THE TRUTH

JAN DE LEEUW

1. Framework

1.1. **Data.** The *data* is a sequence $z = \{z_n\}$ of *observations*, where $z_n \in \mathbb{R}^q$.

Although actual data are always finite sequences, we assume the actual data are actually the first members of an infinite sequence of vectors z. The remaining elements of the sequence are missing data.

1.2. **Statistics.** From the data z we compute an infinite sequence of *statistics* $x = \{x_n\}$, where $x_n \in \mathbb{R}^p$. We do not specify how the statistics are computed.

In most of the examples we have in mind the statistic x_n depends on the first n observations only. For instance, x_n could be the mean of the first n observations, or their covariance matrix, or their histogram. If the z_n are binary indicators, then the means x_n are vectors of proportions.

1.3. **Framework.** The actual data, a finite sequence of vectors, is first imbedded An *observation* $\underline{x} = \{\underline{x}_n\}$ is a sequence of random vectors varying in an open set $\mathcal{S} \subseteq \mathbb{R}^p$. Our observations can be sequences of means, proportions, covariance matrices, and so on. If the z_n are binary indicators, then the means x_n are vectors of proportions.

The basic assumption in this paper is that there is a $\mu \in \mathcal{S}$ such that

$$\lim_{n\rightarrow\infty} \mathbf{E}(\underline{x}_n) = \mu. \tag{1}$$

We call μ the *truth*. Thus our random vectors observe the truth. In general, we do not know the truth, because if we did there would be no reason to observe it[1].

There are many interesting methodological problems with modeling observations as random variables, let alone sequences of random variables, but we ignore these problems in this paper. There are other interesting philosophical problems dealing with the notion of truth. We ignore those as well.

We do assume there is a positive semi-definite matrix S such that

$$\lim_{n\rightarrow\infty} n\mathbf{V}(\underline{x}_n) = \Sigma. \tag{2}$$

We call Σ the *precision*.

In general, we do not know the precision either. In some cases, however, we are able to observe the precision, in the sense that we have a sequence of random matrices $\underline{S} = \{\underline{S}_n\}$, which may or may not depend on $\underline{x}$, such that

$$\lim_{n\rightarrow\infty} \mathbf{E}(\underline{S}_n) = \Sigma. \tag{3}$$

We use $\mathcal{X}(\mu, \Sigma)$ for the set of observations with a given truth μ and a given precision Σ. Two observations $\underline{x}$ and $\underline{z}$ in $\mathcal{X}(\mu, \Sigma)$ are said to be *orthogonal* if

$$n\mathbf{C}(\underline{x}_n, \underline{z}_n) \rightarrow 0. \tag{4}$$

2. Discrepancy

In this paper we want to quantify how far we are from the truth, and we want to study ways of getting nearer to the truth[2].

---

[1] Some of us already know the truth, and do not need observations.

[2] Always keeping in mind what the band of the Titanic played because the ship was going down.

In order to do this, we define a *discrepancy*, which is a function Δ mapping $\mathcal{S} \times \mathcal{S}$ into the non-negative reals. For all $x \in \mathcal{S}$ and $y \in \mathcal{S}$ we have

$$\Delta(x, y) \geq 0, \tag{5a}$$

$$\Delta(x, x) = 0. \tag{5b}$$

We do not assume symmetry or the triangle inequality, because they are not relevant for our purposes.

We choose our discrepancy to be *smooth*, i.e. to be continuously differentiable enough times for the following results to be true. The details are left to others.

Define the partials

$$g(x, y) = \partial_1 \Delta(x, y), \tag{6a}$$

$$h(x, y) = \partial_2 \Delta(x, y). \tag{6b}$$

and

$$A(x, y) = \partial_{11} \Delta(x, y), \tag{7a}$$

$$B(x, y) = \partial_{12} \Delta(x, y), \tag{7b}$$

$$C(x, y) = \partial_{21} \Delta(x, y), \tag{7c}$$

$$D(x, y) = \partial_{22} \Delta(x, y). \tag{7d}$$

For all x and y in $\mathcal{S}$ the matrices $A(x, y)$ are $D(x, y)$ are symmetric, and $B(x, y)$ is the transpose of $C(x, y)$.

