IOM THK75-20

To: WOH

From: T.H. Kerr

[For Course No. EE3372 at Northeastern, please See Sections 1, 1.1, 1.1.1, 1.1.2, 1.2, and 1.3]

Subject: Important Theoretical Limitations Usually Encountered in Parameter Identification as Emphasized by Three Counterexamples

## Introduction

As we get more thoroughly immersed in parameter identification problems, it will be required that attention be focused on certain theoretical details that would normally avoid close scrutiny. Some ideas are so familiar that normally they would not ever be questioned. The point is that many common misconceptions exist. The following short examples are to serve as a reminder of what theoretical limitations are associated with many ideas that currently abound. An overview of what is being presented here may be seen from the elaborated Table of Contents.

## Table of Contents

Page

5.1 For $y = a\tan x$, $a > 0$, with $x$ uniform on $[-\pi, \pi]$, $y$ is Cauchy.

5.2 For $z = \frac{x}{y}$, and $x$ and $y$ are jointly normal zero mean, then $z$ is Cauchy.

1. When Do the <u>Conclusions</u> of the Central Limit Theorem <u>Fail</u>? (i.e., when does the sum of independent selected samples from independently identically distribute populations not go to a Gaussian in the limit as the number of elements in the sum increases without bound?)

## 1.1 Examples of pdf's yielding sums that ~~do~~ go to Gaussian

### 1.1.1 Binomial (discrete)

***Random Variables of Lattice Type***

We now assume that each r.v. of the sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots$ is of lattice type (see page 96), taking values $bk$ ($k = \ldots, -1, 0, 1, \ldots$), forming an arithmetic progression. The densities $f_i(x)$ are sequences of impulses, and their characteristic functions $\Phi_i(\omega)$ are periodic with period $2\pi/b$ (Fig. 8-4).

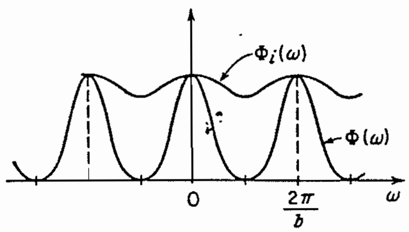



*Fig. 8-4*

The characteristic function of the sum

$$\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_n$$

is given by [see (8-39)]

$$\Phi(\omega) = \Phi_1(\omega) \cdots \Phi_n(\omega)$$

Therefore it is also periodic. From the above, we conclude that the density $f(x)$ of $\mathbf{x}$ is also of lattice type, taking values at the points $bk$. The determination of the probabilities $P\{\mathbf{x} = bk\}$ involves multiple summations. However, for large $n$, these probabilities are approximately equal to the ordinates of a normal curve, with the same mean $\eta$ and variance $\sigma^2$ as $f(x)$:

$$\boxed{P\{\mathbf{x} = bk\} \simeq \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(bk-\eta)^2/2\sigma^2}} \tag{8-104}$$

In other words, $f(x)$ tends to a sequence of equidistant impulses with envelope a normal curve:

$$f(x) \simeq \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\eta)^2/2\sigma^2} \sum_k \delta(x - bk) \tag{8-105}$$

This is the central-limit theorem for lattice type r.v., and can be deduced from (8-101).

*DeMoivre-Laplace Theorem.* The normal approximation

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}}\, e^{-(k-np)^2/2npq} \tag{8-106}$$

[see (3-36)] of the binomial coefficients is a special case of (8-104). Indeed, suppose that the r.v. $\mathbf{x}_i$ take only the value 0 and 1, with

$$P\{\mathbf{x}_i = 1\} = p \qquad P\{\mathbf{x}_i = 0\} = q$$

We then have (see Example 5-29)

$$E\{\mathbf{x}_i\} = p \qquad \sigma_{x_i}^2 = pq \qquad \Phi_i(\omega) = pe^{j\omega} + q$$

The characteristic function of the sum $\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_n$ is given by

$$\Phi(\omega) = (pe^{j\omega} + q)^n = \cdots + \binom{n}{k} p^k q^{n-k} e^{j\omega k} + \cdots$$

Thus $\mathbf{x}$ is of lattice type, and

$$P\{\mathbf{x} = k\} = \binom{n}{k} p^k q^{n-k}$$

But

$$E\{\mathbf{x}\} = np \qquad \sigma_x^2 = npq$$

Substituting the above into (8-104), we obtain (8-106).

