Mini-Review of the Main Points from P598AEM Lecture 11:

Consider carrying out N "experiments" where for ***each*** experiment, we perform a ***single*** measurement x_i of a ***random variable*** x, and where we are trying to "determine" (*i.e.* ***estimate***) a single parameter λ from the ***collection*** (***ensemble***) of the N measurements of x_i.

Later, when we look at practical schemes, we will treat the case of several measurements per experiment and also the (simultaneous) determination of several parameters.

After we have carried out N independent measurements of the random variable x: $x_1, x_1, \ldots, x_N$, we construct a function $S(x_1, \cdots, x_N)$ whose ***value*** is an ***estimator*** of the parameter of interest λ.

The ***numerical value*** of the function S ("the ***estimator***") is ***itself*** a ***random variable***. (*e.g.* repeat the N "experiments" 10^6 times – the value of S for one repetition will be similar, but not identical to those obtained for S in the other 99999 repetitions). Thus, S has a mean and a width , along with higher moments of the S-distribution…

We define the ***bias*** $B(\lambda)$ associated with the ***estimator*** S of the parameter λ as: $B(\lambda) \equiv E[S] - \lambda$

A "***good estimator***" will be ***unbiased***, i.e. have $B(\lambda) \equiv E[S] - \lambda = 0$. Then: $E[S] = \lambda$.

Let $f(x;\lambda)$ be the P.D.F. describing the measurement of the random variable x. λ is the parameter we wish to determine by a series of measurements $x_1, \cdots, x_N$. The joint P.D.F. of this series is $f(x_1, x_2, \cdots, x_N; \lambda) = f(x_1;\lambda) \cdot f(x_2;\lambda) \cdot \ldots \cdot f(x_N;\lambda)$ provided that the N individual measurements of x_i are independent.

Define: $$\mathcal{L}(x_1, x_2, \ldots, x_N; \lambda) \equiv \prod_{i=1}^{N} f(x_i;\lambda)$$

Define: $$\ell(x_1, x_2, \ldots, x_N; \lambda) \equiv \ln \mathcal{L}(x_1, x_2, \ldots, x_N; \lambda) = \sum_{i=1}^{N} \ln f(x_i;\lambda)$$

Define: $$\ell'(x_1, x_2, \ldots, x_N; \lambda) \equiv \frac{d\ell(x_1, x_2, \ldots, x_N; \lambda)}{d\lambda} \equiv \frac{d}{d\lambda} \ln \mathcal{L}(x_1, x_2, \ldots, x_N; \lambda) = \frac{d}{d\lambda}\left\{\sum_{i=1}^{N} \ln f(x_i;\lambda)\right\}$$

Define: $$\ell''(x_1, x_2, \ldots, x_N; \lambda) \equiv \frac{d^2\ell(x_1, x_2, \ldots, x_N; \lambda)}{d\lambda^2} \equiv \frac{d^2}{d\lambda^2} \ln \mathcal{L}(x_1, x_2, \ldots, x_N; \lambda) = \frac{d^2}{d\lambda^2}\left\{\sum_{i=1}^{N} \ln f(x_i;\lambda)\right\}$$

We then showed (for ***unbiased estimators*** S) that a ***lower bound*** existed for the ***variance*** of S:

$$\sigma_S^2 \geq \frac{1}{\sigma_{\ell'}^2} = \frac{1}{E[\ell'^2(\lambda)]} = -\frac{1}{E[\ell''(\lambda)]} = \frac{1}{I(\lambda)}$$

Where: $$E[\ell'^2(\lambda)] = E\left[\left\{\frac{d}{d\lambda}\ln \mathcal{L}(\lambda)\right\}^2\right] = -E[\ell''(\lambda)] = -E\left[\frac{d^2}{d\lambda^2}\ln \mathcal{L}(\lambda)\right] \equiv I(\lambda)$$ ←

$I(\lambda)$ is called the ***Information*** of the sample associated with the parameter λ.