Lemma 2.1. *For all $x \in \mathcal{S}$*

$$g(x, x) = h(x, x) = 0, \tag{8a}$$

$$A(x, x) = -B(x, x) = -C(x, x) = D(x, x). \tag{8b}$$

Proof. Consider $\Delta(x, y)$ as a function of y for fixed x. The function has a minimum at $y = x$, where the partials must vanish. Thus $g(x, x) = 0$. In the same way $h(x, x) = 0$. Differentiating these relations with respect to x we find $A(x, x) + B(x, x) = 0$ and $C(x, x) + D(x, x) = 0$. But $A(x, x)$ and $D(x, x)$ are symmetric, and thus so are $B(x, x)$ and $C(x, x)$. Since

$B(x,x)$ must be the transpose of $C(x,x)$, the two are actually equal, and thus $A(x,x)$ and $D(x,x)$ are equal as well. $\square$

We can use our discrepancy to define another discrepancy $\mathcal{D}$, this time between elements of $\mathcal{X}(\mu,\Sigma)$. Define

$$\mathcal{D}(\underline{x},\underline{y}) = \lim_{n\rightarrow\infty} n\mathbf{E}(\Delta(\underline{x}_n,\underline{y}_n)). \tag{9}$$

With a slight abuse of notation we use μ for the sequence of random variables with all its elements a.s. equal to μ.

Theorem 2.2. *If $\underline{x}$ is in $\mathcal{X}(\mu,\Sigma)$ then*

$$\mathcal{D}(\underline{x},\mu) = \mathcal{D}(\mu,\underline{x}) = \frac{1}{2}\mathbf{tr}\ A(\mu,\mu)\Sigma. \tag{10}$$

If $\underline{x}$ and $\underline{y}$ in $\mathcal{X}(\mu,\Sigma)$ are orthogonal then

$$\mathcal{D}(\underline{x},\underline{y}) = \mathcal{D}(\underline{y},\underline{x}) = \mathbf{tr}\ A(\mu,\mu)\Sigma. \tag{11}$$

Proof. $\square$

3. Statistics

We know that observation $\underline{x}$ observes the truth. But in statistics we do not observe, we *estimate*.

We do not really have any $\underline{x}$, or even a single $\underline{x}_n$. We only have the data $x = \{x_n\}$, a sequence of elements of $\mathcal{S}$, in no sense consisting of random variables.

We link the data to the observations by saying that x_n is a *realization* of $\underline{x}_n$ for all n, or, equivalently, that x is a realization of $\underline{x}$. We invariably get into trouble if we try to tell you what this means, so instead I'll just use the symbol $x \sim \underline{x}$ or $x_n \sim \underline{x}_n$ for it.

In fact, often the data consist of a single *statistic*, i.e. a single element of $\mathcal{S}$, i.e. a single cross table or covariance matrix, and we use a trick to make it into a sequence. We define x_n as the value of our statistic based on the first n observations.

$$n\,\Delta(x_n, \mu) \sim \frac{1}{2}A(x_n, x_n)\,\Sigma_n \sim \frac{1}{2}n\,\Delta(x_n, y_n)$$

4. Estimation

Suppose F is a function mapping $\mathcal{S}$ into $\mathcal{S}$. We suppose that F is smooth. Define the matrix

$$G(x) = \partial F(x)$$

F preserves the truth if $F(\mu) = \mu$. This implies

$$\lim_{n\to\infty} \mathbf{E}(F(\underline{x}_n)) = F(\mu) = \mu$$

and thus $\{F(\underline{x}_n)\}$ observes the truth.

Since we do not know the truth, this does not seem to be a useful concept. But suppose we know that

$$\mu \in \Omega \subseteq \mathcal{S},$$

with Ω a smooth manifold. Define $F(x)$ to be the metric projection of x on Ω, i.e.

$$F(x) = \mathop{\mathbf{argmin}}_{y\in\Omega} \Delta(x, y).$$

If $\mu \in \Omega$ then obviously $F(\mu) = \mu$.

Ω is a *model*. If $\mu \in \Omega$ then the truth is in the model, in other words the model is true.

Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu