DILEMMAS IN TEACHING AN INTRODUCTION TO STATISTICAL INFERENCE TO ECONOMICS STUDENTS

Alona Raviv and Talma Leviatan
Tel Aviv University
Israel

Our approach to teaching statistics to economics students is presented in our 2-volume book. Though written as a textbook for economics students, a broad population who seem to recognize the universalism of the approach is using it. We discuss here the dilemmas encountered in deciding course content and level of mathematical sophistication. The course at Tel Aviv University has undergone several changes, reflecting the changing viewpoints of the "consumer" (Dept. of Economics) and the "supplier" (Dept. of Statistics and OR). Presently, the course is almost as rigorous as the courses offered to statistics majors. Students appreciate the challenges arising in statistical theory, even if they lack sophisticated mathematical reasoning. We illustrate by examples how we teach relatively high-level mathematical concepts in statistical inference, such as maximum likelihood estimation and the Neyman-Pearson Lemma.

INTRODUCTION

The Introduction to Statistics course for economics students has changed considerably since the early days of establishing the Economics Department at Tel Aviv University. For about 15 years, economics students took a very elementary 2-semester course, which included a lot of descriptive statistics, very little probability, and some relatively unsophisticated statistical inference (a so-called "cook book" course, also given to other social sciences majors). Later a course was developed for economics students, in which the special needs and the higher mathematical education of this group were recognized. More time was devoted to probability, and statistical inference was given more accurately. Still, the approach to statistical inference was somewhat elementary, at the level of standard statistical textbooks directed at the social sciences, like Hays (1973). It took several more years for the Economics Department to realize that their students needed to have more statistical knowledge and understanding, to be able to take advanced courses in economics. So, a new syllabus was designed, which included much of what statistics major students have in Statistical Theory. One big obstacle is the fact that this course must be given to first year students who did not yet have enough mathematical background (mainly integration). Consequently, though the statistics course is rigorous, it deals mostly with discrete variables; the only continuous distributions included are the uniform, exponential, and normal ones. For some years we tried using the book by Hodges and Lehmann (1970), but in 1992, we decided to write our own textbook for the 2-semester course Introduction to Statistics for economics students at Tel Aviv University (Leviatan & Raviv, 1995; Raviv & Leviatan, 1994). The book is made up of two volumes: I – Probability, and II – Statistical Inference and includes many "real-life" problems, with detailed solutions. It appears to be of great appeal to a wide range of students. Even Statistics majors in Tel Aviv and in other Israeli universitie, who take the more advanced Statistical Theory course, and also Engineering students (and others), have benefited from this book in understanding the concepts and ideas of statistical inference.

Since publishing the book requests have been made by the Economics and Statistical communities to change the syllabus in various directions. For example, we received a request to expand the probability section, to include continuous bivariate distributions and regression theory (the application of which they get in Econometrics), and emphasize the applied aspects. Such dilemmas are not easy to solve, as the timetable for the course is restricted, and the lack of mathematical knowledge and reasoning of the students at this stage has to be taken into consideration.

In summary, we believe that economics students benefit from the theoretical way in which we teach statistics, and manage very well with it. We introduce the complicated concepts in non-formal, yet accurate ways, using a lot of graphical illustrations to demonstrate the mathematical ideas. The discussions are preceded by simple examples to prepare the ground for the non-trivial issues to come. In this paper we describe how we teach two important and central topics – maximum likelihood estimation, and the Neyman-Pearson Lemma.

MAXIMUM LIKELIHOOD ESTIMATION

In our many years of teaching statistics to economics (and other nonstatistics major) students we have found the topic of estimation, in general, and maximum likelihood estimation, in particular, difficult for students to fully understand. The main problem is the confusion between parameter (state of nature) and statistic (sample function, estimator). We show here how we deal with the notion of the maximum likelihood method. Example 1 explains what we mean by "*how likely is the value of the parameter*," for a specific sample result.

Example 1 (Normal distribution). Measurements were taken from a normal distribution, with unknown mean μ and standard deviation of 2, resulting in 7.3, 5.4, 8.1, 4.2 (chemical content in a popular beverage, in grams, with average $\bar{x} = 6.25$). We present (Figure 1) the sample results ($\bar{x}$ is designated by Δ), together with a few possible normal distributions. In class we show the sample and one normal curve on two separate transparencies, moving the normal curve horizontally along the *x*-axis, to indicate what values of the parameter μ are more "likely" to give the sample actually observed. We do not undertake any computations at this stage. We only want the students to be aware of the "unlikeliness" of values of μ not in the neighborhood of 6 or 7.

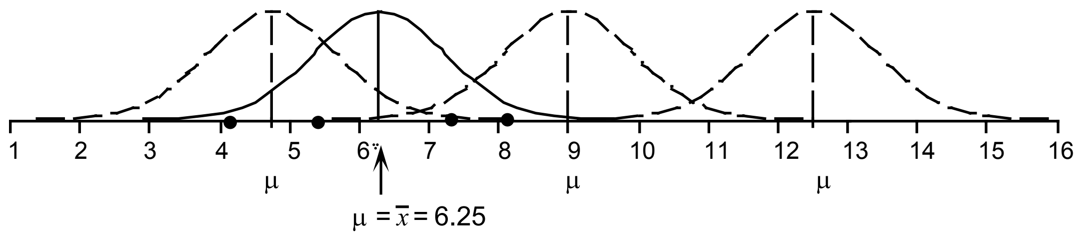

Figure 1. A Sample of 4 Observations, and a Few Normal Density Functions.