Ref. 1

## 1.1.2 Exponential (Continuous)

Using the following properties of characteristic functions:

$$\mathcal{C}_{x_i}(\nu) \triangleq E\left[e^{j\nu x_i}\right] \quad (1$$

$$\mathcal{C}_{ax_i}(\nu) = E\left[e^{j(\nu a)x_i}\right] = C_{x_i}(a\nu) \quad (2$$

$$\mathcal{C}_{ax_i+b}(\nu) = E\left[e^{j\nu(ax_i+b)}\right] = e^{j\nu b}E\left[e^{j\nu a x_i}\right] = e^{j\nu b}\mathcal{C}_{x_i}(a\nu$$

$$\mathcal{C}_{\sum_i^N ax_i+b}(\nu) = \prod_{i=1}^{N} \mathcal{C}_{ax_i+b}(\nu) = e^{j\nu bN}\left[\mathcal{C}_{x_i}(a\nu)\right]^N$$

and the following series expansion:

$$\ln(1+z) = z - \tfrac{1}{2}z^2 + \tfrac{1}{3}z^3 - \tfrac{1}{4}z^4 + \tfrac{1}{5}z^5 - \dots \quad \text{for } |z| \leq 1 \quad (5$$

Eq. 4.1.26 of Ref. 2

in conjunction with the following information on the exponent distribution:

$$E[x_i] = 1 \quad (6$$

$$Var[x_i] = 1 \quad (7$$

$$p_{x_i}(w) = e^{-w} \quad \text{for} \quad 0 \leq w < \infty \quad (8)$$

$$\mathcal{C}_{x_i}(\nu) = \frac{1}{(1 - j\nu)} \quad (9)$$

(Eq. (8) and (9) are Fourier Transform pai

one can form

$$u_i \triangleq \frac{x_i - 1}{\sqrt{N}} = \frac{1}{\sqrt{N}}x_i - \frac{1}{\sqrt{N}} \quad (10$$

and show that $\lim \sum_{i=1}^{N} u_i \longrightarrow N(0,1)$ in distribution as predicted by the Central Limit Theorem. Substituting Eq.(9) and (10) into Eq(3) and (4) yields

$$\mathbb{C}_{\sum_{i=1}^{N}\frac{\left(\frac{x_i-1}{1}\right)}{\sqrt{N}}}(\nu) = \frac{e^{-j\nu\sqrt{N}}}{\left(1-j\frac{\nu}{\sqrt{N}}\right)^N} \qquad (1$$

Taking logarithms of both sides yields

$$\ln \mathbb{C}_{\sum_{i=1}^{N}\frac{\left(\frac{x_i-1}{1}\right)}{\sqrt{N}}}(\nu) = \ln e^{-j\nu\sqrt{N}} - \ln\left(1-j\frac{\nu}{\sqrt{N}}\right)^N \qquad (12-$$

$$= -j\nu\sqrt{N} - N\ln\left(1-j\frac{\nu}{\sqrt{N}}\right) \qquad (12-$$

Using the series expansion of Eq.(4) on Eq.(12-b) results in

$$\ln \mathbb{C}_{\sum_{i=1}^{N}\frac{\left(\frac{x_i-1}{1}\right)}{\sqrt{N}}}(\nu) = -j\nu\sqrt{N} - N\left\{-j\frac{\nu}{\sqrt{N}} - \frac{1}{2}\left(-j\frac{\nu}{\sqrt{N}}\right)^2 + \frac{1}{3}\left(-j\frac{\nu}{\sqrt{N}}\right)^3 - \cdots\right\} \qquad (13)$$

where for each $\nu$ choose N large enough so that

$$\left|\frac{\nu}{\sqrt{N}}\right| < 1 \qquad (14)$$

Now

$$\ln \mathbb{C}_{\sum_{i}^{N}\frac{(x_i-1)}{\sqrt{N}}}(\nu) = \cancel{-j\nu\sqrt{N}} + \cancel{j\nu\sqrt{N}} - \frac{1}{2}\nu^2 - j\frac{1}{3}\frac{\nu^3}{\sqrt{N}} - \frac{1}{4}\frac{\nu^4}{N} \cdots \qquad (15)$$

Now

$$\lim_{N\to\infty} \mathbb{C}_{\sum_{i}^{N}\frac{(x_i-1)}{\sqrt{N}}}(\nu) = e^{\lim_{N\to\infty}\ln \mathbb{C}_{\sum\frac{(x_i-1)}{\sqrt{N}}}(\nu)} = \exp\left\{\lim_{N\to\infty}\left[-\frac{1}{2}\nu^2 - j\frac{1}{3}\frac{\nu^3}{\sqrt{N}} - \frac{1}{4}\frac{\nu^4}{N}\cdots\right.\right. \qquad (1$$

hence

$$\lim_{N\to\infty} \mathbb{C}_{\sum_{i=1}^{N}\frac{(x_i-1)}{\sqrt{N}}}(\nu) = \exp\left\{-\frac{1}{2}\nu^2\right\} \qquad (1$$

where the interchange of Limit and exponentiation are valid by the continuous monotone property of the exponential function. Eq. (17) is recognized to correspond to the characteristic function of a Gaussian. Since the limiting characteristic function of Eq. (17) is a Gaussian, its Fourier transform pair, the pdf also goes to a Gaussian (one definition of convergence in distribution).

