There are several defintions of optimal estimator.

1. Maximum Likelihood Estimator (MLE): An estimator that maximizes the likelihood function, making it the most probable given the observed data.

$$\max_{\hat{x}} \mathbb{P}\left[y = \hat{x}\right]$$

   Usually there is a dataset $\{y^{(i)}\}$ of $i = 1, \ldots, N$ measured y values. In this case, we maximize joint probability

$$\max_{\hat{x}} \mathbb{P}\left[y^{(1)} = \hat{x}, \ldots, y^{(N)} = \hat{x}\right]$$

2. Least Squares Estimator (LSE): An estimator that minimizes the expected value of the squared differences between the estimator and the true parameter value.

$$\min_{\hat{x}} \mathbb{E}\left[(\hat{x} - x)^2\right]$$

   Since expectation value of squares can be decomposed into square of expected value and variance (so–called bais–variance decomposition),

$$\mathbb{E}\left[(\hat{x} - x)^2\right] = \left(\mathbb{E}\left[\hat{x} - x\right]\right)^2 + \mathrm{Var}\left[\hat{x}\right]$$

   minimizing squared errors is equivalent to having no bias $\mathbb{E}\left[\hat{x}\right] = x$ and it has the lowest variance among all unbiased estimators, so it is also called Minimum Variance Unbiased Estimator (MVUE).

These two definitions are not the same. MLEs are not necessarily unbiased. Consider normally ditributed variable $y \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then estimator for μ is the same according to both definitions

$$\hat{\mu} = \hat{\mu}_{\mathrm{MLE}} = \hat{\mu}_{\mathrm{MVUE}} = \frac{1}{N} \sum_{i=1}^{N} y^{(i)}$$

However, estimators for variance are different

$$\hat{\sigma^2}_{\mathrm{MLE}} = \frac{1}{N} \sum_{i=1}^{N} \left(y^{(i)} - \hat{\mu}\right)^2, \qquad \hat{\sigma^2}_{\mathrm{MVUE}} = \frac{1}{N-1} \sum_{i=1}^{N} \left(y^{(i)} - \hat{\mu}\right)^2.$$

Question 1 Consider random variables

$$y_1 = x + n_1,$$
$$y_2 = n_1 + n_2$$

where $n_1 \sim \mathcal{N}(0, \sigma_1^2)$, $n_2 \sim \mathcal{N}(0, \sigma_2^2)$ are independent normally distributed random variables and x is a parameter. What is an expression for optimal estimator $\hat{x}$?

Solution 1

Least Squares Estimator (LSE) Assume that answer is given by linear combination $y_\beta = y_1 + \beta y_2$ with yet undetermined coefficient β. Expected value $\mathbb{E}[y_\beta] = x$, so average value

$$\langle y_\beta \rangle = \frac{1}{N} \sum_{i=1}^{N} y_\beta^{(i)}$$

is an unbaised estimator for x indepedently of value of β.

However, optimal estimator should not only have zero bias, but also have minimal variance.

$$\mathrm{Var}[y_\beta] = \mathbb{E}\left[(y_\beta - x)^2\right] = \mathbb{E}\left[(n_1 + \beta[n_1 + n_2])^2\right],$$
$$\frac{\partial}{\partial \beta} \mathrm{Var}[y_\beta] = \mathbb{E}\left[(n_1 + \beta[n_1 + n_2])(n_1 + n_2)\right] = \sigma_1^2 + \beta(\sigma_1^2 + \sigma_2^2)$$

Minimizing variance over β we find $\beta = -\sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ and finally bias–free and minimal–variance estimator for x would be

$$\hat{x} = \frac{1}{N} \sum_{i=1}^{N} \left[y_1^{(i)} - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2^{(i)} \right], \tag{1}$$

