Mathematical Statistics

[module 2, 2018]

Professor Anatoly Peresetsky
NRU HSE
apereset@nes.ru

Course information

Course Website:	
Instructor's Office Hours:	**tba**
Class Time:	**tba**
Room Number:	**tba**
TAs:	**Sergei Golovan, Alexander Lagoda**

Course description

The course "Mathematical Statistics" is second in the sequence of probability-statistics-econometrics courses in NES.
Main course objects:
Present basic concepts and methods of mathematical statistics;
Practice of application of statistical methods to applied research;
Present basic concepts and methods of multidimensional applied statistical analysis.

Course requirements, grading, and attendance policies

Prerequisites are Calculus, Matrix algebra, Probability theory.
Course consists of 14 lectures and 7 weekly recitation sections. The final grade is based on the final score, which is a weighted average of the home assignments and the final closed-book exam (weights are respectively, 0.2 and 0.8). However, less than 25% on the final exam means failure. At the final exam students may use two format A4 pages with their own notes (handwriting only). Percentage final grade is converted to 5-grade mark basing on curve.

Course contents

1. Sample. Parameter estimators. Properties of estimators. Unbiased, consistent, efficient estimators. (1 lecture)

2. Confidence intervals. Standard confidence intervals for parameters of normal population. Confidence intervals for mean, variance, difference of two means, variances ratio, proportion, difference of two proportions. Sample size. (2 lectures)
3. Tests of statistical hypotheses. I and II type errors. P-value of the test. Tests for parameters of normal population. Tests about mean values, variances, proportions. (1 lecture)
4. Methods of parameters estimation. Method of moments. Maximum likelihood. Information inequality. Delta method. (1 lecture)
5. Critical statistics. Neyman-Pearson lemma. Likelihood ratio test. (1-2 lectures)
6. Goodness-of-fit tests. Contingency tables. Pearson test. Kolmogorov-Smirnov test. (1 lecture)
7. Bayesian methods. Point estimation, credible intervals. (1 lecture)
8. One- and two-factor analysis of variances. (0.5-1 lectures)
9. Introduction to nonparametric methods. Wilcoxon test, run test. Spearman rank correlation. (1 lecture)
10. Methods of classification. Discriminant analysis. Separation of the mixture of distributions. Cluster analysis. Principal components. Factor analysis. (1.5-2 lectures)
11. Sufficient statistics. Minimal sufficient statistics. Rao-Blackwell theorem. Complete statistics. Lehmann-Scheffé theorem. (2 lectures).

Note. Given number of lectures per topic and topic sequence could be adjusted during the course.

Sample tasks for course evaluation

Problem 1 (35 points)

$x_1,\dots,x_n$ is a random sample from the normal distribution $N(\theta,\theta^2)$, where $\theta>0$.

(a) (8 points) Find maximum likelihood estimator $\hat{\theta}_{ML}$ of the parameter θ and its asymptotic distribution.

(b) (4 points) Compare (a) with asymptotic distribution of the estimator $\hat{\theta}_1=\bar{x}$.

(c) (8 points) Compare (a) with asymptotic distribution of the estimator $\hat{\theta}_2=\sqrt{s^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

(d) (7 points) Find optimal (asymptotically efficient) combination of the two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ (consider only constant coefficients). Compare with $\hat{\theta}_{ML}$.

(e) (8 points) Find minimal sufficient statistic for θ. Is it a complete statistic?

Problem 2. (15 points)

The concentration of chemical agent in liquid could be measured with a new measuring device. To find the measuring accuracy, seven pieces of liquid with known concentration were prepared, device was used, and errors were calculated.

They are: 0.4 −1.1 0.2 1.5 −3.1 −2.1 2.

It is known that the mean error EX_k is zero. But accuracy decreases with the number of uses $Var(X_k)=k\cdot\sigma^2$, $k=1,\dots,7$.

(a) (6 points) Estimate σ^2.

(b) (6 points) Construct 90% confidence interval for σ.

(c) (3 points) State all assumptions you used.

Problem 3. (15 points)
An experiment is conducted to study how long different digital camera batteries last. The aim is to find out whether there is a difference in terms of battery life between four brands of batteries using seven different cameras. Each battery was tried once with each camera. The average time the Brand A battery lasted was 43.86 hours. The average times for brands B, C and D were 41.28, 40.86 and 40 hours respectively. The following is the calculated ANOVA table with some entries missing.

