*Adv. Appl. Prob.* **39,** 105–127 (2007)
*Printed in Northern Ireland*


# GENERAL CONVEX STOCHASTIC ORDERINGS AND RELATED MARTINGALE-TYPE STRUCTURES

FRANCISCO VERA,* *National Institute of Statistical Sciences*
JAMES LYNCH,** *University of South Carolina*

**Abstract**

Blackwell (1951), in his seminal work on comparison of experiments, ordered two experiments using a dilation ordering: one experiment, $Y$, is 'more spread out' in the sense of dilation than another one, $X$, if $\mathrm{E}(c(Y)) \geq \mathrm{E}(c(X))$ for all convex functions $c$. He showed that this ordering is equivalent to two other orderings, namely (i) a *total time on test* ordering and (ii) a martingale relationship $\mathrm{E}(Y' \mid X') = X'$, where $(X', Y')$ has a joint distribution with the same marginals as $X$ and $Y$. These comparisons are generalized to balayage orderings that are defined in terms of generalized convex functions. These balayage orderings are equivalent to (i) iterated total integral of survival orderings and (ii) martingale-type orderings which we refer to as *k-mart* orderings. These comparisons can arise naturally in model fitting and data confidentiality contexts.

*Keywords:* Generalized convexity; stochastic ordering; iterated total time on test; $k$-mart; balayage

2000 Mathematics Subject Classification: Primary 60E15; 60G48
Secondary 26B25; 47A20

## 1. Introduction and background

In his seminal work on comparison of experiments, Blackwell (1951) studied the *dilation* ordering, according to which $Y$ is a dilation of $X$ if both random variables have finite means and

$$\mathrm{E}(c(Y)) \geq \mathrm{E}(c(X)) \tag{1.1}$$

for every convex function $c$. He showed that if two experiments are related to each other through (1.1), then there exists a transition kernel between the experiments such that $(X, Y)$ is a one-step *martingale*, i.e.

$$\mathrm{E}(Y \mid X) = X. \tag{1.2}$$

Identity (1.2) has implications for stochastic model building where a principal objective is to choose a simple baseline model that describes the salient features of the population. In particular, from (1.2) we have the stochastic model

$$Y = X + \varepsilon, \tag{1.3}$$

where $Y$ is the population, $X$ is the baseline model, and $\varepsilon$ is an error term. Two features of $X$ in (1.3) are that it is 'fair' and uncorrelated with the error term. Formally,

$$\mathrm{E}(Y) = \mathrm{E}(X) \quad \text{and} \quad \mathrm{cov}(X, \varepsilon) = 0. \tag{1.4}$$

Received 29 June 2005; revision received 28 August 2006.
* Postal address: National Institute of Statistical Sciences, PO Box 14006, Research Triangle Park, NC 27709-4006, USA.
** Postal address: Department of Statistics, University of South Carolina, Columbia, SC 29208, USA.
Email address: lynch@math.sc.edu

In Section 4, the above ideas are extended to more general notions of 'convexity', utilizing a martingale-type structure called a *k-mart*. Background material on generalized convexity defined in terms of Chebyshev systems is given in Section 2.

An operational way to verify (1.1) is through the use of *tail integral of survival* comparisons. Let $F$ denote a cumulative distribution function with nonnegative support and finite mean $\mu$ (i.e. $\mu = \int x \, \mathrm{d}F(x)$). Then the tail integral of survival transform is

$$\bar{F}_1(y) := \int_y^\infty (1 - F(x)) \, \mathrm{d}x =: \int_y^\infty \bar{F}(x) \, \mathrm{d}x.$$

It is well known that (1.1) holds if and only if

$$\mathrm{E}(Y) = \mathrm{E}(X) \quad \text{and} \quad \bar{G}_1 \ge \bar{F}_1, \tag{1.5}$$

where $G$ and $F$ denote the distribution functions of $Y$ and $X$, respectively (see, e.g. Ross (1983) and Shaked and Shanthikumar (1994, Section 2.A.1)). Moreover, (1.5) holds if and only if there exists a joint distribution for $Y$ and $X$ such that (1.2) holds, i.e. $(X, Y)$ has a martingale structure (see, e.g. Blackwell (1951), (1953), Meyer (1966, Section XI.2), Strassen (1965), and references therein).

Blackwell (1951) used a transform similar to the tail integral of survival transform in his work on comparing experiments. More general transformations are presented in Section 3, and are shown to be related to stochastic orderings similar to (1.1), but with more general notions of 'convexity'.

Blackwell's proof of (1.2) is constructive; the required transition kernel is obtained as an iterative procedure where each iteration is a dilation and the iteration converges to the required transition kernel after an infinite number of steps. An alternative proof of Blackwell's result is presented in Section 6. It is also constructive, but the construction is direct, not iterative, with the added benefit that the joint distribution is the one that makes $X$ 'most nearly identical' to $Y$.

Besides the comparison of experiments, $k$-marts and balayages have other applications, for instance in data confidentiality. One way of protecting data is to release a synthetic version of the data rather than the real data. The released data should be similar to the real one, in order that inferences made from the released version of the data be valid (Willenborg and de Waal (2001b)). The synthesization can be achieved through a variety of techniques such as swapping, round off, aggregation or grouping, and adding noise to the data. Here we are interested in developing a theory which addresses aggregation and addition of noise that preserves moment structure in one-way tables or arrays (and the authors are currently developing techniques for the multivariate situation). In Section 5, the moment structure is used to construct and generate $k$-marts which relate the distribution of the original data to the synthetic version.

Dilations can arise quite naturally when fitting mixture models in which the means of the population and the fitted model are the same (Shaked (1980)). Consequently, the joint distribution of $(X, Y)$ can be constructed to have a martingale structure. Higher-order mixture models can often be fitted in a hierarchical way in which the $k$-point fitted model and the population mixture have a $k$-mart structure rather than a martingale structure. These ideas are discussed in Section 4

## 2. Generalized convexity

The concept of convexity with respect to extended complete Chebyshev systems is presented in this section, following the treatment in Karlin and Studden (1966, Chapter I). The following definitions are needed.

**Definition 2.1.** Let $u_0, \dots, u_n$ be real-valued functions on the interval $[a, b]$. The collection $\{u_i\}_0^n$ is said to be a *Chebyshev system*, or a *C-system*, on $[a, b]$ if

$$\det[u_i(x_j)\colon i = 0, \dots, n,\ j = 0, \dots, n] > 0$$

for any choice of $n$ real numbers $x_i$, $a \le x_0 < x_1 < \cdots < x_n \le b$. (Unless otherwise indicated, in the matrix $[U_i(x_j)]$ the first index is for the rows and the second index is for the columns.) If the functions constitute a C-system on any interval $[a, b]$, $a < b$, then we say that they constitute a C-system on $(-\infty, \infty)$.

In the next definition, $u^{(i)}$ denotes the $i$th derivative of the function $u$ and $C^i[a, b]$ denotes the set of real-valued functions on $[a, b]$ with continuous $i$th derivatives (if no interval is specified, the property is assumed to be true on $(-\infty, \infty)$).

**Definition 2.2.** The C-system $\{u_i\}_0^n$ is called an *extended complete Chebyshev system*, or an *ECC-system*, on $[a, b]$ if $u_i \in C^n[a, b]$ and, for $k = 0, \dots, n$, $W(u_0, \dots, u_k) > 0$ on $[a, b]$, where $W(u_0, \dots, u_k)$ denotes the Wronskian of the functions $u_0, \dots, u_k$, i.e.

$$W(u_0, \dots, u_k)(t) = \det[u_i^{(j)}(t)\colon i = 0, \dots, k,\ j = 0, \dots, k]$$

(see Karlin and Studden (1966, Chapter XI, Theorem 1.1)). If the functions constitute an ECC-system on any interval $[a, b]$, $a < b$, then we say that they constitute an ECC-system on $(-\infty, \infty)$ (this is assumed if no interval is specified).

**Remark 2.1.** Without loss of generality, it is assumed that there exists a constant $c \in [a, b]$ for which $u_i^{(j)}(c) = 0$, $i = 1, \dots, n$, $j = 0, \dots, i-1$ (see Karlin and Studden (1966, Chapter XI, Remark 1.2)). Furthermore, the ECC-system $U = \{1, x, \dots, x^n\}$ will be referred to as the classical ECC-system or classical C-system.

**Definition 2.3.** Let $\{u_i\}_0^n$ be an ECC-system. The *Wronskian functions*, $w_0, \dots, w_n$, corresponding to $u_0, \dots, u_n$ are defined by $w_j = D_j u_j$, $j = 0, \dots, n$, where the differential operator $D_j$ is recursively defined as follows:

$$D_0 u = u, \qquad D_j u = \frac{\mathrm{d}}{\mathrm{d}t} \frac{D_{j-1} u}{w_{j-1}}, \qquad j = 1, \dots, n.$$

Here, for $j = n$, $(\mathrm{d}/\mathrm{d}t)(D_{n-1}u/w_{n-1})$ can be replaced by either the right or the left derivative of $D_{n-1}u/w_{n-1}$, and $D_n$ respectively denoted by $D_n^{\mathrm{R}}$ or $D_n^{\mathrm{L}}$.

The next lemma characterizes Wronskians in terms of Wronskian functions.

**Lemma 2.1.** *Let $\{u_i\}_0^n$ be an ECC-system with corresponding Wronskian functions $w_0, \dots, w_n$. Then, for $k = 0, \dots, n$,*

$$W(u_0, \dots, u_k) = w_0^{k+1} w_1^k \cdots w_k.$$

*Proof.* See Karlin and Studden (1966, p. 243 or p. 380) for details.

As an immediate corollary, we have the following result.

**Corollary 2.1.** *The collection $\{u_i\}_0^n$ is an ECC-system if and only if $w_i > 0$, $i = 0, 1, \dots, n$.*

The next lemma characterizes ECC-systems in terms of Wronskian functions and is needed in Section 3. Its proof can be found in Vera and Lynch (2005b, Lemma 2.3).

