10. Physics from Fisher Information.

Motivation: Can the equations of physics be derived from information-theoretic principles?

I. Fisher Information.

Task: To obtain a measure of the accuracy of estimated values of a measureable quantity.

Set-up:

- Let θ = actual value of a measureable quantity.
- Let $(y_1, y_2, ..., y_N)$ be N measurements of the quantity.
- Let $(x_1, x_2, ..., x_N)$ be random errors associated with each measurement.
  - *So*: $y_i = \theta + x_i$.
- Let $p(y_i|\theta)$ be a probability distribution over the y_i's with respect to θ.

- *Recall*: $p(y_i|\theta)$ is the probability of obtaining the value y_i when measuring the quantity, given that the actual value of the quantity is θ.
- *Basic Question*: *What is the amount of information about θ contained in the measurement y_i?*

- Suppose we use the measurements y_i to construct an *estimate* $\hat{\theta}$ of θ.

> *Def. 1.* The *expected error* $E(\hat{\theta}-\theta)$ of $\hat{\theta}$ is given by,
> $$E(\hat{\theta}-\theta) = \int (\hat{\theta}-\theta)\, p(y_i|\theta)\, dy_i$$

> *Def. 2.* The *mean squared error* e^2 of $\hat{\theta}$ is given by,
> $$e^2 = \int (\hat{\theta}-\theta)^2\, p(y_i|\theta)\, dy_i$$

> *Claim*: For "unbiased estimates" for which $E(\hat{\theta}-\theta) = 0$, the mean squared error has a lower bound, $e^2 \geq 1/I[p]$.

> *Def. 3.* The *Fisher Information* $I[p]$ is given by,
> $$I[p] = \int \left(\frac{\partial}{\partial\theta} \ln p(y_i|\theta)\right)^2 p(y_i|\theta)\, dy_i$$

- *What this means:* As the Fisher information $I[p]$ increases, the mean squared error e^2 decreases.
- *So*: The more Fisher information one has about a given probability distribution p, the more precise the estimates of the data will be.

How to derive $e^2 \geq 1/I$.

For an unbiased estimator, $E(\hat{\theta}-\theta) = \int (\hat{\theta}-\theta) p \, dy_i = 0.$

So: $$\frac{\partial}{\partial\theta}\int (\hat{\theta}-\theta) p \, dy_i = -\int p \, dy_i + \int (\hat{\theta}-\theta)\frac{\partial p}{\partial\theta} dy_i = 0.$$

Or: $$\int (\hat{\theta}-\theta)\frac{\partial p}{\partial\theta} dy_i = 1, \qquad \text{since} \quad \int p \, dy_i = 1.$$

Or: $$\int (\hat{\theta}-\theta)\frac{\partial \ln p}{\partial\theta} p \, dy_i = 1, \qquad \text{where} \quad \frac{\partial p}{\partial\theta} = \frac{\partial \ln p}{\partial\theta} p.$$

Or: $$\int \left(\frac{\partial \ln p}{\partial\theta}\sqrt{p}\right)\left((\hat{\theta}-\theta)\sqrt{p}\right) dy_i = 1$$

So: $$\int \left(\frac{\partial \ln p}{\partial\theta}\right)^2 p \, dy_i \int (\hat{\theta}-\theta)^2 p \, dy_i \geq 1 \qquad \text{by the Schwarz Inequality.}$$

- *Now:* Let $N = 1$ and $p(y|\theta) = p(y-\theta) = p(x)$.
- *Then*: I becomes:

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

Fisher Information

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

- *An inverse measure of the mean squared error associated with deviations x in measured values of a physical quantity from its actual value.*

Shannon Information

$$H[p] = -\int \ln[p(x)]p(x)dx$$

- *The maximum amount that messages drawn from a given set of letters x with probability distribution $p(x)$ can be compressed.*
- *The expected value of the information gain, $-\ln[p(x)]$, upon measuring a random variable with values t and probability distribution $p(x)$.*

- *Now*: Let $p(x) = q^2(x)$.
- *Then*: $I[q] = 4\int q'^2 dx$, where $q' = \dfrac{\partial q}{\partial x}$.
- *Note*: $I[q]$ (as well as $H[p]$) is a "functional". It's inputs are *functions* $q(x)$ and its outputs are numbers.
- *Moreover:* I contains the square of a derivative term.

- *What this entails*:

> The function $q(x)$ for which I takes a minimum or maximum (*i.e.*, "extremal") value is the solution to a 2nd-order partial differential equation.

- *Why is this supposed to be important?* The equations of motion of most theories in physics are 2nd-order partial differential equations.

- *So (Friedan 1998)*: Perhaps the equations of motion (of most theories in physics) can be derived by an appropriate reconstrual of Fisher information.

Let's see how this is supposed to work...

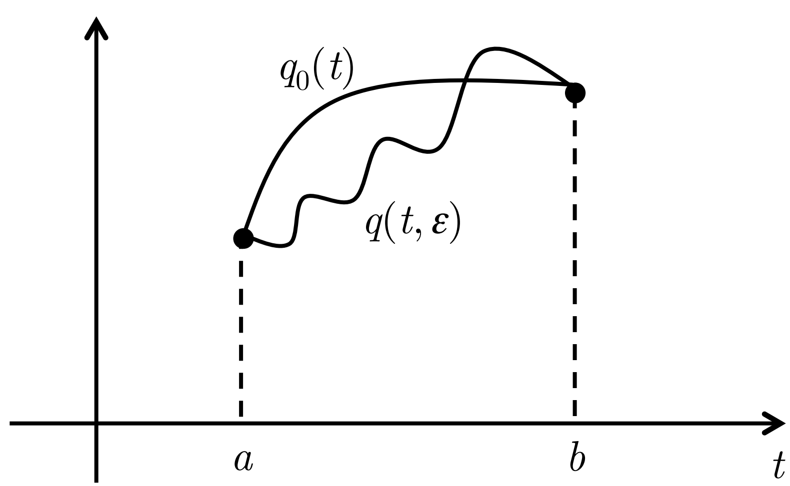
II. Physics and Fisher Information.

- *Suppose:* We're given a functional in the form of a definite integral:

$$S[q] = \int_a^b L(t, q(t), q'(t))dt, \qquad q'(t) = \frac{dq(t)}{dt}$$

- *And:* We want to know the particular function $q(t)$ for which S take an extremal value (minimum or maximum).
- *Now*: Suppose this particular function is given by $q_0(t)$.
- *And*: Consider small variations $q(t,\varepsilon)$ of $q_0(t)$ between the endpoints a, b:

$$q(t,\varepsilon) = q_0(t) + \varepsilon\eta(t), \quad \text{where } \eta(a) = \eta(b) = 0, \text{ and } \frac{\partial q}{\partial \varepsilon} = \eta(t).$$



- For these functions $q(t,\varepsilon)$, $S[q] = S(\varepsilon) = \int_a^b L(t, q(t,\varepsilon), q'(t,\varepsilon))\,dt$.
- *And:* S takes an extremal value where it's slope is zero: $\left.\dfrac{\partial S}{\partial \varepsilon}\right|_{\varepsilon=0} = 0$.
- *Now*: Let $\delta S = \left.\dfrac{\partial S}{\partial \varepsilon}\right|_{\varepsilon=0} d\varepsilon$ be a small variation in S about its extremal value.
- *Then*: The requirement that S take an extremal value, *i.e.*, $\delta S = 0$, entails

$$
\begin{aligned}
0 = \frac{\partial S}{\partial \varepsilon} &= \int_a^b \left(\frac{\partial L}{\partial q}\frac{\partial q}{\partial \varepsilon} + \frac{\partial L}{\partial q'}\frac{\partial q'}{\partial \varepsilon} \right) dt \\
&= \int_a^b \left(\frac{\partial L}{\partial q}\frac{\partial q}{\partial \varepsilon} - \frac{d}{dt}\left(\frac{\partial L}{\partial q'}\right)\frac{\partial q}{\partial \varepsilon} \right) dt + \cancelto{0}{\left.\frac{\partial L}{\partial q'}\frac{\partial q}{\partial \varepsilon}\right|_a^b} \\
&= \int_a^b \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial q'}\right)\right) \eta(t)\,dt
\end{aligned}
$$

Integration by parts:

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du$$

$$u = \frac{\partial L}{\partial q'}, \quad dv = \frac{\partial q'}{\partial \varepsilon}\,dt$$

- Since $\eta(t)$ is arbitrary, this means that the following must hold:

Euler-Lagrange Equation

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial q'}\right) = 0$$

A solution $q(t)$ to the Euler-Lagrange equation is a function that extremizes S!

The function $L(t,\ q,\ q')$ is called the *Lagrangian*.
The functional $S[q] = \int_a^b L(t, q(t), q'(t))dt$, is called the *action*.

<u>*Ex 1*</u>: What function $q(t)$ extremizes the Fisher information $I[q]$?

- For Fisher info, $L(t,\ q,\ q') = 4q'^2$, and the Euler-Lagrange equation is

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial q'}\right) = -\frac{d}{dt}\left(2q'\right) = -8\frac{d^2 q(t)}{dt^2} = 0.$$

2nd-order diff. equ. for q(t)!

<u>*Ex 2*</u>: What function $q(t)$ extremizes an action with $L(t, q, q') = \frac{1}{2}mq'^2 - V(q)$?

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial q'}\right) = -\frac{\partial V}{\partial q} - m\frac{d}{dt}q' = -\frac{\partial V}{\partial q} - m\frac{d^2 q(t)}{dt^2} = 0$$

<u>*Or*</u>: $$m\frac{d^2 q(t)}{dt^2} = -\frac{\partial V}{\partial q}$$

Newton's law of motion for the path q(t) of a particle with mass m interacting with a potential V(q)!

- <u>*Note*</u>: *In the absence of interactions* ($V = 0$), the Newtonian Lagrangian is identical (up to a factor) to the Fisher information Lagrangian.

Two Observations:

(1) The fundamental equations of motion in physics can be derived by extremizing an action S with an appropriately formulated Lagrangian:

- *Classical field theories*: Newton's laws; Maxwell's equations; Einstein equations.
- *Quantum mechanics*: Schrödinger and Dirac equations.
- *Quantum field theories*: QED, QCD, Standard Model.

(2) In the absence of interactions, all of these Lagrangians are schematically identical to the Fisher information Lagrangian (in so far as they all become the square of a derivative term).

- This suggests that, in the absence of interactions, the fundamental equations of motion in physics are the result of extremizing Fisher information; *i.e.*, requiring that $\delta I = 0$.

- *What could this mean ??* In the absence of interactions, the fundamental equations of motion in physics are expressions of how information is extremized (maximized? minimized?) during measurements.

- *But first*: What *about* interactions?

What Frieden Has to Say (1998, pp. 70-71):

"Physical Information"

$K[q] = I[q] - J[q]$

- $I[q]$ = "Intrinsic" information of a physical system obtained during a measurement.
- $J[q]$ = "Bound" information of the system prior to measurement.
- $K[q]$ = "Physical" information of the system.

"Axiom 1"

$\delta I = \delta J$

- "Conservation of information change". (During a measurement, small variations in I equal small variations in J.)

"Variational Principle"

$\delta K = 0$

- When changes in information are conserved during measurements (?), the physical information obtains an extremal value (pg. 71).

"Zero-condition"

$I - \kappa J = 0, \quad \kappa \leq 1$

- Suggested by the constraint $J \geq I$ ("...the information I in the data never exceeds the amount J in the phenomenon").

- "*The variational principle and the zero-condition comprise the overall principle we will use to derive major physical laws.* Since, by Eq. (3.16) [the "variational principle"], physical information K is extremized, the principle is called 'extreme physical information' or EPI." (Friedan, pg. 72.)

- What EPI is supposed to do:

(a) The *variational principle* $\delta K = \delta(I - J) = 0$ is supposed to generate the equations of motion.

- Variation of I provides the "kinetic" (interaction-free) terms.
- Variation of J provides the interaction terms.

(b) The *zero-condition* is supposed to generate the particular form of J for a given phenomenon, with the help of "invariance principles".

***Big Problem*:** *Step (b) doesn't work.* (Lavis & Streater 2002)

- *So*: At most, EPI is a way to *interpret* the Lagrangian formulation of theories in physics in information-theoretic terms.

Three Interpretive Issues:

1. How should the variational principle, $\delta K = \delta(I - J) = 0$, be interpreted?

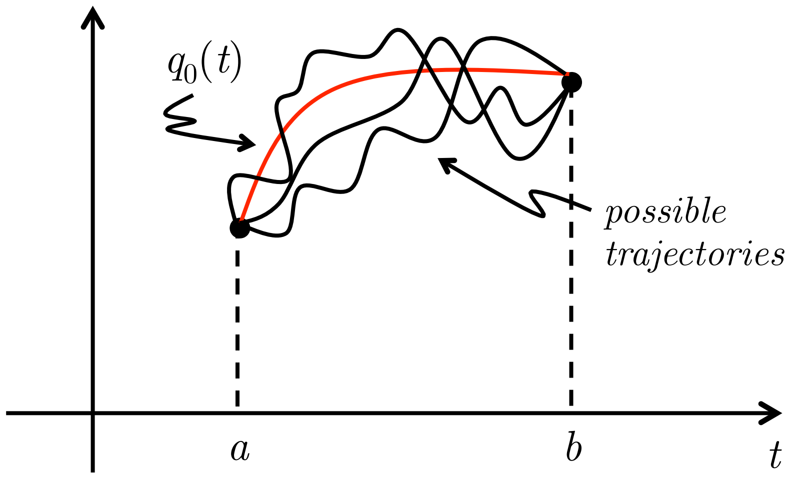
- *Note*: The Lagrangian formulation of physical theories is based on:

Principle of Least Action

- A physical system is described by an action functional $S[q]$, where q represents a possible quantity the system can possess.
- The *actual* quantity the system possesses is that q_0 that extremizes the action: $\delta S[q_0] = 0$.

Example: Massive Newtonian particle

- Described by the Newtonian action
  $S[q(t)] = \int [\frac{1}{2} m q'^{\,2} - V(q)]\, dt$,
  where $q(t)$ is a possible trajectory the particle can possses.
- The actual trajectory $q_0(t)$ is the one for which $\delta S[q_0(t)] = 0$.



How should the Principle of Least Action be interpreted?

(i) A fundamental physical principle?

(ii) A mathematical way of encoding the equations of motion?

Problems with Option (i):

(a) Why do actual physical quantities extremize the action?

- *The action as a final cause of motion?*
- *Contrast with the local differential equations of motion as encoding efficient causes of motion.*

(b) How is the form of the Lagrangian for a given system determined?

- *Standard approach:* *Work backward from the equations of motion for a simple system to the form of its Lagrangian, and then extrapolate to more complicated systems.*
- *But this seems to imply that the equations of motion are more fundamental.*

- *Does EPI offer an information-theoretic foundation for Option (i)?*

Frieden's Game-Theoretic Analogy:

- "Physics is...the science of measurement." (pg. 2.)
- "EPI Theory proposes that all physical theory results from...imperfect observation." (pg. 3.)
- A measurement is a game between observer and Nature in which the observer seeks to maximize info ($\delta K = 0$) and Nature seeks to maximize error (pp. 79-83).

Problems:

(a) $\delta K = 0$ generates *extremal* values of q (not necessarily *maximal* values).

(b) Physics as a science of measurements?

- What about in-principle unmeasureable quantities that figure into physical theories?
- What are measureable quantities?
  - *Properties of the physical system being measured?*
  - *Properties of the measuring device?*
  - *Properties of the observer's perception?*

2. How should the probabilities in EPI be interpreted?

- "EPI is an expression of the 'inability to know' a measured quantity." (pg. 3.)
- EPI is based on concepts (Fisher information) that appear in error analysis and statistics.
- *So*: Perhaps the probabilities are *epistemic*.
  - *In particular*: $p(x)$ is a measure of the observer's degree of belief in the accuracy of a measurement y of a physical quantity with actual value θ and error x.
  - *And*: The (deterministic) equations of motion governing the quantities $p(x)$ (and $q(x)$) indicate how the observer's degrees of belief evolve in time.

Problems

(a) Are physical quantities reducible to measurable/observable quantities?

(b) All things being equal, is an interpretation of physical theory based on a fundamental distinction between observer and phenomena to be preferred over one in which no such distinction appears?

(c) *General consensus*: The probabilities in quantum theories cannot be interpreted as epistemic.

- *So*: Maybe the probabilities are supposed to be *ontic*.
  - "The basic elements of EPI theory are the probability amplitudes q... EPI regards all phenomena as being fundamentally statistical in origin." (pg. 83.)
  - *In particular*: $p(x)$ is a measure of the intrinsically stochastic nature of the phenomena being observed.
  - *And*: The (deterministic) equations of motion governing the quantities $p(x)$ (and $q(x)$) indicate how these stochastic properties of physical phenomena evolve in time.

Problem

- *General consensus*: The processes governed by classical equations of motion (Newtonian) are not inherently stochastic.

3. What is the nature of Fisher Information?

- "This view regards reality as being perpetrated by requests for information. It adds a new creative dimension to the normally passive act of observation." (pg. 108.)
- Wheeler quote (pg. 1):

  "All things physical are information-theoretic in origin and this is a participatory universe... Observer participancy gives rise to information; and information gives rise to physics."

- Recall Timpson (2008):

> *Information* = What is produced by an information source that is required to be reproducible at the receiver if the transmission is to be counted a success.

- *In particular*: Information (in the technical sense) is an "abstract noun".
- "Information" does not refer to a substance (*token*); rather, it refers to a *type*.