# TEACHING STATISTICS THROUGH PARADOXES

**G.J. Székely**
**Eötvös University**
**Budapest, Hungary**

Not to write satire, at least in Juvenalis' opinion, is hard, but not to find paradoxes in mathematical statistics is even harder. Bayesianism and anti-Bayesianism is one of the evergreen fields of controversies. Several recent paradoxes of statistics are due to H. Robbins, Ch. Stein, A. Stuart, D. Basu R.R. Bahadur, D.F. Friedman, D.V. Lindley, J.W. Tukey, L. LeCam, J. Sethuraman, V. Barnett, etc. Most of the classical and several recent paradoxes are contained in the forthcoming book Székely (1986). As the teaching experience of "Budapest Semesters in Mathematics" shows, these paradoxes help the students to perceive new ideas if the courses follow the style of Socrates' dialogues. The earliest paradox of the book comes from the Bible where the story of Jacob and Laban is elucidated from a new, mathematical angle. The following paradoxes have been crystallized during the past years, some of them are results of discussions and debate with my colleague, T.F. Mori.

1. Why do buses run more frequently in the opposite direction in the sense that the expected number of buses passing in the opposite direction while we are waiting for the one we take is bigger than 1/2?

2. M-estimates and L-estimates for location parameters are "almost" incompatible (except some trivial cases as the arithmetic mean of observations in the normal distribution case and the median estimator in the Laplace distribution case).

3. For normal distribution the usual unbiased estimators of the mean and the variance ($\bar{x}$ and $s^2$) are independent (in fact this is a characteristic property of normal distributions) and thus their correlation is 0. This correlation remains 0 if the normal distribution is replaced by any other symmetric distribution having finite variance. Surprisingly, if symmetry is replaced by unimodality then the upper limit of the correlation of $\bar{x}$ and $s^2$ is $\sqrt{15/16}$ ! If not even unimodality is supposed then the upper limit is 1 (which is never attained).

4. Uncorrelatedness of the random variables X and Y does not imply their independence. But (!)

   (i) if X and Y are uncorrelated under restrictions $x_1 \leq X \leq x_2$ and $y_1 \leq Y \leq y_2$, whatever the numbers $x_1 < x_2$ and $y_1 < y_2$ be, then X and Y are independent,

   (ii) if the regression of Y on X and on the regression of X on Y is linear (as in the case of two-dimensional normal distributions) then correlation is the same as maximal correlation thus uncorrelatedness implies independence.

This regression property holds, for example, in the case of two-dimensional beta distribution whose probability density function is proportional to $x^{a-1}y^{b-1}(1-x-y)^{c-1}$ for $x,y,1-x-y \geq 0$, and 0 otherwise ($a,b,c, > 0$). A simple consequence of this property of beta distribution is the following theorem: the correlation of elements of an ordered sample is maximal if the sample comes from a uniform distribution.

5. Several paradoxical situations occur when the maximum likelihood equation has more than one root. An example is the following: the sample elements are normally distributed with expectation $\mu$ and variance proportional to $|\mu|$.

6. Using the analogy of "testing normality" if we want to test "Poissonity" then first we should construct a function g of the sample elements $X_1$, $X_2$, ... $X_n$ such that the distribution of $g(X_1,X_2, \ldots, X_n)$ does not depend on the parameter of the supposed Poisson distribution. Such a function, however, does not exist (except the trivial, constant function)!

7. When apply a two-sided F test, textbooks frequently say that the interval (a,b) where we accept the hypothesis can be calculated from the equations:

$$\int_{-\infty}^{a} dF(x) = \varepsilon/2 \quad \text{and} \quad \int_{b}^{\infty} dF(x) = \varepsilon/2$$

where F is the distribution function of the underlying F-distribution and $\varepsilon$ is the probability of the error of the first kind. However, if both tails of the F-distribution have probabilities $\varepsilon/2$ then the test, in general, is not unbiased. If we prefer unbiased tests then we should require the following conditions:

$$\int_{a}^{b} dF(x) = 1-\varepsilon \qquad \int_{a}^{b} x\,dF(x) = (1-\varepsilon) \int_{-\infty}^{+\infty} x\,dF(x) \quad .$$

Thus we must not use equal probabilities in the tails "for simplicity".

8. Several statistical inferences are based on the limit theory of relative frequencies stating that this frequency k/n tends to the probability p of the observed event. Suppose that after $n_0$ trials the relative frequency is $0 < p < 1$, and from this trial on the probability of events varies after each observation, namely the probability of the next event is just the relative frequency of all previous observations. One can prove that even in this case the relative frequency converges with probability one as the number of the observations n tends to infinity. The distribution of the limit is beta with parameters (p,1-p).

9. Suppose X has an exponential distribution whose parameter has an exponential apriori distribution, too. The parameters of this a priori distribution is also exponentially distributed etc. If we stop after n steps

and the resulting random variable is $X_n$, then the limit distribution of $n^{-1/2}\log X_n$ is normal with expectation 0 and variance $\pi^2/6$!

10. In the last paradox of the talk we discuss the idea of giving statistical solutions for non-statistical problems in number theory, graph theory, etc. On the other hand we can give non-statistical, non-probabilistic solutions for statistical-probabilistic problems using topological algebraic arguments.

    This is the topic of our book Ruzsa-Székely (1987).

## References


Székely, G.J. (1986). Paradoxes in probability theory and mathematical statistics. Reidel, Dordrecht.

Ruzsa, I.Z., & Székely, G.J. (in press). Algebraic probability theory: How to eliminate probabilities from probability theory. Wiley.