Abstract

Let $F$, $G$, and $H$ be probability distributions on the line each having finite variance and suppose $G$ is symmetric. $F$ is "completely asymmetric" (c.a.a.) if the equation $F = C^2H$ implies $C = 0$, i.e., is degenerate. It's proven that $F$ can always be written $F = C^2H$ where $H$ is a.a.a., but this representation may not be unique. Examples of singular and absolutely continuous (with respect to Lebesgue measure) c.a.a. distributions are given. Some extensions of these ideas are mentioned.
1. Introduction and Rareflect

Let \( F, G, \) and \( H \) be distributions (= probability distributions) on \( \mathbb{R} \) (= real line). Let \( c(s) = f^2 \mathbb{W}(ds) \). Say that \( F \) has a finite variance \( \mathbb{V} \) if \( c(s) < \infty \) (e.g. \( G \) is symmetric if \( G(-A) = G(A) \) for every Borel set \( A \in \mathbb{R} \). \( \mathbb{A} = \{ \text{real} \} \). Say that \( F \) is completely symmetric \( (c,s) \) if throughout \( F, G, \) and \( H \) have finite variances, \( G \) is symmetric, and \( F \subseteq G \times \). Then \( c = c_g \). (ex = "convolution", \( \mathbb{W} = \) unit mass at \( x \in \mathbb{R} \).)

The proof of the following employs Hurew's lemma and a martingale argument. It is found in the appendix along with the proof of proposition 2.1.

2. Examples

Some generic examples of c.a.s. distributions are provided by the following

2.1 Proposition:

a) Let \( a,b \in \mathbb{R} \) and \( 1/2 < a < 1 \). Then \( \lambda_a + (1-a) \mu_b \) is c.a.s.

b) Let \( F \) be a distribution with finite variance and support in \( (a, \infty) \), where \( a \) is finite. Suppose \( F \) has a density, \( f \), satisfying

1) \( f \) is upper semicontinuous and finite in \( (a, \infty) \).

2) \( \lim f(a+b) = \infty \)

Then \( F \) is c.a.s.

The proof is found in the appendix. Notice that part (a) shows that the class of c.a.s. distributions isn't closed. E.g., \( \lambda \not\subseteq \frac{1}{2} \lambda_{\frac{1}{2}} + (1-\lambda) \mu_{\frac{1}{2}} \) both weakly and strongly.

An example of the non-uniqueness of the representation (1.1) follows.

2.2 Example: Let \( \mu = (\frac{1}{2}) \lambda_{\frac{1}{2}} + (\frac{1}{2}) \mu_{\frac{1}{2}} \). Then by proposition 2.1
(a) \( \mu \) is o.o.s. Let \( \nu = (1/3)\delta_{-1} + (1/3)\delta_1 \), so \( \nu \) is symmetric, and let \( \eta \) be the symmetric signed measure

\[
\eta = (1/5)\delta_{-2} + (1/10)\delta_{-1} + (4/5)\delta_{0} - (1/10)\delta_{1} + (1/5)\delta_{2}.
\]

Then

\[
\nu^*\eta = (3/25)\delta_{-3} + (1/50)\delta_{-2} + (11/25)\delta_{-1} + (13/50)\delta_{1} + (3/25)\delta_{2} + (2/25)\delta_{3},
\]

and

\[
\nu^*\nu = (1/10)\delta_{-2} + (1/20)\delta_{-1} + (7/20)\delta_{1} + (7/20)\delta_{2} + (1/20)\delta_{3} + (1/10)\delta_{4}.
\]

Notice that both of these are probability distributions, the latter being symmetric. Let \( F = (\nu^*\nu)\delta_{0} = \nu^*(\nu\delta_{0}) = F = (\nu^*\nu)^*\eta \); it is obviously one representation of the type (1.1). But by proposition 1.1, \( \nu^*\mu \) can be written in the form \( \nu^*\nu' \), where \( \nu' \) is symmetric and \( \nu' \) is o.o.s. (I systematically set out to factor \( \nu^*\mu \) in this way and ruled out all non-trivial factorizations. Thus, unless I made a mistake, \( \nu^*\mu \) is actually o.o.s, so \( \nu' = \delta_{0} \).) Thus, \( F = (\nu^*\nu')\delta_{0} \) and \( F = (\nu^*\nu')\mu \) are two decompositions of the type (1.1). If \( \nu^*\nu \) then it would follow by considering Fourier transforms that \( \nu^*\nu' \). But \( \nu' \) is positive while \( \nu \) is not. It follows that \( F \) has at least two distinct decomposition into symmetric and o.o.s. parts. Thus the decompositions (1.1) aren't generally unique.