More Information on *Information*:

The ***Information*** $I(\lambda) \equiv E[\ell'^2(\lambda)] = E\left[\left\{\frac{d}{dl}\ln\mathcal{L}(\lambda)\right\}^2\right] = -E[\ell''(\lambda)] = -E\left[\frac{d^2}{d\lambda^2}\ln\mathcal{L}(\lambda)\right]$

depends (linearly) on N, the number of measurements carried out in the experiment.

We can show that $I(\lambda)$ depends on the P.D.F. $f(x;\lambda)$, but does ***not*** depend on the ***particular*** values of the individual x_i's measured in the experiment.

In P598AEM Lect. Notes 11 (*p.* 6), we defined the quantity: $\phi(x_i;\lambda) \equiv \frac{f'(x_i;\lambda)}{f(x_i;\lambda)}$ and showed that:

$$\ell'(x_1,x_2,\ldots,x_N;\lambda) = \sum_{i=1}^{N}\frac{d}{d\lambda}\ln f(x_i;\lambda) = \sum_{i=1}^{N}\frac{df(x_i;\lambda)/d\lambda}{f(x_i;\lambda)} = \sum_{i=1}^{N}\frac{f'(x_i;\lambda)}{f(x_i;\lambda)} \equiv \sum_{i=1}^{N}\phi(x_i;\lambda)$$

Then:

$$\begin{aligned} E[\ell'^2(\lambda)] &= E\left[\sum_{i=1}^{N}\phi(x_i;\lambda)\sum_{j=1}^{N}\phi(x_j;\lambda)\right] \\ &= E\left[\sum_{i=1}^{N}\{\phi(x_i;\lambda)\}^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\phi(x_i;\lambda)\phi(x_j;\lambda)\right] \\ &= \sum_{i=1}^{N}E\left[\{\phi(x_i;\lambda)\}^2\right] + \sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[\phi(x_i;\lambda)\phi(x_j;\lambda)\right] \\ &= \sum_{i=1}^{N}E\left[\{\phi(x_i;\lambda)\}^2\right] + \sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[\phi(x_i;\lambda)\right]\cdot E\left[\phi(x_j;\lambda)\right] \end{aligned}$$

Since the ***N*** measurements of the $\boldsymbol{x_i}$ are ***independent***.

Next, we show that $E\left[\phi(x_i;\lambda)\right] = 0$. Take $\frac{d}{d\lambda}$ of: $\left\{1 = \int f(x;\lambda)dx\right\}$, which yields:

$$0 = \int f'(x;\lambda)dx = \int \frac{f'(x;\lambda)}{f(x;\lambda)} f(x;\lambda)dx = E\left[\phi(x;\lambda)\right]$$

Thus:

$$\begin{aligned} E[\ell'^2(\lambda)] &= \sum_{i=1}^{N}E\left[\{\phi(x_i;\lambda)\}^2\right] = \sum_{i=1}^{N}\int\phi^2(x_i;\lambda)f(x_i;\lambda)dx \\ &= N\int\phi^2(x_i;\lambda)f(x_i;\lambda)dx = NE\left[\phi^2(x_i;\lambda)\right] \end{aligned}$$

Hence, the ***Information***:

$$I(\lambda) = E[\ell'^2(\lambda)] = -E[\ell''(\lambda)] = NE\left[\sum_{i=1}^{N}\phi^2(x_i;\lambda)\right] = NE\left[\left(\sum_{i=1}^{N}\frac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right)^2\right] = NE\left[\left(\sum_{i=1}^{N}\frac{d}{d\lambda}\ln f(x_i;\lambda)\right)^2\right]$$

simply depends on the ***form*** of the P.D.F. of the measurements $f(x;\lambda)$, and also depends ***linearly*** on the ***number*** of measurements, N.

Finally, for ***unbiased* estimators**, the ***lower bound*** on the ***variance*** of S is:

$$\sigma_S^2 \geq \frac{1}{\sigma_{\ell'}^2} = \frac{1}{I(\lambda)} = \frac{1}{E[\ell'^2(\lambda)]} = -\frac{1}{E[\ell''(\lambda)]} = \frac{1}{NE\left[\sum_{i=1}^{N}\phi^2(x_i;\lambda)\right]}$$

$$= \frac{1}{NE\left[\left(\sum_{i=1}^{N}\frac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right)^2\right]} = \frac{1}{NE\left[\left(\sum_{i=1}^{N}\frac{d}{d\lambda}\ln f(x_i;\lambda)\right)^2\right]}$$

A ***minimum variance* estimator** is a choice of S for which the ***equality*** holds in the above relation – i.e. (in the ***unbiased*** case):

$$\sigma_S^2 = \frac{1}{\sigma_{\ell'}^2} = \frac{1}{I(\lambda)} = \frac{1}{E[\ell'^2(\lambda)]} = -\frac{1}{E[\ell''(\lambda)]} = \frac{1}{NE\left[\sum_{i=1}^{N}\phi^2(x_i;\lambda)\right]}$$

$$= \frac{1}{NE\left[\left(\sum_{i=1}^{N}\frac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right)^2\right]} = \frac{1}{NE\left[\left(\sum_{i=1}^{N}\frac{d}{d\lambda}\ln f(x_i;\lambda)\right)^2\right]}$$

We recall (P598AEM Lect. Notes 4) that the ***covariance inequality*** became an ***equality*** when the variables were ***linearly*** related (***specifically***: $\text{cov}(x,y) = \pm 1$ if $y = ax + b$).

The ***covariance*** here is: $\text{cov}(S,\ell')$, so the ***equality*** will hold if: $\ell'(\lambda) = rS + t$, where r and t are "***constants***", *i.e.* they do ***not*** depend on $x_1, x_2, \ldots, x_N$, (but they ***can/may*** depend on λ).

Recall that (by definition), S ***cannot*** be a function of λ, *i.e.* $S \neq fcn(\lambda)$.

Also, recall that $E[\ell'(\lambda)] = 0$. Thus: $E[\ell'(\lambda) = rS + t] \;\Rightarrow\; 0 = rE[S] + t$, or: $t = -rE[S]$.

Using this, we can rewrite: $\ell'(\lambda) = rS + t = r\left(S - E[S]\right)$

Thus, if we have an ***unbiased* estimator** *S*, for which $\ell'(\lambda) = r(\lambda)\cdot\{S(\lambda) - E[S]\}$ then the ***variance*** of the ***unbiased* estimator** *S* is a ***minimum***.

The condition $\ell' = r(\lambda)\cdot\{S - E[S]\}$ can be shown to be ***necessary*** and ***sufficient*** for the ***variance*** of the ***unbiased* estimator** *S* to be a ***minimum***.

Let us assume that we ***have*** found such an ***unbiased* estimator** *S*, satisfying:

$$\ell'(\lambda) = r(\lambda)\cdot\{S - E[S]\}$$

Then: $$\ell'^2(\lambda) = r^2(\lambda)\cdot\{S - E[S]\}^2$$

And: $$E[\ell'^2(\lambda)] = r^2(\lambda)\underbrace{E\left[\{S - E[S]\}^2\right]}_{\equiv\,\sigma_S^2} = r^2(\lambda)\sigma_S^2 = I(\lambda)$$

Since this is a ***minimum*** ***variance*** ***unbiased*** ***estimator***, then: $\sigma_S^2 = \dfrac{1}{I(\lambda)}$

Hence: $\sigma_S^2 = \dfrac{1}{I(\lambda)} = \dfrac{1}{r^2(\lambda)\sigma_S^2}$ or: $\sigma_S^4 = \dfrac{1}{r^2(\lambda)}$ or: $\sigma_S^2 = \dfrac{1}{r(\lambda)}$

Thus, once we find $\ell'(\lambda) = r(\lambda)\{S - E[S]\}$, we can immediately determine $\sigma_S^2 = 1/r(\lambda)$.

Let us return now to the "weighted mean" example that we worked on in P598AEM Lecture 11, where we had N independent measurements $x_1, x_2, \ldots, x_N$ of the random variable x described by the Gaussian P.D.F. $f(x;\alpha) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\alpha)^2/2\sigma^2}$ and which led to:

$$\ell(\alpha) \equiv \ln \mathcal{L}(\alpha) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - \alpha)^2}{\sigma_i^2} + \text{ constant}$$

$$\ell'(\alpha) \equiv \frac{d\ell(\alpha)}{d\alpha} \equiv \frac{d \ln \mathcal{L}(\alpha)}{d\alpha} = +\sum_{i=1}^{N} \frac{x_i - \alpha}{\sigma_i^2}$$

and the ***estimator*** S (for the value of α which maximizes the above $\ell(\alpha)$), which (here in this case) is the ***weighted mean***:

$$S = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \quad (n.b.\text{ previously, we called this } \alpha^*)$$

Then: $$E[S] = E\left[\frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}\right] = \frac{\sum_{i=1}^{N} \frac{E[x_i]}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^{N} \frac{\hat{x}}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} = \hat{x}\left(\frac{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}\right) = \hat{x} = \alpha$$

Rewrite: $$\ell'(\alpha) = \sum_{i=1}^{N} \frac{x_i - \alpha}{\sigma_i^2} = \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} - \sum_{i=1}^{N} \frac{\alpha}{\sigma_i^2} = \left\{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right\}\left\{\frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} - \alpha\right\} = \left\{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right\}\{S - \alpha\}$$

This ***is*** of the form: $\ell'(\alpha) = r(\alpha)\{S - E[S]\} = \left\{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}\right\}\{S - \alpha\}$

Since: $E[S] = \alpha$, and (here): $\dfrac{1}{\sigma_S^2} = \dfrac{1}{\sigma_\alpha^2} = r(\alpha) = \sum_{i=1}^{N} \dfrac{1}{\sigma_i^2} \quad (\neq fcn(\alpha))$

So we have shown that S is an ***unbiased***, ***minimum*** ***variance*** ***estimator*** of the parameter α.

The ***variance*** of the ***estimator*** *S*: $\sigma_S^2 = \sigma_\alpha^2 = \dfrac{1}{r(\alpha)} = \dfrac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$ or: $\dfrac{1}{\sigma_S^2} = \dfrac{1}{\sigma_\alpha^2} = \sum_{i=1}^{N} \dfrac{1}{\sigma_i^2}$

In the case of $\sigma_i^2 = \sigma^2$ (*i.e.* "equal errors") then: $\dfrac{1}{\sigma_S^2} = \dfrac{1}{\sigma_\alpha^2} = \dfrac{N}{\sigma^2}$, and: $\sigma_S = \sigma_\alpha = \dfrac{\sigma}{\sqrt{N}}$ √√

Actually, we could have arrived at the same result for $1/\sigma_\alpha^2$ simply by doing ***error propagation*** (assuming small uncertainties). But we now know that the ***estimator*** S for finding α when the P.D.F. is a Gaussian, $f(x;\alpha) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\alpha)^2/2\sigma^2}$ is ***the*** best there is!

This ***estimator*** S is ***unbiased***, has ***minimum variance*** and leads to an estimate of α that is closer to the "***true***" value than ***any other*** estimate!

For many calculations, it is useful to re-express the ***Information***:

$$I(\lambda) = E[\ell'^2(\lambda)] = -E[\ell''(\lambda)] = NE\left[\sum_{i=1}^{N} \phi^2(x_i;\lambda)\right] = NE\left[\left(\sum_{i=1}^{N} \frac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right)^2\right] = NE\left[\left(\sum_{i=1}^{N} \frac{d}{d\lambda} \ln f(x_i;\lambda)\right)^2\right]$$

Consider: $\ell(\lambda) = \sum_{i=1}^{N} \ln f(x_i;\lambda)$ and: $\ell'(\lambda) = \sum_{i=1}^{N} \dfrac{d}{d\lambda} \ln f(x_i;\lambda) = \sum_{i=1}^{N} \dfrac{f'(x_i;\lambda)}{f(x_i;\lambda)}$

Then: $\ell''(\lambda) = \sum_{i=1}^{N} \dfrac{d}{d\lambda}\left(\dfrac{f'}{f}\right) = \sum_{i=1}^{N} \dfrac{f\,f'' - f'\,f'}{f^2} = \sum_{i=1}^{N} \dfrac{f''}{f} - \sum_{i=1}^{N}\left(\dfrac{f'}{f}\right)^2$

