

Signal Detection + Estimation

1) For a probability distribution $p(x)$, it can be shown that a reasonable definition of information that satisfies certain desirable properties is

$$I = - \int \ln p(x) dx.$$

The expectation (or average) of this quantity is defined as Entropy.

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

Nature prefers (in some sense) to have distribution that have maximum entropy. Therefore to assign probability distributions requires a constrained optimization solution where Entropy is maximised subject to various constraints that may be placed on the distribution.

1/ for example, if the distribution is normalized
then $\int p(x) dx = 1$

if the mean of the distribution is
known (by measurements say) then

$$\mu = \int x p(x) dx \quad (\text{known})$$

Maximizing entropy subject to these
constraints requires maximizing the cost ^{missing}
function _{term}

$$Q = - \int p \ln p dx - \alpha \left[\int p(x) dx - 1 \right] - \beta \left[\int x p(x) dx - \mu \right]$$

where α and β are Lagrange multipliers
that may be found using the
constraints.

Differentiating Q w.r.t P gives

$$\frac{\partial Q}{\partial p} = - \int_0^{\infty} [\ln p \text{ (circled)} + \alpha + \beta x] dx$$

sign does not agree with defn of Q on previous page

$$\therefore p(x) = e^{1-\alpha} e^{-\beta x}$$

← this solution agrees with

We now use the 2 constants to find α and β

$$\int p(x) dx = \int_0^{\infty} e^{1-\alpha} e^{-\beta x} dx = 1$$

$$\therefore e^{1-\alpha} \int_0^{\infty} e^{-\beta x} dx = 1$$

$$\therefore \beta = e^{1-\alpha}$$

$$\therefore p(x) = \beta e^{-\beta x}$$

Using the second constant

$$\text{mean} = \int_0^{\infty} p(x) dx$$

$$= \int_0^{\infty} \beta e^{-\beta x} dx$$

$$\int e^{-\beta x} dx = \frac{1}{\beta}, \quad \frac{\partial}{\partial \beta} \left(\int e^{-\beta x} dx \right) = - \int x e^{-\beta x} dx$$

$$\therefore -\frac{\partial}{\partial \beta} \left(\frac{1}{\beta} \right) = \int x e^{-\beta x} dx$$
$$= \frac{1}{\beta^2}$$

$$\therefore \text{mean} = \frac{1}{\beta}$$

The distribution having maximum entropy is

$$p(x) = \beta e^{-\beta x}$$

where β^{-1} = mean of the distribution



Writing $p(x) = \lambda e^{-\lambda x}$, the entropy is /5

$$\begin{aligned} H &= - \int_0^{\infty} p(x) \ln p(x) dx \\ &= - \int_0^{\infty} \lambda e^{-\lambda x} (\ln \lambda - \lambda x) dx \\ &= - \lambda \ln \lambda \int_0^{\infty} e^{-\lambda x} dx + \lambda^2 \int_0^{\infty} x e^{-\lambda x} dx \\ &= - \ln \lambda + 1 \end{aligned}$$

The variance is

$$\sigma^2 = \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx$$

since we have shown the mean = $\frac{1}{\lambda}$.

$$\therefore \sigma^2 = \frac{2}{\lambda^2} - \frac{2}{\lambda^2} + \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\therefore H = \ln \sigma + 1 \quad \checkmark$$

2) If the likelihood function is $p(x|\theta)$, the Fisher Information associated with the parameter θ is defined as;

$$I_{\theta} = E \left[\left(\frac{\partial}{\partial \theta} \ln p(x|\theta) \right)^2 \right] = - E \left(\frac{\partial^2 \ln p(x|\theta)}{\partial \theta^2} \right)$$

where the expectation is w.r.t the probability measure $p(x|\theta)$.

For an unbiased estimator $\hat{\theta}$ of θ we have by definition

$$\int (\hat{\theta}(x) - \theta) p(x|\theta) dx = 0$$

Differentiating w.r.t θ we have

$$\int \frac{\partial}{\partial \theta} (\hat{\theta}(x) - \theta) p(x|\theta) dx = 0.$$

$$\int dx \left[\frac{\partial}{\partial \theta} p(x|\theta) \right] (\hat{\theta}(x) - \theta) - \int dx p(x|\theta) = 0.$$

2 cont.)

$$\text{Using } \frac{\partial}{\partial \theta} p(x|\theta) = \frac{\partial}{\partial \theta} \ln p(x|\theta) \cdot p(x|\theta)$$

$$\text{and } \int p(x|\theta) dx = 1 \quad (\text{normalization})$$

$$\text{we have } \int dx \left(\frac{\partial}{\partial \theta} \ln p(x|\theta) \right) p(x|\theta) (\hat{\theta}(x) - \theta) = 0$$

Using the Schwarz inequality

$$\left| \int f(x)^* g(x) dx \right|^2 \leq \int f(x)^* f(x) dx \int g(x)^* g(x) dx$$

with equality iff $g(x) = A f(x)$ ($A = \text{constant}$)

we obtain

$$\int dx p(x|\theta) \left(\frac{\partial}{\partial \theta} \ln p(x|\theta) \right)^2 \geq \int (\hat{\theta}(x) - \theta)^2 p(x|\theta) dx$$

missing power 2

2 cont.

