Lublin, 5.04.2023 r.

Streszczenie w języku angielskim rozprawy doktorskiej mgr Luizy Pańczyk
Temat rozprawy: Wybór optymalnych L–statystyk w estymacji kwantyli

We study the classical statistical problem of the estimation of quantiles by order statistics of the random sample and their linear combinations called L-statistics. For fixed sample size, we determine the single order statistic which is the optimal estimator of a quantile of given order, next we determine the linear combination of two adjacent order statistics. The rules for the choice known in the literature are not precise and based on intuition. Therefore, we propose a totally new approach to the problem introducing new optimality criterion based on the use of nonparametric sharp upper and lower bounds on the bias of the estimation. One of the main achievements of the dissertation is to find explicit analytic formulas for these bounds. Applying this new criterion we derive relatively simple rules of the choice of optimal quantile estimators which can be easily implemented in practise. This is also illustrated and analyzed with many examples of numerical computations.

In Chapter 1 we introduce the basic notions used in the thesis, namely quanitiles, quantile functions, order statistics and L-statistics. We also give numerous examples of known L-statistics used as quantile estimators.

In Chapter 2 we introduce new criterion of optimality of L-statistics as estimators of the quantile of a given order $p \in (0,1)$ with fixed sample size n. The criterion is based on optimal bounds of the bias of the estimation of the quantile by a given L-statistic and it has clear intuitive justification.

In Chapter 3 we determine the explicit form of the bounds of the estimation of quantile functions by a single order statistic. More precisely, we show that for $2 \leq j \leq n-1$ there exists an interval (θ_j, ξ_j) containing the number $\frac{j}{n}$ such that for any distribution function F with the finite variance σ_F^2 and any $p \in (\theta_j, \xi_j)$ we have the inequalities

$$-\frac{F_{j:n}(p)}{\sqrt{p(1-p)}} \leq \frac{\mathbb{E}_F X_{j:n} - F^{\rightarrow}(p)}{\sigma_F} \leq \frac{\mathbb{E}_F X_{j:n} - F^{\leftarrow}(p)}{\sigma_F} \leq \frac{1-F_{j:n}(p)}{\sqrt{p(1-p)}}.$$

Moreover, the bounds are attained for appropriately chosen two-point distributions. For p outside the interval either left- or right-hand side of the inequality has much more complicated form.

In Chapter 4 using the results of the previous chapters we determine the optimal single order statistic as an estimator of the quantile of order $p \in (0,1)$. More precisely, we prove that there exists the unique set of numbers $a_{j,n}$, $1 \leq j < n$, such that if $p \in (a_{j-1,n}, a_{j,n})$, then the optimal choice is $X_{j:n}$, and for $p = a_{j,n}$ both statistics $X_{j:n}$ and $X_{j+1:n}$ are equally good. Moreover, the numbers satisfy the conditions $a_{j,n} < \frac{j}{n} < a_{j+1,n}$ for $1 \leq j < \frac{n}{2}$ and $a_{n-j,n} = 1 - a_{j,n}$. In particular, contrary to the classical choice, the optimal estimator of the quantile $x_{k/n}$ is $X_{k:n}$ if $1 \leq k < \frac{n}{2}$ or $X_{k+1:n}$ if $\frac{n}{2} < k \leq n$.

In this chapter we also consider an equivalent optimality criterion based on the minimization of the maximal bias.

In Chapter 5 we apply the new criterion to the problem of the choice of the best quantile estimator of the form of a linear combination of two adjacent order statistics. We determine the numbers $b_{k,n}$, $1 \leq k \leq n$, such that if $p \in (b_{k,n}, b_{k+1,n})$, then the optimal estimator of the quantile x_p is $(1-\alpha_{n,p})X_{k:n} + \alpha_{n,p}X_{k+1:n}$. The coefficient $\alpha_{n,p}$ is determined uniquely so that $\alpha_{n,p} \in \left(0, \frac{1}{2}\right)$ if and only if $p \in (b_{k,n}, a_{k,n})$, where the numbers $a_{k,n}$ are described above. In particular we obtain the classical definition of the sample median. We underline that the obtained estimator is not a linear interpolation of the quantile function, contrary to most of known estimators.

In the last chapter we study mean-square error of our estimators and we compare them with the estimators known form the literature. Unfortunately, due to complicated analytical expressions, most of the results of this chapter are just examples of numerical computations.