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

**Lemma 2.2.** *For any number $x$, functions $f$ and $g$ of the forms*

$$f(t) = w_0(t) \int_x^t w_1(v_1) \int_x^{v_1} w_2(v_2) \cdots \int_x^{v_{k-1}} w_k(v_k)\,\mathrm{d}v_k \cdots \mathrm{d}v_1 \tag{2.1}$$

*and*

$$g(t) = w_0(t) \int_t^x w_1(v_1) \int_{v_1}^x w_2(v_2) \cdots \int_{v_{k-1}}^x w_k(v_k)\,\mathrm{d}v_k \cdots \mathrm{d}v_1 \tag{2.2}$$

*are polynomials in (i.e. linear combinations of) the functions $u_0, \dots, u_k$ corresponding to $w_0, \dots, w_k$.*

**Definition 2.4.** A function $f$ is said to be *convex with respect to a C-system $U = \{u_i\}_0^n$*, or *$U$-convex*, if

$$\begin{vmatrix} u_0(t_0) & u_0(t_1) & \cdots & u_0(t_{n+1}) \\ u_1(t_0) & u_1(t_1) & \cdots & u_1(t_{n+1}) \\ \vdots & \vdots & \ddots & \vdots \\ u_n(t_0) & u_n(t_1) & \cdots & u_n(t_{n+1}) \\ f(t_0) & f(t_1) & \cdots & f(t_{n+1}) \end{vmatrix} \geq 0$$

for $t_0 < t_1 < \cdots < t_n < t_{n+1}$. If the inequality is strict then we say that the function is *strictly $U$-convex*.

**Definition 2.5.** A function $f$ is said to be *$U$-concave* if $-f$ is $U$-convex.

The next two lemmas, which were proved in Karlin and Studden (1966, Chapter XI), give some characterizations of functions that are convex with respect to ECC-systems.

**Lemma 2.3.** *If $f$ is convex with respect to the ECC-system $\{u_i\}_0^n$ and $n \geq 1$, then $f \in C^{n-1}$.*

**Lemma 2.4.** *The following statements are equivalent.*

(i) *$f$ is convex with respect to the ECC-system $\{u_i\}_0^n$.*

(ii) *$\rho_{\mathrm{R}} f := D_n^{\mathrm{R}} f / w_n$ is a right-continuous, nondecreasing function.*

(iii) *$\rho_{\mathrm{L}} f := D_n^{\mathrm{L}} f / w_n$ is a left-continuous, nondecreasing function.*

A useful representation of $U$-convex functions in terms of the operator $\rho_{\mathrm{R}}$ is presented in the next theorem.

**Theorem 2.1.** *If $f$ is bounded and convex with respect to $\{u_i\}_0^n$, where $u_i^{(j)}(c) = 0$, $i = 1, \dots, n$, $j = 0, \dots, i-1$, for some constant $c$, then*

$$\begin{aligned} f(t) &= w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} w_2(x_2) \cdots \int_c^{x_{n-1}} w_n(x_n) \int_c^{x_n} \mathrm{d}\rho_{\mathrm{R}} f(u)\,\mathrm{d}x_n \cdots \mathrm{d}x_1 \\ &\quad + \frac{f(c)}{w_0(c)} u_0(t) + \frac{D_1 f(c)}{w_1(c)} u_1(t) + \cdots + \frac{D_{n-1} f(c)}{w_{n-1}(c)} u_{n-1}(t) + \frac{D_n^{\mathrm{R}} f(c)}{w_n(c)} u_n(t) \end{aligned}$$

*for $t \geq c$ and*

$$f(t) = (-1)^{n+1} w_0(t) \int_t^c w_1(x_1) \int_{x_1}^c w_2(x_2) \cdots \int_{x_{n-1}}^c w_n(x_n) \int_{x_n}^c \mathrm{d}\rho_{\mathrm{R}} f(u) \, \mathrm{d}x_n \cdots \mathrm{d}x_1 + \frac{f(c)}{w_0(c)} u_0(t) + \frac{D_1 f(c)}{w_1(c)} u_1(t) + \cdots + \frac{D_{n-1} f(c)}{w_{n-1}(c)} u_{n-1}(t) + \frac{D_n^{\mathrm{R}} f(c)}{w_n(c)} u_n(t)$$

*for $t < c$.*

*Proof.* Note first that, for $t \geq c$ and $i = 0, 1, \ldots, n-1$,

$$D_i f(t) = w_i(t) \int_c^t D_{i+1} f(x) \, \mathrm{d}x + \frac{D_i f(c)}{w_i(c)} w_i(t).$$

Applying this identity recursively gives

$$\begin{aligned}
f(t) &= w_0(t) \int_c^t D_1 f(x_1) \, \mathrm{d}x_1 + \frac{f(c)}{w_0(c)} w_0(t) \\
&= w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} D_2 f(x_2) \, \mathrm{d}x_2 \, \mathrm{d}x_1 \\
&\quad + \frac{D_1 f(c)}{w_1(c)} w_0(t) \int_c^t w_1(x_1) \, \mathrm{d}x_1 + \frac{f(c)}{w_0(c)} w_0(t) \\
&\ \ \vdots \\
&= w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} \cdots w_{n-1}(x_{n-1}) \int_c^{x_{n-1}} D_n^{\mathrm{R}} f(x_n) \, \mathrm{d}x_n \cdots \mathrm{d}x_1 \\
&\quad + \frac{D_{n-1} f(c)}{w_{n-1}(c)} w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} \cdots w_{n-2}(x_{n-2}) \int_c^{x_{n-2}} w_{n-1}(x_{n-1}) \, \mathrm{d}x_{n-1} \cdots \mathrm{d}x_1 \\
&\quad + \cdots + \frac{D_1 f(c)}{w_1(c)} w_0(t) \int_c^t w_1(x_1) \, \mathrm{d}x_1 + \frac{f(c)}{w_0(c)} w_0(t) \\
&= w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} \cdots w_{n-1}(x_{n-1}) \int_c^{x_{n-1}} w_n(x_n) \int_c^{x_n} \mathrm{d}\rho_{\mathrm{R}} f(u) \, \mathrm{d}x_n \cdots \mathrm{d}x_1 \\
&\quad + \frac{D_n^{\mathrm{R}} f(c)}{w_n(c)} w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} \cdots \\
&\qquad\qquad \times w_{n-2}(x_{n-2}) \int_c^{x_{n-2}} w_{n-1}(x_{n-1}) \int_c^{x_{n-1}} w_n(x_n) \, \mathrm{d}x_n \cdots \mathrm{d}x_1 \\
&\quad + \frac{D_{n-1} f(c)}{w_{n-1}(c)} w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} \cdots w_{n-2}(x_{n-2}) \int_c^{x_{n-2}} w_{n-1}(x_{n-1}) \, \mathrm{d}x_{n-1} \cdots \mathrm{d}x_1 \\
&\quad + \cdots + \frac{D_1 f(c)}{w_1(c)} w_0(t) \int_c^t w_1(x_1) \, \mathrm{d}x_1 + \frac{f(c)}{w_0(c)} w_0(t) \\
&= w_0(t) \int_c^t w_1(x_1) \int_c^{x_1} \cdots w_{n-1}(x_{n-1}) \int_c^{x_{n-1}} w_n(x_n) \int_c^{x_n} \mathrm{d}\rho_{\mathrm{R}} f(u) \, \mathrm{d}x_n \cdots \mathrm{d}x_1 \\
&\quad + \frac{D_n^{\mathrm{R}} f(c)}{w_n(c)} u_n(t) + \frac{D_{n-1} f(c)}{w_{n-1}(c)} u_{n-1}(t) + \cdots + \frac{D_1 f(c)}{w_1(c)} u_1(t) + \frac{f(c)}{w_0(c)} u_0(t),
\end{aligned}$$

where the last equality follows from Lemma 2.2 of Vera and Lynch (2005b). A similar proof holds for $t < c$.

## 3. Balayages

The concept of balayages is introduced in this section, following the treatments in Meyer (1966) and Lynch (1988). Then a characterization in terms of iterated integrals, similar to the one in Karlin and Studden (1966, Chapter XI, Theorem 5.2), is presented (see also Denuit *et al.* (1998), who did the same thing for classical polynomials).

**Definition 3.1.** Let $U = \{u_i\}_0^n$ be an ECC-system. Let $F$ and $G$ be two finite measures with

$$\int |u_i| \,\mathrm{d}F < \infty, \quad \int |u_i| \,\mathrm{d}G < \infty, \qquad i = 0, \dots, n.$$

We say that $G$ is a *balayage* of $F$, written $G >_U F$, if

$$\int c \,\mathrm{d}G \geq \int c \,\mathrm{d}F$$

for any $U$-convex function $c$ satisfying $|\int c \,\mathrm{d}G| < \infty$ and $|\int c \,\mathrm{d}F| < \infty$. If $U = \{1, x\}$ then (under the same conditions) we say that $G$ is a *dilation* of $F$, written $G >_{\mathrm{d}} F$.

**Remark 3.1.** If $F$ and $G$ have densities $f$ and $g$ with respect to some measure $\nu$, then we write $g >_U f$ if $G >_U F$. Similarly, if $X \sim F$ and $Y \sim G$ then we write $Y >_U X$ if $G >_U F$, i.e. $\mathrm{E}(c(Y)) \geq \mathrm{E}(c(X))$ for any $U$-convex function $c$ satisfying

$$|\,\mathrm{E}(c(X))| < \infty, \qquad |\,\mathrm{E}(c(Y))| < \infty.$$

**Remark 3.2.** Notice that

$$\int u_i \,\mathrm{d}F = \int u_i \,\mathrm{d}G, \qquad i = 0, \dots, n,$$

since both $u_i$ and $-u_i$ are $U$-convex.

