Last update: April 23, 2021

NMST434 – MODERN STATISTICAL METHODS – EXAM

Please note that this document is still in evolution.

The exam will be organized as follows. First, an example will be given and there will be about 50 minutes to solve this example. After handing in this example, the student can make a short break, after which he/she gets two theoretical questions.

1. First part (example)

This part covers: the usage of Δ-theorem, deriving asymptotic normality of moment estimators, maximum likelihood estimators, M-estimators and Z-estimators (including quasi-likelihood), deriving tests based on the maximum likelihood theory, using EM-algorithm, distinguishing between the concepts of missing values (MCAR, MAR, MNAR). Two illustrative examples are given below.

Example 1. Consider independent and identically distributed random vectors $(X_1, Y_1)^{\mathsf{T}}$, …, $(X_n, Y_n)^{\mathsf{T}}$. Suppose that Y_i given X_i has a normal distribution with the mean value $\beta\, X_i$ and variance $2\, X_i$. Further let X_i is uniformly distributed on $[0, 1]$. Derive the asymptotic distribution of $\widehat{\beta}_n$ (the maximum likelihood estimator of β).

Suppose that the normality assumption of the previous paragraph does not hold but the conditional mean Y_i given X_i is still $\beta\, X_i$. Construct a sandwich estimator of the asymptotic variance of the estimator $\widehat{\beta}_n$ in this situation.

Example 2. Suppose we observe independent and identically distributed random variables $X_1, \dots, X_n$ from a 'zero-inflated geometric distribution', that is a mixture of a geometric distribution at the point zero so that

$$\mathsf{P}(X_1 = k) = w\, \mathbb{I}\{k = 0\} + (1 - w)\, p(1 - p)^k, \qquad k = 0, 1, 2, \dots,$$

where $p \in (0, 1)$ and $w \in [0, 1]$ are unknown parameters to be estimated. Let $\widehat{\boldsymbol{\theta}}_n$ be the maximum likelihood estimator of $\boldsymbol{\theta} = (\lambda, w)^{\mathsf{T}}$

(1) Construct a test for testing the null hypothesis $H_0 : (p, w) = (p_0, \frac{1}{2})$ against the alternative $H_1 : (p, w) \neq (p_0, \frac{1}{2})$.
(2) Construct a test for testing the null hypothesis $H_0 : w = \frac{1}{2}$ against the alternative $H_1 : w \neq \frac{1}{2}$.
(3) Construct a test for testing the null hypothesis $H_0 : w = 0$ against the alternative $H_1 : w > 0$.
(4) Show how the EM-algorithm can be used to find the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$.

The (asymptotic) type I error of the tests should be α. Constructing a test means that you:
(1) give a test statistics;
(2) specify a critical region or describe a way how the p-value is calculated (or estimated).

2. Second part (two theoretical questions)

In this part the emphasis will be mainly on describing and explaining the methods and formulating and proving theorems. Only the proofs that were shown during the lectures will be a part of the exam.

But also here the questions can be accompanied with simple examples to find out if the student knows how to apply the method in specific situations. I recommend to go through the examples made at the lectures, exercise classes and homework assignments.

3. Some general comments

Before reading the topics for the exam, please have a look at the following general comments.

- In mathematics when explaining things rigorously one needs to write them down. That is why I recommend to write things down when you are preparing for the exam.
- Be careful about what you are saying and writing down. Everything you say/write should be mathematically correct.
- Some of the proofs or derivations are rather technical and some of the assumptions may be difficult to remember. It is much better to say that now I do not remember this step of the proof than writing down something which is visually similar but which is a mathematical nonsense.
- Please keep in mind that you can **fail the exam** even when you pretty well reproduce all the material from the lecture but you keep saying/writing things that are mathematical nonsense. On the other hand you can pass the exam although you do not remember all technical details but everything you say is mathematically correct.

Moment estimation

Asymptotic normality of moment estimators.