2. Some half-baked ideas

(A detailed exposition of most of the ideas in this section can be found in Hill (1960).)

The original motivation behind this work was to help answer the identifiability question for the following nonparametric regression problem. Let \( X \) be a random variable whose distribution has support \([0,1]\). Let \( f \) be a function on \([0,1]\) and let \( W \) be the distribution on the plane of the vector \((X,f(X))\). Let \( G \) be a distribution on the plane satisfying some symmetry conditions and suppose we observe a random sample from the distribution and \( \eta = \delta_{0} \). The problem is to estimate \( f \) from the random sample.

My original idea for investigating the identifiability of \( f \) was to prove a two-dimensional analogue of proposition 1.1 (easy, I think) and then hope that the decomposition was unique so the identifiability question would be reduced to that of the complete symmetry of \( H \). The latter question is fairly manageable. But, of course, example 2.2 of the last section shows that the decomposition may not be unique.

To get around this problem of lack of uniqueness, I investigated the question of decompositions like (1.1) for objects other than probability distribution, namely for finite signed measures on \( \mathbb{R} \) and for generalized functions on \( \mathbb{R} \). In both cases I assumed analyticity of the Fourier transforms. For those objects the situation is much simpler than it is for probability distributions on \( \mathbb{R} \) and the corresponding decompositions are about as unique as one could hope for. Unfortunately, it appears to be very hard to get the same results on the plane.

Getting away from the original regression problem, here are some
other, related, ideas I've explored. Since the decomposition (1.1) isn't unique, what is the structure of the set of all decompositions? To try and gain some insight into this I considered the following relation. If \( F \) and \( G \), \( H \), and \( K \) are distributions, write \( F \sim G \) if \( H \) and \( K \) are symmetric and \( F \circ H = G \circ K \). \( \sim \) is an equivalence relation. (Note that one can construct random variables \( U,V,X,Y \) with distributions \( R,K,F,G \) resp. such that \( X \) and \( U \) are independent, \( Y \) and \( V \) are independent, \( U \circ X = V \circ Y \), and \( (U,X) \) and \( (V,Y) \) are conditionally independent given \( U \circ X \). This might be a useful tool for studying \( \sim \).)

Let \( (F) = (G: G \sim F) \). Then it's not hard to see that \( (F) \) is convex and closed relative to a number of topologies. Furthermore, if \( \sim \) is a distribution then convolution of elements of \( (F) \) is well-defined and in fact is a group. \((R)\) is the identity and \( (F) \) where \( F(A) = F(-A) \) is the inverse of \( (F) \). Thus, \( \sim \) has a tantalizing structure. Unfortunately, I wasn't able to get any further with these ideas.

Appendix: Proofs.

A.1 Lemma: If \( H \) is a e.e. distribution then \( H^0 \) is also e.e. for every \( x \in R \).

Proof: Suppose \( H \) is e.e. but for some \( x \) \( H^0 \) isn't. Then there are distributions \( G' \) and \( H' \) s.t. \( G' \) is symmetric but \( G \neq G' \) and \( H^0 = G' \circ H' \). Then \( H = G' \circ (H^0 \ominus x) \), contradiction.

q.e.d.

Proof of proposition 1.1: By the lemma, WLOG (without loss of generality) \( F \) is centered: \( \int x F(dx) = 0 \). Let \( H \) denote the set of all distributions, \( x \) on \( R \) such that \( s(H) = \phi \) \( \forall x \). Then there exists a symmetric distribution, \( G \), on \( R \) such that \( s(G) \neq \phi \) and \( F \sim GH \). \( H \) isn't empty since \( F \sim G \).

If \( H \) is a distribution, let \( m(H) = \int x F(dx) \), whenever the latter integral is defined. Let \( S = \{ H: H \) is a distribution and \( s(H) = \phi \} \).