## 1.2 An example of a pdf (Cauchy) that yields a sum that doesn't go to Gaussian

**Table 7.5 Transformations of random variables (distributions specified)** *(Continued)*

| *No.* | *Transformation u* | *Original distribution* | *Final distribution* |
|---|---|---|---|
| 17 | $x_1 + x_2 + \cdots + x_n$ | $x_i$'s all Cauchy distributions with parameters $\theta_i$ and $\alpha_i$ <br> $f(x) = \dfrac{\alpha_i/\pi}{(x-\theta_i)^2 + \alpha_i^2}$ | $h$ = Cauchy <br> $h(u) = \dfrac{\alpha/\pi}{(x-\theta)^2 + \alpha^2} \qquad -\infty < u \le +\infty$ <br> $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ <br> $\theta = \theta_1 + \theta_2 + \cdots + \theta_n$ |

Ref. 3

The Sum of Cauchy variates is Cauchy no matter how many terms are in the sum (even as one goes to the limit). However, all Cauchy random variables have no (precious) moments (a sad life!) as can be seen in the table below.

*Some one-dimensional continuous distribution functions*

| | Name | Domain | Probability Density Function $f(x)$ | Restrictions on Parameters | Mean | Variance | Skewness $\gamma_1$ | Excess $\gamma_2$ | Characteristic function |
|---|---|---|---|---|---|---|---|---|---|
| 26.1.25 | Error function | $-\infty<x<\infty$ | $\frac{h}{\sqrt{\pi}} e^{-h^2x^2}$ | $0<h<\infty$ | 0 | $\frac{1}{2h^2}$ | 0 | 0 | $e^{\frac{-t^2}{4h^2}}$ |
| 26.1.26 | Normal | $-\infty<x<\infty$ | $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$ | $-\infty<m<\infty$ <br> $0<\sigma<\infty$ | $m$ | $\sigma^2$ | 0 | 0 | $e^{imt-\frac{\sigma^2t^2}{2}}$ |
| 26.1.27 | Cauchy | $-\infty<x<\infty$ | $\frac{1}{\pi\beta}\frac{1}{1+\left(\frac{x-\alpha}{\beta}\right)^2}$ | $-\infty<\alpha<\infty$ <br> $0<\beta<\infty$ | not defined | not defined | not defined | not defined | $e^{i\alpha t-\beta\|t\|}$ |

Ref. 2

However, the Gaussian or Normal distribution does have well-defined well behaved moments (as seen in the above table). Therefore the sums of Cauchy distributed variates can never have a character close to Gaussian.

Ref. 3

1.3 The Laplace-Liapounoff conditions, sufficient conditions for the conclusions of the Central Limit Theorem to hold.

**Central Limit Theorem**

A sufficient set of conditions for convergence of the central-limit theorem are the Laplace-Liapounoff conditions, which state that a *sufficient* condition for convergence is that

$$\lim_{n\to\infty} \left[(\operatorname{var} u)^{-\frac{3}{2}} \sum_{i=1}^{n} E(|x_i - \mu_i|^3)\right] = 0$$

provided that each $x_i$ is independent and possesses a third moment. Notice that the criterion involves a ratio. The numerator of the ratio is the sum of quantities which are similar to the third moment about the mean (the absolute-value signs make this different from the third moment). The denominator is the variance of $u$ raised to the $\frac{3}{2}$ power. If the ratio approaches zero as $n \to \infty$, the central-limit theorem holds. Since this is only a sufficient condition, if the test fails, we must investigate further; nothing is certain.

A complete discussion of various necessary and sufficient conditions for the convergence of the central-limit theorem can be found in Munroe.[1]

Ref. 3

(ii) LAPUNOV THEOREM (C.L.T.). *Let $\{X_n\}$, $n = 1, 2, \ldots$ be a sequence of independent random variables. Let $E(X_n) = \mu_n$, $E(X_n - \mu_n)^2 = \sigma_n^2 \neq 0$, $E(x_n - \mu_n)^3 = \alpha_n$ and $E|X_n - \mu_n|^3 = \beta_n$ exist for each $n$. Furthermore let*

$$B_n = \left(\sum_1^n \beta_i\right)^{1/3}, \qquad C_n = \left(\sum_1^n \sigma_i^2\right)^{1/2}.$$

*Then if $\lim (B_n/C_n) = 0$ as $n \to \infty$, the d.f. of*

$$Y_n = \frac{\sum (X_i - \mu_i)}{C_n}$$

*tends to $\Phi(x)$.* (Gaussian)

For a proof of this theorem and other forms of central limit theorems see Gnedenko (1962) and Gnedenko and Kolmogorov (1954).

Gnedenko, B. V. (1962), *Theory of Probability*, Chelsea, New York.

Gnedenko, B. V. and A. N. Kolmogorov (1954), *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading, Mass.

Ref. 4

[1] M. E. Munroe, "The Theory of Probability," McGraw-Hill Book Company, New York, 1951.

2. When Does $E[X] \neq E\{E[X|Y]\}$? (One example is associated with Student's $t$ which has Cauchy as a special case.)

## On the relation $E(X) = E\{E(X|Y)\}$

By PETER ENIS

*State University of New York at Buffalo*

Summary

Attention is drawn to a frequently misstated result concerning a relation between conditional and unconditional expectation. The correct statement of the result is indicated along with a simple example illustrating its need.

*Some key words*: Expectation; Conditional expectation.

In many books on probability and statistics, both elementary and advanced, the reader is provided with the very useful formula for obtaining the expectation of a random variable, *viz*.

$$E(X) = E\{E(X|Y)\}, \tag{1}$$

without being warned against its indiscriminate use. In particular, the requirement that $E(X)$ must exist (Rényi, 1970, p. 260) for (1) to be valid is often omitted, leading the unwary reader to believe that existence of the right hand side guarantees that of the left hand side; this misconception is clearly illustrated by the following simple example.

Let $Y$ be a positive random variable with probability density function

$$g_\nu(y) = \frac{(\frac{1}{2}\nu)^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} y^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}\nu y} \quad (y > 0),$$

where $\nu > 0$. Further, let the conditional distribution of $X$ given $Y = y$ be specified for $y > 0$ by the probability density function

$$f(x|y) = (2\pi)^{-\frac{1}{2}} y^{\frac{1}{2}} e^{-\frac{1}{2}yx^2} \quad (-\infty < x < \infty).$$

Thus, $E(X|Y) = 0$, so that $E\{E(X|Y)\} = 0$. On the other hand, the marginal probability density function of $X$ is easily shown to be

$$h_\nu(x) = \frac{\Gamma\{\frac{1}{2}(\nu+1)\}}{\Gamma(\frac{1}{2}\nu)(\pi\nu)^{\frac{1}{2}}} \left(1+\frac{x^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} \quad (-\infty < x < \infty);$$

i.e. marginally $X$ has the Student $t$ distribution with $\nu$ degrees of freedom. In particular, for $\nu = 1$ the random variable $X$ has a Cauchy distribution for which $E(X)$ does not exist although, as mentioned above, $E\{E(X|Y)\} = 0$.

As the reader can show, it is only for $r \geqslant \nu$ that $E(X^r)$ does not exist, whereas it is only for even $r \geqslant \nu$ that $E\{E(X^r|Y)\}$ does not exist. Furthermore, relation (1) is always true if $X$ is a nonnegative random variable.

Reference

Rényi, A. (1970). *Foundations of Probability*. San Francisco: Holden-Day.

[*Received August* 1972. *Revised November* 1972]

Ref. 5

# 3. When Does No Efficient (i.e., achieving its Cramer-Rao lower bound) Parameter Estimate Exist?

## 3.1 Conditions for being an efficient estimator

**17.15** Now consider an unbiassed estimator, $t$, of some function of $\theta$, say $\tau(\theta)$. This formulation allows us to consider unbiassed and biassed estimators of $\theta$ itself, and also permits us to consider, for example, the estimation of the standard deviation when the parameter is equal to the variance. We thus have

$$E(t) = \int \dots \int t L \, dx_1 \dots dx_n = \tau(\theta). \tag{17.20}$$

We now differentiate (17.20), the result being

$$\int \dots \int t \frac{\partial \log L}{\partial \theta} L \, dx_1 \dots dx_n = \tau'(\theta),$$

which we may re-write, using (17.18), as

$$\tau'(\theta) = \int \dots \int \left\{ t - \tau(\theta) \right\} \frac{\partial \log L}{\partial \theta} L \, dx_1 \dots dx_n. \tag{17.21}$$

