}(-\pi, \pi]$, and hence $\tan(\Theta_1 + u_j) \sim \text{Cauchy}(0, 1)$. The significance of this lemma is that any convex combination of these *dependent* Cauchy(0, 1) variables is still distributed as Cauchy(0, 1). As we shall see below, Theorem 1.1 is a direct consequence of this remarkable result. We first prove Theorem 1.1 when Σ is strictly positive definite, that is, $\Sigma > 0$, and then invoke a limiting argument to cover the cases with $\Sigma \geq 0$.

PROOF OF THEOREM 1.1. When $\Sigma > 0$, we write $\Sigma^{-1} = \{b_{ij}\}$. The joint density of (X, Y) then can be written as

$$f_{X,Y}(x, y) = K \exp\left\{-\frac{1}{2}\left(\sum_{j=1}^{m} b_{jj}(x_j^2 + y_j^2) + 2\sum_{j\neq k} b_{jk}(x_j x_k + y_j y_k)\right)\right\},$$

where $x, y \in \mathbb{R}^m$ and K is a constant that depends only on m and Σ. Let us make the transformation $(X_j, Y_j) = (R_j \sin(\Theta_j), R_j \cos(\Theta_j))$, where $0 \leq R_j < \infty$ and $\Theta_j \in (-\pi, \pi]$. Then the joint density of $R = \{R_1, \ldots, R_m\}$ and $\Theta = \{\Theta_1, \ldots, \Theta_m\}$ is

$$\begin{aligned} &f_{R,\Theta}(r, \theta) \\ &\quad \propto \exp\left\{-\frac{1}{2}\left(\sum_{j=1}^{m} b_{jj} r_j^2 + 2\sum_{j\neq k} b_{jk} r_j r_k \cos(\theta_j - \theta_k)\right)\right\} \prod_{j=1}^{m} r_j, \end{aligned} \tag{3.1}$$

for $r \in [0, \infty)^m$ and $\theta \in (-\pi, \pi]^m$. The term $\prod_{j=1}^m r_j$ in equation (3.1) is the Jacobian of the $(X, Y) \to (R, \Theta)$ transformation.

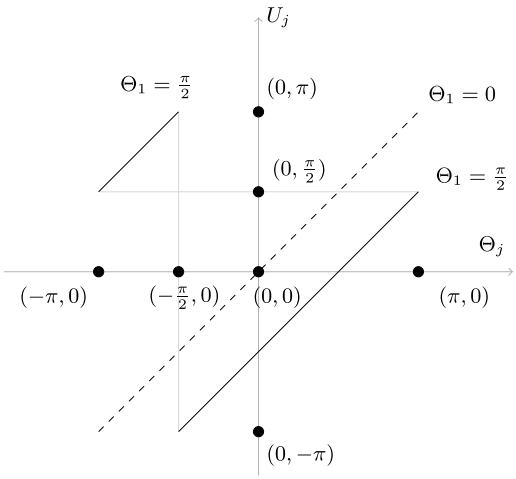


FIG. 1. *For every value of $\Theta_1 \neq 0$, the map $\Theta_j \mapsto U_j$ $(j \geq 2)$ is a disjoint union of two lines, and is one-to-one. The figure shows the graph of (Θ_j, U_j) when $\Theta_1 = \frac{\pi}{2}$ (plotted as two solid lines). When $\Theta_1 = 0$, $U_j = \Theta_j$ (plotted in dashed line).*

We then make a further transformation, $\mathcal{F} : (-\pi, \pi]^m \mapsto (-\pi, \pi]^m$, with $\mathcal{F}(\Theta_1, \ldots, \Theta_m) = (\Theta_1, U_2, \ldots, U_m)$, where

$$U_j = (\Theta_j - \Theta_1) + 2\pi[\mathbf{1}_{\{\Theta_j - \Theta_1 \leq -\pi\}} - \mathbf{1}_{\{\Theta_j - \Theta_1 > \pi\}}], \qquad 2 \leq j \leq m. \tag{3.2}$$

