ESTIMATION, IMPUTATION, RANDOMIZATION AND RISK EQUIVALENCE

Charles Patrick, Statistics Canada

Introduction

The initial research leading up to this paper was motivated by the desire to view the problems of sampling from finite populations from a purely Bayesian decision theoretic view point, and thus to examine what components of practice could be identified explicitly in the model. It is therefore interesting that the theoretical model does offer versions of notions of randomization, imputation, etc.

The problem of decision making under uncertainty has a long and illustrious history, and, generally, has come to be cast in the form of an axiomatic theory of expected utility. Among variants of the theory that are currently available, one of the most brilliant is that put forward by L.J. Savage in his book "Foundations of Statistics" (1954, 1972). As has been usual, for most significant developments throughout the history of science, Savage's theory was not developed in a vacuum. Indeed he acknowledges and draws on the prior ideas of Ramsey (1931), de Finetti (1937), and von Neumann and Morgenstern (1947). A measure of Savage's accomplishment is that his system stood as the most general formulation of the expected utility hypothesis for nearly fifteen years and has yet to be fully superceded. As Savage himself was well aware, his system, notwithstanding its generality, suffers certain structural limitations:

A. The axioms of the system force the "states of the world" to have representation as an infinite set.

B. The system presumes that any logically possible decision is available.

C. The system does not reflect the realistic possibility that the choice of a decision, prior to realizing its consequences, might cause the probability distribution over "the states of the world" to be changed.

All three limitations strain intuition, though A and B have the flavour of mathematical idealization, and may not seem as disconcerting as C.

In 1971, Luce and Krantz, following earlier work by Fishburn (1964) and Pfanzagl (1967, 1968), developed a theory of conditional expected utility, which in its most general form more or less removes all of the constraints A through C, for a price. A complete discussion of this theory and some comparisons with some other theories may be found in Chapter Eight of "Foundations of Measurement" by Luce, Krantz, Suppes, and Tversky (1971).

The purpose of the present paper is to take the Luce-Krantz theory in its most abstract form and entertain some interpretations, an interpretation amounting to a possible model for conditional statistical decision theory.

The main body of the paper is organized into sections as follows:

Section I: develops notation and introduces a version of the Luce-Krantz theory in its most general form.

Section II: introduces a notion of "compromise", which is applied to a collection of conditional expected utility structures to yield a composed utility structure, which is not necessarily an expected utility structure.

Section III: examines interpretations of the abstract Luce-Krantz theory in a statistical setting. The pivotal assumption made here is that the choice of an experiment is a component of the decision.

I. A formulation of the Abstract Luce-Krantz Theory

This section presents the abstract Luce-Krantz model for decision making in its most summary form; for further details see Krantz, Luce, Suppes, and Tversky (1971).

Following most models for qualitative probability, the primitives are three: A measurable space $(X,\mathcal{E})$, the elements of $\mathcal{E}$ having the obvious interpretation as events. A sub-family $\mathcal{N}$ of $\mathcal{E}$, those that are a priori perceived as highly unlikely. A set $\mathcal{C}$, which in most interpretations is the set of consequences.

Based on the sets $\mathcal{E}$ and $\mathcal{C}$ are constructed functions, called conditional decisions, as follows: For $A\varepsilon\mathcal{E}$, any function

$$f_A: A \to \mathcal{C}$$

is a decision conditional on A, which to each x in A assigns a consequence, $f_A(x)$, from $\mathcal{C}$, but, for x not in A, remains undefined.

If A and B are disjoint events and if f_A and g_B are two conditional decisions, then $f_A \cup g_B$ is the decision conditional on $A\cup B$ defined by

$$\begin{aligned}(f_A \cup g_B)\ (x) &= f_A\ (x) \quad \text{if } x \varepsilon A \\ &= g_B\ (x) \quad \text{if } x \varepsilon B \ .\end{aligned}$$

If A and B are two events such that $A\subset B$, and if f_A and g_B are two conditional decisions, then f_A is said to be the restriction of g_B to A if

$$f_A\ (x) = g_B\ (x) \quad \text{for } x \varepsilon A \ .$$

If A is an event, conditional decision f_A is said to be conditionally constant if there is a $c \varepsilon \mathcal{C}$ such that

$$f_A\ (x) = c \qquad \text{for } x \varepsilon A \ .$$

A constant decision conditional on A is often denoted c_A.

