Introduction to EPI
(Extreme Physical Information)

Presented in the Embryo Physics Course
http://www.embryophysics.org
October 20, 2010
By

B.Roy Frieden, Optical Science, Univ. of Arizona, Tucson, Arizona 85721
email: roy.frieden@optics.arizona.edu

INTRODUCTION TO EPI

EPI = *Extreme physical information* is a systematic approach to physics and, more generally, all unknown statistical systems. Its aim is to find:

(i) the system's *amplitude functions and associated PDFs*; or

(ii) its deterministic *law of data formation*. The working premise is that, in general, measurements are inaccurate. Thus, EPI builds on the Cramer-Rao inequality.

CRAMER-RAO INEQUALITY (DERIVATION), SFI pp. 29-30)

(Genl. Refs: Van Trees, "Detection, Estimation & Modulation" or Frieden "SFI")

(Note: Refs. "SFI" are "Science from Fisher Information" book)

Simple Derivation, with truly Sweeping Consequences

Only assumption: **Given data** $\mathbf{y} \equiv \mathbf{y}_1, \ldots, \mathbf{y}_N$ **about an unknown parameter value** θ, **consider class of** unbiased **estimator functions** $\hat{\theta}(\mathbf{y})$,

$$< \hat{\theta}(\mathbf{y}) > \equiv \theta.$$

Averaging is w.r.t. the 'likelihood PDF' $P(\mathbf{y}|\theta)$. **Then unbiasedness means**

$$\int d\mathbf{y} \; \hat{\theta}(\mathbf{y}) P(\mathbf{y}|\theta) \equiv \theta.$$

Equivalently

$$\int d\mathbf{y} \left[\hat{\theta}(\mathbf{y}) - \theta \right] P(\mathbf{y}|\theta) \equiv 0.$$

Differentiate $\partial/\partial\theta$,

$$\int dy \left[\hat{\theta}(y) - \theta\right]\frac{\partial P}{\partial\theta} - \int dy\, P \equiv 0, \qquad P \equiv P(y|\theta).$$

Use the identities

$$\frac{\partial P}{\partial\theta} \equiv P\frac{\partial \ln P}{\partial\theta}, \qquad \int dy\, P \equiv 1 \ \ (\textbf{normaliz.}).$$

Straightforward result is

$$\int dy \left[\hat{\theta}(y) - \theta\right]P\frac{\partial \ln P}{\partial\theta} = 1.$$

GENERALIZED CRAMER-RAO INEQUALITY (CONTINUED)

Factor the preceding integrand, insert arbitrary function $\mathbf{f}(\mathbf{y} - \theta)$ so that it cancels,

$$\int d\mathbf{y}[(\hat{\theta}(\mathbf{y}) - \theta)\mathbf{P}^{1/2}\mathbf{f}^{1/2}(\mathbf{y} - \theta)][\mathbf{f}^{-1/2}(\mathbf{y} - \theta)\mathbf{P}^{1/2}\frac{\partial \ln \mathbf{P}}{\partial \theta}] = 1$$

This is effectively a dot-, or inner-, product. Take its modulus-square. Then using the Schwarz inequality gives

$$\int d\mathbf{y}\left|\hat{\theta}(\mathbf{y}) - \theta\right|^2 \mathbf{P}\mathbf{f}(\mathbf{y} - \theta) \int d\mathbf{y}\left|\frac{\partial \ln \mathbf{P}}{\partial \theta}\right|^2 \mathbf{P}\mathbf{f}^{-1}(\mathbf{y} - \theta) \geq 1.$$

Notice how function $\mathbf{f}(\mathbf{y} - \theta)$ persists in each left-hand factor.

Next, for *further* generality, multiply the first integral by b = const. and the second by 1/b.

Then we define a weighted mean-square error from the 1ˢᵗ integral,

$$e^2 \equiv b \int dy \left| \hat{\theta}(y) - \theta \right|^2 Pf(y - \theta)$$

and from the 2ⁿᵈ integral a generalized Fisher information

$$I \equiv b^{-1} \int dy \left| \frac{\partial \ln P}{\partial \theta} \right|^2 Pf^{-1}(y - \theta).$$

Then the Schwarz inequality said these obey

$$e^2 I \geq 1,$$

called the 'generalized Cramer- Rao' inequality.

Cases of interest: $f(x) = 1,\ b = 1$: **Gives ordinary Cramer- Rao inequality, with**

$$I = \int dy \left| \frac{\partial \ln P}{\partial \theta} \right|^2 P \Rightarrow \int dx \frac{(dP/dx)^2}{P}$$

for 1 datum-case, where the PDF P obeys shift invariance and $x \equiv y - \theta$ is the 'noise.'