Source	Sum of Squares	df	Mean square	*F*-stat
Batteries				
Cameras			26	
Error				
Total	343			

(a) (7 points) Complete the table using the information provided above.
(b) (3 points) Is there a significant difference between the performance of different battery brands?
(c) (5 points) Construct a 90% confidence interval for the difference between brands A and D. Would you say there is a difference?

Problem 4. (10 points)
Two students made measurements of the content of carbon in the lake of Moscow. They produced 10 measurements altogether using the same device. Since they have to submit separate home works, first student used first 8 observations $x_1,\dots,x_8$ to get sample mean of $\bar{x}$ and sample variance of s_x^2; the second student used last 6 observations $x_5,\dots,x_{10}$ to get sample mean of $\bar{y}$ and sample variance of s_y^2. Assume that observations follow normal distribution $N(\mu,\sigma^2)$. If you are given $\bar{x}=80,\ \bar{y}=70,\ s_x^2=10,\ s_y^2=8$,
Find the most efficient unbiased estimator of μ from the family of $\alpha\bar{x}+\beta\bar{y}$.

Problem 5 (30 points)
Let $X_1,\dots,X_n$ is a random sample from Poisson distribution with mean λ $(0<\lambda<\infty)$. The point of interest is estimation of the parameter $\tau(\lambda)=P(X=0)$.
(a) (5 points) Using the information inequality find low boundary for the variance of unbiased estimator W of the parameter $\tau(\lambda)$.
(b) (5 points) Show that $U=\sum_{i=1}^{n}X_i$ is sufficient statistic for λ.
(c) (5 points) Show that $U=\sum_{i=1}^{n}X_i$ is complete statistic for λ.
(d) (5 6 points) Find UMVUE for $\tau(\lambda)$.
(e) (5 points) Find variance of the estimator (d) and compare it with the low boundary (a).
(f) (5 points) Find ML estimator for $\tau(\lambda)$ and its asymptotic distribution.

Problem 6 (10 points)
Let $X_1,\dots,X_8$ be a sample of size 8 from the normal distribution $N(\mu_X,\sigma_X^2)$, $\bar{x}=30$, $s_X=5$. Let $Y_1,\dots,Y_3$ be an independent sample of size 3 from the normal distribution $N(\mu_Y,\sigma_Y^2)$, $\mu_Y=0$, and

$$\sum_{i=1}^{3}Y_i^2=84.$$

(a) (7 points) At 10% significance level test null hypothesis that population variances of the distributions are the same.

(b) (3 points) Find 90% confidence interval for the variance σ_X^2 .

Problem 7 (15 points)

Random variable X has p.d.f. $f(x) = \dfrac{2x}{\theta^2}$, $0 < x \le \theta$, and 0 elsewhere.

Prior information is: $\theta \in [0.5, 4.5]$. Given the sample of size $n = 2$, $\{X_i\} = \{1, 2\}$, find Bayes estimate of the parameter θ , corresponding for the quadratic loss function.

Course materials

Required textbooks and materials

Hogg R.V., Tanis E.A., Zimmerman D. (2014). *Probability and statistical inference*, 9th edition. Pearson.

Johnston A.R. and Bhattacharyya G.K. (2014). *Statistics. Principles and methods*. 7th edition, Wiley.

Hogg R.V., McKean J.W. and Craig Allen T. (2018). *Introduction to mathematical statistics*, 8th edition, Pearson Prentice Hall.

Casella G., and Berger R.L. (2012). *Statistical inference*. 2nd edition. Duxbury.

Additional materials

Wackerly D.D., Mendenhall W. III, Scheaffer R.L. (2008). *Mathematical statistics with applications*. 7th edition, Thompson.

Айвазян С.А., Мхитарян В.С. (2001). *Прикладная статистика и основы эконометрики*. (2-е издание). Том 1: Теория вероятностей и прикладная статистика. М.: ЮНИТИ.

Айвазян С. А., Мхитарян В. С. (2001). *Прикладная статистика в задачах и упражнениях*. М.: ЮНИТИ.

Гмурман В. Е. (2014). *Теория вероятностей и математическая статистика*. 12-е издание.. М., ЮРАЙТ.

Шведов А.С. (2005). *Теория вероятностей и математическая статистика*. Москва. ГУ ВШЭ.

Boos D. D., and L. A. Stefanski (2013). *Essential statistical inference: Theory and methods* (*Springer Texts in Statistics*)

Academic integrity policy

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.