Given $\mathcal{D}$ is a set of conditional decisions, we assume it is equipped with a weak order (a transitive, connected binary relation) denoted by $\succsim$; $f\succsim g$ will have interpretation, g is not preferred to f. When both $f\succsim g$ and $g\succsim f$, we will say that f and g are equally preferred, and denote this circumstance by $f\sim g$. We wish to calibrate the numerical representation in an intrinsic fashion, using a device called a standard sequence, which may be defined as follows: Take $A\varepsilon\mathcal{E}-\mathcal{N}$ and let N be a sequence of consecutive integers. A set of decisions $\{f^i_A: i\varepsilon N, f^i_A \varepsilon \mathcal{D}\}$ is a standard sequence if for some $B\varepsilon\mathcal{E}-\mathcal{N}$, $A\cap B = \emptyset$, and g^0_B, $g^1_B\varepsilon\mathcal{D}$ with $g^0_B \not\sim g^1_B$, then for all i, i+1 ε N

$$f^i_A \cup g^1_B \sim f^{i+1}_A \cup g^0_B \ .$$

A set of decisions $\{f^t: t\varepsilon T, f^t\varepsilon\mathcal{D}\}$ is said to be strictly bounded, if there exist two decisions g, $h\varepsilon\mathcal{D}$ such that $g\succ f^t\succ h$ for all $t\varepsilon T$. We now have enough notation to give expression to the main definition and representation theorem.

Definition 1

Suppose $\mathcal{E}$ is an algebra of subsets of a set X, $\mathcal{N}$ is a subset of $\mathcal{E}$, $\mathcal{C}$ is a set, $\mathcal{D}$ is a set of functions whose domains are elements of $\mathcal{E}$ and ranges are subsets of $\mathcal{C}$, and $\succsim$ is a binary relation on $\mathcal{D}$. Then $(X,\mathcal{E},\mathcal{N},\mathcal{C},\mathcal{D},\succsim)$ is a conditional decision structure if and only if for all A, $B\varepsilon\mathcal{E}-\mathcal{N}$, R, $S\varepsilon\mathcal{E}$, and all f_A, f^i_A, $f_{A\cup R}$, g_B, g^i_B, h^i_A, k^i_B ε $\mathcal{D}$ where $i\varepsilon N$, the following nine axioms are satisfied.

1. Closure:
   (i) If $A\cap B = \emptyset$, then $f_A\cup g_B$ $\varepsilon\mathcal{D}$.
   (ii) If $B\subset A$, the restriction of f_A to B is in $\mathcal{D}$.
   (iii) For every $c\varepsilon\mathcal{C}$, c_X is in $\mathcal{D}$.
2. Weak Order: $\succsim$ is a weak ordering of $\mathcal{D}$.
3. Union indifference: If $A\cap B = \emptyset$ and $f_A\sim g_B$, then $f_A\cup g_B\sim f_A$.
4. Independence: If $A\cap B = \emptyset$, then $f^1_A\succsim f^2_A$ iff $f^1_A\cup g_B\succsim f^2_A\cup g_B$.
5. Compatibility: If $\{f^i_A: i\varepsilon N\}$ and $\{h^i_B: i\varepsilon N\}$ are two standard sequences such that, for some j, $j+1$ ε N, $f^j_A\sim h^j_B$ and $f^{j+1}_A\sim h^{j+1}_B$, then for all $i\varepsilon N$ $f^i_A\sim h^i_B$.
6. Archimedean: Any strictly bounded standard sequence is finite.
7. Nullity:
   (i) If $R\varepsilon\mathcal{N}$ and $S\varepsilon R$ then $S\varepsilon\mathcal{N}$.
   (ii) $R\varepsilon\mathcal{N}$ iff, for all $f_{A\cup R}\varepsilon\mathcal{D}$ and $A\cap R=\emptyset$ $f_{A\cup R}\sim f_A$, where f_A is the restriction of $f_{A\cup R}$ to A.
8. Non-triviality:
   (i) $\mathcal{E}-\mathcal{N}$ has at least three pairwise disjoint elements.
   (ii) $\mathcal{D}/\sim$ has at least two distinct equivalence classes.
9. Restricted Solvability:
   (i) If A and g_B are given, then there exists $h_A\varepsilon\mathcal{D}$ such that $h_A\sim g_B$.
   (ii) If $A\cap B=\emptyset$ and $h^1_A\cup g_B\succsim f_{A\cup B}\succsim h^2_A\, g_B$, then there exists $h_A\varepsilon D$ such that $f_{A\cup B}\sim h_A\cup g_B$.

Theorem 1

Suppose that $(X,\mathcal{E},\mathcal{N},\mathcal{C},\mathcal{D},\succsim)$ is a conditional decision structure. Then there exist real-valued functions u on $\mathcal{D}$ and P on $\mathcal{E}$ such that $(X,\mathcal{E},P)$ is a finitely additive probability space, and for all A, $B\varepsilon\mathcal{E}-\mathcal{N}$, $R\varepsilon\mathcal{E}$, f_A, $g_B\varepsilon\mathcal{D}$

(i) $R\varepsilon\mathcal{N}$ iff $P(R) = 0$
(ii) $f_A\succsim g_B$ iff $u(f_A) \geq u(g_B)$
(iii) If $A\cap B = \emptyset$, then
$$u(f_A\cup g_B)=u(f_A)\ P(A|A\cup B) + u(g_B)P(B|A\cup B).$$

Moreover, P is unique, and u is unique up to a positive linear transformation. [Note: $(PA|A\cup B)=P(A)/P(A\cup B)$]

Note that the foregoing theorem assigns utility values only to decisions, and not to consequences.