Next we present a discrete Poisson problem.

Example 2 (Poisson distribution). Suppose we have the results 1, 4, 2, 1, 7 of a sample from a Poisson distribution with parameter λ (number of cars entering the university parking lot in a 5-minute time period). In this case we cannot draw a single probability distribution that can be moved along the *x*-axis. Thus, for presentation in class, we put a few Poisson distributions separately on the sample data, to examine which value of the parameter λ best suits the sample results. Figure 2 gives only two such distributions. These figures show that the value λ = 3 is far more "likely" to give the observed sample than the other proposed values of λ.

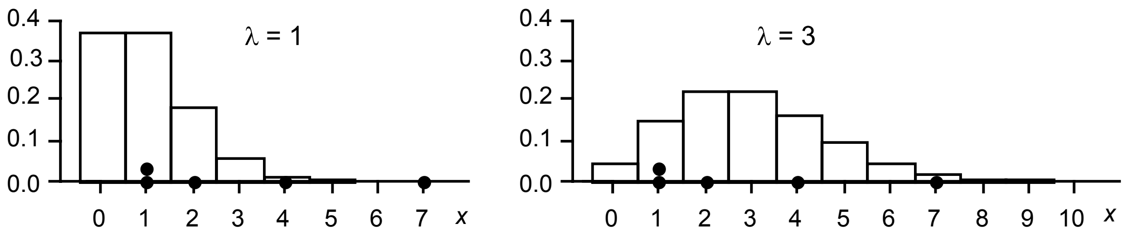

Figure 2. A Sample drawn from a Poisson Distribution, with Two Distribution Functions.

Now we can write the exact likelihood function (probability of the observed sample as a function of the parameter λ) for our specific sample, and the logarithm of the likelihood function:

$$\text{(1)}\quad L(\lambda) = P_\lambda(X_1 = 1, X_2 = 4, X_3 = 2, X_4 = 1, X_5 = 7) = e^{-5\lambda}\frac{\lambda^{15}}{1!4!2!1!7!} = \frac{e^{-5\lambda}\lambda^{15}}{48(7!)} \quad \text{for } \lambda > 0$$

$$\text{(2)}\quad \log L(\lambda) = \log\frac{e^{-5\lambda}\lambda^{15}}{48(7!)} = 15\log\lambda - 5\lambda - \log[48(7!)] \quad \text{for } \lambda > 0$$

A graph of the likelihood, Eqn. (1), is shown in Figure 3a, and a graph of the logarithm of the likelihood, Eqn. (2), is shown in Figure 3b. By means of the latter figure we remind the students that maximizing the logarithm is equivalent to maximizing the original likelihood (due to the increasing monotonicity of the logarithm function).

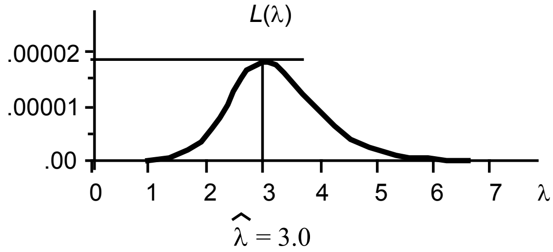

Figure 3a. The Likelihood Function

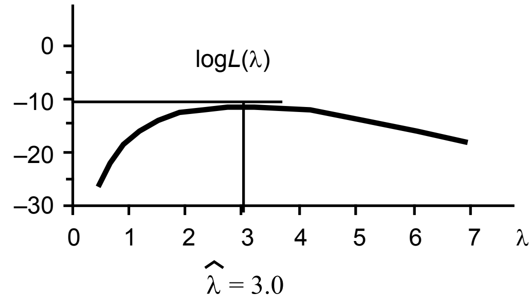

*Figure 3b.*The Logarithm of the Likelihood Function.

MOST POWERFUL TESTS: THE NEYMAN-PEARSON LEMMA

In our book we describe in a simple way the theoretical, not so easy to understand, Neyman-Pearson principle for finding most powerful tests. The principle is presented after defining the basic concepts of hypothesis testing, i.e., rejection region, type I and type II errors, significance level, and power of a test. First, we define the *likelihood ratio* concept for two simple hypotheses, as a function of the sample results. This ratio expresses how much larger the probability of an outcome under a specific alternative P_1 is, compared to the probability under a specific null model P_0.

Below we show how to find a "good" α-level test for two simple hypotheses, i.e., keeping the type I error less than a specific value α, and having a maximum power (or low type II error). As always, we start with an example.