**Definition 3.2.** Let $\{u_i\}_0^n$ be an ECC-system, with corresponding Wronskian functions $w_0, \dots, w_n$, such that $u_i^{(j)}(c) = 0,\ i = 1, \dots, n,\ j = 0, \dots, i-1$. Let $F$ denote a finite measure. The *lower* and *upper iterated integrals* of $F$ with respect to $\{u_i\}_0^n$ are functions recursively defined as follows:

$$F_0(t) = \int_{-\infty}^{t} w_0(u) \,\mathrm{d}F(u), \qquad F_i(t) = \int_{-\infty}^{t} w_i(u) F_{i-1}(u) \,\mathrm{d}u, \quad i = 1, \dots, n,$$
$$\bar{F}_0(t) = \int_{t}^{\infty} w_0(u) \,\mathrm{d}F(u), \qquad \bar{F}_i(t) = \int_{t}^{\infty} w_i(u) \bar{F}_{i-1}(u) \,\mathrm{d}u, \quad i = 1, \dots, n.$$

The next few results give useful properties of the iterated integrals defined above. In preparation, let $\{u_i\}_0^n$ be an ECC-system, with corresponding Wronskian functions $w_0, \dots, w_n$, such that $u_i^{(j)}(c) = 0,\ i = 1, \dots, n,\ j = 0, \dots, i-1$, and let $F$ denote a finite measure for which $\int |u_i| \,\mathrm{d}F < \infty,\ i = 0, \dots, n$.

The following proposition characterizes $U$-moments in terms of iterated integrals. Its proof can be found in Vera and Lynch (2005b, Proposition 3.1).

**Proposition 3.1.** $(-1)^i F_i(c) + \bar{F}_i(c) = \int u_i \,\mathrm{d}F$.

Now, similar to (2.1) and (2.2), define

$$f_{t,i}(u) = w_0(t) \int_t^u w_1(v_1) \int_t^{v_1} w_2(v_2) \cdots \int_t^{v_{i-1}} w_i(v_i)\, \mathrm{d}v_i \cdots \mathrm{d}v_1$$

and

$$g_{t,i}(u) = w_0(t) \int_u^t w_1(v_1) \int_{v_1}^t w_2(v_2) \cdots \int_{v_{i-1}}^t w_i(v_i)\, \mathrm{d}v_i \cdots \mathrm{d}v_1.$$

Note that by Lemma 2.2 both $f_{t,i}$ and $g_{t,i}$ are polynomials in the $u_i$. The next proposition relates conditional expectations of $f_{t,i}$ and $g_{t,i}$ with the iterated integrals from Definition 3.2.

**Proposition 3.2.** *Assume that $F$ is a probability distribution and let $\bar{F} = 1 - F$ (not to be confused with the upper and lower iterated integrals of $F$). Also, let $X \sim F$. Then*

$$\mathrm{E}(f_{t,i}(X) \mid X > t) = \frac{\bar{F}_i(t)}{\bar{F}(t)}$$

*and*

$$\mathrm{E}(g_{t,i}(X) \mid X \le t) = \frac{F_i(t)}{F(t)}.$$

*Proof.* A simple application of Fubini's theorem shows that

$$\bar{F}_i(t) = \int_t^\infty f_{t,i}(u)\, \mathrm{d}F(u).$$

The first result follows by dividing both sides of this equation by $\bar{F}(t)$. The proof of the second result is analogous.

**Remark 3.3.** For the classical ECC-system, $f_{t,i}(x) = (x-t)^i$ and $g_{t,i}(x) = (t-x)^i$. In reliability theory, $\mathrm{E}((X-t)^j \mid X > t)$ is the $j$th moment of the residual life.

**Remark 3.4.** For the classical ECC-system, if $F_0$ is the empirical distribution over the $n$ points $x_i$, $x_1 < \cdots < x_n$, then Proposition 3.2 helps us to see that

$$\bar{F}_i(x_i) = \frac{1}{n}((x_{i+1} - x_i)^i + \cdots + (x_n - x_i)^i).$$

**Remark 3.5.** Proposition 3.2 tells us that

$$\bar{F}_i(t) = \mathrm{E}(f_{t,i}(X)1_{\{X>t\}}) \quad \text{and} \quad F_i(t) = \mathrm{E}(g_{t,i}(X)1_{\{X\le t\}}),$$

where $1_{\{\cdot\}}$ denotes the indicator function. For the classical ECC-system, these expectations are equivalent to those used in Denuit *et al.* (1998, Theorem 3.2).

The next lemma tells us that, except for a sign, the difference of upper integrals is the same as the difference of lower integrals.

**Lemma 3.1.** *Let $F$ and $G$ be as defined in Definition 3.1. If*

$$\int u_i \,\mathrm{d}F = \int u_i \,\mathrm{d}G \tag{3.1}$$

*for $i = 0, \ldots, n$, then*

$$\bar{G}_i - \bar{F}_i = (-1)^{i+1}(G_i - F_i). \tag{3.2}$$

*Proof.* An induction argument is used here. Let $i = 0$. Then

$$\begin{aligned}
\bar{G}_0(t) - \bar{F}_0(t) &= \int_t^{\infty} w_0(u)\,\mathrm{d}(G(u) - F(u)) \\
&= \int_{-\infty}^{\infty} w_0(u)\,\mathrm{d}(G(u) - F(u)) - \int_{-\infty}^{t} w_0(u)\,\mathrm{d}(G(u) - F(u)) \\
&= 0 - (G_0(t) - F_0(t)) \\
&= -(G_0(t) - F_0(t)),
\end{aligned}$$

where the last equality follows from (3.1). Next, suppose that (3.2) holds for $i = 0, \ldots, k$. Then, for $t < c$,

$$\begin{aligned}
\bar{G}_{k+1}(t) - \bar{F}_{k+1}(t) &= \int_t^{\infty} w_{k+1}(u)(\bar{G}_k(u) - \bar{F}_k(u))\,\mathrm{d}u \\
&= \int_t^{c} w_{k+1}(u)(\bar{G}_k(u) - \bar{F}_k(u))\,\mathrm{d}u + \int_c^{\infty} w_{k+1}(u)(\bar{G}_k(u) - \bar{F}_k(u))\,\mathrm{d}u \\
&= \int_t^{c} w_{k+1}(u)(-1)^{k+1}(G_k(u) - F_k(u))\,\mathrm{d}u + \bar{G}_{k+1}(c) - \bar{F}_{k+1}(c) \\
&= \bar{G}_{k+1}(c) - \bar{F}_{k+1}(c) + (-1)^{k+1}\int_{-\infty}^{c} w_{k+1}(u)(G_k(u) - F_k(u))\,\mathrm{d}u \\
&\quad - (-1)^{k+1}\int_{-\infty}^{t} w_{k+1}(u)(G_k(u) - F_k(u))\,\mathrm{d}u \\
&= \bar{G}_{k+1}(c) - \bar{F}_{k+1}(c) + (-1)^{k+1}(G_{k+1}(c) - F_{k+1}(c)) \\
&\quad + (-1)^{k+2}\int_{-\infty}^{t} w_{k+1}(u)(G_k(u) - F_k(u))\,\mathrm{d}u \\
&= \int u_{k+1}\,\mathrm{d}(G - F) + (-1)^{k+2}(G_{k+1}(t) - F_{k+1}(t)) \\
&= (-1)^{k+2}(G_{k+1}(t) - F_{k+1}(t)),
\end{aligned}$$

where the penultimate equality follows from Proposition 3.1 and the last equality follows from (3.2). A similar proof holds for $t > c$.

The next theorem is a generalization of a result of Denuit *et al.* (1998, Theorem 3.3) related to tail integral of survival transforms, and is similar to a result of Karlin and Studden (1966, Chapter XI, Theorem 5.1) (also, see Vera and Lynch (2005a, Section 3) for a problem in which a nonclassical C-system arises in a mixed distribution setting).

**Theorem 3.1.** *If (3.1) holds then $G >_U F$ if and only if $\bar{G}_n \geq \bar{F}_n$.*

*Proof.* The proof relies on the representation given in Theorem 2.1. First assume that $\bar{G}_n - \bar{F}_n \geq 0$, and suppose that $f$ is $U$-convex with $|\int f\,\mathrm{d}G| < \infty$ and $|\int f\,\mathrm{d}F| < \infty$. Then, with $a_i = D_i f(c)/w_i(c)$,

$$\begin{aligned}
\int f\,\mathrm{d}(G-F) &= \int_{-\infty}^{c} f(t)\,\mathrm{d}(G(t)-F(t)) + \int_{c}^{\infty} f(t)\,\mathrm{d}(G(t)-F(t)) \\
&= (-1)^{n+1}\int_{-\infty}^{c} w_0(t)\int_{t}^{c} w_1(x_1)\int_{x_1}^{c}\cdots w_n(x_n)\int_{x_n}^{c}\mathrm{d}\rho_{\mathrm{R}} f(u)\,\mathrm{d}x_n\cdots\mathrm{d}x_1\,\mathrm{d}(G(t)-F(t)) \\
&\quad + \int_{c}^{\infty} w_0(t)\int_{c}^{t} w_1(x_1)\int_{c}^{x_1}\cdots w_n(x_n)\int_{c}^{x_n}\mathrm{d}\rho_{\mathrm{R}} f(u)\,\mathrm{d}x_n\cdots\mathrm{d}x_1\,\mathrm{d}(G(t)-F(t)) \\
&\quad + \sum_{i=0}^{n}\biggl(\int_{-\infty}^{c} a_i u_i(t)\,\mathrm{d}(G(t)-F(t)) + \int_{c}^{\infty} a_i u_i(t)\,\mathrm{d}(G(t)-F(t))\biggr) \\
&= (-1)^{n+1}\int_{-\infty}^{c}\int_{-\infty}^{u} w_n(x_n)\int_{-\infty}^{x_n}\cdots w_1(x_1) \\
&\qquad\qquad \times\int_{-\infty}^{x_1} w_0(t)\,\mathrm{d}(G(t)-F(t))\,\mathrm{d}x_1\cdots\mathrm{d}x_n\,\mathrm{d}\rho_{\mathrm{R}} f(u) \\
&\quad + \int_{c}^{\infty}\int_{u}^{\infty} w_n(x_n)\int_{x_n}^{\infty}\cdots w_1(x_1)\int_{x_1}^{\infty} w_0(t)\,\mathrm{d}(G(t)-F(t))\,\mathrm{d}x_1\cdots\mathrm{d}x_n\,\mathrm{d}\rho_{\mathrm{R}} f(u) \\
&\quad + \sum_{i=0}^{n} a_i\int u_i d(G-F) \\
&= (-1)^{n+1}\int_{-\infty}^{c}(G_n(u)-F_n(u))\,\mathrm{d}\rho_{\mathrm{R}} f(u) + \int_{c}^{\infty}(\bar{G}_n(u)-\bar{F}_n(u))\,\mathrm{d}\rho_{\mathrm{R}} f(u),
\end{aligned}$$

where the third equality follows from Fubini's theorem and the last equality follows from (3.1). Note that, from (3.2), $\bar{G}_n - \bar{F}_n \geq 0$ is equivalent to $(-1)^{n+1}(G_n - F_n) \geq 0$. Hence, $\int f\,\mathrm{d}(G-F) \geq 0$.