If \( H \in S \) and \( F = GH \), then \( G \times S \) \( \subseteq \), \( G \leq \phi \) s.t. \( H = G \circ H \). Notice that if \( H_1 \leq H_2 \) and \( H_2 \leq H_3 \) then \( H_1 \leq H_3 \). Hence, \( G \times S \times S \) \( = S \) (just consider variances) so by symmetry of \( G_1 \) and \( G_2 \), \( G_2 = G_1 \circ G_2 \), \( G_2 = G_1 \). Thus, \( H_1 = G_1 \circ G_2 \). Hence, \( G \times S \times S \) \( = S \) (just consider variances) so by symmetry of \( G_1 \) and \( G_2 \), \( G_2 = G_1 \). Thus, \( H_1 = H_2 \). It follows that \( S \) is a partial ordering.

The proposition amounts to saying \( H \) contains a minimal element w.r.t. (with respect to) \( S \). I'll prove this using Zorn's lemma. Let \( A = \{ m(H) \leq 1 \} \) \( \subseteq \) \( S \) can be a totally ordered w.r.t. \( \leq \). Then it suffices to show that \( A \) has a lower bound in \( H \). Let \( v = \inf \{ s(H): m(H) \leq 1 \} \). Choose a countable sequence \( H_n \) such that \( s(H_n) > v \) \( \forall n \) and write \( H_n = H_n(x) \) \( \ni x = 1,2, \ldots \). Then \( m > a \) implies \( s(H_n) < s(H) \leq s(H_n) \) and since \( A \) is totally ordered it follows that \( H_n \leq H \).

Claim: Any lower bound for \( (H_n) \) in \( H \) (if there are any) is a lower bound for \( A \). For suppose not. Then there exists \( a < 1 \) and \( H \in S \) such that \( H_n \leq H \) \( \forall n \) but \( m(H) < m(H) \). This implies \( v < s(H) \leq s(H) \) for all \( n \). But this contradicts the way \( (H_n) \) was
chosen.

Thus, to find a lower bound for $a$, I need only find one for $\gamma_n$. I'll do this by constructing a reversed martingale whose marginal distributions are $\gamma_n$, $n \geq 1$, and then applying a martingale limit theorem. If $\beta \in \mathbf{R}$, $\beta \leq \gamma_{n-1} (\gamma_n = P)$ so there exists a symmetric distribution $\gamma_n$ s.t. $\beta = \gamma_n \mathcal{Q}_P$. For each $n$, let $X_{n-1} \cdots \cdots X_{n,n}$ be a r.v. (random variable) with distribution $\gamma_n$ and let $Y_{n, n}$ be independent r.v.'s independent of $X_{n,n}$ s.t.

$Y_{n,n}$ has distribution $\gamma_n$. Let

$$X_{n,n} = Y_{n,n} + \frac{1}{m} Y_{n,n-1}, \quad n \geq 1.$$  

Then $X_{n,n}$ has distribution $\gamma_n$. Note that for each $n$, $X_{n,n}$ is independent of $X_{n-1,n} \cdots X_{n,n}$, Let $\gamma_n$ denote the joint distribution of $X_{n,n} \cdots X_{n,n}$.

By the Kolmogorov extension theorem there exists a process $X_n$, $X_{n-1} \cdots \cdots P_n$ is the joint distribution of $X_0 \cdots X_n$. Thus, $X_n$ has distribution $\gamma_n$, $n \geq 0$.

Let $Y_{n,k} = X_{n,k} - X_{n-1,k} = Y_{n,k}$. Then for $k \geq 1$,

1) $Y_{n,k}$ has distribution $\gamma_n$ (in particular $X_{0,0} = 0$).
2) $X_{n-1}, X_{n-1} \cdots$ are independent, and
3) $Y_{n,k}$ is independent of $X_{n-1}, \quad k \geq 0$.

It follows that $(X_n, n \geq 0)$ is a martingale relative to its own internal history, $P_{n-1} = o(X_{n-1} = \gamma_n), n \geq 0$. Since $s(\mathbf{R}_n) < \infty$, for

all $a$, $(X^2_{n-1}, n \geq 0)$ is a submartingale relative to $P_{n-1}$. It follows from theorem 9.4.7 in Chung (1974) that $X_n$ converges q.s. and in $L^2$ to a r.v. $X_n$.