To determine the sense in which an assignment of utilities to consequences seems to happen, recall that for every $c\varepsilon\mathcal{C}$, $c_X\varepsilon\mathcal{D}$. Thus by the closure axiom, for every $c\varepsilon\mathcal{C}$ and every $A\varepsilon\mathcal{E}$, $c_A\varepsilon\mathcal{D}$ but it may nonetheless be true that there are $A,B\varepsilon\mathcal{E}$ such that $c_A\succ c_B$. Now for each $c\varepsilon\mathcal{C}$ define $v(c,A) = u(c_A)\ P(A)$.

By theorem 1 (iii) for $A\cap B=\emptyset$ $v(c,A\cup B)=v(c,A) + v(c,B)$ so that for each $c\varepsilon\mathcal{C}$ $v(c,\cdot)$ is a finitely additive set function on $\mathcal{E}$ such that $v(c,A)=0$ whenever $P(A)=0$. If P were countably additive, **then the usual Radon-Nikodym Theorem would guarantee the existence for each** $c\varepsilon\mathcal{C}$ of a function on X, say $v(c,x)$ such that $v(c,A)=\int A(x)v(c,x)P(dx)$ so that $u(c_A)=(1/[P(A)])\int A(x)v(c,x)P(dx)$ where $A(x)=1$ if $x\varepsilon A$ and $=0$ otherwise.

It is interesting that this integral representation can be extended beyond constant decisions in the usual integration theoretic manner. First, to gamble, where we define a gamble to be a conditional decision f_A whose range is a finite set, and is of the form

$$f_A = \bigcup\{c_{A_i}^{(i)}: \ i=1,\ldots,n\}$$

which means that

$$u(f_A) = \sum_{i=1}^{n} u(c_{A_i}^{(i)})P(A_i| \bigcup_{i=1}^{n} A_i)$$
$$= \sum_{i=1}^{n} v(c^{(i)}, A_i)/ \sum_{i=1}^{n} P(A_i)$$
$$= \sum_{i=1}^{n}\int A_i(x)\ v(c^{(i)},x)P(dx)/ \sum_{i=1}^{n}\int A_i(x)P(dx)$$
$$= \int v(f_A(x),x)\ P(dx|\cup A_i)\,.$$

Thus the utility of a gamble has an integral representation in terms of the functions $v(c,\cdot)$. This representation can be extended to more general decisions under a suitable definition of approximation, to yield for each $f_A\varepsilon\mathcal{D}$ the integral representation $u(f_A) = \int v(f_A(x),x)P(dx|A)$.

II. Combining Opinions

The previous section outlines a reasonably general decision structure for a single decision maker. However, there are numerous occasions when a choice must satisfy more than one decision maker. In this section, we explore one procedure by which the preferences of a group may be amalgamated yielding a composed decision structure. The group of potential participants in the decision process will be represented by a set Ξ, with individuals denoted ξ. At any implementation of a decision process the group of actual participants may be a subset α of Ξ, and the class of such groups form an algebra $\mathcal{A}$. The basic constraints that will be observed are that each element α of $\mathcal{A}$ sees essentially the same decision problem in the sense that the basic measurable space $(X,\mathcal{E})$, the consequence space $\mathcal{C}$, and the decision space $\mathcal{D}$ will be the same for all $\alpha\varepsilon\mathcal{A}$. However, each α will perceive its own set of null events $\mathcal{N}(\alpha)$ and impose a weak order $\succsim(\alpha)$ on $\mathcal{D}$. Thus for each α in $\mathcal{A}$ we define a decision structure to be

$$(X,\mathcal{E},\mathcal{N}(\alpha),\ \mathcal{C},\mathcal{D},\succsim(\alpha))$$

Notwithstanding the notational similarity, we do not assume that this decision structure satisfies the axioms of Definition 1 of section I, but do assume that $\succeq(\alpha)$ is numerically representable by a utility function $u(\ |\alpha)$ for every $\alpha\varepsilon A$. Given that the indexed set of utility functions is available, a natural question is how might one define $u(\ |\alpha\cup\beta)$ so that it represents a "compromise" between α and β and is computable on the basis of $u(\ |\alpha)$ and $u(\ |\beta)$. This we accomplish by introducing the notion of composable: A family $\{u(\ |\alpha):\ \alpha\varepsilon A\}$ of utility functions is said to be composable, if for each $f_A\varepsilon\mathcal{D}$ and whenever α,β in A are such that $\alpha\cap\beta=\emptyset$

$$u(f_A|\alpha\cup\beta) = F[u(f_A|\alpha),\ u(f_A|\beta),\ \pi(\alpha),\ \pi(\beta)]$$

where F is a positive-real-valued function defined on R^4, and may vary with $f_A\varepsilon\mathcal{D}$, π is a positive-real-valued function on A, and may also vary with $f_A\varepsilon\mathcal{D}$, satisfying:

1. $F[u(f_A|\alpha),\ u(f_A|\beta),\ \pi(\alpha),\ \pi(\beta)]$
   $= F[u(f_A|\beta),\ u(f_A|\alpha),\ \pi(\beta),\ \pi(\alpha)].$
2. $F[u(f_A|\alpha),\ F[u(f_A|\beta),\ u(f_A|\gamma),\ \pi(\beta),\pi(\gamma)],\ \pi(\alpha),\ \pi(\beta\cup\gamma)].$
   $= F[F[u(f_A|\alpha),\ u(f_A|\beta),\ \pi(\alpha),\ \pi(\beta)],\ u(f_A|\gamma),\ \pi(\alpha\cup\beta),\ \pi(\gamma)].$
3. If $u(f_A|\alpha) \leq u(f_A|\beta)$ then
   $u(f_A|\alpha) \leq F[u(f_A|\alpha),\ u(f_A|\beta),\ \pi(\alpha),\ \pi(\beta)] \leq u(f_A|\beta).$
4. If $u(f_A|\alpha) < u(f_A|\beta)$ and $\pi(\alpha) = \pi(\beta)$, then
   $F[u(f_A|\alpha),\ u(f_A|\gamma),\ \pi(\alpha),\ \pi(\gamma)] < F[u(f_A|\beta),\ u(f_A|\gamma),\ \pi(\beta),\ \pi(\gamma)].$
5. If $u(f_A|\alpha) = u(f_A|\beta) < u(f_A|\gamma)$ and $\pi(\alpha) < \pi(\beta)$ then
   $F[u(f_A|\alpha),\ u(f_A|\gamma),\ \pi(\alpha),\ \pi(\gamma)] > F[u(f_A|\beta),\ u(f_A|\gamma),\ \pi(\beta),\ \pi(\gamma)].$
6. $F[u(f_A|\alpha),\ u(f_A|\beta),\ \pi(\alpha),\ \pi(\beta)]$
   $= F[u(f_A|\alpha),\ u(f_A|\beta),\ k\pi(\alpha),\ k\pi(\beta)]$ whenever $k>0$.

Note that 4,5, and 6 are a detailing of the "majority principle" embodied in our concept of compromise. Conditions 1-6 almost characterise what are generally called quasi-linear weighted means. If we assume that π is an additive set function, then 1-6 are equivalent to the six conditions in Aczel (1966) that characterise the notion of a quasi-linear mean, where it is also shown that

$$F[u(f_a|\alpha),\ u(f_A|\beta),\ \pi(\alpha),\ \pi(\beta)] = \phi^{-1}\left[\phi[u(f_A|\alpha)]\frac{\pi(\alpha)}{\pi(\alpha)+\pi(\beta)} + \phi[u(f_A|\beta)]\frac{\pi(\beta)}{\pi(\alpha)+\pi(\beta)}\right].$$

with ϕ a positive-real-valued strictly increasing continuous function on R_+, which in accordance with the foregoing may depend on f_A. When necessary, to emphasize the potential dependence, we will often write

$$\phi[u(f_A|\alpha)] \text{ as } \phi[u(f_A|\alpha),\ f_A].$$

This means that, whenever $\alpha\cap\beta=\emptyset$ and for every $f_A\varepsilon\mathcal{D}$.

$$\phi[u(f_A|\alpha\cup\beta),\ f_A]\ \pi(\alpha\cup\beta|f_A) = \phi[u(f_A|\alpha),\ f_A]\ \pi(\alpha|f_A) + \phi[u(f_A|\beta),\ f_A]\ \pi(\beta|f_A)$$

Thus for each $f_A\varepsilon\mathcal{D}$

$$\tau(\ |f_A) = \phi[u(f_A|\ ),\ f_A]\ \pi(\ |f_A)$$

is an additive set function on A such that

$$\tau(\alpha|f_A) = 0 \text{ whenever } \pi(\alpha|f_A) = 0.$$

Thus the notion of a compromise is effected by the choice of two functions $\phi:R_+ \times \mathcal{D} \to R_+$, strictly increasing in the first argument, and $\pi:A \times \mathcal{D} \to R_+$, additive in the first argument.

As indicated earlier the composite utility functions do not lend themselves to the expected utility representation without some additional assumptions.