Next assume that $G >_U F$, and define a function $f_t$ as follows:

$$f_t(x) = \begin{cases} w_0(x)\displaystyle\int_{t}^{x} w_1(x_1)\int_{t}^{x_n}\cdots w_{n-1}(x_{n-1})\int_{t}^{x_n} w_n(x_n)\,\mathrm{d}x_n\cdots\mathrm{d}x_1, & x \geq t, \\ 0, & x < t. \end{cases}$$

It is then easy to show that $\rho_{\mathrm{R}} f_t(x)$ equals 1 for $x \geq t$ and 0 for $x < t$. Hence, by Lemma 2.4, $f$ is $U$-convex. An application of Fubini's theorem gives $\bar{G}_n(t) = \int f_t\,\mathrm{d}G$ and $\bar{F}_n(t) = \int f_t\,\mathrm{d}F$. Thus,

$$\bar{G}_n(t) - \bar{F}_n(t) = \int f_t\,\mathrm{d}(G-F) \geq 0.$$

## 4. *k*-marts

In this section, $k$-mart structures are introduced and defined as follows.

**Definition 4.1.** Let $U = \{1, u_1, \dots, u_n\}$ be a C-system and, for $k = \lfloor (n+2)/2 \rfloor$ (where $\lfloor\cdot\rfloor$ denotes the integer part of its argument), let $(\boldsymbol{X}, Y) \equiv (X_1, \dots, X_k, Y)$ be jointly distributed random variables with $X_1, \dots, X_k$ independent and identically distributed. We say

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

that $(X_1, \ldots, X_k, Y)$ *has a k-mart structure (or is a k-mart)* if, for $j = 1, \ldots, n$,

$$\mathrm{E}(u_j(Y) \mid X_1, \ldots, X_k) = \frac{u_j(X_1) + \cdots + u_j(X_k)}{k} =: \bar{u}_j(\boldsymbol{X}) \tag{4.1}$$

**Remark 4.1.** A 1-mart with respect to the C-system $\{1, x\}$ is a one-step martingale.

Blackwell (1951) proved that a dilation is related to a one-step martingale (see also Strassen (1965)). A similar result relating balayages and $k$-marts is presented in the next theorem.

**Theorem 4.1.** *Let $U = \{1, u_1, \ldots, u_n\}$ be a C-system of continuous functions on a finite interval $[a, b]$ and let $F$ and $G$ be two distributions under which $\mathrm{E}(u_j) < \infty$. Then $G >_U F$ if and only if there exists a k-mart structure $(X_1, \ldots, X_k, Y)$ with $X_1 \sim F$ and $Y \sim G$.*

Theorem 4.1 is a direct consequence of the next two lemmas, where $E_{\boldsymbol{X}}$ is a distribution placing equal masses $1/k$ at $X_i$, $i = 1, \ldots, k$, and $(Y \mid \boldsymbol{X})$ is the conditional distribution of $Y$ given $\boldsymbol{X}$.

**Lemma 4.1.** *Let $U = \{1, u_1, \ldots, u_n\}$ be a C-system of continuous functions on a finite interval $[a, b]$ and let $F$ and $G$ be two distributions under which $\mathrm{E}(u_j) < \infty$. Then $G >_U F$ if and only if there exist jointly distributed random variables $(X_1, \ldots, X_k, Y)$ with $X_1 \sim F$ and $Y \sim G$ such that $(Y \mid X_1, \ldots, X_k) >_U E_{\boldsymbol{X}}$.*

*Proof.* See Lynch (1988, Theorem 4.1) for details.

**Lemma 4.2.** *The collection of random variables $(X_1, \ldots, X_k, Y)$ is a k-mart if and only if $(Y \mid X_1, \ldots, X_k) >_U E_{\boldsymbol{X}}$.*

*Proof.* If $(Y \mid X_1, \ldots, X_k) >_U E_{\boldsymbol{X}}$ then it is trivial to see that (4.1) holds since both $u_i$ and $-u_i$ are $U$-convex for $i = 0, \ldots, n$. To prove the converse, suppose that (4.1) holds. Then, by Theorem 2.1 of Lynch (1988), $\mathrm{E}(c(Y) \mid X_1, \ldots, X_k) \geq \mathrm{E}(c(Z))$ for any $U$-convex function $c$, where $Z \sim E_{\boldsymbol{X}}$; thus, $(Y \mid X_1, \ldots, X_k) >_U E_{\boldsymbol{X}}$.

Theorem 4.1 can be generalized to C-systems of continuous functions on an infinite interval, with some mild restrictions. For the C-system $U = \{1, u_1, \ldots, u_n\}$, suppose that there exists a positive, continuous function $w$ for which

$$\lim_{|t| \to \infty} \frac{u_i(t)}{w(t)} = l_i$$

exists and is finite for $i = 1, \ldots, n$. Let $S$ be a strictly increasing function from $[-\infty, \infty]$ onto $[0, 1]$, and define

$$v_i(t) = \begin{cases} \dfrac{u_i(S^{-1}(t))}{w(S^{-1}(t))}, & t \in (0, 1), \\ l_i, & t = 0, 1, \end{cases}$$

for $i = 1, \ldots, n$. The next few results characterize the collection $V = \{1, v_1, \ldots, v_n\}$ and its relationship with $U$.

**Proposition 4.1.** *If $U$ is a C-system of continuous functions on $(-\infty, \infty)$, then $V$ is a C-system of continuous functions on* $[0, 1]$.

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

*Proof.* Note that, for $0 \le x_0 < x_1 < \cdots < x_n \le 1$,

$$\begin{aligned}\det[v_i(x_j)\colon i = 0, \dots, n,\ j = 0, \dots, n] \\ = \frac{\det[u_i(S^{-1}(x_j))\colon i = 0, \dots, n,\ j = 0, \dots, n]}{w(x_1) \cdots w(x_n)}.\end{aligned}$$

The result follows from the facts that $U$ is a C-system and $S^{-1}$ is a strictly increasing function.

**Proposition 4.2.** *If $c(\cdot)$ is $V$-convex then $c(S(\cdot))$ is $U$-convex. Similarly, if $c(\cdot)$ is $U$-convex then $c(S^{-1}(\cdot))$ is $V$-convex.*

*Proof.* The proof is similar to that of Proposition 4.1, but uses the determinant displayed in Definition 2.4 (with the function $c$ replacing the function $f$).

**Lemma 4.3.** *Let $F$ and $G$ be distributions on $[0, 1]$ with no mass at 0 or 1, and let $\mu(t) = F(S^{-1}(t))$ and $\nu(t) = G(S^{-1}(t))$ be distributions on $[-\infty, \infty]$. Then $G >_U F$ if and only if $\nu >_V \mu$.*

*Proof.* Let $c(\cdot)$ be $V$-convex. Then, by Proposition 4.2, $c(S(\cdot))$ is $U$-convex. Therefore,

$$\mathrm{E}_\nu(c(X)) = \mathrm{E}_G(c(S(X))) \ge \mathrm{E}_F(c(S(X))) = \mathrm{E}_\mu(c(X)).$$

The proof of the converse result is analogous.

**Theorem 4.2.** *Let $F$ and $G$ be two distributions on $(-\infty, \infty)$ under which $\mathrm{E}(u_j) < \infty$. Then $G >_U F$ if and only if there exists a $k$-mart $(X_1, \dots, X_k, Y)$ with $X_1 \sim F$ and $Y \sim G$.*

*Proof.* Suppose that $(X_1, \dots, X_k, Y)$ has a $k$-mart structure. Then, for a convex function $c$,

$$\mathrm{E}(c(Y)) = \mathrm{E}(\mathrm{E}(c(Y) \mid X_1, \dots, X_k)) \ge \mathrm{E}\left(\frac{c(X_1) + \cdots + c(X_k)}{k}\right) = \mathrm{E}(c(X_1)),$$

where the last equality follows from the fact that $X_1, \dots, X_k$ are identically distributed. Therefore, $G >_U F$.

To prove the converse result, suppose that $G >_U F$. Then $\nu >_V \mu$, by Lemma 4.3. By Lemma 4.1, there exists a random vector $(W_1, \dots, W_k, Z)$ with a $k$-mart structure and with $W_1 \sim \mu$ and $Z \sim \nu$. Define $X_i = S^{-1}(W_i)$ for $i = 1, \dots, k$ and define $Y = S^{-1}(Z)$. Then, for a $U$-convex function $c$,

$$\begin{aligned}\mathrm{E}(c(Y) \mid X_1, \dots, X_k) &= \mathrm{E}(c(S^{-1}(Z)) \mid W_1, \dots, W_k) \\ &\ge \frac{c(S^{-1}(W_1)) + \cdots + c(S^{-1}(W_k))}{k} \\ &= \frac{c(X_1) + \cdots + c(X_k)}{k} \\ &= \bar{c}(\boldsymbol{X}),\end{aligned}$$

where the inequality follows from the fact that $c(S^{-1}(\cdot))$ is $V$-convex.

The previous developments can be used to build complex models similar to (1.3) from a baseline distribution using the basic operations of mixtures and convolutions.

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

Definition 4.1 suggests that a population $Y$, with distribution $G$, can be modeled in terms of a $k$-fold convolution of a distribution $F$ plus an error term, i.e.

$$Y = \frac{X_1 + \cdots + X_k}{k} + \varepsilon, \tag{4.2}$$

where the error term, $\varepsilon$, must now have more structure than in (1.4); enough that $(X_1, \dots, X_k, Y)$ is a $k$-mart with respect to the classical C-system $U = \{1, x, \dots, x^n\}$. The next proposition describes such a $k$-mart structure.