Case $f(x) = 1,\ b = -1$, giving a Negative information

$$I = - \int dy \left| \frac{\partial \ln P}{\partial \theta} \right|^2 P .$$

This corresponds to the case where unknown parameter θ is *pure imaginary*. (This proved useful in deriving quantum mechanics, where the time is indeed an imaginary parameter ict_0.) Then the error e^2 in its estimate should be negative. And it is, from last defining equation for e^2.

SPECIAL FORMS of FISHER I : Single-datum Cases (SFI p. 29)

Special Cases of Shift-invariance and $N = 1$ variable: Here

$P(y|\theta) = P(y|\theta) = p(y - \theta) \equiv p(x), \ x \equiv y - \theta.$ **Then**

$$I = \int dx \, p\left(\frac{\partial \ln p}{\partial x}\right)^2 \equiv \int dx \, \frac{(\partial p/\partial x)^2}{p}.$$

Obvious question: At x for which $p(x) \to 0$, does I blow up?

Q-Form of I : **Let** $p(x) \equiv q^2(x)$ **define a new, real function** $q(x).$ **Directly**

$$I = 4 \int dx \, \left(\frac{\partial q}{\partial x}\right)^2.$$

In words: I is proportional to the gradient content of the amplitude content of the PDF. There is no longer a problem at points where continuously $p(x) \to 0$. I remains finite.

If instead $p(x) = 0$ discontinuously, as at a step at the edge of the domain of x, the step is avoided by taking a 'principal value integral'. However if the step is within the domain of x, it is included in the integration, giving infinity for I.

Wootter's quantum form: The preceding has the added significance of being a 'quantum information' as prescribed by W.K. Wootters (1981).

I AS A MEASURE OF *DISORDER* (SFI pgs. 41-45)

The preceding integral shows that the more spread out q (and therefore the PDF p) is the smaller is I. This says that the more disordered the system is the smaller is I. It measures disorder.

Like an entropy, I also tends to increase with time,

$$dI/dt \leq 0.$$

(Fisher I-theorem). This was the first prediction of EPI. It was confirmed by Plastino (1996).

Fisher as basis for thermodyn: Predicted in SFI (pgs. **47,415**) and being confirmed by Plastino et al.

(**Fisher temperature**). Predicted to obey $1/T_F = -k(dI/d\theta)$, $k = k(\theta)$.

SOURCE INFO. J (SFI pgs. 3, 75-76, 80-82, 86-87)

The C-RI suggests presence of an information channel

$$J \to I, \quad \text{where } I = \kappa J, \ \kappa = \text{const.} = (0,1)$$

Any source function (charge, current, economic production function,...) is assumed known. This can help define the functional J. Information $I - J$ called the 'physical information.' Since $\kappa \leq 1$ from preceding, $I - J$ represents a general loss of information, in line with Plato-Kant views of reality (difference between a ground-truth noumenon and its observed phenomenon).

Demanding that the information loss be a minimum (more generally, an extremum) actually allows one to solve for the amplitudes $\psi_n(x)$ of the channel! In summary, the universe *COOPERATES* with the observer. This is in two ways:

(I) It permits received information I that is optimally close to the maximum possible amount J.

(II) Knowing this, the observer can use it to find the unknown amplitudes of the system!

AXIOMS OF EPI

EPI is acronym for *Extreme physical information*, a method of finding the amplitudes $\psi_n(x)$ and PDFs of the channel.

(1) Activating the measurement process perturbs the system, therefore perturbing I and J.

(2) The two perturbations are equal, $\delta I = \delta J$, or equivalently $\delta(I - J) = 0$, implying

$$I - J = \text{extrem.} \qquad \left(\text{EPI variational principle}\right)$$

(Identically true if J defines the information in a space that is unitary to that of the measurements.)

(3) From above,

$$I = \kappa J, \text{ or equivalently } I - \kappa J = 0 \quad \left(\text{EPI zero principle}\right)$$

The latter in the *local* sense of the integrands $i(x) - \kappa j(q, x) = 0$, where $\int dx\, i(x) \equiv I$ and $\int dx\, j(q, x) \equiv J$.

FINDING INFO. J (SFI pgs. 3, 86-87)

Info. I is generic (data are data). Hence it is always known as one of the above forms for I (or equivalent ones in curved space).

On the other hand, κ and J are specific to the given noumenon. In general, there are 3 approaches to finding these. These approaches are outgrowths of 3 levels of prior knowledge about the unknown system (named by the American philosopher Chas. S. Peirce, circa 1897).

