II.1
Variational Characterization of Bayesian Inference

$(\mathbb{X}, \mathfrak{X})$.

$H : \mathbb{X} \longrightarrow (-\infty, \infty]$.

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

$\tilde{P}_x, \hat{P}_x \in \mathcal{P}(\mathfrak{X})$

$\tilde{p}_x, \hat{p}_x$ densities

$$h(\tilde{P}_x | \hat{P}_x) = \int_X \left(\log \frac{\tilde{p}_x}{\hat{p}_x}\right) \tilde{p}_x \, dx \quad (1)$$

$$(2) \quad i(\tilde{H}) = -\log\left(\int_X \exp(-\tilde{H}) \, dP_x\right.$$

$(P_x :$ prior$)$.

$$-\log\left(\frac{d\tilde{P}_X}{d\hat{P}_X}\right) : \text{"Shannon Information"}$$

$h(\tilde{P}_X \mid \hat{P}_X)$ = Average Reduction in the degree of surprise in the outcome arising from acceptance of $\tilde{P}_X$ as the distribution rather than $\hat{P}_X$

<u>Interpret</u>: $\exp(-\tilde{H})$ as a <u>likelihood</u> for X, associated with some unspecified observation, then $\tilde{H}(x)$ = residual degree of surprise in that "observation" if we already know $X = x$

II.3

$$i(\tilde{H}) = -\log\left(\int_X \exp(-\tilde{H})\, dP_X\right)$$

= Total degree of surprise in that Observation i.e. the Information in the unspecified observation if all we know: P_X prior.

$\tilde{H}(X)$: X-conditional information

$i(\tilde{H})$: Inf. in that observation $\quad P_{XY}$

$\qquad\qquad\qquad\qquad\qquad\qquad P_X \quad P_Y$

__Th:__

(1) $i(H(\cdot, y)) = \underset{\tilde{P}_X}{\mathrm{Min}} \left[h(\tilde{P}_X | P_X) + \langle H(\cdot, y), \tilde{P}_X \rangle \right]$

(2) $h\left(P_{X|Y}(\cdot, y) \middle| P_X\right) = \underset{\tilde{H}}{\mathrm{Max}} \left\{ i(\tilde{H}) - \langle \tilde{H}, P_{X|Y}(\cdot, y) \rangle \right\}$

IV.

(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.

$$H^*(X) = H(X, y) + K$$

<u>Conceptualization</u>

Information Processing over & above that in prior P_X

In (1): Source of additional Information in $Y = y$

Bayes Formula: Extracts inf. pertinent to X $h(P_{X|Y}(\cdot, y) \mid P_X)$ and leaves

<u>residual</u> $\langle H, P_{X|Y} \rangle$.

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 is Posterior Distribution $P_{X|Y}(\cdot, y)$. The aim was

is to postulate an observation

i.e. a likelihood function which $\exp(-\tilde{H})$ gives rise to this observation.

Input Information

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

is <u>merged</u> with the residual information of the postulated observation

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

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

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

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

= Inf. in Postulated Obs. compatible with $P_{X|Y}(\cdot, y)$

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