Parameter estimation

Alexander Khanov

PHYS6260: Experimental Methods is HEP
Oklahoma State University

September 20, 2023

Estimators

- As a result of an experiment, we have data (measurements), which are a set of N numbers $\mathbf{x} = (x_1, \ldots, x_N)$
- We have a theory describing (we think) our data, which depends on a number of parameters $\mathbf{p} = (p_1, \ldots, p_n)$
- We want to find the best estimates for the theory parameters and also figure out if the theory is good or bad. For that, we construct a function of measurements $\mathbf{P}(\mathbf{x})$ called an estimator
- We want $\mathbf{P}$ to be consistent with $\mathbf{p}$, meaning that as the number of measurements N increases, the probability of $\mathbf{P}$ to be near the true values of $\mathbf{p}$ approaches 1 (this is called "convergence in probability"):

$$\forall \varepsilon > 0 : \lim_{N \to \infty} \mathrm{Prob}(|\mathbf{P} - \mathbf{p}| < \varepsilon) = 1$$

- Since the measurements are subject to fluctuations (i.e. x_i are distributed according to their p.d.f.'s), so are the estimators

Properties of estimators

- Bias b: the difference between the expectation value of the estimator and the true value of the parameter, $b = \langle P \rangle - p$
  - in general, we want an "unbiased" estimator for which $b = 0$
  - bias depends on the choice of parameters
    - e.g. if P is an unbiased estimator of p it doesn't mean that $P' = P^2$ is an unbiased estimator of p^2
  - if P is biased with bias b then $P' = P - b$ is unbiased
- Efficiency ε: the ratio of the minimum possible variance of any estimator to the variance of a given estimator, $\varepsilon = \sigma^2_{\min}/\sigma^2$
  - an estimator is called efficient if $\varepsilon = 1$
  - in many cases there is a trade-off between bias and efficiency – the unbiased estimator may have very low efficiency, e.g. due to large systematic uncertainties

Estimator for mean

$$X = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- If $x_1, \ldots, x_N$ are unbiased measurements of the same unknown quantity x with a common mean x_0 and variance σ^2, then

  $$\langle X \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle x_i \rangle = \frac{1}{N} N x_0 = x_0 \qquad (X \text{ is unbiased})$$

- The variance of the estimator

  $$\mathrm{Var}(X) = \left\langle (X - \langle X \rangle)^2 \right\rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2 = \frac{\sigma^2}{N}, \text{ since}$$

  $$\left\langle X^2 \right\rangle = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle x_i x_j \rangle, \quad \langle x_i x_j \rangle = \begin{cases} 0, \; i \neq j \\ \left\langle x^2 \right\rangle = \sigma^2 + x_0^2, i = j \end{cases}$$

Estimator for variance

$$V = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - X)^2$$

- If $x_1, \ldots, x_N$ come from a normal distribution with a common mean x_0 and variance σ^2, then $\dfrac{V(N-1)}{\sigma^2} = \sum_{i=1}^{N}\dfrac{(x_i - X)^2}{\sigma^2}$ is distributed as χ^2 with $N-1$ degrees of freedom:
$\left\langle \dfrac{V(N-1)}{\sigma^2} \right\rangle = N-1, \quad \mathrm{Var}\left(\dfrac{V(N-1)}{\sigma^2}\right) = 2(N-1)$, therefore
$\langle V \rangle = \sigma^2$ (V is unbiased), $\mathrm{Var}(V) = \dfrac{2\sigma^4}{N-1}$
- If x_0 is known, then the unbiased estimator is $\dfrac{1}{N}\sum_{i=1}^{N}(x_i - x_0)^2$

Mean-squared error

- Mean-squared error $\mathrm{MSE} = \left\langle (P-p)^2 \right\rangle$
  $b = \langle P \rangle - p \quad \sigma^2 = \left\langle P^2 \right\rangle - \langle P \rangle^2$
  $\langle P \rangle = b + p \quad \left\langle P^2 \right\rangle = \sigma^2 + (b+p)^2$
  $\mathrm{MSE} = \left\langle P^2 \right\rangle - 2\langle P \rangle p + p^2 = a^2 + (b+p)^2 - 2(b+p)p + p^2 = \sigma^2 + b^2$
- MSE characterizes both bias and variance

Mean-squared error: example

- Consider a different estimator for variance: $V' = \dfrac{1}{N}\sum_{i=1}^{N}\left\langle (x_i - X)^2 \right\rangle$

$$\left\langle V' \right\rangle = \frac{N-1}{N}\left\langle V \right\rangle = \frac{N-1}{N}\sigma^2,$$

$$\mathrm{Var}(V') = \left(\frac{N-1}{N}\right)^2 \qquad \mathrm{Var}(V) = \frac{2(N-1)\sigma^4}{N^2}$$