This is a form of $U_j = (\Theta_j - \Theta_1) \bmod (2\pi)$, but with the assurance that the support of U_j is $(-\pi, \pi]$ regardless of the value of Θ_1, and that $U_j - U_k = (\Theta_j - \Theta_k) \bmod (2\pi)$, and $\Theta_j = (\Theta_1 + U_j) \bmod (2\pi)$. The map $\mathcal{F}$ is one-to-one as shown in Figure 1. Furthermore, the points where the map $\mathcal{F}$ is *not* differentiable is contained in the set

$$\{\Theta \in (-\pi, \pi]^m : \Theta_j - \Theta_1 \in \{-\pi, \pi\} \text{ for some } j \geq 2\},$$

as can be seen from Figure 1. Clearly, this set has Lebesgue measure zero. Outside this set, we have $\frac{\partial U_j}{\partial \Theta_j} = 1$. Thus, the Jacobian of the map $\mathcal{F}$ is 1 for all $\Theta \in (-\pi, \pi]^m$ except for the above measure zero set.

Set $U_1 \equiv 0$ and denote $U = (U_1, U_2, \ldots, U_m)$. Since $\cos(W_1) = \cos(W_2)$ for any $W_1 = W_2 \bmod (2\pi)$, we can write the joint density in the new coordinates as

$$f_{R,\Theta_1,U}(r, \theta_1, u) \propto \exp\left\{-\frac{1}{2}\left(\sum_{j=1}^m b_{jj} r_j^2 + 2\sum_{j \neq k} b_{jk} r_j r_k \cos(u_k - u_j)\right)\right\} \prod_{j=1}^m r_j$$

with $r \in [0, \infty)^m$, $\theta_1 \in (-\pi, \pi]$, $u_1 = 0$ and $u_2, \ldots, u_m \in (-\pi, \pi]$. The only observations we need from the above line are: (i) Θ_1 is independent of U and (ii) $\Theta_1 \sim \text{Unif}(-\pi, \pi]$. But Z in (1.2) can be written as

$$Z = \sum_{j=1}^m w_j \frac{X_j}{Y_j} = \sum_{j=1}^m w_j \tan(\Theta_j) = \sum_{j=1}^m w_j \tan(\Theta_1 + U_j),$$

because $\tan(W_1) = \tan(W_2)$ for any $W_1 = W_2 \bmod (2\pi)$. Since U is independent of Θ_1, conditional on U, Lemma 3.1 yields that $Z \sim \text{Cauchy}(0, 1)$. It follows immediately that Z is also marginally distributed as Cauchy$(0, 1)$, completing the proof when $\Sigma > 0$.

When we relax the assumption $\Sigma > 0$ to $\Sigma \geq 0$, Σ^{-1} may not exist. However, for any $n \in \mathbb{N}$, $\Sigma^{(n)} = \Sigma + n^{-1}\mathrm{I}_{m\times m} > 0$. Let $X^{(n)} = (X_1^{(n)}, \ldots, X_m^{(n)})$ and $Y^{(n)} = (Y_1^{(n)}, \ldots, Y_m^{(n)})$ be i.i.d. from $\mathrm{N}(0, \Sigma^{(n)})$. As $n \to \infty$, we have $(X^{(n)}, Y^{(n)}) \rightsquigarrow (X, Y)$, where "$\rightsquigarrow$" indicates convergence in distribution. Next, the mapping $\zeta : \mathbb{R}^{2m} \mapsto \mathbb{R}$ defined by

$$\zeta(x, y) = \sum_{j=1}^{m} w_j \frac{x_j}{y_j}$$

is continuous, except when $y \in B$ with

$$B = \Big\{(y_1, \ldots, y_m) \in \mathbb{R}^m : \min_{1\leq j\leq m} |y_j| = 0\Big\}.$$

