Basics tutorial (2019) solutions & hints

Background estimation: Where would you measure next?
The question can be broken down in the following steps:

1. Given that the data are Gaussian, what is the maximum-likelihood estimator for the offset c?
2. What is the variance of that estimator?
3. Depending on where I add a data point, how does the variance decrease?
4. In which option does the variance decrease the most? This is where we should measure again.

The solution runs as follows.

1. Our data are Gaussian, the standard deviations σ are given. Their squares σ^2 are the variances. We generalize the problem, and assume we have A data points in box 1 where the standard deviation is σ_A, and B points in box 2 with standard deviation σ_B, and C points with σ_C in box 3.

   Our data vector is $\mathbf{x} = (x_1, x_2, ..., x_n)$. Our model vector is $\boldsymbol{\mu}^\top = (c, c, c, ..., c)$, since we attempt to measure a constant offset c. To be more transparent, we write the estimator with a hat:

$$\boldsymbol{\mu}^\top = (\hat{c}, \hat{c}, ..., \hat{c}). \tag{0.1}$$

   The maximum likelihood estimator for $\hat{c}$ is where χ^2 is minimal. We have

$$\begin{aligned} \chi^2 &= (\mathbf{x} - \boldsymbol{\mu})^\top \mathsf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \\ &= \sum_{i=1}^{n} \frac{(x_i - \hat{c})^2}{\sigma_i^2}, \end{aligned} \tag{0.2}$$

   where the second line arises since the data points are uncorrelated. Otherwise, mixed terms would appear.

   We now have to solve

$$\frac{\mathrm{d}\chi^2}{\mathrm{d}\hat{c}} \stackrel{!}{=} 0. \tag{0.3}$$

   We have

$$\begin{aligned} \frac{\mathrm{d}\chi^2}{\mathrm{d}\hat{c}} &= \sum_{i=1}^{n} \frac{-2(x_i - \hat{c})}{\sigma_i^2} \\ &= \sum_{i=1}^{n} \left(\frac{-2x_i}{\sigma_i^2} + \frac{2\hat{c}}{\sigma_i^2} \right). \end{aligned} \tag{0.4}$$

   Solving the last line for $\hat{c}$, whe have that the maximum-likelihood estimator is

$$\hat{c} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}. \tag{0.5}$$

2. The variance of this estimator is its averaged scatter around the true value c_t

$$
\begin{aligned}
\mathrm{var}(\hat{c}) &= \langle (c_t - \hat{c})(c_t - \hat{c}) \rangle \\
&= \langle c_t^2 - 2c_t\hat{c} + \hat{c}^2 \rangle \\
&= c_t^2 - 2c_t\langle \hat{c} \rangle + \langle \hat{c}^2 \rangle \\
&= c_t^2 - 2c_t^2 + \langle \hat{c}^2 \rangle \\
&= \langle \hat{c}^2 \rangle - c_t^2.
\end{aligned} \tag{0.6}
$$

Points to be explained here is that all non-random variables can be taken out of averages, since they do not scatter. This is why the true value c_t is taken out. The linearity of the average (used in line 1) is also to be explained.

We therefore see that we have to first work out $\langle \hat{c}^2 \rangle$, and then add the last data point such that the new $\langle \hat{c}^2 \rangle$ is minimized.

We have

$$
\mathrm{var}(\hat{x} + \hat{y}) = \hat{\mathrm{x}} + \hat{\mathrm{y}}, \quad \mathrm{var}(a\hat{x}) = a^2 \mathrm{var}(\hat{x}). \tag{0.7}
$$

Applied to $\hat{c}$, we therefore have

$$
\begin{aligned}
\langle \hat{c}^2 \rangle &= \frac{\left\langle \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} \right)^2 \right\rangle}{\left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^2} \\
&= \frac{\left\langle \left(\sum_{i=1}^n \frac{x_i^2}{\sigma_i^4} \right) + 2 \left(\sum_{i \neq j} \frac{\cancel{x_i x_j}}{\sigma_i^2 \sigma_j^2} \right) \right\rangle}{\left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^2}
\end{aligned} \tag{0.8}
$$

The second term cancels on average, since there is no correlation between x_i and x_j. Continuing, we have

$$
\begin{aligned}
\langle \hat{c}^2 \rangle &= \frac{\sum_{i=1}^n \frac{\sigma_i^2}{\sigma_i^4}}{\left(\sum_{i=1}^n \frac{1}{\sigma_i^2} \right)^2} \\
&= \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}
\end{aligned} \tag{0.9}
$$

3. In the case at hand, we have hence

$$
\begin{aligned}
\langle \hat{c}^2 \rangle &= \frac{1}{\sum_{i=1}^A \frac{1}{\sigma_A^2} + \sum_{i=1}^B \frac{1}{\sigma_B^2} + \sum_{i=1}^C \frac{1}{\sigma_C^2}} \\
&= \frac{1}{A/\sigma_A^2 + B/\sigma_B^2 + C/\sigma_C^2}
\end{aligned} \tag{0.10}
$$

For the plot given in the exercise, we have $A = 2, \sigma_A^2 = 1$, and $B = 2, \sigma_B^2 = 4$ and $C = 1, \sigma_C^2 = 2$. The current scatter of $\hat{c}$ is therefore

$$
\langle \hat{c}_{A,B,C}^2 \rangle = 0.33333 \tag{0.11}
$$

And we can now test how it decreases when adding one more data point into any of the bins

$$\begin{aligned} \langle \hat{c}^2_{A+1,B,C} \rangle &= 0.25 \\ \langle \hat{c}^2_{A,B+1,C} \rangle &= 0.3077 \\ \langle \hat{c}^2_{A,B,C+1} \rangle &= 0.2857. \end{aligned} \tag{0.12}$$

We therefore see that it would be best to spend our money on measuring once more in the first bin, where the variance is smallest.