Discussion of
"A Model of Scientific Communication"

by Isaiah Andrews and Jesse M. Shapiro
February (2020)

James J. Heckman

THE UNIVERSITY OF CHICAGO

NBER Summer Institute – Labor Studies
July 22nd, 2020

Table of Contents

“Knowledge is useful if it helps to make the best decisions.”

- Jacob Marschak (1953)

“Economic Measurements for Policy and Predictions.”
Cowles Monograph 13.

- “Statistical decision theory, following Wald (1950), is the dominant theory of **optimal** estimation in econometrics.”

- Many economists interested in describing what has been found, not making any specific decision: inference vs. decision (Birnbaum Likelihood Principle, 1962).

- Discovery of unknown phenomena: not optimal estimation of a parameter of a distribution describing the data or useful in some decision.

Return to Table of Contents

Useful Backdrop:
Bayesian Statisticians and Remote Clients

$$f(x,\omega) = g(x|\omega)m(\omega) = h(\omega|x)n(x) \qquad (1)$$

- *x* is an observation (usually multi-dimensional) from one of a family of populations denoted by $g(x|\omega), \omega \in \Omega, x \in X$
- $g(x|\omega)$ is, for fixed ω, a density function for *x*

- ω is a set of parameters characterizing the distribution of x, sometimes called an "aspect of the world";
- $m(\omega)$ is a client's prior distribution for ω.
- This distribution is determined temporally prior to the client's decision but not necessarily prior to statistical analysis.

- The responsibility for choosing $\mathrm{m}(\omega)$ rests with the client though he may choose a function that the statistician or some other agent formulates or helps formulate.

- $h(\omega|x)$ is the conditional density for ω given x;

- $n(x)$ is the marginal density for *x*;

- $f(x, \omega)$ is the marginal density for *x, ω*;

$$u = \bar{u}(z) \qquad (2)$$

- Where *z* is a substantive outcome of a decision-making process (e.g., net revenue, total profit and loss statement and balance sheet, votes, mortality experience);
- *u* is a client's realized utility corresponding to outcome z;
- $k(z|a, \omega, \lambda)$ is a density for outcomes;

- a is a decision or action taken by client;
- λ represents factors beyond the control of the client that affect outcome but not observation-another "aspect of the world";
- $r(\lambda)$ is the client's subjective distribution of λ; and
- X, Ω, A, Z and Λ are sets of possible values of x, ω, a, z and λ.

- Let x_0 be a particular value of *x* and let $[u|a, x_0]$ denote the client's expected utility corresponding to act a after x_0 has bee observed, viz.,

$$[a \mid a, x\_0] = \int \Lambda \int \Omega \int z \, u\tilde{}(z) k(z \mid a, \omega, \lambda) h(\omega \mid x\_0) r(\lambda) \;\; dz d\omega d\lambda \quad (3)$$

- Define

$$[u\^{} * \mid A, x\_0] = \underset{a \in A}{\max} [u|a, x\_0] \quad (4)$$

- Let x_0 be a particular value of *x* and let $[u|a, x_0]$ denote the client's expected utility corresponding to act a after x_0 has bee observed, viz.,

- $a_{x_0}^*$ will denote an act which maximizes the integral in equation (3).

- $a_{x_0}^*$ and $[u^*|A, x_0]$ will be said to be associated with a *complete* Bayesian solution to the indicated decision problem.

Parcels of Information

A. The data, x_0.

B. The full likelihood function based on the observations, $g(x_0|\omega)$or on a sufficient statistic, $g(y_0|\omega)$ where $y_0 = \tilde{y}(x_0)$ and $\tilde{y}$ is a sufficient statistic.

C. A partial likelihood function, $\ddot{g}(\omega_0|\omega)$ where $\omega_0 = \widetilde{\omega}(x_0)$ and $\widetilde{\omega}$ is a nonsufficient statistic.

D. A distribution free likelihood function.

E. Posterior distributions for representative prior distributions.

F. Solutions to representative decision problems.

G. Contour maps or systems of snug regions in the parameter space.

Return to Table of Contents

This Paper

- Observe data $\mathcal{X} \in \mathcal{X}$ for $\mathcal{X}$: a finite sample space, $|\mathcal{X}| < \infty$.

- Distribution of X governed by a parameter $\theta \in \Theta$, with $X|\theta \sim F_\theta$, for Θ a finite parameter space.

- F_θ has support equal to $\mathcal{X}$ for all $\theta \in \Theta$.

- The analyst publicly commits to rule $c\colon \mathcal{X} \times [0,1] \to \Delta(\mathcal{D})$.