Each of the functions, for each $\alpha\varepsilon A$

$$u_1^*(f_A|\alpha) = u(f_A|\alpha)$$
$$u_2^*(f_A|\alpha) = \phi[u(f_A|\alpha),\ f_A]$$
$$u_3^*(f_A|\alpha) = \phi[u(f_A|\alpha),\ f_A]\ \pi(\alpha|f_A)$$

define weak order on $\mathcal{D}$, denoted $\succeq(\alpha,1)$, $\succeq(\alpha,2)$, $\succeq(\alpha,3)$ respectively. Thus if we assume that for each $\alpha\varepsilon A$, the weak order $\succeq(\alpha,j)$ (j=1,2 or 3) is a conditional decision structure in the sense of Section I, we get, for each $\alpha\varepsilon A$, positive real valued functions $u_j^*(\ |\alpha)$ and $P_j(\ |\alpha)$ on $\mathcal{D}$ and E, such that $(X,E,P_j(\ |\alpha))$ is a finitely additive probability space, and for every $A,B\varepsilon\ E\text{-}N(\alpha)$, $R\varepsilon E$, $f_A,g_B\varepsilon\mathcal{D}$

(i) $R\varepsilon N(\alpha)$ iff $P_j(R|\alpha) = 0$

(ii) $f_A\succeq(\alpha,j)g_B$ iff $u_j^*(f_A|\alpha) \geq u_j^*(g_B|\alpha)$

(iii) If $A\cap B = \emptyset$ then

$$u_j^*(f_A\cup g_B|\alpha) = u_j^*(f_A|\alpha)\ P_j(A|A\cup B,\alpha) + u_j^*(g_B|\alpha)\ P_j(B|A\cup B,\alpha).$$

The case j=3 is the most general, since j=2 is a special case under the assumption $\pi(\alpha|f_A) \equiv \pi(\alpha)$, which causes the weak-orders $\succeq(\alpha,3)$ and $\succeq(\alpha,2)$ to become indistinguishable. The case j=1 is a further special case under the additional assumption that ϕ is independent of its second argument, causing the three weak-orders to become indistinguishable. In what follows we shall generally work with the case j=2 (i.e., π independent of f_A).

III. Some Interpretations

In this section we transcribe sections I and II into more familiar statistical terms: Denote the space of "ideal" observations by the measurable space (Y,F), the parameter space by (Θ,B), the space of "estimates" by (K,H) and, recognizing the possible occurrence of "errors" in observation, the space of observations with error by $(\mathcal{Z},\mathcal{G})$. For any measurable space (X,A), denote the convex set of finitely additive probabilities on A by $M(X,A)$.

A conditional experiment or randomization over Y is defined as a map, P_A, with domain A, an element of $\mathcal{B}$, and range a subset of $M(Y,\mathcal{F})$, subject to the proviso that for each $F\varepsilon\mathcal{F}$, the real-valued map, $P_A(F)(\cdot)$, is measurable in θ in the usual sense (i.e., is weakly measurable). An I-estimator is an $(\mathcal{F},\mathcal{H})$ - measurable map on Y into K, an E-estimator is a $(G,\mathcal{H})$ - measurable map on $\mathcal{Z}$ into K.

In an error-free setting, a typical decision is an ordered pair (P,f), P a randomization over Y, f an I-estimator. Associated with each (P,f) is a randomization Q(=fP) over K, the image of P under f, and for most practical purposes (f,P) may be confounded with Q. Thus, recalling section I, identify $(X,\mathcal{E},\mathcal{N})$ with $(\Theta,\mathcal{B},\mathcal{N}^*)$, $\mathcal{N}^*\subset\mathcal{B}$, $\mathcal{C}$ with $M(K,\mathcal{H})$, $\mathcal{D}$ with a set of conditional randomizations over K, and assume

$$(\Theta,\ \mathcal{B},\ \mathcal{N}^*,\ M(K,\ \mathcal{H}),\ \mathcal{D},\succsim)$$

is a conditional decision structure in the sense of section I. Then following the reasoning contained there, under suitable regularity conditions, we are assured of the existence of a finitely additive probability μ on $\mathcal{B}$, and a positive-real-valued map v on $\mathcal{D}\times\Theta$ such that

$$u(Q_A) = \int v(Q_A(\theta),\theta)\ \mu(d\theta|A).$$

Since, in general, for each $\theta\varepsilon\Theta$, the map $v(Q_A(\theta),\theta)$ on $M(K,\mathcal{H})$ is non-linear, u is a non-linear functional over randomizations. Some conditions under which linearity obtains are discussed in Fishburn (1970), and lead to the following simplification

$$\begin{aligned} u(Q_A) &= \int v(Q_A(\theta),\theta)\ \mu(d\theta\ |A) \\ &= \int w(x,\theta)\ Q_A(dx|\theta)\ \mu(d\theta|A) \\ &= \int w(x,\theta)\ \tau(dx,d\theta|K\times A) = u^*(\tau) \end{aligned}$$