Maximum likelihood method

Asymptotic tests - with/without nuisance parameters.
Asymptotic efficiency of maximum likelihood estimators
Confidence ellipsoids and intervals
Profile likelihood

Conditional and marginal likelihood

Maximum likelihood estimators in non i.i.d. settings. Random design vs. fixed design

Several independent samples

M-estimators when the function $\rho(x; t)$ is convex in t. Derivation of the asymptotic distribution of these estimators in specific situations (e.g. a sample quantile, Huber M-estimator)

M- and Z-estimators

The definitions and examples

Maximum likelihood (MLE), least squares (LS) and least absolute deviation (LAD) estimators viewed as M- and/or Z-estimators.

Parameters identified by the M- (Z)-estimators

Consistency and asymptotic normality of Z-estimators.

The estimation of the asymptotic variance.

Likelihood under model misspecification. Introducing the idea.

Sandwich-estimator of the variance (general construction as well as the construction in specific examples - linear model, logistic regression, Poisson regression)

White's heteroscedasticity-consistent estimator

M- and Z-estimators in statistics

Idea of the breakdown point

Robust estimation of location - Huber estimator when compared to the sample mean and median, studentization of Huber estimator

Robust estimation in linear models - Huber estimator when compared to the least squares method and LAD method, studentization of Huber estimator.

Quantile regression

The key lemma on identification of quantiles

Regression quantiles - what is being estimated, interpretation of the parameters, transformation of the response

Some notes about the inference in regression quantile models

Asymptotic normality of sample quantiles

EM-algorithm

Description of the algorithm

Theoretical results

Rate of convergence of EM-algorithm

Missing data

Taxonomy of missing data - MCAR, MAR, MNAR

Methods for dealing with missing data - complete case analysis, available case analysis, observed likelihood methods, imputation (simple, multiple), reweighting.

Example question: Suppose you observe independent random pairs $(Y_1, X_1)^\mathsf{T}, \dots, (Y_n, X_n)^\mathsf{T}$ such that Y_i given X_i has a Bernoulli distribution with a probability of success given by

$$\mathsf{P}(Y_i = 1|X_i) = \tfrac{\exp\{\beta_0+\beta_1 X_i\}}{1+\exp\{\beta_0+\beta_1 X_i\}}, \qquad i = 1, \dots, n,$$

where β_0 and β_1 are unknown parameters. Further suppose that X_i is uniformly distributed on $[-1, 1]$. Finally X_i is always observed and Y_i is missing if only if $X_i > 0$.

Is the missing mechanism MCAR, MAR or MNAR?

Describe briefly how the complete case analysis and direct likelihood method would work in this setting. Are both methods appropriate for estimation of the parameters β_0 and β_1?

And what about estimation of $\mathsf{P}(Y_i = 1)$? Shall we make use of complete case analysis or prefer an appropriate imputation method?

Bootstrap

Monte Carlo principle

Bootstrap as a Monte Carlo principle combined with plug-in principle

Standard nonparametric bootstrap

Theoretical results for the sample mean

Confidence intervals derived with the help of bootstrap

Limits of the standard nonparametric bootstrap

Bootstrap and variance estimation

Parametric bootstrap and its applications to GOF testing

Bootstrap in linear models

Testing and bootstrap

Permutation tests

Example question: Describe the standard nonparametric bootstrap and formulate theorem(s) that justifies its use. Describe the general idea of using bootstrap for testing. Consider the following example.

Suppose we observe a random sample $X_1, \ldots, X_n$ from an unknown distribution F. Suggest how the bootstrap can be used to test for the null hypothesis that the variance of the distribution F is one.

Kernel density estimation

Naive estimator of density

Bochner's lemma

Asymptotic unbiasedness of kernel estimator of density

Pointwise consistency of kernel estimator of density

Pointwise asymptotic normality of kernel estimator of density

MSE (mean squared error), MISE (mean integrated squared error) and their asymptotic approximations (AMSE, AMISE)

Asymptotically optimal bandwidth choice

Normal reference rule

Least-squares cross-validation

Kernel regression

Local polynomial regression

Local linear and local constant estimators

Asymptotically optimal bandwidths

Rule of thumb for bandwidth selection

Cross-validation

Nearest-neighbour bandwidth choice

Robust locally weighted regression (LOWESS)