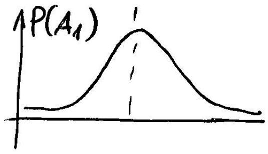
COMBINING DATA

Suppose we have two methods to determine a given mean value $\langle A \rangle$
For instance we have two runs at β_1 and β_2 and we can use both of them to estimate $\langle A \rangle$ at a value of β: we obtain A_1 from the run at β_1 and A_2 from the run at β_2.

WHAT IS THE <u>OPTIMAL</u> WAY TO COMBINE THE RESULTS?

FORMALIZING THE QUESTION

Method 1: gives an estimate A_1 of $\langle A \rangle_\pi$
if we repeat the simulations several times we obtain a distribution



We neglect bias and assume $\langle A_1 \rangle_{MC} = \langle A \rangle_\pi$

$$\langle (A_1 - \langle A \rangle_\pi)^2 \rangle_{MC} = \sigma_1^2 \leftarrow \left(\text{square of the error}\right)$$

Method 2: same as for method 1:

$$\langle A_2 \rangle_{MC} = \langle A \rangle_\pi$$

$$\langle (A_2 - \langle A \rangle_\pi)^2 \rangle_{MC} = \sigma_2^2$$

We consider the combined estimator

$$A_x = x A_1 + (1-x) A_2$$

The estimator A_x is correct for any x.

$$\langle A_x \rangle_{MC} = x \langle A_1 \rangle_{MC} + (1-x) \langle A_2 \rangle_{MC}$$

$$= x \langle A \rangle_\pi + (1-x) \langle A \rangle_\pi = \langle A \rangle_\pi$$

The error is

$$\langle (A_x - \langle A \rangle_\pi)^2 \rangle_{MC} = \langle (x(A_1 - \langle A \rangle_\pi) + (1-x)(A_2 - \langle A \rangle_\pi))^2 \rangle_{MC}$$

$$= x^2 \langle (A_1 - \langle A \rangle_\pi)^2 \rangle_{MC} + (1-x)^2 \langle (A_2 - \langle A \rangle_\pi)^2 \rangle_{MC}$$

$$+ 2x(1-x) \langle (A_1 - \langle A \rangle_\pi)(A_2 - \langle A \rangle_\pi) \rangle_{MC}$$

Because of the independence of the data

$$\langle (A_1 - \langle A \rangle_\pi)(A_2 - \langle A \rangle_\pi) \rangle_{MC} =$$

$$\langle A_1 - \langle A \rangle \rangle_{MC} \times \langle A_2 - \langle A \rangle_\pi \rangle_{MC} = 0$$

Therefore

$$\sigma_x^2 = x^2 \sigma_1^2 + (1-x)^2 \sigma_2^2$$

[This is the usual independent-error formula]

Optimization

We look for the minimum of σ_X^2

$$\frac{\partial \sigma_X^2}{\partial x} = 2x(\sigma_1^2+\sigma_2^2) - 2\sigma_2^2 = 0 \qquad x = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2\sigma_2^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}$$

$$x = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} \qquad \text{optimal } \boxed{A_X = \frac{\frac{A_1}{\sigma_1^2}+\frac{A_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}}$$

$$\boxed{\frac{1}{\sigma_X^2} = \frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} \quad \text{for the optimal case}$$

In practical applications, σ_1 and σ_2 are also estimated from the data

A limiting case [EXAMPLE]

Suppose we have two estimates

(a) one very precise est: A_1 $\sigma_1 = 1$ est of $\sigma_1 = 1.1$

(b) one very imprecise est: A_2 $\sigma_2 = 5$ est of $\sigma_2 = 2$

We use

$$X = \frac{\frac{1}{\sigma^2_{1,est}}}{\frac{1}{\sigma^2_{1est}} + \frac{1}{\sigma^2_{2,est}}} = \frac{\frac{1}{1.1^2}}{\frac{1}{1.1^2} + \frac{1}{4}} = 0.768 \; , \; 1 - X = 0.232$$

$$A_X = 0.768\, A_1 + 0.232\, A_2$$

$$\sigma^2_X = 0.768^2 \cdot 1 + 0.232^2 \cdot 5^2 = 1.94$$

$$\sigma_X = 1.39$$

A_X has a <u>larger</u> error than A_1
We have added <u>noise</u>, not signal, to A_1

It is dangerous to use the formula in the presence of a very accurate and a very inaccurate estimate of the same quantity