which is a linear functional on $M(K\times\Theta,\ \mathcal{H}\times\mathcal{B})$. Thus an optimal solution need only be sought among the extreme points of the convex set of probabilities satisfying the following constraints on margins and disintegrations:

$$\tau(K\times B\ |K\times A) = \mu(B|A),\text{ and}$$

$$\tau(dx,d\theta|K\times A) = Q_A(dx|\theta)\ \mu(d\theta|A)$$

where, according to Fishburn (1970), the available randomizations, Q_A, should be convex set. If we further assume that Q_A is derived from a single predetermined, P_A, then the choice reduces to that of an I-estimator, and is accomplished by the usual Bayes inversion:

$$\begin{aligned} u(Q_A) &= \int w(x,\theta),\ Q_A(dx|\theta)\ \mu(d\theta|A) \\ &= \int w(f(y),\theta)\ P_A(d_y|\theta)\ \mu(d\theta|A) \\ &= \int w(f(y),\theta)\ \mu(d\theta|y,A)\ P(dy)\ . \end{aligned}$$

Now turning to the more realistic setting, where we assume that observations are not error-free, and so are elements of $\mathcal{Z}$. In this circumstance the decision process can proceed in one of two ways. The first alternative would be to choose a strategy as if no "error" were expected, and upon receipt of the actual outcome z (i.e., signal plus noise!), guess at which y should have obtained (i.e., guess at the signal!) and then apply the chosen I-estimator f to the guessed y.

This procedure is analogous to what has come to be known as "imputation". The second alternative is more direct in defining a decision to be an ordered pair (P,g), where P, as before, is a randomization over Y, and g is an E-estimator. Thus (P,g) is a map with domain $A\times\mathcal{Z}$, $A\varepsilon\mathcal{B}$, and range in $M(Y,\mathcal{F})\times K$. Identify $(X,\mathcal{E},\mathcal{N})$, with $(\Theta\times\mathcal{Z},\mathcal{B}\otimes\mathcal{G},\mathcal{N}^*)$, $\mathcal{N}^*\subset\mathcal{B}\otimes\mathcal{G}$, $\mathcal{C}$ with $M(Y,\mathcal{F})\times K$, $\mathcal{D}$ with a set of ordered pairs as defined above and assume

$$(\Theta\times\mathcal{Z},\ \mathcal{B}\otimes\mathcal{G},\mathcal{N}^*,\ M(Y,\mathcal{J})\times K,\ \mathcal{D},\succsim)$$

is a conditional decision structure in the sense of section I. As before, under suitable regularity conditions, we are assured of a finitely additive probability μ on $\mathcal{B}\otimes\mathcal{G}$, and a positive-real-valued map v on $\mathcal{D}\times K\times\Theta\times\mathcal{Z}$, such that

$$\begin{aligned} u(P_A,g) &= \int v(P_A(\theta),g(z),\theta,z)\ \mu(d\theta,dz|A\times\mathcal{Z}) \\ &= \int v(P_A(\theta),g(z),\theta,z)\ \nu(dz|\theta)\mu(d\theta|A) \end{aligned}$$

Thus, for each θ, we have two probability functions, $P(|\theta)$ on $(Y,\mathcal{F})$, and $\nu(|\theta)$ on $(\mathcal{Z},\mathcal{G})$. P is well-defined as one of the available randomizations, but what is ν? In one sense, ν represents the "irreducible" noise in the "decision system", which can be decomposed according to the randomization P chosen. Also P and ν are aspects of the partially controllable relationship between "ideal" and "real" observations. A particularly simple example of this connection is that of being the marginal randomizations of a joint randomization over $Y\times\mathcal{Z}$. The "design" of an experiment P would then amount to the choice of a randomization over $Y\times\mathcal{Z}$ from among those having marginal randomization over $\mathcal{Z}$ fixed at ν, and marginal randomization over Y belonging to the set of available randomizations over Y. Under appropriate regularity assumptions, this choice can be exercised through the selection of a disintegration $P(dy|\theta,z)$ of $P(dy|\theta)$ with respect to $\nu(dz|\theta)$, so that

$$P(dy|\theta,z)\ \nu(dz|\theta) = \nu(dz|y,\theta)\ P(dy|\theta)$$

where $\nu(dz|y,\theta)$ represents the "noise" in the "decision system" under experimental conditions P.