Thus: $E[\ell''(\lambda)] = E\left[\sum_{i=1}^{N} \dfrac{f''}{f}\right] - E\left[\sum_{i=1}^{N}\left(\dfrac{f'}{f}\right)^2\right] = \underbrace{E\left[\sum_{i=1}^{N} \dfrac{f''}{f}\right]}_{=0} - E[\ell'^2(\lambda)] = -E[\ell'^2(\lambda)]$

We can show that the first term on the RHS is zero:

Start with: $1 = \int f(x;\lambda)\,dx$

Take $\dfrac{d}{d\lambda}$: $0 = \int f'\,dx$

Take another $\dfrac{d}{d\lambda}$: $0 = \int f''\,dx = \int \left[\dfrac{f''}{f}\right] f\,dx = E\left[\dfrac{f''}{f}\right]$

$\therefore$ $I(\lambda) = E[\ell'^2(\lambda)] = -E[\ell''(\lambda)]$

For example, for N ***independent*** measurements of a Gaussian-distributed ***random variable*** *x*, we saw that the ***log likelihood*** was:

$$\ell(\alpha) = -\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\alpha)^2}{\sigma_i^2} \quad + \quad \text{constant}$$

$$\ell'(\alpha) = +\sum_{i=1}^{N}\frac{x_i-\alpha}{\sigma_i^2}$$

$$\ell''(\alpha) = -\sum_{i=1}^{N}\frac{1}{\sigma_i^2}$$

Thus: $$I(\alpha) = E[\ell'^2(\alpha)] = -E[\ell''(\alpha)] = \sum_{i=1}^{N}\frac{1}{\sigma_i^2}$$

And: $$\sigma_\alpha^2 = \frac{1}{I(\alpha)} = \frac{1}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}}$$ as we have already seen.

It is interesting to note, from the above, that for Gaussian-distributed measurements of a ***random variable*** *x*, the ***Information*** $I(\alpha) = \sum_{i=1}^{N}\frac{1}{\sigma_i^2}$, is ***constant***, independent of the *x*'s or their ***expectation value***, and depends ***only*** on the individual σ 's.

Let us now look at measurements described by the ***Poisson Distribution***:

$$P(n;\nu) = \frac{\nu^n}{n!}e^{-\nu} = \text{ probability of observing } n \text{ events when } \nu \text{ are expected.}$$

We have already seen that: $E[n] = \nu$ and: $\sigma_n^2 = \nu$.

Suppose that we are studying some phenomenon which we believe is ***Poisson distributed***. Let us consider, for example, an "experiment" in which we count for one minute, and that *x* = *n*, the number of counts recorded. We repeat the experiment N times, and then calculate the ***likelihood***:

$$\mathcal{L} = \prod_{i=1}^{N}P(x_i;\nu) = \prod_{i=1}^{N}\frac{\nu^{x_i}}{(x_i)!}e^{-\nu}$$

$$\ell(\nu) = \ln\mathcal{L}(\nu) = \sum_{i=1}^{N}\left\{-\nu + x_i\ln\nu - \ln(x_i!)\right\}$$

$$\ell'(\nu) = \frac{d\ell(\nu)}{d\nu} = \frac{d\ln\mathcal{L}(\nu)}{d\nu} = \sum_{i=1}^{N}\left\{-1+\frac{x_i}{\nu}\right\}$$

We solve for the value of $v = v^*$, at which $\ell'(v^*) = \dfrac{d\ell(v^*)}{dv} = 0$:

$$0 = \ell'(v)\Big|_{v=v^*} = \left.\frac{d\ell(v)}{dv}\right|_{v=v^*} = \sum_{i=1}^{N}\left\{-1+\frac{x_i}{v^*}\right\} = -N + \frac{1}{v^*}\sum_{i=1}^{N} x_i = -N + \frac{1}{v^*}N\,\overline{x} = N\left(\frac{\overline{x}}{v^*}-1\right)$$

which yields the ***estimator***: $S = v^* = \dfrac{1}{N}\sum_{i=1}^{N} x_i = \overline{x}$, the ***sample mean*** – *i.e.* ***simple/arithmetic mean***.