The approaches give rise, respectively, to *THREE LEVELS OF ACCURACY* in the solutions. In descending order of accuracy these are:

(A) **ABDUCTION**, or absolute truth. That is, $I = J$ is achieved. I is an L^2 length, so $I = J$ is achieved by a *length-preserving* or UNITARY transformation. (For example it is momentum space in quantum mech. or 'investment momentum' space in econ.). It can be shown that for a unitary transformation *the EPI variational principle holds identically* (without need for axiom 2). Also, $\kappa = 1$, so that the result is exact. The information in the phenomenon actually *achieves* its level in the noumenon - contra Kant but pro Spinoza!

(B) DEDUCTION, or secondary (inferred) truth. This is knowledge of an *invariance principle* that holds in measurement space. (For example, it is 'continuity of flow' in deriving both Maxwell's e.m. eqs. and Einstein's eqs. of general relativity.) The unknowns κ and J are solved for as simultaneous solutions to the 2 EPI requirements (extremum and zero conditions). Results are slightly erroneous, e.g. they are non $-$ quantum in nature. As a corrob., $\kappa < 1$, in fact often exactly 1/2. Half of the noumenal information does not reach the phenomenon.

(C) **INDUCTION**, or empirical (sampled) truth. This is knowledge of one or a few data values from the system. This may also be called a 'Bayesian' estimate, in that the choice of data is arbitrary. Results are least accurate, but do give smooth curves $\psi_n(x)$ because of *the local nature* of I. Nevertheless, results are still quite accurate in both thermod. and economics. Here κ is never found and J is left undefined, replaced by Lagrange constraint terms expressing the data. Examples of (A),(B),(C) follow.

Ex. of DEDUCTION: **FINDING CANCER GROWTH LAW (SFI pgs. 398-405)**

Data scenario : **A cancer is identified in situ (on the person) at a time τ, where**

$$\tau = \theta + t,$$

$\theta =$ **time at which cancer began growing, and** $t =$ **elapsed time since onset.**

Question: **What is** $m(t)$, **the biomass growth law of the cancer?**

Connection with a PDF: By "law of large numbers," $m(t) \propto p(t)$ **the probability of randomly detecting the cancer at the time** t. **Therefore, can use EPI to establish** $p(t)$ **and then the growth law.**

Type of EPI Solution Chosen :

An abduction solution, type (A), would be preferred since it is exact. However, currently no unitary transform space for cancer growth is known, so abduction (A) is ruled out. Consequently we try the next lower level of accuracy:

(B) DEDUCTION. This requires knowledge of invariances, and seeks the EPI solution that simultaneously satisfies both the EPI variational- and zero- principles.

Invariances that are 'deduced', to be obeyed by the solution:

(1) Cancerous tissue does not function, only reproduces. Also, it lacks telomeres which indicate *age*. Therefore it exhibits *minimum* I *about its age* t. The minimum value is the invariant.

(2) Many cancer cells independently radiate poisonous lactic acid upon a nearby healthy cell. The effective message is 'die!'. By additivity of such information, I can now exceed the J from a single cancer cell. That is, can now have $\kappa > 1$.

(3) Boundary values at t = 0 : source info. is zero and the cancer is massless, $j(0) = 0$ and $p(0) = 0$.

GETTING INFORMATIONS I,J (SFI pg. 400)

Assume shift invariance, $p(\tau|\theta) = p(t) \equiv q^2(t)$ (same growth characteristics of cancer, indep. of absolute age). The information functional I is accordingly

$$I \equiv 4 \int_0^T dt\, \dot{q}^2(t), \quad \dot{q}(t) \equiv dq/dt.$$

Represent J as a general integral

$$J \equiv 4 \int dt\, j[q,t],$$

where cancer source info. $j[q,t]$ is some unknown function of q and t. The problem is to find it, and κ.

The approach (as above) is to satisfy *both* EPI conditions with one function $q(t)$. The two are the EPI variational principle condition

$$\frac{d}{dt}\left(\frac{\partial(\dot{q}^2 - j[q, t])}{\partial \dot{q}}\right) = \frac{\partial(\dot{q}^2 - j[q, t])}{\partial q},$$

and the EPI zero condition

$$\dot{q}^2 - \kappa j[q, t] = 0.$$

Note that the latter is the microlevel version (the integrand) of the requirement $I - \kappa J = 0$. In type (B) solutions, the zero condition is always satisfied on the microlevel. The rest is algebra.