**Proposition 4.3.** *Let $U_n$ be the classical C-system $\{1, x, \dots, x^n\}$, and let $X_1, \dots, X_k$ be independent and identically distributed with $X_1 \sim F$ for some distribution $F$ with respect to which the expectations of the functions in $U_n$ are finite. For some random variable $\varepsilon$ with $\mathrm{E}(|\varepsilon|^j) < \infty$, define $Y$ as in (4.2). Then $(X_1, \dots, X_k, Y)$ has a $k$-mart structure if and only if*

$$\mathrm{E}(\varepsilon^j \mid X_1, \dots, X_k) = \frac{(X_1 - \bar{X})^j + \cdots + (X_k - \bar{X})^j}{k} \tag{4.3}$$

*for $j = 1, \dots, n$, where $\bar{X} = (X_1 + \cdots + X_k)/k$.*

*Proof.* Assume that $(X_1, \dots, X_k, Y)$ has a $k$-mart structure. Then

$$\begin{aligned}
\mathrm{E}(\varepsilon^j \mid X_1, \dots, X_k) &= \mathrm{E}((Y - \bar{X})^j \mid X_1, \dots, X_k) \\
&= \mathrm{E}\bigg(\sum_{l=0}^{j} \binom{j}{l} (-\bar{X})^{j-l} Y^l \mid X_1, \dots, X_k\bigg) \\
&= \sum_{l=0}^{j} \binom{j}{l} (-\bar{X})^{j-l}\, \mathrm{E}(Y^l \mid X_1, \dots, X_k) \\
&= \sum_{l=0}^{j} \binom{j}{l} (-\bar{X})^{j-l} \frac{1}{k} \sum_{i=1}^{k} X_i^l \\
&= \frac{1}{k} \sum_{i=1}^{k} \sum_{l=0}^{j} \binom{j}{l} (-\bar{X})^{j-l} X_i^l \\
&= \frac{1}{k} \sum_{i=1}^{k} (X_i - \bar{X})^j,
\end{aligned}$$

as required. Proving that $(X_1, \dots, X_k, Y)$ has a $k$-mart structure given that (4.3) holds is similar.

The next few results give properties of $k$-marts and various useful relationships.

**Proposition 4.4.** *The collection of random variables $(X_1, \dots, X_k, Y)$ is a $k$-mart with respect to the C-system $\{1, u_1, \dots, u_n\}$ if and only if $(\bar{u}_j(\boldsymbol{X}), u_j(Y))$ is a one-step martingale for $j = 1, \dots, n$.*

*Proof.* The proof follows straightforwardly from Definition 4.1.

**Proposition 4.5.** *Let $U$ be the classical C-system $\{1, x, \dots, x^n\}$. If $(X_1, \dots, X_k, Y)$ is a $k$-mart with respect to $U$, then $(ZX_1, \dots, ZX_k, ZY)$ and $(Z + X_1, \dots, Z + X_k, Z + Y)$, where $Z$ is independent of $X_i$, $i = 1, \dots, k$, and $Y$, are also $k$-marts.*

*Proof.* The result is proved here for the multiplicative case. The proof for the additive case is similar. By direct calculation,

$$\begin{aligned}
\mathrm{E}((ZY)^j \mid ZX_1, \dots, ZX_k) &= \mathrm{E}(Z^j Y^j \mid ZX_1, \dots, ZX_k) \\
&= \mathrm{E}(\mathrm{E}(Z^j Y^j \mid Z, X_1, \dots, X_k) \mid ZX_1, \dots, ZX_k) \\
&= \mathrm{E}(Z^j \,\mathrm{E}(Y^j \mid Z, X_1, \dots, X_k) \mid ZX_1, \dots, ZX_k) \\
&= \mathrm{E}(Z^j \,\mathrm{E}(Y^j \mid X_1, \dots, X_k) \mid ZX_1, \dots, ZX_k) \\
&= \mathrm{E}\left(Z^j \frac{X_1^j + \cdots + X_k^j}{k} \,\middle|\, ZX_1, \dots, ZX_k\right) \\
&= \mathrm{E}\left(\frac{(ZX_1)^j + \cdots + (ZX_k)^j}{k} \,\middle|\, ZX_1, \dots, ZX_k\right) \\
&= \frac{(ZX_1)^j + \cdots + (ZX_k)^j}{k},
\end{aligned}$$

as required.

**Proposition 4.6.** *For the classical C-system $U$, if $Y >_U X$ and $Z$ is independent of $X$ and $Y$, then $ZY >_U ZX$ and $Z + Y >_U Z + X$.*

*Proof.* This result is a direct consequence of Lemma 4.1 and Proposition 4.5.

Proposition 4.6 can be used to study how, for scale and location families, balayage structures for mixing distributions are transmitted to the mixed distribution. Let $F_\theta(x) = F(x/\theta)$, where $F$ is a cumulative distribution function and $\theta \in \Theta$ is a scale parameter, and let $\mu$ and $\nu$ be two distributions on $\theta$. Let $X$, $Y$, and $Z$ be independent with $X \sim \mu$, $Y \sim \nu$, and $Z \sim F$. Then $XZ$ and $YZ$ are random variables with respective mixed distributions $F_\mu = \int F_\theta \,\mathrm{d}\mu(\theta)$ and $F_\nu = \int F_\theta \,\mathrm{d}\nu(\theta)$.

More generally, a hierarchical balayage arises naturally when fitting more general mixtures using the 'method of moments'. To see this, for $r = 1, 2, \dots$ let $W_r$ denote the probability distribution that places the masses $w_{r1}, \dots, w_{rr}$ at the respective positions $\theta_{r1}, \dots, \theta_{rr} \in \Theta$, where this is done in such a way that $W_{r+1} >_{U_{2r-1}} W_r$. Then define ${}_rF$ as the distribution of a mixture over the family $F_\theta$, $\theta \in \Theta$, such that

$${}_rF = \int F_\theta \,\mathrm{d}W_r(\theta) = \sum_{i=1}^{r} w_{ri} F_{\theta_{ri}}.$$

Assume that $F_\theta$ is a distribution with density $f_\theta$ with respect to some measure $\lambda$, and that $f_\theta(x)$ is *totally positive* in $\theta$ and $x$ (see Karlin (1968) for the definitions and its consequences). Then, by the variation diminishing theorem (Karlin (1968)), ${}_{k+r}F - {}_kF$ has at most $2k - 1$ sign changes, since the difference

$${}_{k+r}f - {}_kf = \int f_\theta(\mathrm{d}W_{k+r}(\theta) - \mathrm{d}W_k(\theta))$$

has at most $2k$ sign changes. Assume also that

$$\tilde{U}_{2r-1} = \left\{\tilde{u}_j : \tilde{u}_j(\theta) = \int u_j(x)\,\mathrm{d}F_\theta(x),\ j = 0, \dots, 2r-1\right\}$$

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

is a C-system for $r = 1, 2, \ldots$. Then, by Theorem 3.1 of Lynch (1988),

$$_{k+r}F >_{U_{2r-1}} {}_kF.$$

Hence, by Lemma 4.1, there exist jointly distributed random variables with a $k$-mart structure and marginal distributions $_kF$ and $_{k+r}F$.

Vera and Lynch (2005a) used these ideas to fit a binomial mixture.

## 5. Synthetic data

The ideas presented in the previous sections can be applied to problems of statistical disclosure control, which is defined as 'the discipline concerned with the modification of statistical data, containing individual information about entities…, in order to prevent third parties working with these data to recognize individuals in the data' (Willenborg and de Waal (2001a)).

Statistics such as the mean and the variance are not enough for many statistical analyses, and the original data may be required. Nevertheless, because of disclosure limitations, sometimes it is not possible to release the original version of the data to the parties doing the statistical analysis. It could be possible, however, to release a data set similar to the original one, keeping some of its properties but concealing any sensitive information (see, e.g. Dalenius and Reiss (2002)).

In this section, some techniques for data disclosure are presented which use the ideas of balayages and $k$-marts. In particular, a modification of microaggregation resulting in a balayage structure is proposed in Subsection 5.1, and the generation of synthetic data through the generation of $k$-marts is discussed in Subsection 5.2.

### 5.1. Moment-preserving microaggregation

One technique of data disclosure is microaggregation, which consists in grouping the data and replacing the values within each group by their mean (Defays and Anwar (1995)). If there were only one group then all of the original data would be replaced by one value, the mean. If each group were to contain one data point, then the disclosed data would be the original data.

Suppose that the original data are divided into groups of size $n$, let $x_1, \ldots, x_n$ denote the original values in one of the groups, and let $y_1, \ldots, y_n$ denote the released values for this group. Then, using microaggregation, $y_i = \bar{x}$, $i = 1, \ldots, n$, i.e. the released version replaces the empirical distribution within this group by a distribution that places mass 1 at $z_1 := \bar{x}$. While this construction preserves the first moment of the group, it does not preserve higher-order moments; in particular, the released values for the group will have the smallest second moment among all those distributions having the same first moment. Moreover, the data in the group is a balayage of the released data with respect to the C-system $\{1, x\}$.

An alternative way of releasing the data within a group is by considering a distribution that matches the first $2k - 1$ moments of the group. Here we consider the *narrow* representation of those moments, which is a distribution with the specified $2k - 1$ moments that minimizes the $2k$th moment among all those distributions that match the first $2k - 1$ moments (Denuit *et al.* (2000)). A method of finding such a distribution, similar to that of Lindsay (1989, Theorem 2C), is presented in the following results, restricted only to moments of the classical C-system.

In preparation for the next proposition, the polynomial $Q$ is defined as

$$Q(z) := (z - z_1) \cdots (z - z_k) =: z^k + a_{k-1}z^{k-1} + \cdots + a_1 z + a_0. \tag{5.1}$$

Also, $H$ denotes a distribution whose support, $\{z_1, \ldots, z_k\}$, is such that $z_1 < \cdots < z_k$.

