Exercise Sheet #5

due on Tuesday, 08 December 2020, 12 noon sharp

1. **EXAMple** The notions of *bias* and *mean square error (MSE)* were defined in the lecture. Show that, for an estimator $\hat{\theta}_n$ defined on a dataset $\boldsymbol{x}$ drawn iid. from the likelihood $p(\boldsymbol{x} \mid \theta)$, it holds that
$$\mathrm{MSE}(\hat{\theta}_n) = \mathrm{bias}^2(\hat{\theta}_n) + \mathbb{V}_{p(\boldsymbol{x}|\theta)}(\hat{\theta}_n),$$
where $\mathbb{V}_{p(\boldsymbol{x}|\theta)}(\hat{\theta}_n)$ denotes the sampling variance of the estimator.

2. **Theory Question:** This question is a classic example[1] for the shortcomings of maximum likelihood estimation. Assume you are given the numbers
$$\boldsymbol{X} = [101, 1, 93, 78, 239, 185, 65, 202, 12, 125]$$
These are assumed to be iid. samples $X_1, \dots, X_n \sim \mathbb{U}(0, \theta)$, drawn uniformly from the interval $(0, \theta)$, with unknown upper limit $\theta \in \mathbb{R}_+$.
   (a) What is the *likelihood* of a single observation $p(X_i \mid \theta)$? What is the joint likelihood $p(\boldsymbol{X} \mid \theta)$?
   (b) What is the *maximum likelihood estimator*?
   (c) Make a plot of the full likelihood. Normalize it (numerically, by calling `likelihood /= likelihood.sum()`) to get the associated posterior
$$p(\theta \mid \mathbf{X}) = \frac{p(\boldsymbol{X} \mid \theta)}{\int p(\boldsymbol{X} \mid \theta)\, d\theta}$$
   under an uninformative prior. Find the 25th, 50th and 75th percentile of the estimate.

3. **Practical Question:** You can find this week's sheet on Ilias as `Exercise_05.ipynb`

[1] A slightly more elaborate variant of this model was used by the Western Allies to estimate the number of German Panzer V (Panther) tanks prior to the invasion in Normandy, based on the consecutive mold numbers imprinted in wheels of captured or destroyed tanks. The analysis worked famously well, and the estimate has hiterto become known as "the German tank problem" (although it was also used to estimate the stock of other equipment, including V1 and V2 rockets). The estimated number of tanks was 270. The correct number was apparently 276, which is also the number used to generate the data $\boldsymbol{X}$ above. Source: R. Ruggles & H. Brodie. *An Empirical Approach to Economic Intelligence in World War II*. J of the American Statistical Assoc. (JASA) **42**/237 (3/1947), pp. 72-91. The numbers are on page 83, top.