PD Dr. Torsten Enßlin
Margret Westerkamp
Johannes Harth-Kitzerow

Exercise sheet 5

Note: *This exercise sheet will be discussed in the tutorials from the 5th of June on.*

Exercise 5 - 1

Consider the following coin toss experiment:

- A large number n of coin tosses are performed and the results are stored in a data vector $d^{(n)} = (d_1, \ldots, d_n) \in \{0,1\}^n$, where 0 and 1 represent the possible outcomes head and tail.
- Individual tosses are independent from each other.
- All tosses are done with the same coin with an unknown bias $f \in [0,1]$; i.e., $\mathcal{P}(d_i|f) = f^{d_i}(1-f)^{1-d_i}$.

Assume that a fraction $\bar{f}$ out of the n coin tosses yielded head.

a) Derive the Gaussian approximation of the PDF $\mathcal{P}(f|d^{(n)})$ around its maximum. – You can use a saddle point approximation; i.e., identify the maximum, and taylor-expand the (negative) logarithm of $\mathcal{P}(f|d^{(n)})$ around it up to second order in order to identify the variance of the Gaussian (3 points).

b) Use this Gaussian approximation to derive an approximation for $\mathcal{P}(d_{n+1}|d^{(n)})$.
Hint: You can assume that the Gaussian distribution is narrow enough such that the integration boundaries $[0,1]$ can be replaced by $(-\infty, \infty)$ (2 points).

c) Now calculate the exact posterior mean for $\bar{f}$ and the exact expression for $\mathcal{P}(d_{n+1}|d^{(n)})$ (2 points).

Note: $\int_0^1 \mathrm{d}x\, x^\alpha (1-x)^\beta = \Gamma(\alpha+1)\Gamma(\beta+1)/\Gamma(\alpha+\beta+2)$, where Γ is the Gamma function.

Exercise 5 - 2

A strictly positive quantity x is bound from above, say by $x \leq 1$. You learn that its natural logarithm is typically the negative number $-l$.

a) How much information did you gain about x (1 point)?

b) What is the least informative value for l you could have gotten (1 point)?

c) Use these insights to explain **Benford's law** B, under which a set of positive numbers exhibits the leading digit $d \in \{1, \ldots 9\}$ with a probability $P(d|B) = \log_{10}\left(1 + \frac{1}{d}\right)$. Such distributions are observed in many human made, measured, and even mathematical sets of numbers. Benford's law is used to identify potential tax frauds (2 points).

Exercise 5 - 3

Let $P(s) = \mathscr{G}(s, S)$ with $s = (s_1, ..., s_n)^t$ be a real multivariate zero-centered Gaussian with Covariance $\langle ss^t \rangle = S$. We would like to fit another Gaussian distribution $P'(s) = \mathscr{G}(s, S')$ to it that has a diagonal covariance matrix $S'_{ij} = \delta_{ij}\sigma_i$. Here δ_{ij} denotes the Kronecker delta.

a) What is the optimal approximating Gaussian $P'(s)$ to $P(s)$, as parameterized by σ, obtained through minimizing the loss

$$\sigma = \arg\min_\sigma \mathrm{KL}(P(s), P'(s)) \text{ ?}$$

(2 points)

PD Dr. Torsten Enßlin
Margret Westerkamp
Johannes Harth-Kitzerow

b) What is the least updating fit of $P'(s)$ to $P(s)$, as obtained through

$$\sigma = \arg\,\min_\sigma \mathrm{KL}(P'(s), P(s))\ ?$$

(2 points)

c) Let

$$S = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0.2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Visualize the Gaussian $\mathscr{G}(s, S)$ as well as the two different ways to fit a diagonal Gaussian to it introduced above. To visualize them, use a computer plotting samples of each of the three distributions. You can draw a sample from a multivariate Gaussian distribution by applying the square root of its covariance matrix to a white noise sample. (optional)

This exercise sheet will be discussed in the tutorials.
group A, Wednesday 16:00 - 18:00 Theresienstr. 39 - room B 101,
group B, Thursday, 10:00 - 12:00, Geschw.-Scholl-Pl. 1 (B) - room B 015
group C, Thursday, 16:00 - 18:00, Theresienstr. 37 - room A 449
backup slot, Wednesday, 8:00 - 10:00, Theresienstr. 37 - A 450
(to be used whenever Thursday is a vacation day)
www.mpa-garching.mpg.de/∼ensslin/lectures