**Proposition 5.1.** *Let $Z \sim H$. Then*

$$\mathrm{E}(Z^j Q(Z)) = 0 \tag{5.2}$$

*for $j = 0, 1, \ldots, 2k-1$, provided that the expectations exist.*

*Proof.* The result follows from simply noting that $Q$ is equal to 0 when evaluated at the support points of $H$.

Proposition 5.1 can be used to find the support points, $z_1, \ldots, z_k$, of any distribution $H$, given the first $2k-1$ moments of $H$, i.e.

$$\mu_j := \int z^j \, \mathrm{d}H(z), \qquad j = 0, 1, \ldots, 2k-1, \tag{5.3}$$

by simply finding the roots of $Q$.

The next proposition tells us how to find the polynomial $Q$ (see (5.1)) from the moments $\mu_j$.

**Proposition 5.2.** *The coefficients of $Q$, namely $a_0, \ldots, a_{k-1}$, are the solution to the system of equations*

$$\mu_j a_0 + \mu_{j+1} a_1 + \cdots + \mu_{j+k-1} a_{k-1} = -\mu_{k+j}, \qquad j = 0, \ldots, k-1.$$

*Proof.* The result is immediate from (5.2) and (5.3).

**Remark 5.1.** Note that the polynomial $Q$ is similar to the polynomial defined in Denuit *et al.* (2000, Equation 4.4). However, all the coefficients of $Q$ can be found by solving one system of equations, which is numerically more efficient than calculating a determinant for each coefficient.

The next proposition shows how to determine the masses $\alpha_1, \ldots, \alpha_k$ corresponding to $z_1, \ldots, z_k$ for the distribution $H$.

**Proposition 5.3.** *Let $Z \sim H$. Then*

$$\alpha_i = \mathrm{E}\biggl(\frac{(Z - z_1) \cdots (Z - z_{i-1})(Z - z_{i+1}) \cdots (Z - z_k)}{(z_i - z_1) \cdots (z_i - z_{i-1})(z_i - z_{i+1}) \cdots (z_i - z_k)}\biggr).$$

*Proof.* The result follows from simply noting that the function inside the expectation is equal to 0 at all points of support except $z_i$, where it is equal to 1.

Note that the function inside the expectation can be evaluated quickly from $Q$ using synthetic division.

Propositions 5.1, 5.2, and 5.3 yield an algorithm to find the principal representation (Karlin and Studden (1966, Chapter II)) of a finite moment sequence in the case of the classical C-system, when $n$ is odd. This algorithm has been implemented by the authors, using the computer language R. A description of this algorithm is as follows.

**Algorithm 5.1.**

- *Input* $\mu_1, \mu_2, \ldots, \mu_{2k-1}$.
- *Solve for* $a_0, \ldots, a_{k-1}$ *using*

$$\mu_j a_0 + \mu_{j+1} a_1 + \cdots + \mu_{j+k-1} a_{k-1} = -\mu_{k+j}, \qquad j = 0, \ldots, k-1.$$

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

- *Find the roots,* $z_1, \dots, z_k$*, of* $Q(z) = z^k + a_{k-1}z^{k-1} + \cdots + a_1 z + a_0 = 0$.
- *Perform synthetic division:*

$$Q_i'(z) = \frac{Q(z)}{z - z_i} = z^{k-1} + b_{i,k-2}z^{k-2} + \cdots + b_{i1}z_1 + b_{i0}, \qquad i = 1, \dots, k.$$

- *Let*

$$\alpha_i = \frac{\mu_{k-1} + b_{i,k-2}\mu_{k-2} + \cdots + b_{i1}\mu_1 + b_{i0}}{Q_i'(z_i)}, \qquad i = 1, \dots, k.$$

- *Output* $z_1, \dots, z_k, \alpha_1, \dots, \alpha_k$.

### 5.2. Generating *k*-marts

One way to generate synthetic data is through the use of $k$-marts. Suppose that $X_1, \dots, X_k$ is a sample from the original data. Let $Y$ be a random variable that is jointly distributed with the $X_i$ in such a way that $(X_1, \dots, X_k, Y)$ is a $k$-mart. Then in the released data we could substitute the original values of this group with values generated from the distribution of $Y$ given $X_1, \dots, X_k$.

We next present two ways of generating $k$-marts. The first is based on the construction of a discrete distribution for $Y$ given $X_1, \dots, X_k$. In the second, the distribution of $Y$ given $X_1, \dots, X_k$ is constructed to be 'most random' in the sense of *maximum entropy*. These two approaches are discussed in Sections 5.2.1 and 5.2.2, respectively.

Throughout this section, $U = \{1, u_1, \dots, u_n\}$ is assumed to be a C-system and, for $k = \lfloor n + 2/2 \rfloor$, $X_1, \dots, X_k$ are assumed to be independent and identically distributed.

5.2.1. *The discrete-distribution approach.* Here we construct a $k$-mart $(\boldsymbol{X} = (X_1, \dots, X_k), Y)$ where $Y \mid X_1, \dots, X_k$ has a discrete distribution, say $G_{\mathrm{d}}$. If $Y_{\mathrm{d}} \sim G_{\mathrm{d}}$ (for some random variable $Y_{\mathrm{d}}$), it is then necessary that

$$\mathrm{E}(u_j(Y_{\mathrm{d}})) = \bar{u}_j(\boldsymbol{X}), \qquad j = 1, \dots, n. \tag{5.4}$$

One option would be to set the support points for $Y_{\mathrm{d}}$, say $y_1, \dots, y_k$, to equal $X_1, \dots, X_k$, and to set the probability masses all to equal $1/k$. In such a case, it is trivial to prove that $Y$ is equal in distribution to $X_1$ and that $(X_1, \dots, X_k, Y)$ is a $k$-mart.

Another option is to add an extra support point for $Y_{\mathrm{d}}$, say $y_{k+1}$, and to set the support points' respective probability masses to equal $p_1, \dots, p_k, p_{k+1}$. This distribution is chosen such that (5.4) is satisfied, i.e. such that

$$p_1 u_j(y_1) + \cdots + p_k u_j(y_k) + p_{k+1} u_j(y_{k+1}) = \frac{u_j(X_1) + \cdots + u_j(X_k)}{k}, \qquad j = 1, \dots, n.$$

The support of a distribution $G_{\mathrm{d}}$, $\{y_1, \dots, y_k, y_{k+1}\}$, satisfying

$$\mu_j = \int y^j \,\mathrm{d}G_{\mathrm{d}}(y) = \frac{X_1^j + \cdots + X_k^j}{k}, \qquad j = 1, \dots, 2k-1,$$

can be found using Propositions 5.1 and 5.2. However, $2k+1$ moments are needed.

Let $\theta_j = \int y^{2k-1+j}\,\mathrm{d}G_{\mathrm{d}}(y)$, $j = 1, 2$, be parameters associated with $G_{\mathrm{d}}$. The space, $\Theta$, of possible values for the pair $(\theta_1, \theta_2)$ was studied in detail for C-systems by Karlin and Studden (1966, Chapter II, Section 7). Lindsay (1989, Theorem 2A) gave a simple solution for $\Theta$ in the

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

case of classical C-systems. According to his result, $\theta_2$ can be any scalar but $\theta_1$ must satisfy the determinant condition

$$\begin{vmatrix} 1 & \mu_1 & \cdots & \mu_{k-1} & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_k & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k-1} & \theta_1 \end{vmatrix} > 0.$$

Note that the above condition determines a lower bound, $L$, for $\theta_1$. Karlin and Studden (1966, Chapter II, Section 6) proved that $L$ is the $2k$th moment of the principal representation of $(1, \mu_1, \dots, \mu_{2k-1})$. In this case, the principal representation of these moments is the distribution putting equal masses $1/k$ at $X_1, \dots, X_k$.

The above discussion gives the following result.

**Lemma 5.1.** *For the moment sequence* $(1, \mu_1, \dots, \mu_{2k-1})$*, the space,* $\Theta$*, of possible values of* $\mu_{2k}$ *and* $\mu_{2k+1}$ *is*

$$\Theta = \{(\theta_1, \theta_2) \colon \theta_1 > L,\ \theta_2 \in \mathbb{R}\}$$

*with*

$$L = \frac{X_1^{2k} + \cdots + X_k^{2k}}{k},$$

*where* $X_1, \dots, X_k$ *is the principal representation of* $(1, \mu_1, \dots, \mu_{2k-1})$.

The procedure presented above can be used to model a population that is a balayage of a baseline model, by adding one or two parameters for the $2k$th and $(2k+1)$th moments of the distribution of $Y \mid X_1, \dots, X_k$. This model may be closer to the real population, while still keeping the salient features of the population in terms of the baseline model.

5.2.2. *A maximum-entropy approach.* A distribution for $Y \mid X_1, \dots, X_k$, say $G_{\mathrm{e}}$, can be constructed in such a way that $(X_1, \dots, X_k, Y)$ is a $k$-mart and $G_{\mathrm{e}}$ is most random or most uncertain in the sense of information entropy. We focus on C-systems with $u_0 \equiv 1$. For a distribution $\mu$ with density $f_\mu$ with respect to some measure $\nu$, the *entropy* of $\mu$ is defined as

$$\operatorname{ent}(\mu) = -\int \log(f_\mu)\, \mathrm{d}\mu.$$

It is well known (see Jaynes (1957a), (1957b)) that if, for some given moment sequence $\eta_1, \dots, \eta_n$, the entropy is maximized over all distributions $\mu$ satisfying

$$\int u_j(y)\, \mathrm{d}\mu(y) = \eta_j, \qquad j = 1, \dots, n,$$

then the maximum is attained at $\mu^*$, where

$$\mathrm{d}\mu^*(y) = \exp\biggl\{-\sum_{j=1}^{n} \lambda_j u_j(y) + \psi(\lambda_1, \dots, \lambda_n)\biggr\}\, \mathrm{d}\nu(y).$$

Here $\psi(\lambda_1, \dots, \lambda_n)$ is a normalizing constant (such that the density integrates to 1) and $\lambda_1, \dots, \lambda_n$ are the solutions to the system of equations

$$\frac{\partial \psi(\lambda_1, \dots, \lambda_n)}{\partial \lambda_j} = \eta_j, \qquad j = 1, \dots, n.$$