EPI SOLUTION

Combining the two equations gives (ultimately)

$$\sqrt{j}\,\frac{\partial j}{\partial q}(1+\kappa) + \sqrt{\kappa}\,\frac{\partial j}{\partial t} = 0.$$

Seek a separable solution

$$j[q,t] \equiv j_{Q}(q)j_{T}(t).$$

In the usual way, this results in two equations, one in $j_{Q}(q)$ and the other in $j_{T}(t)$. Their solutions give

$$j[q,t] = \left(\frac{\frac{Aq}{1+\kappa} + B}{\frac{At}{\sqrt{\kappa}} + C}\right)^{2}, \quad A, B, C = \text{constants}.$$

On the other hand, the square-root of the zero-condition equation gives

$$\dot{q} = \sqrt{\kappa}\,\sqrt{j[q,t]}\,.$$

Combining the last two equations gives

$$\dot{q} = \left(\frac{\frac{Aq}{1+\kappa} + B}{\frac{At}{\sqrt{\kappa}} + D} \right), \quad D = C/\sqrt{\kappa}\,.$$

This is easily integrated, giving

$$q(t) = \left(\frac{1+\kappa}{A} \right)\left(\frac{At}{\kappa} + D \right)^{\kappa/(\kappa+1)} - (1+\kappa)\frac{B}{A}\,.$$

This is a power $-$ law in t, plus a constant. Imposing invariance (3) above (boundary conditions) gives

$$q(t) = \left(\frac{1+\kappa}{A} \right)\left(\frac{At}{\kappa} \right)^{\alpha}, \quad \alpha \equiv \left(\frac{\kappa}{\kappa+1} \right).$$

Note: This solution does check in satisfying both EPI requirements.

WHAT IS THE VALUE OF THE POWER α?

Using invariance (1) above that I = min., let α be such that $I = I(\alpha)$ is min. This requires form of $I(\alpha)$.

Differentiating the preceding eq. to get $\dot{q}(t)$, and also normalizing $q^2(t)$, gives

$$I \equiv 4 \int_0^T dt \dot{q}^2(t) = \frac{4}{T^2}\left(\frac{2\alpha+1}{2\alpha-1}\right)\alpha^2 = I(\alpha).$$

Setting $dI(\alpha)/d\alpha = 0$ gives a solution

$$\alpha = \tfrac{1}{4}(1 \pm \sqrt{5}).$$

The negative root gives a negative I, and so is rejected. The final solution is

$$p(t) \equiv q^2(t) = Kt^{2\alpha} = Kt^{\phi}, \quad \phi = \frac{1}{2}(1 + \sqrt{5}) = 1.618034\ldots,$$

the Fibonacci **golden mean**!

Conclusion: In situ cancer grows relatively slowly, obeying a power law with power less than 2. This is much slower than in vitro (test tube) cancer growth, which instead grows exponentially in time.

WHAT IS THE ERROR IN KNOWLEDGE OF THE ONSET TIME τ OF THE CANCER? (SFI p. 407)

We had

$$e_{min} = \frac{1}{\sqrt{I}} = \left[\frac{4}{T^2} \left(\frac{2\alpha + 1}{2\alpha - 1} \right) \alpha^2 \right]^{-1/2} = T \sqrt{\frac{2}{11 + 5\sqrt{5}}} \approx 0.3T.$$

Therefore the rms error in knowledge of the onset time is considerable – *30% of the total period* T *(many months) of observations*.

The power-law prediction is to be compared with the results of six clinical studies. These showed power-law behavior, with the power ϕ at empirical values

$$1.72, 1.69, 1.47, 1.75, 2.17, \text{and } 1.61.$$

These have a sample mean of 1.73 and standard deviation of 0.23, well-agreeing with our theoretical value of $1.618\ldots$

Also, a comparison is made of our theoretical growth law $t^{1.618}$ (re-expressed as occurrence of tumor size vs size) with the corresponding empirical law using simultaneously all the data from the six studies. This is shown in Fig. 15.3. Again, there is good agreement. The theoretical approach is confirmed.

In the above economic problems we minimized the single information I subject to constraints only, gaining a benefit of smoothness. By comparison, here we minimized the dual information quantity $I - J$, gaining both smoothness (due to I) and extra 'groundtruth' (due to J). Thus, *EPI at prior knowledge level (B) tends to give not only smooth results but, in fact, correct results.*

Fig. 15.3. Tumor growth dynamics based on the information-degradation model of carcinogenesis as predicted by EPI (solid straight line) compared with experimental data from six studies: points △ from Tabar *et al.* (1992); ⊡ from Fagerberg *et al.* (1985); ⊙ from Thomas *et al.* (1984); ⊠ from De Koning *et al.* (1995); ◎ from Peer *et al.* (1994); and ⊗ from Burhenne *et al.* (1992). Figure from Gatenby and Frieden (2002).