STATISTICAL INFERENCE
AND
STATISTICAL MECHANICS

SANJOY K. MITTER, MIT

OCTOBER 2004
ISINS,
BEIJING AND XI’AN, CHINA
→ **INTERCONNECTIONS OF CONTROL AND COMMUNICATION SYSTEMS**

→ **WHAT ARE THEIR FUNDAMENTAL LIMITATIONS?**

→ **RAPPROCHEMENT OF PARTIALLY OBSERVED STOCHASTIC CONTROL WITH INFORMATION THEORY**

→ **FUNDAMENTAL CONCEPTUAL DIFFICULTY: INFORMATION THEORY USES PASSING TO THE THERMODYNAMIC LIMIT (BLOCK LENGTHS GOING TO INFINITY) TO PROVE THE NOISY CHANNEL CODING THEOREM**

→ **UNACCEPTABLE IN A CONTROL CONTEXT MUST TAKE INTO ACCOUNT DELAYS**

→ **DYNAMICAL VIEW OF INFORMATION THEORY NEEDED**
We need to understand how information flows from sensors to estimators to controllers over communication channels, in the sense: laws of conservation (dissipation) of information flows?

Close ties to nonequilibrium statistical mechanics (Ruelle, Gallavotti, . . .)
THIS LECTURE

INFORMATION FLOW IN THE KALMAN FILTER

WE SHOW:


LECTURE IS IN TWO PARTS

(1) VARIATIONAL INTERPRETATION OF BAYESIAN ESTIMATION

(2) INFORMATION FLOW IN THE KALMAN FILTER

Note: Conceptual View is one of Interconnection of Systems (for example, Interconnection of State Dynamics, Observation and the Filter)

More General Research Program: Dynamical View of Network Information Theory
Variational Characterization of Bayesian Inference

\((X, \mathcal{X})\)

\(\mathcal{H} : X \rightarrow (-\infty, \infty)\)

\(\mathcal{P}(X) = \text{Set of Prob. measures on } (X, \mathcal{X})\)

\(\tilde{P}_X, \hat{P}_X \in \mathcal{P}(X)\)

\(\tilde{p}_X, \hat{p}_X \) densities

\[
h(\tilde{P}_X|\hat{P}_X) = \int_X \left( \log \frac{\tilde{p}_X}{\hat{p}_X} \right) \tilde{p}_X \, dx \tag{1}\]

\[
i(\tilde{H}) = - \log \int_X \exp(-\tilde{H}) \, dP_x \tag{2}\]

\((P_X : \text{prior})\)
\[- \log \left( \frac{d\tilde{P}_X}{d\hat{P}_X} \right) : \text{“Shannon Information”} \]

\[ h(\tilde{P}_X|\hat{P}_X) = \text{Average Reduction} \]

in the degree of surprise in the outcome arising from acceptance of \( \tilde{P}_X \) as the distribution rather than \( \hat{P}_X \).

**Interpret:** \( \exp(-\tilde{H}) \) as a *likelihood* for \( X \), associated with some unspecified observation, then \( \tilde{H}(X) = \text{residual degree of surprise in that “observation” if we already know } X = x. \)

\[ i(\tilde{H}) = - \log \left( \int_{x} \exp(-\tilde{H})dP_X \right) = \]

Total degree of surprise in that Observation, i.e. the Information in the unspecified observation if all we know: \( P_X \) prior.
\[ \tilde{H}(X) : \text{X-conditional information} \]

\[ i(\tilde{H}) : \text{Inf. in that observation} \]

**Theorem 1**

(i) 
\[ i((H(\cdot, y)) = \min_{\tilde{P}_X} [h(\tilde{P}_X|P_X) + \langle H(\cdot, y), \tilde{P}_X \rangle] \]

(ii) 
\[ h(P_{X|Y}(\cdot, y)|P_X) = \max_{\tilde{H}} \left\{ i(\tilde{H}) - \langle \tilde{H}, P_{X|Y}(\cdot, y) \rangle \right\} \]

(iii) \( P_{X|Y}(\cdot, y) \) is the unique minimizer in (1)

(iv) If \( H^* \) is a maximizer in (2), then \( \exists K \in \mathbb{R} \) s.t.
\[ H^*(X) = H(X, y) + K \]
Conceptualization

Information Processing over and above that in prior $P_X$

In (1): Source of additional information is $Y = y$

Bayes Formula: Extracts info. pertinent $h(P_{X|Y}(\cdot, y)|P_X)$ and leaves residual $\langle H, P_{X|Y} \rangle$.

Input information is held in likelihood $\exp(-H(\cdot, y))$ and extracted information in $P_{X|Y}(\cdot, y)$

Arbitrary Information procedure that postulates $\tilde{P}_X$ as post-obs. distribution has access to additional information. Hence: the notion Apparent Information.

In (2): Source of additional information in Posterior Distribution $P_{X|Y}(\cdot, y)$. The aim now is to postulate an observation, i.e. a likelihood
function $\exp(-\tilde{H})$ which gives rise to this observation.

Input Information

$$h \left( P_{X|Y}(\cdot, y) | P_X \right)$$

is *merged* with the residual information of the postulated observation

$$\langle \tilde{H}, P_{X|Y}(\cdot, y) \rangle :$$

Result $\geq i(\tilde{H})$

With equality $\Leftrightarrow$ Obs. is compatible with $P_{X|Y}$

$$i(\tilde{H}) - \langle \tilde{H}, P_{X|Y}(\cdot, y) \rangle$$

$= \text{Inf. in Postulated Obs.}$

compatible with $P_{X|Y}(\cdot, y)$

Compatible Inf. of $\exp(-\tilde{H})$