Note that ***this*** S ***is*** an ***unbiased*** estimator, since: $E[S] = \hat{x} = v^*$

We also have:

$$\ell'(v) = \frac{d\ell(v)}{dv} = \sum_{i=1}^{N}\left\{\frac{x_i}{v}-1\right\} = \frac{1}{v}\sum_{i=1}^{N}\{x_i - v\} = \frac{1}{v}\left\{\sum_{i=1}^{N} x_i - Nv\right\} = \frac{1}{v}\{N\,\overline{x} - Nv\} = \frac{N}{v}\{\,\overline{x}-v\} = N\left\{\frac{\overline{x}}{v}-1\right\}$$

But from above: $S = v^* = \dfrac{1}{N}\sum_{i=1}^{N} x_i = \overline{x}$, the ***sample mean***, and: $E[S] = v$

Thus: $\ell'(v) = \dfrac{N}{v}\{\,\overline{x}-v\} = \dfrac{N}{v}\{v^* - v\} = \dfrac{N}{v}\{S - E[S]\}$

This ***is*** of the form: $\ell'(v) = r(v)\{S - E[S]\}$,

and is ***the necessary*** and ***sufficient*** condition such that ***this*** S ***is*** a ***minimum variance estimator***!

Furthermore, we see from comparing the above $\ell'(v)$ relations, that: $\sigma_{v^*}^2 = \dfrac{1}{r(v^*)} = \dfrac{v^*}{N}$.

Now let us return to the **Maximum Likelihood Method (M.L.M.)**

We will show that this completely general method is ***unbiased*** and also gives ***minimum variance*** for the case of $N \to \infty$.

The M.L.M. takes $\ell(x_1, x_2, \ldots, x_N; \lambda) \equiv \ln \mathcal{L}(x_1, x_2, \ldots, x_N; \lambda)$ and evaluates it for measured random variables $x_1, x_2, \ldots, x_N$.

Then $\ell'(\lambda) = \dfrac{d}{d\lambda}\ell(\lambda) = \dfrac{d}{d\lambda}\ln\mathcal{L}(\lambda)$ is set to zero.

Let us assume that the solution for $\ell'(\lambda) = 0$ has only a single root, at: $\lambda = \lambda^*$.

Then the ***estimator*** S is just λ^*. Note that this ***defines*** the ***estimator*** S ***implicitly***, rather than as an ***explicit*** function $S(x_1, x_2, \ldots, x_N)$.

In some cases (as in the previous examples, above), S turns out to be an ***explicit*** function $S(x_1, x_2, \ldots, x_N)$, but this will ***not*** be true in general.

In order to handle this, we expand $\ell'(\lambda)$ in a Taylor series about $\lambda = \lambda^*$:

$$\ell'(\lambda) = \underbrace{\ell'(\lambda^*)}_{=0} + (\lambda - \lambda^*)\ell''(\lambda^*) + \ldots = (\lambda - \lambda^*)\ell''(\lambda^*) + \ldots \quad \text{since } \ell'(\lambda^*) = 0.$$

Now: $\dfrac{\ell''(\lambda^*)}{N} = \dfrac{1}{N}\sum_{i=1}^{N}\dfrac{d}{d\lambda}\left\{\dfrac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right\}\Bigg|_{\lambda=\lambda^*}$

If N is ***large enough***, we ***can*** replace the ***sample mean*** $\frac{1}{N}\sum_{i=1}^{N} Q_i$ by the ***expectation value*** $E[Q]$.