By the Cauchy–Schwarz inequality, we have from (17.21)

$$\{\tau'(\theta)\}^2 \leqslant \int \dots \int \{t - \tau(\theta)\}^2 L \, dx_1 \dots dx_n . \int \dots \int \left(\frac{\partial \log L}{\partial \theta}\right)^2 L \, dx_1 \dots dx_n,$$

which, on rearrangement, becomes

$$\operatorname{var} t = E\{t - \tau(\theta)\}^2 \geqslant \{\tau'(\theta)\}^2 / E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right]. \tag{17.22}$$

This is the fundamental inequality for the variance of an estimator, often known as the Cramér–Rao inequality, after two of its several discoverers (C. R. Rao (1945) ; Cramér (1946)) ; it was apparently first given by Aitken and Silverstone (1942). Using (17.19), it may be written in what is often, in practice, the more convenient form

$$\operatorname{var} t \geqslant -\{\tau'(\theta)\}^2 / E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right). \tag{17.23}$$

We shall call (17.22) and (17.23) the minimum variance bound (abbreviated to MVB) for the estimation of $\tau(\theta)$. An estimator which attains this bound for all $\theta$ will be called a MVB estimator.

It is only necessary that (17.18) hold for the MVB (17.22) to follow from (17.20).

$$E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right] = -E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right). \tag{17.19}$$

If (17.19) also holds, we may also write the MVB in the form (17.23).

Ref. 6

**17.16** In the case where $t$ is estimating $\theta$ itself, we have $\tau'(\theta) = 1$ in (17.22) and for an *unbiassed* estimator of $\theta$

$$\operatorname{var} t \geqslant 1/E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right] = -1/E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right). \tag{17.24}$$

In this case the quantity $I$ defined as

$$I = E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right] \tag{17.25}$$

is sometimes called the *amount of information* in the sample, although this is not a universal usage.

Important →

**17.17** It is very easy to establish the condition under which the MVB is attained. The inequality in (17.22) arose purely from the use of the Cauchy–Schwarz inequality, and the necessary and sufficient condition that the Cauchy–Schwarz inequality becomes an equality is (cf. **2.7**) that $\{t-\tau(\theta)\}$ is proportional to $\frac{\partial \log L}{\partial \theta}$ for all sets of observations. We may write this condition

$$\frac{\partial \log L}{\partial \theta} = A.\{t-\tau(\theta)\}, \tag{17.26}$$

where $A$ is independent of the observations but may be a function of $\theta$. Thus (17.26) becomes

$$\frac{\partial \log L}{\partial \theta} = A(\theta)\{t-\tau(\theta)\}. \tag{17.27}$$

Further, from (17.27) and (17.18),

$$\operatorname{var}\left(\frac{\partial \log L}{\partial \theta}\right) = E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right] = \{A(\theta)\}^2 \operatorname{var} t, \tag{17.28}$$

and since in this case (17.22) is an equality, (17.28) substituted into it gives

$$\operatorname{var} t = |\tau'(\theta)/A(\theta)|. \tag{17.29}$$

We thus conclude that if (17.27) is satisfied, $t$ is a MVB estimator of $\tau(\theta)$, with variance (17.29), which is then equal to the right-hand side of (17.23). If $\tau(\theta) \equiv \theta$, $\operatorname{var} t$ is just $1/A(\theta)$, which is then equal to the right-hand side of (17.24).

Equivalent Conditions: for achieving Cramer-Rao Lower Bound →

**17.19** From (17.27) we have on integration the necessary form for the Likelihood Function (continuing to write $A(\theta)$ for the integral of the arbitrary function $A(\theta)$ in (17.27))

$$\log L = tA(\theta)+P(\theta)+R(x_1, x_2, \ldots, x_n),$$

which we may re-write in the frequency-function form

$$f(x\,|\,\theta) = \exp\{A(\theta)B(x)+C(x)+D(\theta)\}, \tag{17.30}$$

where $t = \sum_{i=1}^{n} B(x_i)$, $R(x_1, \ldots, x_n) = \sum_{i=1}^{n} C(x_i)$ and $P(\theta) = nD(\theta)$. (17.30) is often called the exponential family of distributions.

