Exercises for the lecture on

Statistical Methods

Prof. Dr. Vladimir Spokoiny
Vladimir.Spokoiny@wias-berlin.de
Dr. Thorsten Dickhaus
Thorsten.Dickhaus@wias-berlin.de
www.math.hu-berlin.de/~dickhaus/

Humboldt-University Berlin
Department of Mathematics
Winter term 2013 / 2014

Sheet 1

Solution are due on Monday, October 21st, 2013, 3:15pm.
Every completely and correctly solved exercise gives 4 points.

Exercises

1. Formalize (mathematically) statistical models for the following three experiments. Find out if the respective models are parametric or non-parametric.

   (a) We consider the German lottery „6 aus 49“ and suppose that we have uncertainty about whether each of the 49 numbers has the same probability of being drawn. Therefore, we observe all 520 Saturday drawings in the decade $2011 - 2020$ and make a tally chart of how often each number is drawn. We do not have any doubts about the independence of the drawings.

   (b) In „Survival Analysis“ one is interested in the time until the first occurrence of a specified target event (failure of a technical device, disease of a bacterial culture, etc.). The distribution of such a time (regarded as a random variable) can often be modeled well by an exponential distribution. We check n randomly chosen light bulbs coming from one and the same large production line and note how long each of them glows until filament damage. We are interested in the mean glowing time until failure.

   (c) Assume that association between the type I diabetes risk in humans and gender is of interest. To this end, integers n_1 and n_2 are fixed and n_1 randomly chosen women and n_2 randomly chosen men from a specified target population (for instance, the inhabitants of Berlin) perform an oral glucose tolerance test (OGTT). We note for every of the $n = n_1 + n_2$ study participants the OGTT result (type I diabetes yes / no). For ease of argumentation we assume than an OGTT is a perfect tool for diagnosing type I diabetes. Furthermore, we neglect all other potential factors that may influence type I diabetes manifestation.

2. **Game theory.** For determining who has to wash the dishes next weekend, you and your opponent play the following game: The opponent chooses a number $\vartheta \in \{0, 1\}$. You have to guess this number. If you guess wrong, you have to wash the dishes; if you guess correctly, you do not have to wash the dishes (0-1-loss). Your opponent helps you as follows. He covertly flips a coin. If the coin shows head, he tells you ϑ. If the coin shows tail, he performs a second covert coin flip and tells you 1 in case of tail and 0 in case of head (in this second coin flip). Let the random variable X represent the number that is told to you by the opponent and x its realization.

   (a) Model this game as a statistical decision problem. In particular, derive $\mathbb{P}_\vartheta$ for $\vartheta \in \{0, 1\}$.

   (b) Consider the set of concurring decision rules $\mathcal{M} = \{\delta_1, \delta_2\}$ with $\delta_1(x) = x$ and $\delta_2(x) = 1$ and show

      (i) δ_1 and δ_2 are both admissible in $\mathcal{M}$.

(ii) δ_1 is minimax in $\mathcal{M}$.

3. **Conjugate distributional classes.**
Show that the following conjugation relationships hold.

    (a) Let $n \in \mathbb{N}$ be a fixed sample size and $X = (X_1, \ldots, X_n)^\top$ a vector of n real-valued, stochastically independent, identically $\mathcal{N}(\mu, \sigma^2)$-distributed random variables. In this, assume that the variance σ^2 is known and, consequently, that the parameter of interest is the expectation μ. The family of normal distributions on $\mathbb{R}$ for this parameter μ is conjugate to the family $((\mathcal{N}(\mu, \sigma^2))^n)_{\mu \in \mathbb{R}}$ of normal distributions for the observables.

    (b) The family of *gamma distributions* for λ is conjugate to the family of *Poisson distributions* with intensity parameter λ. (We only make one single observation!)

    Hint: Bayes formula for densities! In part (a), consider first only the case $n = 1$.

4. **Multiple Select.** Which of the following statements are true and which are false?
Please give reasons for your respective decisions (one short sentence each is sufficient).

    1. The quadratic risk for estimating the expectation of a distribution by the arithmetic mean of an independent and identically distributed sample drawn from this distribution must not necessarily exist (in $\mathbb{R}$).
    2. If two classes $\mathcal{M}_1$ and $\mathcal{M}_2$ of decision rules for the same statistical decision problem are given and $\mathcal{M}_1 \subset \mathcal{M}_2$, then the minimax risk over $\mathcal{M}_1$ can never exceed the minimax risk over $\mathcal{M}_2$.
    3. For two decision rules $\delta_1 \neq \delta_2$ for the same statistical decision problem, it always either holds that δ_1 is better than δ_2 or that δ_2 is better than δ_1.
    4. If the prior distribution for a given statistical parameter $\vartheta \in \Theta$ is the Dirac distribution with point mass 1 in one particular element ϑ^* (say) of Θ, then also the posterior distribution has no mass outside of ϑ^*, no matter the data.