$\therefore$ for ***large*** N: $\qquad \dfrac{\ell''(\lambda^*)}{N} \Rightarrow E\left[\sum_{i=1}^{N}\dfrac{d}{d\lambda}\left\{\dfrac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right\}\Bigg|_{\lambda=\lambda^*}\right]$

Since the set of N measurements all originate from the ***same*** P.D.F., the ***expectation value*** on the RHS is the ***same*** for all i:

$$\ell''(\lambda^*) \Rightarrow NE\left[\sum_{i=1}^{N}\frac{d}{d\lambda}\left\{\frac{f'(x_i;\lambda)}{f(x_i;\lambda)}\right\}\Bigg|_{\lambda=\lambda^*}\right] = NE\left[\sum_{i=1}^{N}\frac{d^2}{d\lambda^2}\ln f(x_i;\lambda)\Bigg|_{\lambda=\lambda^*}\right]$$

In fact, the expression on the right is just: $E[\ell''(\lambda)|_{\lambda=\lambda^*}] = E[\ell''(\lambda^*)] = -I(\lambda^*)$

Thus, for N ***large enough***, we can expand $\ell'(\lambda)$ in a Taylor series about $\lambda = \lambda^*$:

$$\ell'(\lambda) = (\lambda - \lambda^*)\ell''(\lambda^*) + \cdots \;\simeq\; -I(\lambda^*)(\lambda - \lambda^*) + \cdots$$

We assume that higher orders can be neglected (for small $|\lambda - \lambda^*|$) and so:

$$\frac{d}{d\lambda}\ell(\lambda) = \ell'(\lambda) \;\simeq\; -I(\lambda^*)(\lambda - \lambda^*) \quad (\textit{n.b. } I(\lambda^*) \text{ is just a } \underline{\textbf{number}}.)$$

Solving this 1st-order linear differential equation for $\ell(\lambda)$:

$\therefore \qquad \ell(\lambda) = -\tfrac{1}{2}I(\lambda^*)(\lambda - \lambda^*)^2 + \text{constant} = \ln\mathcal{L}(\lambda)$

And thus: $\qquad \mathcal{L}(\lambda) = Ce^{-(\lambda-\lambda^*)^2/(2/I(\lambda^*))} = Ce^{-(\lambda-\lambda^*)^2/2\sigma_\lambda^2}$, with: $\sigma_\lambda^2 = \dfrac{1}{I(\lambda^*)}$.

So for ***large*** N, we see that the ***likelihood function*** $\mathcal{L}(\lambda)$ ***is*** Gaussian near $\lambda = \lambda^*$ (the mean), and has a ***variance*** $\text{var}(\lambda) \equiv \sigma_\lambda^2 = 1/I(\lambda^*)$.

We are ***estimating*** λ, so we choose the ***estimator*** $S = \lambda^*$.

Since $\lambda^* = \hat{\lambda}$, we see that S is an ***unbiased estimator*** of λ.

Since $\ell'(\lambda) \simeq -I(\lambda^*)(\lambda - \lambda^*) = I(\lambda^*)\{S - E[S]\}$, which ***is*** of the general form:

$$\ell'(\lambda) = r(\lambda)\{S - E[S]\}, \quad \text{hence: } \sigma_\lambda^2 = \frac{1}{r(\lambda)} = \frac{1}{I(\lambda^*)}.$$

Thus, we also see that $S = \lambda^*$ ***is*** a ***minimum variance estimator*** of λ.

Summary:

In the limit $N \to \infty$, the Maximum Likelihood Method (M.L.M.) gives an ***estimate*** of the parameter λ which is ***unbiased*** and of ***minimum variance***. The ***likelihood function*** $\mathcal{L}(\lambda)$ ***is*** Gaussian. The ***mean*** of the parameter λ is λ^*, the ***variance*** of the parameter λ is given by:

$$\sigma_\lambda^2 = \sigma_{\lambda^*}^2 = \frac{1}{I(\lambda^*)} = -\frac{1}{E[\ell''(\lambda^*)]} = \frac{1}{E[\ell'^2(\lambda^*)]}$$

Warning:

In practice, the M.L.M. is often used for N ***very*** much less than ∞. No ***general*** rules exist to determine whether ***all*** of these results hold for ***finite*** N, *e.g.* how much ***Bias*** an ***estimator*** S may have, how far from ***minimum variance*** the solution will be, *etc*. These details will depend on the particular $f(x;\lambda)$ as well as N. ***Proceed with due caution in such situations***!