Maximum Likelihood Estimator (MLE) Alternatively, we could use likelihood maximum as criterion for optimal estimator. Let us consider $y_3 = y_1 - y_2 = x - n_2$. Probability distribution of datastream $\{y_1^{(i)}, y_3^{(i)}\}$ happening is

$$P\left(\{y_1^{(i)}, y_2^{(i)}\}\right) \propto \prod_{i=1}^{N} \exp\left[-\frac{(y_1^{(i)} - x)^2}{2\sigma_1^2}\right] \exp\left[-\frac{(y_3^{(i)} - x)^2}{2\sigma_2^2}\right]$$

Finding maximum of log–likelihood with respect to x we come to

$$\log P = \sum_{i=1}^{N} \left[-\frac{(y_1^{(i)} - x)^2}{2\sigma_1^2} - \frac{(y_3^{(i)} - x)^2}{2\sigma_2^2} \right] + \mathrm{const}$$
$$\left.\frac{\partial \log P}{\partial x}\right|_{x=\hat{x}} \propto \frac{\langle y_1 \rangle - \hat{x}}{\sigma_1^2} + \frac{\langle y_3 \rangle - \hat{x}}{\sigma_2^2} = 0,$$
$$\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) \hat{x} = \frac{\langle y_1 \rangle}{\sigma_1^2} + \frac{\langle y_3 \rangle}{\sigma_2^2},$$

Finally,

$$\hat{x} = \frac{\sigma_2^2 \langle y_1 \rangle}{\sigma_1^2 + \sigma_2^2} + \frac{\sigma_1^2 \langle y_3 \rangle}{\sigma_1^2 + \sigma_2^2} = \langle y_1 \rangle - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \langle y_2 \rangle.$$

Remebering that $\langle y_3 \rangle = \langle y_1 \rangle - \langle y_2 \rangle$ we see that MLE is the same as MVUE (1) above.

Question 2 Consider random variables $y_1 = x + n_1$ and $y_2 = n_1 + n_2$, where $n_1 \sim \mathcal{N}(0, \sigma_1^2)$, $n_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $x \sim \mathcal{N}(x_0, \sigma_0^2)$ are independent random variables. What is the optimal estimator for x_0?

Solution 2

Least Squares Estimator (LSE) Again, let $y_\beta = y_1 + \beta y_2$, average $\langle y_\beta \rangle$ is a bias–free estimator for x_0 for any value of β since $\mathbb{E}[y_\beta] = \mathbb{E}[x] = x_0$. Let's minimize variance of our estimator.

$$\begin{aligned}
\mathrm{Var}[y_\beta] &= \mathbb{E}\left[(y_\beta - x_0)^2\right] = \mathbb{E}\left[(n_1 + x - x_0 + \beta[n_1 + n_2])^2\right], \\
\frac{\partial}{\partial \beta}\mathrm{Var}[y_\beta] &= \mathbb{E}\left[(n_1 + x - x_0 + \beta[n_1 + n_2])(n_1 + n_2)\right] = \sigma_1^2 + \beta(\sigma_1^2 + \sigma_2^2) = 0 \\
\beta &= \frac{-\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.
\end{aligned}$$

So the answer is same as before (1)

Maximum Likelihood Estimator (MLE) Let's introduce $n_0 \sim \mathcal{N}(0, \sigma_0^2)$, then $y_1 = x_0 + n_1 + n_0$ and $y_2 = n_1 + n_2$. Probability of given dataset is