Example 3. A person keeps his 5 keys on a ring. When he returns home at night after having a few drinks, he tries to open the door by trying the keys one by one. Let X be the number of trials until opening the door. Suppose that the null hypothesis is H_0: The person is so drunk that he tries the keys with replacement; the alternative hypothesis is H_1: The person is only slightly drunk, and thus tries the keys without replacement. (In the book we use this example with 10 keys.) In terms of X's distribution, H_0: $X \sim$ Geometric(1/5), H_1: $X \sim$ Unif(1, 5).

Suppose that the drunken person succeeds in opening the door with the 4th trial. The probability of getting that result under H_0 is $P_0(4) = P_{H_0}(X = 4) = [4/5]^3[1/5] = 0.1024$, while under H_1 it is $P_1(4) = P_{H_0}(X = 4) = 1/5 = 0.2$. The ratio between the two likelihoods (the likelihood ratio) is $\frac{1/5}{4^3/5^4} = [5/4]^3 \cong 1.953$. We see then that the result X = 4 is about twice as likely to happen under the alternative hypothesis than under the null model. In other words, if we decide to reject H_0 for this result, we "gain" about twice the power than what we "pay" for the type I error. The Neyman-Pearson Lemma states that in order to decide which model is more plausible for a specific result, the most powerful decision rule is the one that rejects the null hypothesis when the likelihood ratio is "too large".

The likelihood ratio for a general sample point $X = x$ in our example is

(3) $\lambda(x) = \frac{P_1(x)}{P_0(x)} = \frac{1/5}{[4/5]^{x-1}[1/5]} = [5/4]^{x-1}$ for x = 1, 2, …, 5, and $\lambda(x) = 0$ for $x > 5$.

The likelihood ratio principle: Arrange all the possible sample points in descending order of their likelihood ratio. This order determines the order in which the points enter the rejection region. The larger the likelihood ratio for a point is, the earlier it enters the rejection region.

Table 1 shows all x values arranged by descending size of their likelihood ratio, together with their probabilities and cumulative probabilities under the two models.

Table1
The Values of X in Likelihood Ratio Descending Order

x	order	$P_0(x)$	$P_1(x)$	$\lambda(x)$	$P_0(x \le X \le 5)$	$P_1(x \le X \le 5)$
5	1	.082	.2	2.441	.082	.2
4	2	.102	.2	1.953	.184	.4
3	3	.128	.2	1.563	.312	.6
2	4	.160	.2	1.250	.472	.8
1	5	.200	.2	1.000	.672	1.0
6	6	.066	.0	0.000	.738	1.0
⋮						

The idea of the likelihood ratio principle is neither to reject the null hypothesis for points having small probabilities under H_0, nor for points having high probabilities under H_1, but to reject those for which the *ratio* between the two probabilities is high, i.e., *a higher "profit" for less "expenses"*. Using the likelihood ratio principle, there is a whole class of possible rejection regions (tests). Three of them are presented in Table 2.

Table 2
Rejection Regions for Example 3 that are based on the Likelihood Principle

C	$P_0(C)$	$P_1(C)$
$C_1 = \{5\}$	.082	.2
$C_2 = \{4, 5\}$	.184	.4
$C_3 = \{3, 4, 5\}$	.312	.6

To use the likelihood ratio principle for a given significance level α, we arrange the possible outcomes in descending order of their likelihood ratio. We enter the first outcomes of this list into the rejection region one by one. The decision of where to stop, namely, which likelihood ratios are considered "too large", is determined solely by the required significance level α; we are allowed to enter outcomes into the rejection region as long as the probability of the region, under H_0, does not exceed the size α. In our example, if we allow only $\alpha = .05$, then no rejection region is available. If we allow $\alpha = .10$, the only appropriate test has the rejection region $C_1 = \{5\}$, with a maximal power of $\pi = .2$. For $\alpha = .20$, we use the rejection region $C_2 = \{4, 5\}$, with a maximal power of $\pi = .4$. The next region is already too large, having a level of $\alpha = .312$.

THE NEYMAN-PEARSON LEMMA

We present in class the Neyman-Pearson Lemma, omitting the formal proof (in the book a detailed proof is given for discrete distributions as an appendix). The lemma shows us how to get the class of most powerful tests, out of which we select the best one according to the given α.A scatter diagram of pairs (α, π) for all the possible tests is a nice way to demonstrate that the class of likelihood ratio tests yields indeed the most powerful tests.Many other non trivial issues in statistical inference can be taught using charts and graphs. For example, we use graphs to explain the meaning of a confidence interval, as well as proving the relationship between a confidence interval and 2-sided hypothesis testing.

REFERENCES

Hays, W.L. (1973). *Statistics for the Social Sciences*. London: Holt, Rinehart and Winston.

Hodges, J.L., & Lehmann, E.L. (1970*). Basic Concepts of Probability and Statistics* (2nd edn.). San Francisco: Holden-Day.

Leviatan, T., & Raviv, A. (1995*). Introduction to probability and statistics: Probability* (2nd edn.). Tel Aviv: Amihai. (In Hebrew).

Raviv, A., & Leviatan, T. (1994). *Introduction to probability and statistics: Statistical inference*. Tel Aviv: Amihai. (In Hebrew).