Statistical methods – an introduction (SS 2016)

Problem set 9

Problem 1 [4 points] *Systematic error*

Assume that you can measure a certain length with a yardstick, with random reading error $\pm\Delta r$, and that there is a systematic shift of the zero-point, $z = z_0 \pm \Delta z$, where Δz is the uncertainty of this shift.

I.e., if you would try to measure a certain (actual) length a', your measurement would be $a = a' \pm \Delta r + z$.

Now assume that you measure two lengths, a and b, and that you are interested in the ratio of the *actual* lengths, a'/b'. Calculate the corresponding error as a function of the above quantities.

Finally, adopt the following values: $a = 40$ cm, $b = 20$ cm, $\Delta r = 0.5$ cm, $z_0 = 6$ cm and $\Delta z = 1$ cm. Provide your result for the ratio of the actual lengths, and the corresponding error.

Problem 2 [2 points] *Consistency of an estimator*

Verify the following statement: In order to prove that an estimator S is consistent, one has 'only' to prove that (i) $Var(S)$ decreases monotonically (at least at large N) with increasing sample size, N, towards zero, and that (ii) the estimator is asymptotically unbiased.

Hint: Consider, e.g., the Tchebychev inequality

Problem 3 [2 points] *Efficiency of estimators*

x_1, x_2 and x_3 are the elements of a sample drawn independently from a continuous population of unknown expectation value $\hat{x}$ but known variance σ^2.

a) Show that

$$\begin{aligned} S_1 &= \frac{1}{2}x_1 + \frac{1}{5}x_2 + \frac{3}{10}x_3 \\ S_2 &= \frac{1}{5}x_1 + \frac{2}{3}x_2 + \frac{2}{15}x_3 \\ S_3 &= \frac{1}{5}x_1 + \frac{9}{20}x_2 + \frac{7}{20}x_3 \end{aligned}$$

are unbiased estimators of $\hat{x}$.

b) Calculate the variances $\sigma^2(S_1), \sigma^2(S_2), \sigma^2(S_3)$.

c) Show that the arithmetic mean $\bar{x} = 1/3 \sum_{i=1}^{3} x_i$ has the smallest variance of all estimators of the type $S = \sum_{i=1}^{3} a_i x_i$ with the constraint $\sum_{i=1}^{3} a_i = 1$. Compute this variance and compare with the variances obtained in b).

Problem 4 [4 points] *Elections!*

Compute the probability density for the fraction of voters, r, that will vote with 'yes' during a poll, if within a sample of 8 voters 3 of them vote with 'yes'. Determine the probability, $P(r > 0.5)$, that more than 50 % of the voters will vote with 'yes'.

Hint 1: Use Bayes' theorem, and assume that all $r \in [0, 1]$ are equally probable, i.e., use a uniform prior.

Hint 2a: At some point, you will need the so-called beta-function,

$$\int_0^1 x^m (1-x)^n dx = \beta(m+1, n+1) = \frac{\Gamma(m+1)\Gamma(n+1)}{\Gamma(m+n+2)}.$$

Hint 2b: The integral ratio

$$\frac{\int_0^a x^m (1-x)^n dx}{\int_0^1 x^m (1-x)^n dx} = \mathrm{ibeta}(m+1, n+1, a)$$

defines the so-called incomplete beta-function (normalized, i.e. ibeta $= 1$ for $a = 1$), and can be found both in `idl` and `gnuplot` under `ibeta` and in `python` under `scipy.special.betainc`.

Have fun, and much success!