4 Bayesian Learning

The 'score' function from [Seeger et al., 2010] is the mutual information between the desired measurement $\tilde{y}$ and the image u conditioned on existing measurements y.

$$\mathcal{S} = I(U;\tilde{Y}|Y) = H(U|Y) - H(U|\tilde{Y},Y)$$

This has a very intuitive interpretation. The current measure of uncertainty in the parameter recovery (reconstruction) is $H(U|Y)$ for the given data, Y. The measure of uncertainty in the parameter recovery (reconstruction) for a new measurement, $\tilde{Y}$, is $H(U|Y,\tilde{Y})$. The measurement, $\tilde{Y}$, is 'optimal' when is reduction in uncertainty in the parameter recovery (reconstruction) is maximized.

We will assume that the joint probability between (measurement, acquisition) pairs is jointly normal. The joint probability of signal measurements as a function of design parameters is assumed to be indepdent.

$$p(y_1,y_2|u,k_1,k_2) = C\exp\begin{bmatrix}y_1-\mu_1\\y_2-\mu_2\end{bmatrix}^\top\begin{bmatrix}P_{11}^{-1} & 0\\0 & P_{22}^{-1}\end{bmatrix}\begin{bmatrix}y_1-\mu_1\\y_2-\mu_2\end{bmatrix} = \underbrace{C_1\exp\|y_1(u,k_1)-\mu_1\|_{P_{11}}}_{p(y_1|u,k_1)}\underbrace{C_2\exp\|y_2(u,k_2)-\mu_2\|_{P_{22}}}_{p(y_2|u,k_2)}$$

For simplicity of notation, assume that two measures arise from two distinct acquisition parameters such that the dependence on k is assumed: $p(y_1,y_2|u) = p(y_1|u)p(y_2|u)$. The joint distribution for the measurements is generally correlated through the model.

$$p(y_1,y_2) = \int_U p(y_1,y_2,u)du = \int_U p(y_1,y_2|u)p(u)du = \int_U p(y_1|u)p(y_2|u)p(u)du \qquad p(u|y_i) = \frac{p(u,y_i)}{p(y_i)} = \frac{p(y_i|u)p(u)}{p(y_i)}$$

$$p(y_2|y_1) = \frac{p(y_1,y_2)}{p(y_1)} = \frac{1}{p(y_1)}\int_U p(y_1|u)p(y_2|u)p(u)du = \int_U \frac{p(y_1|u)}{p(y_1)}p(y_2|u)p(u)du = \int_U p(u|y_1)p(y_2|u)du$$

$$p(y_i) = \int_U p(y_i|u)p(u)du$$

Under these assumptions, the model for probability of parameters condition on the joint probability of the data reduces to the probabulity of the parametery recovery (reconstruction) with respect to the data multiplied by the probability of the new measurements given the parametery recovery (reconstruction) [Seeger et al., 2010].

$$p(u|\tilde{y},y) = \frac{p(u,\tilde{y},y)}{p(\tilde{y},y)} = \frac{p(\tilde{y},y|u)p(u)}{p(\tilde{y},y)} = \frac{p(\tilde{y}|u)p(y|u)p(u)}{p(\tilde{y},y)} = \frac{p(\tilde{y}|u)p(y|u)p(u)}{p(\tilde{y},y)} = p(\tilde{y}|u)p(u|y)\frac{p(y)}{p(\tilde{y},y)} \propto p(u|y)p(\tilde{y}|u)$$

The final information gain may be computed as:

$$\begin{aligned}\mathcal{S} = I(U;\tilde{Y}|Y) &= H(U|Y) - H(U|\tilde{Y},Y)\\ &= -\int_y dy\, p(y)\int_u du\, p(u|y)\log p(u|y) + \int_{\tilde{y}} d\tilde{y}\int_y dy\, p(\tilde{y},y)\int_u du\, p(u|\tilde{y},y)\log p(u|\tilde{y},y)\\ &= -\int_y dy\, p(y)\int_u du\, p(u|y)\log p(u|y) + \int_{\tilde{y}} d\tilde{y}\int_y dy\, p(\tilde{y}|y)p(y)\int_u du\, p(u|\tilde{y},y)\log p(u|\tilde{y},y)\\ &= \int_y dy\, p(y)\left(-\int_u du\, p(u|y)\log p(u|y) + \int_{\tilde{y}} d\tilde{y}\, p(\tilde{y}|y)\int_u du\, p(u|\tilde{y},y)\log p(u|\tilde{y},y)\right)\end{aligned}$$

We have two options to compute the information gain $I(U;\tilde{Y}|Y)$ or scoring function $\mathcal{S}$ at this point:

1. Assume that the measurements are completely predicted by the model. ie

$$p(y) = \int_U p(y|u)p(u)du$$

2. Use Gaussian distribution around actual signal measurements y^* and ignore the model predicted measurements

$$p(y) = \mathcal{N}(y^*,\sigma) = \exp\left(\frac{y-y^*}{\sigma}\right) \neq \int_U p(y|u)p(u)du$$

@dmitchell412 how do we interpret this ? how does modeling errors affect this choice ? which is 'better' ? should we further separate into predicted measurement vs actual measurement ? ie Kalman Filter [Maybeck, 1979].

$$z = Hy + \nu \qquad \Rightarrow \qquad I(U;\tilde{Y}|Z)$$

4.1 Information theory identities

Key ideas of active Bayesian learning follows from the definition of conditional mutual information [Cover and Thomas, 2012] and *repeeated* application of the definition of conditional probability and conditional entropy.

$$p(x, y|z) \equiv \frac{p(x, y, z)}{p(z)} = \frac{p(x|y, z)p(y, z)}{p(z)} = p(x|y, z)p(y|z)$$

$$H(Y|X) \equiv E_x \left[H(Y|X = x)\right] = \sum_x p(x)H(Y|X = x) = -\sum_x p(x) \sum_y p(y|x) \log p(y|x)$$

$$\begin{aligned}
I(X; Y|Z) &\equiv \sum_x \sum_y \sum_z p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} = \sum_x \sum_y \sum_z p(x, y, z) \log \frac{p(x|y, z)p(y|z)}{p(x|z)p(y|z)} = \sum_x \sum_y \sum_z p(x, y, z) \log \frac{p(x|y, z)}{p(x|z)} \\
&= -\sum_x \sum_y \sum_z p(x, y, z) \log p(x|z) + \sum_x \sum_y \sum_z p(x, y, z) \log p(x|y, z) \\
&= -\sum_x \sum_z p(x, z) \log p(x|z) + \sum_x \sum_y \sum_z p(x|y, z)p(y, z) \log p(x|y, z) \\
&= -\sum_z p(z) \sum_x p(x|z) \log p(x|z) + \sum_y \sum_z p(y, z) \sum_x p(x|y, z) \log p(x|y, z) = H(X|Z) - H(X|Y, Z)
\end{aligned}$$