Maximum Likelihood Estimation - Constant DC component estimation

Andersen Ang

First created: 2014. Last update : 2017

1 The estimation problem

Consider to estimate a DC signal in the addictive white Gaussian noise environment $x[n] = A + w[n]$ over N samples ($n = 0, 1, ..., N-1$).

- A is a constant, but unknown. It is the parameter to be estimated : $\theta = A$. So in this case $p\left(x;\theta\right) = p\left(x;A\right)$
- w[n] is a addictive white Gaussian noise $w[n] \sim \mathcal{N}\left(0, \sigma^2\right)$.

2 The estimation for $N = 1$: Single observation

Since there is only 1 observed data $(N = 1)$, so what we have is

$$x[0] = A + w[0]$$

The conditional PDF of the model for single obervation is

$$p\left(x[0];A\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\Big(x[0]-A\Big)^2\right\}$$

The log-likelihood function will be

$$\ln p\left(x[0];A\right) = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\Big(x[0]-A\Big)^2$$

Take$\dfrac{\partial}{\partial A}$

$$\frac{\partial \ln p\left(x[0];A\right)}{\partial A} = \frac{x[0]-A}{\sigma^2}$$

Take $\dfrac{\partial}{\partial A}$ again

$$\frac{\partial^2 \ln p\left(x[0];A\right)}{\partial A^2} = \frac{-1}{\sigma^2}$$

The Fisher Information will be

$$I\left(A\right) = -\mathbb{E}\left[\frac{\partial^2 \ln p\left(x[0];A\right)}{\partial A^2}\right] = \frac{1}{\sigma^2}$$

Thus the variance of the estimrator will be

$$Var\left(\hat{A}\right) = \frac{1}{I\left(\theta\right)} = \sigma^2$$

Which is the variance of the addictive noise. Notice that this is exactly the Cramer-Rao Lower Bound of this estimator, that means

The lower bound of the variance of the estimator $\hat{A} = \sigma^2$

It's physical meaning is that, if the variance of the addictive noise is small, then the variance of the estimator, (on average) will be as small as the variance of the noise. This is natural, because in the estimation model $x[0] = A + w[0]$, only noise $w[0]$ corrupt our measurement of the A, and therefore a samller noise variance should make the measurement of the DC component easier (having smaller estimator variance).

3 The estimation for N observation

Now instead of 1 observation, we have N observations

$$\begin{aligned} x[0] &= A + w[0] \\ x[1] &= A + w[1] \\ &\vdots \\ x[N-1] &= A + w[N-1] \end{aligned}$$

That means

$$x[n] = A + w[n] \; n = 0, 1, ..., N-1$$

Each data point will have a PDF

$$p\left(x[k]\,;\,A\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x[k] - A\right)^2\right\}$$

By the multiplication law of probability, we have

$$p\left(x[0], x[1], ..., x[N-1]\,;\,A\right) = p\left(x[0]; A\right) p\left(x[1]; A\right) ... p\left(x[N-1]; A\right)$$

Denote $p\left(x[0], x[1], ..., x[N-1]\,;\,A\right)$ as $p(x; A)$, the PDF of the model will be

$$p\left(x; A\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x[0] - A\right)^2\right\}\right) ... \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x[N-1] - A\right)^2\right\}\right)$$

Using compact notation

$$p\left(x; A\right) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \prod_{n=0}^{N-1} \exp\left\{-\frac{1}{2\sigma^2}\left(x[n] - A\right)^2\right\}$$

Since $\exp\alpha \exp\beta = \exp(\alpha + \beta)$, the likelihood function is

$$p(x; A) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left(x[n] - A\right)^2\right\}$$

The log-likelihood function will be

$$\ln p\left(x; A\right) = -\frac{N}{2} \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left(x\left(n\right) - A\right)^2$$

Take $\dfrac{\partial}{\partial A}$ and $\dfrac{\partial^2}{\partial A^2}$

$$\frac{\partial \ln p\left(x; A\right)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(x\left(n\right) - A\right) = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} x\left(n\right) - \frac{NA}{\sigma^2}$$

$$\frac{\partial^2 \ln p\left(x; A\right)}{\partial A^2} = \frac{-N}{\sigma^2}$$

The Fisher Information Matrix is

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln p(x; A)}{\partial A^2}\right] = \frac{N}{\sigma^2}$$

Thus the variance of the estimrator will be

$$Var\left(\hat{A}\right) = \frac{1}{I(\theta)} = \frac{\sigma^2}{N}$$

Which is also the Cramer-Rao Lower Bound.

Now for N-sample estimation, the variance is scaled by the factor $\frac{1}{N}$. It means the more data you have (larger N), the more accurate your estimator will be (lower variance)