Module 4F6 2004. ✓

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 expection (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

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

Maximizing entropy subject to these constraints requires maximizing the cost function

$$Q = -\int p \ln p \, dx - \alpha \int p(x)\, dx - \beta \int x\, p(x)\, dx$$

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

Differentiating Q w.r.t p gives

$$\frac{\partial Q}{\partial p} = -\int_0^\infty \left[\ln p - 1 + \alpha + \beta x\right] dx$$

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

We now use the 2 constants to find α and β

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

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

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

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

1/ Using the second constraint

$$\text{mean} = \int x \, p(x) \, dx$$

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

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

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

$$= \frac{1}{\beta^2}.$$

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

The distribution having maximum entropy is

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

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

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

$$H = -\int_0^\infty p(x) \ln p(x)\, dx$$

$$= -\int \lambda e^{-\lambda x} (\ln\lambda - \lambda x)\, dx.$$

$$= -\lambda \ln\lambda \int e^{-\lambda x} dx + \lambda^2 \int x e^{-\lambda x} dx$$

$$= -\ln\lambda + 1.$$

The variance is

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

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

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

$$\therefore H = \ln\sigma + 1.$$

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 of θ we have by definition

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

Differentiating w.r.t θ we have

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

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

2 cont.)

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

and $\int p(x|\theta)\, dx = 1$ (normalization)

we have

$$\int dx\, \frac{\partial}{\partial\theta} \ln p(x|\theta) \cdot p(x|\theta) \left(\hat{\theta}(x) - \theta\right) = 1$$

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 = constant)

we obtain

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

$$\geq 1$$

2 cont.

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

we have

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

The condition for an unbiased *efficient* estimator is found when the above inequality is an equality (ie $g(x) = A f(x)$)

In the above,

$$g(x) \triangleq \frac{\partial}{\partial \theta} \ln p(x|\theta) \cdot \sqrt{p(x|\theta)}.$$

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

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

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

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

2 cont)

Taking expectations, we have

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

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

for an efficient unbiased estimator.

Integrating gives

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

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

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

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 error 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 $\max_{H_j} \left(p(H_j \mid y)\right)$.

$$p(H_j \mid y) = \frac{p(y \mid H_j)\, p(H_j)}{p(y)} \quad \text{(Bayes)}.$$

$\therefore$ choose H_i corresponding to

$$\boxed{\max_{H_j} \left\{ p(y \mid H_j)\, p(H_j) \right\} = p(y \mid H_i)\, p(H_i)}$$

The average error probability is

$$P_e = \sum_{i=1}^{m} p(D_i \mid H_1)\, p(H_1) - p(D_1 \mid H_1)\, p(H_1)$$

$$+ \sum p(D_i \mid H_2)\, p(H_2) - p(D_1 \mid H_2)\, p(H_2)$$

$$+ \cdots \sum_{i=1}^{m} p(D_i \mid H_m)\, p(H_m) - p(D_1 \mid H_m)\, p(H_m).$$

3/ but $\sum_{i=1}^{m} p(D_i \mid H_i) = 1$

$$\therefore P_e = \sum p(H_j) - \sum p(D_i \mid H_i)\, p(H_i)$$

$$\therefore P_e = 1 - \sum_{i=1}^{m} p(D_i \mid H_i)\, p(H_i)$$

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

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

and

$$p(y \mid H_1) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2}(y-A)^2}.$$

Then we have

$$\frac{p(y \mid H_1)}{p(y \mid H_0)} = e^{-\frac{1}{2\sigma^2}(-2Ay + A^2)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \quad -(1$$

Where λ is a threshold.

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

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

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

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

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

where y_T is given by the equality 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

$$P(D_1|H_1) = \int_{y_T}^{\infty} p(y|H_1)\,dy \qquad (\text{det. prob})$$

$$P(D_1|H_0) = \int_{y_T}^{\infty} p(y|H_0)\,dy \qquad (\text{F.A. prob})$$

Slope of ROC curve is

$$\text{slope} = \frac{dP(D_1|H_1)}{dP(D_1|H_0)} = \frac{dP(D_1|H_1)}{dy_T} \cdot \frac{dy_T}{dP(D_1|H_0)}$$

but $\frac{dP(D_1|H_1)}{dy_T} = -p(y_T|H_1)$, $\frac{dP(D_1|H_0)}{dy_T} = -p(y_T|H_0)$

3

4/ $\therefore \dfrac{dP(P_1|H_1)}{dP(D_1|H_0)} = \dfrac{p(y_T|H_1)}{p(y_T/H_0)} = \lambda.$

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

—//—

$$p_1(y) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - 1}{\sigma}\right)^2\right)$$

$$p_0(y) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2}\sum\left(\frac{y_i}{\sigma}\right)^2\right).$$

$\therefore$ decision rule is : choose H_1 if

$$\exp\left(\frac{1}{2}\sum_{i=1}^{N}\left(\frac{2y_i - 1}{\sigma}\right)\right) \geq \lambda.$$