*PD Dr. Torsten Enßlin*
*Vishal Johnson, Andreas Popp*

# Exercise sheet 8

## Exercise 8 - 1

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)?

## Exercise 8 - 2

You know that a quantity $x$ is in $[1, M]$, where $M \gg b > 1$ is a large positive number and $b > 1$ is a certain base in which digits of numbers are to be expressed. The logarithm $z = \log_b x$ of $x$ in the chosen base $b$ gives you an idea of the order of magnitude of $x$ in this base.

**a)** Without any additional information, what is the least informative PDF for $x$? What distribution does this imply for $z$? (1 point)

**b)** Now instead assume that $z$ is uniformly distributed. What conclusion can you draw on the least informative distribution of $x$? (1 point)

**c)** Assume that the base $b = 10$, use these insights to explain **Benford's law** $B$, under which a set of positive numbers exhibits the leading digit $d \in \{1, \dots 9\}$ 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, when these numbers are distributed across many different length scales. Benford's law is used to identify potential tax frauds (3 points).

*hint*: Find the probability of a generic interval $I_k$ such that $10^k d \leq x < 10^k(d+1)$ for a certain $k \in \mathbb{N}$. Then sum over different $k$s for fixed $d$. Finally, consider the limit for large $M \to \infty$.

## Exercise 8 - 3

The goal of this exercise is to evaluate the integral

$$I(x) = \int_{-\infty}^{\infty} \mathrm{d}k \frac{e^{-ikx}}{k^2 + m^2} \tag{1}$$

**a)** For $k \in \mathbb{C}$, the integrand can diverge. Identify the poles of the integrand and sketch them in the complex plane, sketch the integration path of $k \in (-\infty, \infty)$ as well (2 points).

**b)** Draw a contour line that closes the integral, connecting $k = \infty$ to $k = -\infty$. The integral of the added contour line should be 0. Illustrate why it is 0 by comparing orders in $k$ (2 points).

Hint: You may assume $x > 0$.

**c)** The closed integral encloses one pole. Identify its order and evaluate the integral $I(x)$ using the residue theorem,

$$\int_\Gamma f = (2\pi\, i) \sum_{a \in R_f} \mathrm{ind}_\Gamma(a)\ \mathrm{Res}_a f, \tag{2}$$

where $\Gamma$ is a closed contour, $R_f$ are the poles of $f$, and $\mathrm{ind}_\Gamma(a)$ is the number of counter-clockwise windings of $\Gamma$ around $a$, and the residue formula,

$$\mathrm{Res}_a f = \frac{1}{(n-1)!} \lim_{z \to a} \frac{\partial^{n-1}}{\partial z^{n-1}} \left[(z-a)^n f(z)\right], \tag{3}$$

where $n$ is the order of the pole $a$ (2 points).

Hint: A clockwise winding is a negative counter-clockwise winding.

**Exercise 8 - 4**

You are interested in three numbers, $s = (s_1, s_2, s_3) \in \mathbb{R}^3$. Your measurement device, however, only measures three differences between the numbers, according to

$$d_1 = s_1 - s_2 + n_1 \tag{4}$$
$$d_2 = s_2 - s_3 + n_2 \tag{5}$$
$$d_3 = s_3 - s_1 + n_3 \tag{6}$$

with some noise vector $n \in \mathbb{R}^3$. Assume a Gaussian prior $\mathcal{P}(s) = \mathcal{G}(s, S)$ for $s$ and a Gaussian PDF for the noise, $\mathcal{P}(n) = \mathcal{G}(n, N)$, with $N_{ij} = \sigma^2 \delta_{ij}$.

**a)** Assume that the prior is degenerate, i.e., $S^{-1} \equiv 0$. Write down the response matrix, try to give the posterior $\mathcal{P}(s|d)$, and explain why this is problematic (2 points).

**b)** Now assume that $S_{ij} = \sigma^2 \delta_{ij}$. Work out the posterior $\mathcal{P}(s|d)$ in this case (1 point).

Note: Using a computer algebra system, e.g., SAGE (`http://www.sagemath.org/`), for the matrix operations is okay.

**Exercise 8 - 5**

The numbers quantifying the degree of industrialization $\tilde{\iota}$ of a society, its fertility rate $\tilde{f}$, and the stork population $\tilde{s}$ on its territory are random variables assumed to belong to a joint three-dimensional Gaussian distribution. Consider the fluctuations around the respective mean values $\iota = \tilde{\iota} - \langle\, \tilde{\iota}\, \rangle_{(\tilde{\iota})}$, $f = \tilde{f} - \langle\, \tilde{f}\, \rangle_{(\tilde{f})}$, and $s = \tilde{s} - \langle\, \tilde{s}\, \rangle_{(\tilde{s})}$. It is known that both the stork index $s$ and the fertility index $f$ are anticorrelated with the degree of industrialization. The normalized correlation coefficients are $c_{s\iota} = -0.85$ and $c_{f\iota} = -0.70$, where

$$c_{ab} = \frac{\langle ab \rangle_{(a,b)}}{\sqrt{\langle aa \rangle_{(a)} \langle bb \rangle_{(b)}}}. \tag{7}$$

Assume further that there is no direct correlation between $f$ and $s$, i.e., $\mathcal{P}(s|f,\iota) = \mathcal{P}(s|\iota)$ and $\mathcal{P}(f|s,\iota) = \mathcal{P}(f|\iota)$. Derive an expression for $\langle sf \rangle_{(s,f,\iota)}$. Use this to calculate the normalized correlation coefficient $c_{sf}$ (2 points).

---

*This exercise sheet will be discussed during the exercises.*
*Group 01, Wednesday 18:00 - 20:00, Theresienstr. 37, A 449,*
*Group 02, Thursday, 10:00 - 12:00, Theresienstr. 37, A 249,*

*https://wwwmpa.mpa-garching.mpg.de/ ensslin/lectures/lectures.html*