As a consequence of this, for each $n \geq 0$

$$\sum_{k=1}^{\infty} Y_{n,k}$$

converges q.s. to r.v. $\sum_{k=1}^{\infty} Y_{n,k}$ s.t. $\sum_{k=1}^{\infty} Y_{n,k}$ is symmetric, $s \{ Y_{n,k} \} < \infty,$ and $X_n + \sum_{k=1}^{\infty} Y_{n,k} = X_n$. Furthermore, for each $n \geq a$,

$$\sum_{k=1}^{\infty} Y_{n,k}$$

is independent of $X_n$. Therefore, $\sum_{k=1}^{\infty} Y_{n,k}$ is also independent of $X_n$.

(just consider characteristic functions.)

But this means that if $\mathbf{R}_n$ is the distribution of $X_n$, then $\mathbf{R}_n = H$ and $\mathbf{R}_n \vee \mathbf{R}_n = \mathbf{R}_n = P$. The proposition follows.

q.e.d.

Proof of proposition 3.1: a) By lemma 3.1 we can assume $-a$. Let $1/2 < \lambda < 1$ and let $P = \lambda \mathbf{H}_a + (1-\lambda) \mathbf{E}_a$. If $\gamma = 0$ then we already know that $P$ is q.s. Assume that $\gamma \neq 0$ and suppose $F = a \mathbf{H}_a$, where $a$ is symmetric. The support of the convolution of two probability distributions is the sum of their support. It follows that either

(*) $\gamma$ has support on two points and $H$ on one, or
(\*) vice versa.
Suppose (9) holds, then it easily follows that $V$ is symmetric, which it isn't. Thus, (10) must hold and (a) is proved.

b) The proof is a little involved but the basic idea is as follows. Write $V = e^a g$ where $g$ is symmetric. It will turn out that $g$ has bounded support. If $2b \geq 0$ is the width of that support, we'll see that $f$ must blow up at $a \pm 2b$, contradicting (i) unless $b = 0$, i.e. unless $F$ is o.o.s. By Lemma A.1 we can assume $a = 0$. By (ii) inf supp $F = 0$. Since supp $F$ = supp $a$ + supp $g$ it follows that if $b = b = \inf $ supp $a$, then $b \rightarrow -a$ and $0 = \inf $ supp $g$. But $g$ is symmetric so $b = \inf $ supp $g \geq 0$.

In the remainder of the proof it will be convenient for $a$ and $g$ to have densities. Since $a$ and $g$ may not have densities, I change the distributions of $a$ and $g$ (and, hence, of $V$) so that they do have densities. I do this by convolving them with absolutely continuous distributions concentrated about 0.

Let $V_a$ be a symmetric unimodal probability density with support $[0,1]$. Suppose $V_a$ is everywhere differentiable with bounded derivative. Then $V_a(n) > 0$ if $-1 < n < 1$, $V_a(n) \leq 0$ if $n \leq 0$, and $V_a(n) \geq 0$ if $n \geq 0$. Let $g_a = k_a^2 V_a$. Then $g$ has support $[-1,1]$. Using the dominated convergence theorem, the mean value theorem, and the fact that $g$ has a bounded derivative, it's easy to see that the convolution of $g$ with any distribution, $g$, also has a bounded derivative, $F^{(2)}(x-y)(dy)$. In particular, $g_a$ has a bounded derivative. For $a > 0$, let $V_a(n) = V_a^a V_a(n/a)$, $g_a(n) = V_a^a V_a(n/a)$. Thus $V_a$ and $g_a$ have supports $[-a,a]$, $[-2a,2a]$, resp. and $g_a = k_a^2 V_a$.

Let $b_a(n) = b_a(n) dx$ and $g_a(n) dx = a_a(n) dx = b_a^a g_a(n) dx$.

Assume $F$ isn't o.o.s. This amounts to assuming that $b > 0$. Let $F_a = F_a^a$, $F_a = F_a^a$, $F_a = F_a^a F_a = F_a^a$. Then $F_a$, $F_a$, and $F_a$ have differentiable densities $V_a = e^a V_a$, $g_a = e^a g_a$, $V_a = e^a V_a$, resp.

(A.1) $f_a(2b) \to = a \to 0$.

I'll show that this implies $f(2b) = a \to 0$. Part (b) then follows.

$$f_a(2b) = \int g_a(2b-y)(y)dy = \int g_a(2b-y)(y)dy = \int_{-a}^a g_a(2b-y)(y)dy = \int_{-2a}^{2a} g_a(2b-y)(y)dy \to 2b-2a$$

so by (A.1) this supremum $\to = a \to 0$. On the other hand, $f$ is upper semicontinuous so

$$\sup_{2b-2a} f(y) \to f(2b).$$
(A.1) is a consequence of two further facts:

(A.2) $f_a(-a) \to 0$ as $a \to 0$, and

(A.3) $f_a(2b) \geq f_a(-a)$ for all $a > 0$.