$$\begin{aligned}
P(y_1, y_2) &= \int \frac{dn_0 dn_1 dn_2}{(2\pi)^{3/2}\sigma_0\sigma_1\sigma_2} \exp\left[-\frac{n_0^2}{2\sigma_0^2} - \frac{n_1^2}{2\sigma_1^2} - \frac{n_2^2}{2\sigma_2^2}\right] \delta(n_0 + n_1 + x_0 - y_1)\delta(n_1 + n_2 - y_2) \\
&= \int \frac{dn_1}{(2\pi)^{3/2}\sigma_0\sigma_1\sigma_2} \exp\left[-\frac{(y_1 - x_0 - n_1)^2}{2\sigma_0^2} - \frac{n_1^2}{2\sigma_1^2} - \frac{(y_2 - n_1)^2}{2\sigma_2^2}\right] \\
&= \int \frac{dn_1}{(2\pi)^{3/2}\sigma_0\sigma_1\sigma_2} \exp\left[-\frac{n_1^2}{2\bar{\sigma}^2} + \frac{(y_1 - x_0)n_1}{\sigma_0^2} + \frac{y_2 n_1}{\sigma_2^2}\right] \exp\left[-\frac{(y_1 - x_0)^2}{2\sigma_0^2} - \frac{y_2^2}{2\sigma_2^2}\right] \\
&= \frac{\bar{\sigma}}{2\pi\sigma_0\sigma_1\sigma_2} \exp\left[\frac{\bar{\sigma}^2}{2}\left(\frac{(y_1 - x_0)}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right)^2\right] \exp\left[-\frac{(y_1 - x_0)^2}{2\sigma_0^2} - \frac{y_2^2}{2\sigma_2^2}\right]
\end{aligned}$$

Here we use notation

$$\frac{1}{\bar{\sigma}^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \qquad\qquad \frac{1}{\sigma_{12}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Condition for optimal estimator $\partial \log P / \partial x_0 = 0$ gives

$$\begin{aligned}
\frac{\bar{\sigma}^2}{\sigma_0^2}\left(\frac{(y_1 - x_0)}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right) &= \frac{y_1 - x_0}{\sigma_0^2} \\
\frac{y_1 - x_0}{\sigma_0^2} + \frac{y_2}{\sigma_2^2} &= \frac{y_1 - x_0}{\bar{\sigma}^2} \\
\frac{y_2}{\sigma_2^2} &= \frac{y_1 - x_0}{\sigma_{12}^2}
\end{aligned}$$

Finally,

$$x_0 = y_1 - \frac{\sigma_{12}^2}{\sigma_2^2} y_2 = y_1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y_2$$

We come to the result that estimator is the same as before (1).

Question 3 Consider random variables $y_1 = x + n_1$ and $y_2 = n_1 + n_2$, where $n_1 \sim \mathcal{N}(0, \sigma_1^2)$, $n_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $x \sim \mathcal{N}(0, \sigma_0^2)$ are independent random variables. What is the optimal estimator for σ_0^2?

Solution 3

Maximum Likelihood Estimator (MLE) Using calculation from before

$$P(y_1, y_2) = \frac{\bar{\sigma}}{2\pi\sigma_0\sigma_1\sigma_2} \exp\left[\frac{\bar{\sigma}^2}{2}\left(\frac{y_1}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right)^2\right] \exp\left[-\frac{y_1^2}{2\sigma_0^2} - \frac{y_2^2}{2\sigma_2^2}\right]$$

Optimal σ_0 satisfies

$$\frac{\partial \log P}{\partial \sigma_0^2} = \frac{1}{2}\left(\frac{y_1}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right)^2 \frac{\partial \bar{\sigma}^2}{\partial \sigma_0^2} - \bar{\sigma}^2\left(\frac{y_1}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right)\frac{y_1}{\sigma_0^4} + \frac{y_1^2}{2\sigma_0^4} - \frac{1}{2\sigma_0^2} + \frac{1}{2\bar{\sigma}^2}\frac{\partial \bar{\sigma}^2}{\partial \sigma_0^2} = 0$$

Let's remeber that

$$\frac{\partial \bar{\sigma}^2}{\partial \sigma_0^2} = -\bar{\sigma}^4 \frac{\partial}{\partial \sigma_0^2}\frac{1}{\bar{\sigma}^2} = \frac{\bar{\sigma}^4}{\sigma_0^4}.$$

So we end up with

$$\frac{1}{2}\left(\frac{y_1}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right)^2 - \left(\frac{y_1}{\sigma_0^2} + \frac{y_2}{\sigma_2^2}\right)\frac{y_1}{\bar{\sigma}^2} + \frac{y_1^2 - \sigma_0^2}{2\bar{\sigma}^4} + \frac{1}{2\bar{\sigma}^2} = 0$$

It is a cubic equation in σ_0^{-2}, so I can't solve it analytically.