For $\boldsymbol{x} = (x_1, \dots, x_k)$, define $\lambda(\boldsymbol{x})$ to be the solution to the normal equations

$$\frac{\partial \psi(\lambda_1, \dots, \lambda_n)}{\partial \lambda_j} = \bar{u}_j(\boldsymbol{x}), \qquad j = 1, \dots, n. \tag{5.5}$$

Suppose now that $(X_1, \dots, X_k)$ has a distribution $F$ with support $\mathcal{S}$, and that (5.5) has a solution for every $\boldsymbol{x} \in T \subset \mathcal{S}$. Define $G_{\mathrm{e}}$, the distribution of $Y \mid X_1, \dots, X_k$, by

$$\mathrm{d}G_{\mathrm{e}}(y) = \begin{cases} \exp\Big\{-\sum_{j=1}^{n} \lambda_j(\boldsymbol{X}) u_j(y) + \psi(\lambda(\boldsymbol{X}))\Big\} \,\mathrm{d}\nu(y), & \boldsymbol{X} \in T, \\ \dfrac{1}{k} I_{X_1} + \cdots + \dfrac{1}{k} I_{X_k}, & \boldsymbol{X} \in T^{\mathrm{c}}, \end{cases}$$

where $I_X$ denotes the density (with respect to $\nu$) of a distribution that places mass 1 at the point $X$. For $\boldsymbol{X} \in T$, $G_{\mathrm{e}}$ is the most random or most uncertain distribution among all those distributions $G$ satisfying

$$\int u_j(y)\,\mathrm{d}G(y) = \bar{u}_j(\boldsymbol{X}), \qquad j = 1, \dots, n.$$

Also, the marginal distribution of $Y$ is a balayage of the marginal distribution of $X_1$.

When fitting a model to a population, this approach allows us to improve the fit without adding any extra parameters. Moreover, if (5.5) has a solution for every $\boldsymbol{x} \in \mathcal{S}$, then a model fitted in this way maximizes the entropy of the conditional distribution of $Y \mid \boldsymbol{X}$, thus accounting for a ‘worst-case scenario’, while matching several moments of the target population.

## 6. Construction of a 1-mart with ‘most nearly identical’ elements

Theorems 4.1 and 4.2 imply the existence of a transition kernel from $(X_1, \dots, X_k)$ to $Y$, given the marginal distributions $F$ and $G$. Blackwell (1951) gave a construction of such a transition kernel for the C-system $\{1, x\}$ where the desired kernel is obtained as an iterative procedure each iteration of which is a dilation and which converges to the required transition kernel after an infinite number of steps. In this section, we construct such a transition kernel for the C-system $\{1, u(x)\}$ which is not iterative, but is direct, with the added benefit that the joint distribution is that which makes $X$ ‘most nearly identical’ to $Y$, i.e. maximizes $\mathrm{P}(X = Y)$.

The ideas in this section can be used to fit stochastic models in which the observed variable is a dilation of a latent variable that is of interest to the researcher, yet the probability that the two variables are equal is maximized. We require the following definition.

**Definition 6.1.** The jointly distributed random variables $(X, Y)$ are said to be a 1-mart structure if $\mathrm{E}(u(Y) \mid X) = u(X)$.

Let $F$ and $G$ be absolutely continuous with respect to the Lebesgue measure, $\sigma$, and let $f$ and $g$ be their respective densities. Assume that $g - f$ has a finite number of sign changes and that these sign changes occur at $\zeta_1, \dots, \zeta_{2m}$ for some $m \in \{1, 2, \dots\}$ (according to Vera and Lynch (2005b, Corollary 3.1), the number of sign changes is even).

Define $p$ by

$$p = \frac{1}{2} \int |f - g|\,\mathrm{d}\sigma = \int (f - g)^+\,\mathrm{d}\sigma,$$

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

and let $\bar{p} = 1 - p$; thus defined, $p$ is just the Scheffé distance between $F$ and $G$. Also, let

$$h_0 = \frac{f \wedge g}{\bar{p}}, \qquad h_1 = \frac{(f-g)^+}{p}, \qquad h_2 = \frac{(g-f)^+}{p},$$

where the binary operator '$\wedge$' denotes the minimum of its arguments and $(\cdot)^+$ returns its argument when it is positive and 0 when its argument is negative.

The next proposition gives a mixture representation for $f$ and $g$ in terms of $h_i$, $i = 0, 1, 2$.

**Proposition 6.1.** *The densities $f$ and $g$ can be represented as mixtures of $h_0$, $h_1$, and $h_2$ as follows:*

$$f = \bar{p}h_0 + ph_1, \qquad g = \bar{p}h_0 + ph_2.$$

*Proof.* To prove the result for $f$, simply note that $ph_1 = (f-g)^+$ and $\bar{p}h_0 = f \wedge g$. Therefore,

$$\bar{p}h_0 + ph_1 = f \wedge g + (f-g)^+ = f.$$

The proof for $g$ is similar.

The following proposition gives a necessary and sufficient condition in terms of $h_i$, $i = 1, 2$, for $g$ to be a balayage of $f$ (with respect to $U = \{1, u(x)\}$).

**Proposition 6.2.** *For the densities $f$ and $g$ and their mixture representation in terms of $h_0$, $h_1$, and $h_2$, we have $g >_U f$ if and only if $h_2 >_U h_1$.*

*Proof.* For any function $c$,

$$\int c(g-f)\,\mathrm{d}\sigma = p\int c(h_2 - h_1)\,\mathrm{d}\sigma.$$

Therefore,

$$\int cg\,\mathrm{d}\sigma - \int cf\,\mathrm{d}\sigma \geq 0 \quad \Longleftrightarrow \quad \int ch_2\,\mathrm{d}\sigma - \int ch_1\,\mathrm{d}\sigma \geq 0.$$

Thus, if $c$ is $U$-convex then

$$\int cg\,\mathrm{d}\sigma \geq \int cf\,\mathrm{d}\sigma \quad \Longleftrightarrow \quad \int ch_2\,\mathrm{d}\sigma \geq \int ch_1\,\mathrm{d}\sigma,$$

which completes the proof.

Propositions 6.1 and 6.2 suggest a method to construct jointly distributed random variables $X$ and $Y$ with respective marginals $F$ and $G$. In the case in which $g - f$ has a finite number of sign changes, say $2m$, this method is as follows.

Define $\zeta_0 = -\infty$ and $\zeta_{2m+1} = \infty$, let $A_j = (\zeta_{2j}, \zeta_{2j+1})$, $j = 0, \dots, m$, and let $B_j = [\zeta_{2j-1}, \zeta_{2j}]$, $j = 1, \dots, m$. Notice that, by Vera and Lynch (2005b, Corollary 3.1), $g - f$ is positive on $A_j$, $j = 0, \dots, m$, and is negative on $B_j$, $j = 1, \dots, m$. Therefore, the supports of $h_1$ and $h_2$ are $B_1 \cup \dots \cup B_m$ and $A_0 \cup \dots \cup A_m$, respectively.

Let

$$\beta_j = \int_{B_j} h_1\,\mathrm{d}\sigma \quad \text{and} \quad h_{1j}(x) = \frac{h_1(x)}{\beta_j}1_{\{x \in B_j\}}, \tag{6.1}$$

https://doi.org/10.1239/aap/1175266471 Published online by Cambridge University Press

for $j = 1, \dots, m$, and note that $\beta_1 + \cdots + \beta_m = 1$ and

$$h_1 = \beta_1 h_{11} + \cdots + \beta_m h_{1m}. \tag{6.2}$$

Also, let

$$\begin{aligned} \alpha_{j\mathrm{L}} &= \int_{A_0 \cup \cdots \cup A_{j-1}} h_2 \,\mathrm{d}\sigma, & \alpha_{j\mathrm{R}} &= \int_{A_j \cup \cdots \cup A_m} h_2 \,\mathrm{d}\sigma, \\ h_{2j\mathrm{L}}(x) &= \frac{h_2(x)}{\alpha_{j\mathrm{L}}} 1_{\{x \in A_0 \cup \cdots \cup A_{j-1}\}}, & h_{2j\mathrm{R}}(x) &= \frac{h_2(x)}{\alpha_{j\mathrm{R}}} 1_{\{x \in A_j \cup \cdots \cup A_m\}}. \end{aligned} \tag{6.3}$$

Note that $\alpha_{j\mathrm{L}} + \alpha_{j\mathrm{R}} = 1$ and that

$$h_2 = \alpha_{j\mathrm{L}} h_{2j\mathrm{L}} + \alpha_{j\mathrm{R}} h_{2j\mathrm{R}} \tag{6.4}$$

for $j = 1, \dots, m$. Finally, let $\mu_{1j}$, $\mu_{2j\mathrm{L}}$, and $\mu_{2j\mathrm{R}}$ be the respective expectations of $u(Z)$ when $Z$ is distributed according to $h_{1j}$, $h_{2j\mathrm{L}}$ and $h_{2j\mathrm{R}}$, for $j = 1, \dots, m$, and let

$$h_{2j} = \frac{\mu_{2j\mathrm{R}} - \mu_{1j}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{L}} + \frac{\mu_{1j} - \mu_{2j\mathrm{L}}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{R}}. \tag{6.5}$$

We now generate $(X, Y)$ jointly according to the following construction.

**Construction 6.1.** (Joint distribution for $(X, Y)$.)

1. *Generate $U$ from a uniform random variable in the interval* $(0, 1)$.
2. *If $U > p$ (i.e. with probability $\bar{p}$),*
   - *generate $X$ from $h_0$ and*
   - *set $Y = X$.*
3. *If $U \le p$ (i.e. with probability $p$),*
   - *generate $X$ from $h_1$ and,*
   - *for $X \in B_j$, generate $Y$ from the mixture*

$$\frac{\mu_{2j\mathrm{R}} - u(X)}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{L}} + \frac{u(X) - \mu_{2j\mathrm{L}}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{R}}.$$

This construction produces a 1-mart $(X, Y)$ where $X$ and $Y$ are most nearly identical and have respective marginals $F$ and $G$. This is stated formally in the next theorem. For its proof, the following result (see Lemma 2.1 and Equation (2.5) of Sethuraman (2002)) is needed.