Proof of (A.3): Since $g_+\eta$ is symmetric

$$f_a(-a) = \int_0^a g_a(x+y)f(y)\,dy$$

$$= \int_0^a g_a(I-a) f(ax)\,da.$$  

$$\leq \inf f(x) \int_0^a g_a(I-x)\,dx.$$  

Now, $\int_0^a g_a(I-x)\,dx > 0$ and $\inf f(x) > 0$ by (ll)  

$$= \inf f(y) \to 0 \text{ as } a \to 0 \text{ by (ll)}.$$  

Proof of (A.5):

$$f_a(-a) = \left(n_a^0 m_0\right)(-a) = \int_{-a}^{-a-y} n_a\,m_0(y)\,dy$$  

$$= \int_{-a}^{-a-y} n_a\,m_0\,dy.$$  

On the other hand,

$$f_a(2b) = \int_{-a}^{-a-2b} n_a\,m_0\,dy$$

$$\geq \int_{-a}^{-a-2b} n_a\,m_0\,dy.$$  

Combining this with (A.4) yields (making the change of variables $y = 2b-x$),

$$f_a(2b) - f_a(-a) \geq \int_{-a}^{-a-2b} n_a\,m_0\,dy$$  

Now, if $y \in (-b/2, b)$, then $-b > y - 2b > -y$ and $b > y > 2b - y$. Since $n_a$ and $m_0$ are increasing on $(-\infty, -b)$ and $(-b, \infty)$, respectively, it follows that the last integral in (A.5) is nonnegative, i.e. (A.3) holds.

q.e.d.

Remark: Notice that in the proof of proposition 2.1, no use whatsoever was made of any moment properties of the distributions involved. Thus, proposition 2.1 holds for a stronger definition of c.s.s. in which no moment requirements are made.
References

Academic, New York.

Ellis, S.P. (1980) “Nonparametric regression with errors in both
variables; decomposition of probability measures, summable sequences,
and generalized functions into symmetric and completely asymmetric
parts,” unpublished manuscript.
On the Representation of Probability Distributions as the Convolution of Symmetric and Completely Asymmetric Parts

Steven P. Ellis

Massachusetts Institute of Technology
Statistics Center
Cambridge, MA 02139

Office of Naval Research
Statistics and Probability Code 436
Arlington, VA 22217

Convolution, completely asymmetric distribution, symmetric distribution

Let F, G, and H be probability distributions on the line each having finite variance and suppose G is symmetric. If $F = G^2H$ implies $G = e^\Delta_0$, i.e. is degenerate. It's proven that F can always be written $F = G^2H$ where H is c.a.s., but this representation may not be unique. Examples of singular and absolutely continuous (with respect to Lebesgue measure) c.a.s. distributions are given. Some extensions of these ideas are mentioned.
Errata in Ellis (1984) "On the representation..."

This sheet lists some misprints in the paper which appear thru my own negligence. (It’s the misprints, not the paper which appears thru my negligence.) Most of the misprints have the following origin. The paper was typed on a word processor whose character set does not include certain mathematical symbols. I had intended to go thru and insert these symbols by hand but forgot to. So on p. 3, 4th line on the first paragraph, the symbol $\subseteq$ should be inserted between A and R. On p. 7 a $\exists$ should be inserted before the expression "$([F]: F is a distribution)". Also in that paragraph it is $\mathcal{G}$ that is well-defined and a group. In the parenthetical statement in the same paragraph $[F], \tilde{F}(A) = F(-A)$, is the inverse of $[F]$.

The symbol $\mathbb{B}$ should be replaced by $\subseteq$ wherever it is found:

Third paragraph on p. 8 and in the first paragraph on p. 12.

A couple more conventional misprints are as follows. In the first paragraph beginning on p. 9, "$\mathbb{N} n" should be replaced by "$\mathbb{N}_n" in the 6th line. In the first line on p. 14 the limits, b-s and b, of integration for the integral are in the wrong spot.

I found a few more misprints, but they shouldn’t cause any confusion.