Under certain circumstances, z may be perceived as being approximately sufficient for θ in the sense that $P(dy|\theta,z)\simeq P(dy|z)$, under an appropriate concept of "$\simeq$". Thus any "guess" of y, using $P(dy|\theta,z)$ would depend primarily on z and very little on θ. Thus the choice of $P(dy|\theta)$ and an E-estimator g can be viewed as interchangeable with a choice of $P(dy|z)$ and an I-estimator f, with the foregoing outlining the circumstances under which they are risk-equivalent. A particularly simple example, assuming all symbols well-defined, is given by the equation

$$\int f(y)\ P(dy|z) = g(z)$$

This is also a particularly simple example of possible relationships between I- and E- estimators. In general then, the foregoing interpretations of the general decision model confirms the intuitive view of imputation as a device that provides a link between I- and E- estimation, and go somewhat further in indicating when imputation is a valid tactic contributing to the implementation of an estimation strategy.

The foregoing formulation of the abstract model seems to encompass both the general statistic model as well as a general (usual?) model for finite population surveys. The remainder of this section is devoted to examining the latter assertion.

Denote the finite population of identifiable units by

$$u = \{1,2,\ldots,N\}.$$

For simplicity, assume each unit in u carries an $\mathcal{R}$-valued characteristic of interest, so the population of characteristics can be represented as an array

$$\theta = (\theta_1, \theta_2,\ldots,\theta_N) \quad \text{in } \mathcal{R}^N.$$

Even though, in principle, a sample is any subset of u, we will adopt a somewhat more convention-bound representation, by representing a sample as a non-empty "ordered" subset of θ in the following sense.

$$s = (i_1,i_2,\ldots,i_n) \quad , i_1 < i_2 <\ldots< i_n$$

with the set of all samples denoted by $\mathcal{S}$. It is easily seen that there is a one-to-one correspondence between the set of "unordered" samples and the set of "ordered" samples. Each s in $\mathcal{S}$ can be viewed as a "projection" from $\mathcal{R}^N$ onto $R^{n(s)}$, where n(s) is the size of the sample s. The ordered set representation preserves identifiability by specifying exactly which subspace of $\mathcal{R}^N$ the projection is onto. The range $s(\mathcal{R}^N)$ can be symbolically represented $\{s\}\times\mathcal{R}^N$, which has typical element (s,θ). We can now define the ideal-observation space as

$$Y = \cup\{ \{s\} \times \mathcal{R}^N: \; s\varepsilon\mathcal{S} \}$$

so that each $s\varepsilon\mathcal{S}$ is a map on $\mathcal{R}^N$ into Y. It is particularly useful to examine the dual view, namely that each $\theta\varepsilon\mathcal{R}^N$ is a map on $\mathcal{S}$ unto Y. Recalling that a sample design is a probability on the discrete space $\mathcal{S}$, usually denoted p, each $\theta\varepsilon\mathcal{R}^N$ carried p onto a discrete probability on Y, and the correspondence

$$\theta \rightarrow \theta(p) = P(\theta)$$

defines a map on $\mathcal{R}^N$ into $M(Y,\mathcal{F})$, whose range is a subset of discrete probabilities. So that P is nothing more than a randomization in the sense defined above. In this ideal setting, the usual finite population estimator

$$e : Y \rightarrow K$$

corresponds precisely to the notion of an I-estimator, which carries the randomization P over Y onto a randomization Q over K. In the present context, Q is no more than the trace of p, under the composite map $e\cdot\theta$ as θ varies over $\mathcal{R}^N$. The correspondence with the error-free general statistical model is now obvious and

$$u(Q) = \int v(Q(\theta),\theta)\, \mu(d\theta)$$
$$= \int v(e(\theta(p)),\theta)\, \mu(d\theta)$$
$$= \int v^*(p,e,\theta)\, \mu(d\theta) = u^*(p,e)$$

where v* is, in general non-linear in p. (The linear case has already seen substantial attention in the literature.)

To accommodate the more realistic observations with error model requires a real-observation space which we define as

$$\mathcal{Z} = \cup\{ \{s\} \times \mathcal{Q}^N : s\varepsilon\mathcal{S} \}$$

where $\mathcal{Q} = \mathcal{R}\cup\mathcal{M}$, $\mathcal{M}$ a set of special symbols to accommodate unanticipated "observations" (e.g. variety of non-response, etc.) A typical element of $\mathcal{Z}$ will be denoted (s,θ^*). Now each $\theta^*\varepsilon\mathcal{Q}^N$ can be viewed as a map on $\mathcal{S}$ into $\mathcal{Z}$, carrying the design p onto a discrete probability on $\mathcal{Z}$, giving rise to randomization P* over $\mathcal{Z}$. Since the actual observation is an element of $\mathcal{Z}$, the appropriate type of estimator is an E-estimator

$$d: \; \mathcal{Z} \rightarrow K$$

which, in turn carries P* onto a randomization Q* over K. Thus a decision is an ordered pair (P,Q*), which is a map on $\mathcal{R}^N\times\mathcal{Q}^N$ into $M(Y,\mathcal{F}) \times M(K,\mathcal{H})$, and employing the now familiar reasoning based on the results of section I, with the usual caveats, we get

$$u(P,Q^*) = \int v(P(\theta),Q^*(\theta^*),\, \theta,\, \theta^*)\, \mu(d\theta,d\theta^*)$$
$$= \int v^*(p,d,\theta,\theta^*)\, \mu(d\theta,d\theta^*)$$
$$= \quad u^*(p,d)$$

where, again, v* is not necessarily linear in p. The linear case has seen some sporadic discussion in the literature, usually under the guise of imputation strategies.