**Efficiency**

**17.28** So far, our discussion of MV estimation has been exact, in the sense that it has not restricted sample size in any way. We now turn to consideration of large-sample properties. Even if there is no MV estimator for each value of $n$, there will often be one as $n$ tends to infinity. Since most of the estimators we deal with are asymptotically normally distributed in virtue of the Central Limit theorem, the distribution of such an estimator will depend for large samples on only two parameters—its mean value and its variance. If it is a consistent estimator it will commonly be asymptotically unbiassed—cf. **17.9**. This leaves the variance as the means of discriminating between consistent, asymptotically normal estimators of the same parametric function.

Among such estimators, that with MV in large samples(*) is called an *efficient* estimator, or simply *efficient*, the term being due to Fisher (1921a). It follows from the result of **17.26** that efficient estimators tend asymptotically to equivalence.

Ret. 6

## 3.2 One example where no efficient estimate exists (Cauchy).

*Example 17.7*

To estimate $\theta$ in

$$dF(x) = \frac{1}{\pi}\frac{dx}{\{1+(x-\theta)^2\}}, \qquad -\infty \leqslant x \leqslant \infty.$$

We have

$$\frac{\partial \log L}{\partial \theta} = 2\Sigma \frac{x-\theta}{\{1+(x-\theta)^2\}}.$$

This cannot be put in the form (17.27). Thus there is no MVB estimator in this case.

Ref. 6

# 4 Interesting Properties of the Cauchy Distribution.

## 4.1 Cauchy is preserved under addition.

**Table 7.5 Transformations of random variables (distributions specified)** *(Continued)*

| No. | Transformation $u$ | Original distribution | Final distribution |
|---|---|---|---|
| 17 | $x_1 + x_2 + \cdots + x_n$ | $x_i$'s all Cauchy distributions with parameters $\theta_i$ and $\alpha_i$ $f(x) = \dfrac{\alpha_i/\pi}{(x-\theta_i)^2 + \alpha_i^2}$ | $h$ = Cauchy $h(u) = \dfrac{\alpha/\pi}{(x-\theta)^2+\alpha^2} \quad -\infty < u \le +\infty$ $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ $\theta = \theta_1 + \theta_2 + \cdots + \theta_n$ |

Ref. 3

## 4.2 Cauchy is preserved under division.

**Table 7.5 Transformations of random variables (distributions specified)** *(Continued)*

| No. | Transformation $u$ | Original distribution | Final distribution |
|---|---|---|---|
| 18 | $\dfrac{1}{x}$ | $x$ Cauchy with parameters $\theta, \alpha$ | $u$ = Cauchy $h(u) = \dfrac{\alpha'/\pi}{(x-\theta')^2+\alpha'^2} \quad -\infty < u \le +\infty$ $\theta' = \dfrac{\theta}{\alpha^2+\theta^2} \quad \alpha' = \dfrac{\alpha}{\alpha^2+\theta^2}$ |

Ref. 3

As shown:

2. $g(\mathrm{x}) = a/\mathrm{x}$ (hyperbolic). With

$$\mathrm{y} = \frac{a}{\mathrm{x}}$$

the equation $y = a/x$ has a single solution

$$x = \frac{a}{y}$$

for every $y$ (Fig. 5-14). Since

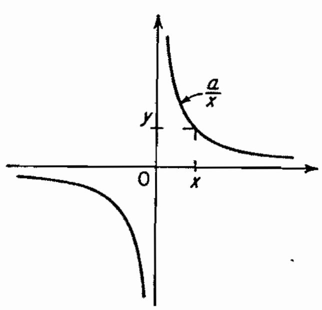



*Fig. 5-14*

$$g'(x) = -\frac{a}{x^2} = -\frac{y^2}{a}$$

we conclude from (5-6) that the density of $y$ is given by

$$f_y(y) = \frac{|a|}{y^2} f_x\left(\frac{a}{y}\right) \qquad (5\text{-}8)$$

*Special Case.* If x has a Cauchy density

$$f_x(x) = \frac{\alpha/\pi}{x^2 + \alpha^2}$$

with parameter $\alpha$ and $\mathrm{y} = a/\mathrm{x}$, then

$$f_y(y) = \frac{|a|}{y^2}\,\frac{\alpha/\pi}{a^2/y^2 + \alpha^2} = \frac{|a|/\alpha\pi}{y^2 + a^2/\alpha^2}$$

is also a Cauchy density with parameter $|a|/\alpha$.

Ref. 1

4.3 Cauchy is unimodal, has a mode and median, but no mean (no expected value or other higher moments) yet does have a characteristic function.