Now the result follows from the continuous mapping theorem. Indeed, since $(X^{(n)}, Y^{(n)}) \rightsquigarrow (X, Y)$, the continuous mapping theorem yields that

$$Z^{(n)} = \zeta\big(X^{(n)}, Y^{(n)}\big) \rightsquigarrow \zeta(X, Y) = Z \tag{3.3}$$

as $n \to \infty$, provided the points of discontinuity of ζ belong to a zero-measure set. However, since $Y_j \sim \mathrm{N}(0, \sigma_{jj})$ where $\sigma_{jj} > 0$ by our assumption, we have

$$\mathbb{P}(Y \in B) = \mathbb{P}\Big(\min_{1\leq j\leq m} |Y_j| = 0\Big) \leq \sum_{j=1}^{m} \mathbb{P}\big(|Y_j| = 0\big) = 0,$$

verifying (3.3). By our previous argument, we know that $Z^{(n)} \sim \text{Cauchy}(0, 1)$ for all $n \in \mathbb{N}$, and hence (3.3) implies that $Z \sim \text{Cauchy}(0, 1)$. □

Acknowledgements. We thank Alexei Borodin, Joseph Blitzstein, Carlo Beenakker, Mathias Drton, Sukhada Fadnavis, Steven Finch, Gérard Letac, Luke Miratrix, Carl Morris, Christian Robert, Aaron Smith, Alex Volfovsky, Nanny Wermuth and Robert Wolpert for helpful comments and encouragement. We also thank the Editor, Associate Editor and the referees for catching an oversight in our proof in the initial submission, and NSF and ONR for partial financial support.

REFERENCES

[1] Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **116**. Cambridge Univ. Press, Cambridge. MR2512800

[2] Cohen, M. P. (2012). Sample means of independent standard Cauchy random variables are standard Cauchy: A new approach. *Amer. Math. Monthly* **119** 240–244. MR2911439

[3] Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics*. Birkhäuser, Basel. MR2723140

[4] DRTON, M. and XIAO, H. (2016). Wald tests of singular hypotheses. *Bernoulli* **22** 38–59.
[5] FERGUSON, T. S. (1962). A representation of the symmetric bivariate Cauchy distribution. *Ann. Math. Stat.* **33** 1256–1266. MR0143281
[6] GAFFKE, N., HEILIGERS, B. and OFFINGER, R. (2002). On the asymptotic null-distribution of the Wald statistic at singular parameter points. *Statist. Decisions* **20** 379–398. MR1960710
[7] GAFFKE, N., STEYER, R. and VON DAVIER, A. A. (1999). On the asymptotic null-distribution of the Wald statistic at singular parameter points. *Statist. Decisions* **17** 339–358. MR1745064
[8] GLONEK, G. F. V. (1993). On the behaviour of Wald statistics for the disjunction of two regular hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** 749–755. MR1223941
[9] HARMAN, H. H. (1960). *Modern Factor Analysis*. Univ. Chicago Press, Chicago, IL. MR0159393
[10] HINKLEY, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika* **56** 635–639. MR0254946
[11] JAMALIZADEH, A. and BALAKRISHNAN, N. (2008). On a generalization of bivariate Cauchy distribution. *Comm. Statist. Theory Methods* **37** 469–474. MR2392340
[12] LETAC, G. (1977). Which functions preserve Cauchy laws? *Proc. Amer. Math. Soc.* **67** 277–286. MR0584393
[13] MARSAGLIA, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *J. Amer. Statist. Assoc.* **60** 193–204. MR0178490
[14] MCCULLAGH, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79** 247–259. MR1185127
[15] PITMAN, E. J. G. and WILLIAMS, E. J. (1967). Cauchy-distributed functions of Cauchy variates. *Ann. Math. Stat.* **38** 916–918. MR0210166
[16] WILLIAMS, E. J. (1969). Cauchy-distributed functions and a characterization of the Cauchy distribution. *Ann. Math. Stat.* **40** 1083–1085. MR0243657

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1 OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138
USA
E-MAIL: pillai@fas.harvard.edu
meng@fas.harvard.edu