In general, the finite population survey model has more structure than the general statistical model (since it is a special case!), and this can be taken advantage of in the observations with error case to yield a result which looks different (and simpler?) than that obtained in the general case.

The situation for more than one decision maker is not any more complicated, and each of the foregoing interpretations can be easily cast thus. Here we content ourselves with casting the last of our interpretations thus:

Recalling the notation developed in Section II, each $\xi\varepsilon\Xi$ is a decision maker, and we assume, for each ξ

$$u(P,Q^*|\xi) = \int v^*(p,d,\theta,\theta^*|\xi)\, \mu(d\theta,d\theta^*|\xi)$$
$$= u^*(p,d|\xi)$$

i.e.,each ξ has a distinctly personal utility and prior. Further we suppose the compromise must be effected for a "committee" $\alpha\varepsilon\Xi$, and following a very specialized reasoning based on Section II, define

$$\phi[u(P,Q^*|\alpha)]$$
$$= \int \phi[v^*(p,d,\theta,\theta^*|\xi)]\, \mu(d\theta,d\theta^*|\xi)\, \tau(d\xi|\alpha)$$
$$= \int \phi[v^*(p,d,\theta,\theta^*|\alpha)]\, \mu(d\theta,d\theta^*|\alpha)$$
$$= \quad \phi[u^*(p,d|\alpha)]$$

where φ is a positive-real-valued strictly increasing continuous function on the positive reals and τ is a finitely additive probability on $(\Xi,\mathcal{A})$. Thus $v^*(\;|\alpha)$ and $\mu(\;|\alpha)$ represent the compromise utility and prior respectively, for the "committee" α. Now, even if $v^*(\;|\xi)$ is linear in p for each $\xi\varepsilon\alpha$, $v^*(\;|\alpha)$, and thus $u^*(\;|\alpha)$, will, in general, be fairly spectacularly non-linear in p. The last few results can be viewed as abstract analogues of the strategic considerations discussed in Patrick (1973).

Conclusions

The foregoing account suggests a number of things:

First, that imputation, as it is commonly understood, is not more general or robust than conventional estimation. And, under appropriate circumstances, imputation can be made to serve as a useful implementational device linking the presumably more complicated observations-with-error situations with the presumably more simple observations-without-error situations. Second, when design is a part of the decision strategy, life can get very complicated because of inherent nonlinearities, which, in the context of finite population survey models, means that purposive sampling should be the exception rather than the rule. Does this last mean randomization has a Bayes justification? Yes, with the qualification that the Fisherian averaging-out-view does not seem to have found a Bayes expression!

REFERENCES

1. Aczel, J. (1966) Lectures on Functional Equations and Their Applications, Academic Press, New York.
2. de Finetti, B. (1937) La prévision: ses lois logiques, ses sources subjectives. Translated by H.E. Kyburg in Kyburg and Smokler (1964).
3. Fishburn, P.C. (1964) Decision and Value Theory, Wiley, New York.
4. Fishburn, P.C. (1970) Utility Theory for Decision Making, Wiley, New York.
5. Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971) Foundations of Measurement, Vol. I., Academic Press, New York.
6. Kyburg Jr., H.E., and Smokler, H.E., (Eds) (1964) Studies in Subjective Probability, Wiley, New York.
7. Luce, R.D., and Krantz, D.H. (1971) Conditional Expected Utility, Econometrica, 39, pp. 253-271.
8. Patrick, C.A. (1973) A Bayesian Model for General Survey Situations - presented at the 1st session of the I.A.S.A. in Vienna, Austria.
9. Pfanzagl, J. (1967) Subjective probability derived from the Morgenstern-von Newmann utility concept. In M. Shubik (Ed.), Essays in mathematical economics in honor of Oskar Morgenstern, Princeton University Press, Princeton, New Jersey.
10. Pfanzagl, J. (1968) Theory of Measurement, Wiley, New York.
11. Ramsey, F.P. (1931) Truth and Probability - reprinted in Kyburg and Smokler (1964).
12. Savage, L.J. (1954, 1972) Foundation of Statistics, Dover, New York.
13. von Neumann, J., and Morgenstern, O. (1947) Theory of games and economic behavior, 2nd edition, Princeton University Press, Princeton, New Jersey.