**Lemma 6.1.** *Let $F$ and $G$ be two distribution functions and let $p$ be the Scheffé distance between $F$ and $G$. Then*

$$\inf \mathrm{P}(X \neq Y) = p,$$

*where the infimum is over all jointly distributed random variables $X$ and $Y$ with respective marginals $F$ and $G$.*

**Theorem 6.1.** *If $X$ and $Y$ are generated using Construction 6.1, then* $\mathrm{E}(u(Y) \mid X) = u(X)$ *and* $\mathrm{P}(X = Y)$ *is maximized among all jointly distributed random variables $X$ and $Y$ with respective marginals $F$ and $G$.*

The above theorem is a consequence of the following propositions. By $q$ we denote the density of the conditional distribution of $Y$ given $X$, given by the second step of Construction 6.1.

**Proposition 6.3.** *If* $(X, Y)$ *is generated from Construction 6.1, then*

$$\mathrm{E}(u(Y) \mid X) = u(X).$$

*Proof.* By direct calculation,

$$\begin{aligned}
\mathrm{E}(u(Y) \mid X) &= \mathrm{E}(\mathrm{E}(u(Y) \mid X, U)) \\
&= \mathrm{E}(\mathrm{E}(u(Y)(1_{\{U>p\}} + 1_{\{U\le p\}}) \mid X, U) \mid X) \\
&= \mathrm{E}(\mathrm{E}(u(X)1_{\{U>p\}} \mid X, U) \mid X) + \mathrm{E}(\mathrm{E}(u(Y)1_{\{U\le p\}} \mid X, U) \mid X) \\
&= \mathrm{E}(u(X)1_{\{U>p\}} \mid X) + \mathrm{E}\biggl(\mathrm{E}\biggl(u(Y)1_{\{U\le p\}}\sum_{j=1}^{m} 1_{\{X\in B_j\}} \biggm| X, U\biggr) \biggm| X\biggr) \\
&= \mathrm{E}(1_{\{U>p\}})\,\mathrm{E}(u(X) \mid X) + \mathrm{E}\biggl(\sum_{j=1}^{m} \mathrm{E}(u(Y)1_{\{U\le p\}}1_{\{X\in B_j\}} \mid X, U) \biggm| X\biggr) \\
&= \bar{p}u(X) \\
&\quad + \mathrm{E}\biggl(1_{\{U\le p\}}\sum_{j=1}^{m}\biggl(\frac{\mu_{2j\mathrm{R}} - u(X)}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}}\mu_{2j\mathrm{L}} + \frac{u(X) - \mu_{2j\mathrm{L}}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}}\mu_{2j\mathrm{R}}\biggr)1_{\{X\in B_j\}} \biggm| X\biggr) \\
&= \bar{p}u(X) + \mathrm{E}(1_{\{U\le p\}})\,\mathrm{E}\biggl(\sum_{j=1}^{m} u(X)1_{\{X\in B_j\}} \biggm| X\biggr) \\
&= \bar{p}u(X) + p\,\mathrm{E}(u(X) \mid X) \\
&= \bar{p}u(X) + pu(X) \\
&= u(X),
\end{aligned}$$

as required.

**Proposition 6.4.** *The density* $h_2$ *can be represented as a mixture of the densities* $h_{2j}$, $j = 1, \dots, m$, *as follows, where* $\beta_j$ *is as defined in (6.1) and* $h_{2j}$ *is as defined in (6.3):*

$$h_2 = \sum_{j=1}^{m} \beta_j h_{2j}.$$

*Proof.* Let us try to represent $h_2$ as a mixture of $h_{21L}, h_{21R}, \dots, h_{2mL}, h_{2mR}$. By (6.4), there are an infinite number of ways to do this, since this mixture is not identifiable. Let us suppose that

$$h_2 = \sum_{j=1}^{m} \beta_j(\gamma_{j\mathrm{L}} h_{2j\mathrm{L}} + \gamma_{j\mathrm{R}} h_{2j\mathrm{R}}), \tag{6.6}$$

where $\gamma_{j\mathrm{L}} + \gamma_{j\mathrm{R}} = 1,\ j = 1, \dots, m$. Note that if (6.6) holds then

$$\mu_2 = \sum_{j=1}^{m} \beta_j(\gamma_{j\mathrm{L}}\mu_{2j\mathrm{L}} + \gamma_{j\mathrm{R}}\mu_{2j\mathrm{R}}).$$

Recall that $\mu_2 = \mu_1$ since $h_2 >_U h_1$. Also, from (6.2) it follows that

$$\mu_2 = \mu_1 = \sum_{j=1}^{m} \beta_j \mu_{1j}.$$

Therefore,

$$\sum_{j=1}^{m} \beta_j \mu_{1j} = \sum_{j=1}^{m} \beta_j (\gamma_{j\mathrm{L}} \mu_{2j\mathrm{L}} + \gamma_{j\mathrm{R}} \mu_{2j\mathrm{R}}).$$

One way for this equality to hold is for $\mu_{1j} = \gamma_{j\mathrm{L}} \mu_{2j\mathrm{L}} + \gamma_{j\mathrm{R}} \mu_{2j\mathrm{R}}$ to hold. This gives

$$\gamma_{j\mathrm{L}} = \frac{\mu_{2j\mathrm{R}} - \mu_{1j}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} \quad \text{and} \quad \gamma_{j\mathrm{R}} = \frac{\mu_{1j} - \mu_{2j\mathrm{L}}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}},$$

for $j = 1, \dots, m$. This, together with (6.5), completes the proof.

**Proposition 6.5.** *The variable $Y$ generated using Construction 6.1 has the required marginal distribution, i.e. $Y \sim G$.*

*Proof.* By Proposition 6.1, all we need to prove is that $Y \sim h_2$ if $X \sim h_1$. To do so, note that

$$\begin{aligned}
\int q(y|x) h_1(x)\,\mathrm{d}x &= \sum_{j=1}^{m} \beta_j \int_{B_j} \left( \frac{\mu_{2j\mathrm{R}} - u(x)}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{L}}(y) + \frac{u(x) - \mu_{2j\mathrm{L}}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{R}}(y) \right) h_{1j}(x)\,\mathrm{d}x \\
&= \sum_{j=1}^{m} \beta_j \left( \frac{\mu_{2j\mathrm{R}} - \mu_{1j}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{L}}(y) + \frac{\mu_{1j} - \mu_{2j\mathrm{L}}}{\mu_{2j\mathrm{R}} - \mu_{2j\mathrm{L}}} h_{2j\mathrm{R}}(y) \right) \\
&= \sum_{j=1}^{m} \beta_j h_{2j}(y) \\
&= h_2(y).
\end{aligned}$$

The results presented above can be generalized to the case in which the number of sign changes is countable.

## Acknowledgements

This research was partially supported by the NSF, grant no. DMS 0243594. We would like to thank the referee for comments on earlier versions of this paper and additional references. This version was written while both authors were at the National Institute of Statistical Sciences and the Statistical and Applied Mathematical Science Institute, where the first author was a postdoctoral fellow and the second was on sabbatical. The authors would like to thank the institutes for a stimulating environment and the support they received while there.

## References

Blackwell, D. (1951). Comparison of experiments. In *Proc. 2nd Berkeley Symp. Math. Statist. Prob.*, University of California Press, Berkeley, CA, pp. 93–102.

Blackwell, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24,** 265–272.

Dalenius, T. and Reiss, S. P. (2002). Data-swapping: a technique for disclosure control. *J. Statist. Planning Infer.* **6,** 73–85.

DEFAYS, D. AND ANWAR, N. (1995). Microaggregation: a generic method. In *Proc. 2nd Internat. Symp. Statist. Confidentiality*, Office for Official Publications of the European Community, Luxembourg, pp. 69–78.
DENUIT, M., LEFÈVRE, C. AND SHAKED, M. (1998). The $s$-convex orders among real random variables, with applications. *Math. Inequalities Appl.* **1,** 585–613.
DENUIT, M., LEFÈVRE, C. AND SHAKED, M. (2000). On $s$-convex approximations. *Adv. Appl. Prob.* **32,** 994–1010.
JAYNES, E. T. (1957a). Information theory and statistical mechanics. *Phys. Rev.* **106,** 620–630.
JAYNES, E. T. (1957b). Information theory and statistical mechanics. II. *Phys. Rev.* **108,** 171–190.
KARLIN, S. (1968). *Total Positivity*. Stanford University Press.
KARLIN, S. AND STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. John Wiley, New York.
LINDSAY, B. G. (1989). Moment matrices: application in mixtures. *Ann. Statist.* **17,** 722–740.
LYNCH, J. (1988). Mixtures, generalized convexity and balayages. *Scand. J. Statist.* **15,** 203–210.
MEYER, P. A. (1966). *Probability and Potentials*. Blaisdell, London.
ROSS, S. M. (1983). *Stochastic Processes*. John Wiley, New York.
SETHURAMAN, J. (2002). Some extensions of the Skorohod representation theorem. *Sankhyā A* **64,** 884–893.
SHAKED, M. (1980). On mixtures from exponential families. *J. R. Statist. Soc.* **42,** 192–198.
SHAKED, M. AND SHANTHIKUMAR, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, San Diego, CA.
STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36,** 423–439.
VERA, F. AND LYNCH, J. (2005a). $K$-mart stochastic modeling using iterated total time on test transforms. In *Modern Statistical and Mathematical Methods in Reliability* (Ser. Qual. Reliab. Eng. Statist. **10**), World Scientific, Singapore, pp. 395–409.
VERA, F. AND LYNCH, J. (2005b). Generalized convex stochastic orderings and related martingale-type structures. Tech. Rep., Department of Statistics, University of South Carolina. Available at http://www.stat.sc.edu/~veraf/.
WILLENBORG, L. AND DE WAAL, T. (2001a). *Elements of Statistical Disclosure Control* (Lecture Notes Statist. **155**). Springer, New York.
WILLENBORG, L. AND DE WAAL, T. (2001b). *Introduction to Biostatistics*. Freeman and Company, San Francisco, CA.