- Maps from realizations of the data X and the public random variable V into a distribution over decisions $d \in \mathcal{D}$, for $\mathcal{D}$ a finite space.

- Let $\mathcal{C}$ denote the set of all such rules.

- $c(X, V) \in \mathcal{D}$ denotes the random realization from a given rule $c \in \mathcal{C}$.

- Rules are evaluated by their performance with respect to a closed set $\mathcal{A} \subseteq \Delta(\Theta)$ of priors on the parameter space: ***audience***.

- The ordering of rules under the risk function $\rho(\cdot,\cdot)$ may depend on the prior $a \in \mathcal{A}$.

Definition 1. *For a given risk function $\rho(\cdot,\cdot)$ and audience $\boldsymbol{\mathcal{A}}$, a rule $c^* \in \mathcal{C}$ is*

- **admissible** *if there exists no rule $c \in \mathcal{C}$ such that $\rho(c,a) \leq \rho(c^*,a)$ for all, with strict inequality for at least one $a \in \boldsymbol{\mathcal{A}}$.*

- ω**-optimal** *if for a probability measure ω with support equal to* $\boldsymbol{\mathcal{A}}$.

$$\int_{\mathcal{A}} \rho(c^*,a)d\omega(a) = \inf_{c \in \mathcal{C}} \int_{\mathcal{A}} \rho(c,a)d\omega(a). \qquad (1)$$

- ***minimax*** *if*

$$\sup_{a \in \mathcal{A}} \rho(c^*,a) = \inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} \rho(c,a).$$

Definition 2. *Fix a loss function* $L: \mathcal{D} \times \Theta \to \mathbb{R}_{\geq 0}$*. Then:*

- *The* ***communication model*** *takes* $\mathcal{A} = \Delta(\Theta)$

$$\rho(c, a) = R_a^*(c) = E_a\left[\min_d E_a[L(d, \theta)|c(X, V), V]\right],$$

- $R_a^*(c)$ *the* ***communication risk*** *of rule* $c \in \mathcal{C}$ *for prior* $a \in \Delta(\Theta)$ *and* E_a *the expectation under* a*.*

- *The **decision model** takes* $\mathcal{A} = \Delta(\Theta)$ *and*

$$\rho(c, a) = R_a(c) = E_a[L(c(X, V), \theta)]$$

for $R_a(c)$ *the **decision risk** of rule* $c \in \mathcal{C}$ *for prior* $a \in \Delta(\Theta)$.

- *The classical model takes $\mathcal{A}$ to be the vertices of $\Delta(\Theta)$, such that each prior $a \in \mathcal{A}$ places probability 1 on some $\theta(a) \in \Theta$, and takes*

$$\rho(c, a) = R_{\theta(a)}(c) = E_{\theta(a)}[L(c(X, V), \theta(a))]$$

*for $R_{\theta(a)}(c)$ the **frequentist risk** of rule $c \in \mathcal{C}$ for the parameter $\theta(a) \in \Theta$.*

- **<u>Definition 4</u>.** *Let $\mathcal{P}$ be the set of partitions of $\mathcal{X}$, with generic element $P \in \mathcal{P}$. Let $\mathcal{P}^*$ denote the subset of $\mathcal{P}$ such that for every cell $\mathcal{X}_p \in P \in \mathcal{P}^*$, each agent has at least one decision $d \in \mathcal{D}$ that is optimal for every $X \in \mathcal{X}_p$. That is,*

$$\mathcal{P}^* = \left\{P \in \mathcal{P}: \left\{\cap_{x \in \mathcal{X}_p} \underset{d \in \mathcal{D}}{\arg\min} E_a[L(d,\theta)|X]\right\} \neq \emptyset \; for \; all \; \mathcal{X}_p \in P, a \in \Delta(\Theta)\right\}.$$

- *The **effective size of the sample space $\mathcal{X}$**, denoted $N(\mathcal{X})$, is the minimal size of a partition in $\mathcal{P}^*$.*

Corollary 1.

If (i) there exists a decision $d \in \mathcal{D}$ *that is dominated in loss, and*

(ii) $\mathcal{N}(\mathcal{X}) \geq |\mathcal{D}|$*,*

then any rule $c \in \mathcal{C}$ *that is admissible under the classical model is inadmissible under the communication model, and vice versa.*

Corollary 4.

If (i) there exists a decision $d \in \mathcal{D}$ *that is dominated in loss, and*

(ii) $\mathcal{N}(\mathcal{X}) \geq |\mathcal{D}|$*; then any rule* $c^* \in \mathcal{C}$ *that is minimax and admissible under the decision (or classical) model is minimax and inadmissible under the communication model.*

[Return to Table of Contents](#)