- V' is obviously biased: $b = \left(\dfrac{N-1}{N} - 1\right)\sigma^2 = -\dfrac{\sigma^2}{N}$
- However, this estimator has better MSE than V:

$$\mathrm{MSE}(V') = \frac{2(N-1)\sigma^4}{N^2} + \frac{\sigma^4}{N^2} = \frac{(2N-1)\sigma^4}{N^2}, \quad \mathrm{MSE}(V) = \frac{2\sigma^4}{N-1}$$

$$\frac{\mathrm{MSE}(V')}{\mathrm{MSE}(V)} = \frac{(N-1/2)(N-1)}{N^2} < 1$$

Efficient estimators

- How one can figure out that an estimator is efficient (has the least possible variance)?
- Cramér-Rao bound: for any unbiased estimator P of parameter p that depends on N independent measurements of a variable x distributed according to the probability density function $f(x, p)$,

$$\mathrm{Var}(P) \geq \frac{1}{N\, I(p)},$$

where $I(p)$ is Fisher information:

$$I(p) = \left\langle \left(\frac{\partial}{\partial p} \ln f(x, p) \right)^2 \right\rangle = \int \left(\frac{\partial}{\partial p} \ln f(x, p) \right)^2 f(x, p)\, dx$$

Robust estimators

- An estimator P is robust for a given distribution if small deviations from the distribution result in small changes of the estimator
- Sensitivity curve: effect of adding a datapoint

$$SC(x,N) = \frac{P\{x_1,\ldots,x_{N-1},x\} - P\{x_1,\ldots,x_{N-1}\}}{1/N}$$

- Influence curve: $IC(x) = \lim_{N\to\infty} SC(x,N)$
    - example: mean $P_N\{x_1,\ldots,x_N\} = \frac{1}{N}\sum_{i=1}^{N} x_i$, $\lim_{N\to\infty} P_N = p$

$$SC(x,N) = \frac{\frac{(N-1)P_{N-1}+x}{N} - P_{N-1}}{1/N} = x - P_{N-1}, IC(x) = x - p$$

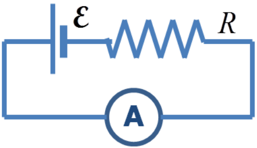
- An estimator is robust if its SC is a bounded function
    - mean is not robust, median is

Estimators for mean and variance: different variances

- If $x_1, \ldots, x_N$ are unbiased measurements of the same unknown quantity x with different variances $\sigma_1^2 \ldots \sigma_N^2$, then a commonly used estimator is the weighted average:
  $X = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}$, where $w_i = \frac{1}{\sigma_i^2}$ is unbiased ($\langle X \rangle = x$) with variance
  $\sigma_X^2 = \frac{1}{\sum_{i=1}^N w_i}$

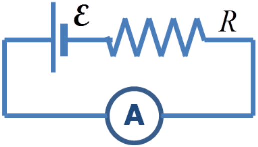
Resistance measurement: an example

- Suppose we want to measure the value of a resistor.
- We have three voltage supplies - 5 V, 10 V, and 20 V, and an ammeter



- The results of the measurements are 5.1 mA, 9.9 mA, 19.2 mA for the voltage of 5 V, 10 V, 20 V, respectively
- We assume the Ohm's law $I = V/R$
- R is the parameter of the theory we want to estimate

Resistance measurement: solution 1



V [V]	5	10	20
I [mA]	5.1	9.9	19.2
R [kΩ]	0.98	1.01	1.04

- We need to make some assumptions about the errors

Model 1: errors are unknown but scale with the measurement
In this case all measurements of R will have the same error:

$$R = \frac{0.98 + 1.01 + 1.04}{3} = 1.01 \text{ k}\Omega$$

$$\Delta R = \sqrt{\frac{1}{3-1}(0.03^2 + 0^2 + 0.03^2)} = 0.03 \text{ k}\Omega$$

Resistance measurement: solution 2

Model 2: the current has a fixed measurement error 0.3 mA

V [V]	5	10	20
I [mA]	5.1	9.9	19.2
R [kΩ]	0.98	1.01	1.04
$\sigma(R)$ [kΩ]	0.058	0.031	0.016

$$R = \frac{\frac{0.98}{0.058^2} + \frac{1.01}{0.031^2} + \frac{1.04}{0.016^2}}{\frac{1}{0.058^2} + \frac{1}{0.031^2} + \frac{1}{0.016^2}} = 1.032 \text{ k}\Omega$$

$$\Delta R = \sqrt{\frac{1}{\frac{1}{0.058^2} + \frac{1}{0.031^2} + \frac{1}{0.016^2}}} = 0.014 \text{ k}\Omega$$

- why the result is shifted towards the last measurement?
- why the error is much smaller – $\langle\sigma\rangle = 0.035$ kΩ?