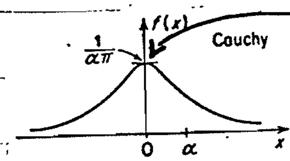
The *median*, $\check{X}$, on the other hand, is any value of $X$ such that one-half of the values are above and one-half below it (it divides the area of the histogram in half).* The *mode* is that value of $X$ which appears most frequently. If the data are grouped we usually choose as the mode the midvalue of the interval with the highest frequency. The mode is not used at all commonly, because it can be difficult to define or interpret.

Ref. 7

**Cauchy** (Fig. 4-13)

$$f(x) = \frac{\alpha/\pi}{\alpha^2 + x^2} \tag{4-55}$$

Ref. 1




mode and median value is $x = 0$

*Fig. 4-13*

Ref. 1

*Some one-dimensional continuous distribution functions*

| | Name | Domain | Probability Density Function $f(x)$ | Restrictions on Parameters | Mean | Variance | Skewness $\gamma_1$ | Excess $\gamma_2$ | Characteristic function |
|---|---|---|---|---|---|---|---|---|---|
| 26.1.25 | Error function | $-\infty<x<\infty$ | $\frac{h}{\sqrt{\pi}} e^{-h^2x^2}$ | $0<h<\infty$ | 0 | $\frac{1}{2h^2}$ | 0 | 0 | $e^{\frac{-t^2}{4h^2}}$ |
| 26.1.26 | Normal | $-\infty<x<\infty$ | $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$ | $-\infty<m<\infty$ $0<\sigma<\infty$ | $m$ | $\sigma^2$ | 0 | 0 | $e^{imt-\frac{\sigma^2t^2}{2}}$ |
| 26.1.27 | Cauchy | $-\infty<x<\infty$ | $\frac{1}{\pi\beta} \frac{1}{1+\left(\frac{x-\alpha}{\beta}\right)^2}$ | $-\infty<\alpha<\infty$ $0<\beta<\infty$ | not defined | not defined | not defined | not defined | $e^{i\alpha t-\beta\lvert t\rvert}$ |

Ref. 2

# 5 How May the Cauchy Distribution Physically Arise?

## 5.1 For $y = a \tan x$, $a > 0$, with $x$ uniform on $[-\pi, \pi]$, $y$ is Cauchy.

7. $g(\mathbf{x}) = a \tan \mathbf{x}$. The case

$$\mathbf{y} = a \tan \mathbf{x} \qquad a > 0$$

can be handled similarly. The equation $y = a \tan x$ has infinitely many solutions (Fig. 5-25$a$):

$$x_n = \arctan \frac{y}{a} \qquad n = \ldots, -1, 0, 1, \ldots$$

for any $y$. Since

$$g'(x) = \frac{a}{\cos^2 x} = \frac{a^2 + y^2}{a}$$

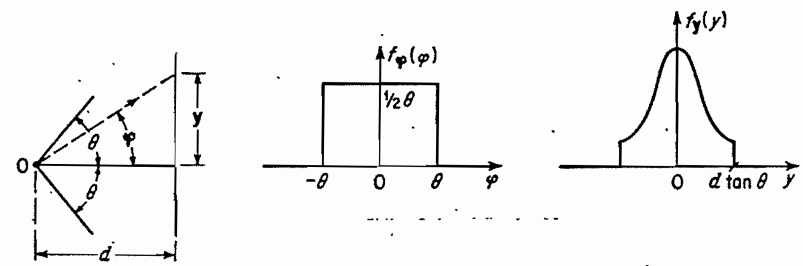
we obtain from (5-6)

$$f_y(y) = \frac{a}{a^2 + y^2} \sum_{n=-\infty}^{\infty} f_x(x_n) \tag{5-17}$$

*Special Case.* If $\mathbf{x}$ is uniformly distributed in the interval $(-\pi,\pi)$, then only two terms in (5-17) are different from zero and

$$f_y(y) = \frac{a}{a^2 + y^2} \frac{2}{2\pi} = \frac{a/\pi}{a^2 + y^2} \tag{5-18}$$

Thus the r.v. $\mathbf{y}$ has a Cauchy density as in Fig. 5-25$b$.




($a$) ($b$)

***Fig. 5-26***

**Example 5-15.** A particle leaves the origin in a free motion, forming an angle $\boldsymbol{\varphi}$ with the horizontal axis. With $\mathbf{y}$ as in Fig. 5-26$a$, we have

$$\mathbf{y} = d \tan \boldsymbol{\varphi}$$

Assuming that $\boldsymbol{\varphi}$ is a r.v. uniformly distributed in the interval $(-\theta,\theta)$, we shall determine the density of $\mathbf{y}$. Reasoning as in (5-18), we find

$$f_y(y) = \begin{cases} \dfrac{d/2\theta}{d^2 + y^2} & |y| < d \tan \theta \\ 0 & |y| > d \tan \theta \end{cases}$$

Thus $f_y(y)$ is a truncated Cauchy density (Fig. 5-26$b$).