Using the definition of the Fisher information ^(panel 2) and the mean square error $\varepsilon^2 \triangleq \int (\hat{\theta}(x) - \theta)^2 p(x|\theta) dx$

we have

$$\varepsilon^2 \geq \frac{1}{I_{\theta}}$$

The condition for an unbiased efficient estimator is found when the above inequality is an equality (i.e. $g(x) = A(\theta)$)

In the above,

$$g(x) \triangleq \frac{d}{d\theta} \ln p(x|\theta) = \frac{1}{\sqrt{p(x|\theta)}}$$

$$f(x) \triangleq (\hat{\theta}(x) - \theta) \sqrt{p(x|\theta)}$$

$$\therefore \boxed{\frac{d}{d\theta} \ln p(x|\theta) = A(\theta) (\hat{\theta}(x) - \theta)} \quad \text{--- (1)}$$

To find $A(\theta)$, differentiate (1)

$$\frac{d^2}{d\theta^2} \ln p(x|\theta) = -A(\theta) + (\hat{\theta}(x) - \theta) \frac{dA(\theta)}{d\theta}$$

2 cont)

4

Taking expectations, we have

$$- E \left(\frac{\partial^2 \ln p(x|\theta)}{\partial \theta^2} \right) = A(\theta) \triangleq I_{\theta}$$

$$\frac{\partial \ln p(x|\theta)}{\partial \theta} = I_{\theta} (\hat{\theta}(x) - \theta)$$

for an efficient unbiased estimator.

Integrating gives

$$\ln p(x|\theta) = \int I(\theta') (\hat{\theta}(x) - \theta') d\theta' + \ln h(x)$$

where $\ln h(x)$ is an arbitrary function of x .

$$\therefore p(x|\theta) = h(x) g(T(x), \theta)$$

which is the Neyman Fisher factorization theorem and $T(x)$ is a sufficient statistic for θ .

3 / In detection theory, there are essentially two types of errors that can be made.

If we say that a signal is present when in fact it is not then we make an error of the first kind — also known as the "false alarm error".

If we say that a signal is not present when in fact it is, then we make an error of the second kind.

The log likelihood ratio allows us to rank one hypothesis versus another.

This can be used for either model selection or for detection. If the ratio is > 1 this allows us to pick the "better" hypothesis.

3 cont

MAP detection chooses the hypothesis which has the largest a-posteriori probability.

Choose H_i corresponding to $\max_{H_j} (P(H_j|y))$.

$$P(H_j|y) = \frac{P(y|H_j) P(H_j)}{P(y)} \quad (\text{Bayes})$$

\therefore choose H_i corresponding to

$\max_{H_j} \{P(y H_j) P(H_j)\}$ $= P(y H_i) P(H_i)$
--

The average error probability is

$$P_e = \sum_{i=1}^m P(D_i|H_1) P(H_1) - P(D_i|H_1) P(H_1)$$

$$+ \sum_{i=1}^m P(D_i|H_2) P(H_2) - P(D_i|H_2) P(H_2)$$

$$+ \dots + \sum_{i=1}^m P(D_i|H_m) P(H_m) - P(D_i|H_m) P(H_m)$$

3

but

$$\prod_{i=1}^m p(D_i | H_i) = 1$$

$$\therefore P_e = \sum_{j=1}^7 p(H_j) - \sum_{i=1}^7 p(D_i | H_i) p(H_i)$$

$$\therefore P_e = 1 - \sum_{i=1}^7 p(D_i | H_i) p(H_i)$$

3

4/ If we wish to compare the two following hypotheses,

$$P(y | H_0) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2} y^2}$$

and

$$P(y | H_1) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2} (y-A)^2}$$

Then we have.

$$\frac{P(y | H_1)}{P(y | H_0)} = e^{-\frac{1}{2\sigma^2} (-2Ay + A^2)} \underset{H_0}{\underset{H_1}{>}} \lambda \quad (1)$$

Where λ is a threshold.

The N-P hypothesis test requires that the false alarm probability be

$$P(D_1 | H_0) = \int_{y \in R_1} P(y | H_0) dy = \alpha$$

$$\int_{y_T}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-y^2/2\sigma^2} dy = \alpha$$

4 / let $\frac{y}{\sqrt{2}\sigma} = u.$

$$\alpha = \int_{u_T}^{\infty} e^{-u^2} du.$$

where y_T is given by the equation of (1).

$$\therefore \alpha = \frac{1}{2} \operatorname{erfc} \left(\frac{1}{\sqrt{2}} \left(\frac{1}{2} \left(\frac{A}{\sigma} \right) + \frac{\sigma}{A} \log \lambda \right) \right)$$

The detection and false alarm probabilities are

$$P(D, | H_1) = \int_{y_T}^{\infty} P(y | H_1) dy \quad (\text{det. prob})$$

$$P(D, | H_0) = \int_{y_T}^{\infty} P(y | H_0) dy \quad (\text{F.A. prob})$$

Slope of ROC curve is

$$\text{slope} = \frac{dP(D, | H_1)}{dP(D, | H_0)} = \frac{dP(D, | H_1)}{dy_T} \cdot \frac{dy_T}{dP(D, | H_0)}$$

but $\frac{dP(D, | H_1)}{dy_T} = -P(y_T | H_1), \quad \frac{dP(D, | H_0)}{dy_T} = -P(y_T | H_0)$

$$\frac{4}{\therefore} \frac{dP(D, H_1)}{dP(D, H_0)} = \frac{P(y_T | H_1)}{P(y_T | H_0)} = \lambda.$$

$\therefore \lambda$ is the slope of the ROC curve at the required F.A. prob α .