Ref. 1

5.2 For $z = \frac{x}{y}$, and $x$ and $y$ are jointly normal zero mean then $z$ is Cauchy

**Table 7.5 Transformations of random variables (distributions specified)**

| No. | Transformation $u$ | Original distribution | Final distribution |
|---|---|---|---|
| 6 | $\frac{x_1}{x_2}$ | $x_1, x_2$ have a joint normal distribution with zero mean, variances $\sigma_1$ and $\sigma_2$, and correlation $\rho$ | $u$ = Cauchy $h(u) = \dfrac{1}{\dfrac{\pi\sigma_1\sigma_2}{\sqrt{1-\rho^2}}\left[\left(\dfrac{u}{\sigma_1} - \dfrac{\rho}{\sigma_2}\right)^2 + \dfrac{\sqrt{1-\rho^2}}{\sigma_1^2}\right]}$ $\quad -\infty < u \le +\infty$ |

Ref. 3

**Example 7-9.** We assume that the r.v. $\mathbf{x}$ and $\mathbf{y}$ are jointly normal with

$$f_{xy}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)} \tag{7-24}$$

and shall determine the density of their ratio

$$\mathbf{z} = \frac{\mathbf{x}}{\mathbf{y}}$$

From (7-23) we have

$$f_z(z) = \frac{2}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^\infty y e^{-\frac{y^2}{2(1-r^2)}\left[\frac{z^2}{\sigma_1^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right]} dy$$

But

$$\int_0^\infty y e^{-y^2/2a^2} dy = a^2 \int_0^\infty e^{-w} dw = a^2$$

Hence, after some simple computations,

$$f_z(z) = \frac{\sqrt{1-r^2}\,\sigma_1\sigma_2/\pi}{\sigma_2^2(z - r\sigma_1/\sigma_2)^2 + \sigma_1^2(1-r^2)} \tag{7-25}$$

Thus: If $\mathbf{x}$ and $\mathbf{y}$ are jointly normal with zero mean, then their ratio $\mathbf{z} = \mathbf{x}/\mathbf{y}$ has a *Cauchy* density (Fig. 7-14) centered at $z = r\sigma_1/\sigma_2$.

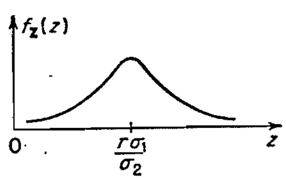



*Fig. 7-14*

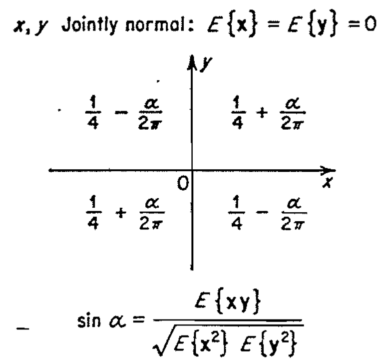



*Fig. 7-15*

The corresponding distribution is found by integration:

$$F_z(z) = \int_{-\infty}^{z} f_z(\alpha)\,d\alpha = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{\sigma_2 z - r\sigma_1}{\sigma_1\sqrt{1-r^2}} \tag{7-26}$$

If $\mathbf{x}$ and $\mathbf{y}$ are independent, then $r = 0$ and (7-25) becomes

$$f_z(z) = \frac{\sigma_1/\pi\sigma_2}{z^2 + \sigma_1^2/\sigma_2^2} \tag{7-27}$$

Ref. 1

## References


1. Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, Inc., New York, 1965.
2. Abramowitz, M. and Stegun, I.A., Ed., *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series 55,
3. Shooman, M.L., *Probabilistic Reliability: An Engineering Approach* McGraw-Hill Book Company, New York, 1968.
4. Rao, C.R., *Linear Statistical Inference and Its Applications*, John Wiley and Sons, Inc., New York, 1965.
5. Enis, P., "On the relation $E[X] = E\{E[X|Y]\}$," *Biometrika*, Vol. 60, No. 2, August 1973.
6. Kendall, M., and Stuart, A., *The Advanced Theory of Statistics*, Vol Hafner Publishing Co., New York, 1967.
7. Johnson, N.L., and Leone, F.C., *Statistics and Experimental Design*, Vol. 1, John Wiley and Sons, Inc., New York, 1964.