THE USE OF COUNTEREXAMPLES IN LEARNING PROBABILITY AND STATISTICS

Jordan Stoyanov
Institute of Mathematics
Bulgarian Academy of Sciences
Sofia, Bulgaria

Introductory Notes

Many of the lecturers here expressed their opinion that the primary task of statistics departments is to produce truly professional statisticians. However at least a part of the statisticians must have a serious mathematical background especially in probability theory, stochastic processes and mathematical statistics. The new term *stochastics* is very suitable for these three disciplines. Scientists working in the field of stochastics will be called *stochasticians.* It is obvious that the stochasticians are educated at the mathematics departments of universities. So, the main problem is how to organize the teaching process for the students in order to produce good statisticians and stochasticians who would be able to solve problems arising in the real world as well as other purely theoretical problems.

Counterexamples: Meaning and Use

Since the curriculum of the mathematics departments includes courses in stochastics, and we have to teach such courses, the goal is to do this in the best possible way. On one hand we use the best existing textbooks, the classical books written by A. Kolmogorov, H. Cramér, P. Lévy, W. Feller, J. Doob, B. Gnedenko, M. Loeve, K.L. Chung, E. Lehmann, as well as many other recent textbooks and lecture notes. On the other hand for our lectures and seminars we can use different approaches, ideas and schemes.

Let us describe one possible general scheme which we could follow in teaching courses in stochastics:

THEORY EXAMPLES COUNTEREXAMPLES APPLICATIONS

We use the term *counterexample* in its standard meaning as accepted in mathematics. The counterexample is a statement which in a sense is opposite to another statement. If we have proved a statement under conditions which are necessary and sufficient, then everything is clear: any change of the conditions implies a change of the statement, and any new statement, even close to the given one, requires new conditions. So, our attention will be focussed on statements involving conditions which are only necessary or only sufficient.

The counterexamples help us to understand better the main results in stochastics as well as to discover the relationship between near notions and concepts. The systematic use of counterexamples makes the courses and seminars much more interesting. It is a nice and very natural way to stimulate our students to think deeper and to search for answers to nontrivial questions. There is no doubt that the learning of stochastics can be done much more effectively if we use counterexamples. The role of such examples in forming the so–called stochastic thinking is essential.

Let us mention finally that not only the so–called positive results and examples, but also the counterexamples define the power, the wideness, the depth, the degree of nontriviality, and of course, the beauty of the theory.

List of Selected Counterexamples

Now we shall present a short list of specific counterexamples. They concern the basic probabilistic objects: random events, random variables and stochastic processes. We are not going to consider here counterexamples from mathematical statistics.

1. Introducing a probability measure P we assume that P is either additive or σ–additive. Every σ–additive measure is additive, but not conversely. There are probability measures which are additive but not σ–additive.

2. For random events we consider the notions of a mutual independence and a pairwise independence. There are sets of random events which are pairwise independent but not mutually independent.

3. Suppose A, B, C are random events satisfying the relation:

$$P(ABC) = P(A)P(B)P(C).$$

Question: Does it follow that A, B, C are mutually independent? Answer: No.

4. If Ω is the space of the outcomes of some experiment and P is a probability measure on the set of random events, then it is not always true that in Ω do there exist nontrivial independent events.

5. Let X, Y, Z be random variables. Suppose X and Y are identically distributed: $X \stackrel{d}{=} Y$. Does it follow from here that $X \cdot Z \stackrel{d}{=} Y \cdot Z$? The answer is negative.

6. If $F(x_1,\cdots,x_n)$, $(x_1,\cdots,x_n) \in R^n$ is the distribution function (d.f.) of the random vector $(X_1,\cdots,X_n)$ then the marginal d.f.s $F_1(x_1),\cdots,F_n(x_n)$ of $X_1,\cdots,X_n$, respectively, are determined uniquely. But not conversely. For given marginals $F_1,\cdots,F_n$ we can find even infinitely many n–dimensional distributions having $F_1,\cdots,F_n$ as marginal distributions.

7. Let f be a unimodal probability density function and $\star$ means the convolution operation. Then in general $f \star f$ need not be unimodal.

8. One of the important properties of the expectation E is the linear property: $E[X_1+X_2] = EX_1 + EX_2$, $E[X_1+X_2+X_3] = EX_1 + EX_2 + EX_3$, etc. Question: Does this linear property always hold? Answer: No.

9. Suppose X is a random variable (r.v.) with a symmetric distribution. Then all odd order moments of X vanish: $E[X^{2k+1}] = 0$, $k = 0,1,2,\cdots$ But not conversely. There are r.v.'s whose odd order moments vanish and despite this fact the distributions of these variables are nonsymmetric.

10. If X and Y are independent r.v.'s and $g_1(x)$, $g_2(x)$, $x \in R^1$ are arbitrary (Borel) functions, then $g_1(X)$ and $g_2(Y)$ are r.v.s which are independent. However the converse statement is not always true. In particular, X^2 and Y^2 can be independent even if X and Y are dependent.

11. Let X be a r.v. with characteristic function (ch.f.) $\phi(t)$, $t \in R^1$. If $E|X| < \infty$, then $\phi'(0)$ exists. But not conversely. The derivative $\phi'(0)$ can exist even if $E|X| = \infty$.

12. If $(X_1,\cdots,X_n)$ is a random vector with n–dimensional normal distribution then each of $X_1,\cdots,X_n$ and any subset of them has a normal distribution. However the converse statement is not true. We can present an example of non–normal vector $(X_1,\cdots,X_n)$ such that any k of its components, $X_{i_1},\cdots,X_{i_k}$, $k = 1,2,\cdots,n-1$ are mutually independent and each X_j is normally distributed.

13. Let F be a d.f. and $\{\alpha_1, \alpha_2,\cdots\}$ be the set of its moments:

$$\alpha_k = \int x^k dF(x), \quad k = 1,2,\cdots .$$

If F is the only d.f. with these moments, we say that the moment problem for F is determined. Otherwise the moment problem is undetermined. It can be shown that there are absolutely continuous distributions and discrete distributions which are not uniquely determined by their moments.

14. Usually we introduce and study the following kinds of convergence of sequences of r.v.s:

convergence in distribution ($\xrightarrow{d}$), convergence in probability ($\xrightarrow{P}$),

convergence in L^r–sense ($\xrightarrow{L^r}$), and almost sure convergence ($\xrightarrow{a.s.}$).

Then $\xrightarrow{a.s.}$ implies $\xrightarrow{P}$; $\xrightarrow{L^r}$ implies $\xrightarrow{P}$; $\xrightarrow{P}$ implies $\xrightarrow{d}$. It can be shown that all these implications are strong.

15. For sequences of r.v.s $\{X_n, n \geq 1\}$ we consider two laws of large numbers, the weak law (WLLN) and the strong law (SLLN). If $\{X_n\}$ satisfies the SLLN then it satisfies also the WLLN. But not conversely. There are sequences $\{X_n\}$ obeying the WLLN but not the SLLN.

16. It is well–known how important is the central limit theorem (CLT) for sequences of r.v.'s $\{X_n, n \geq 1\}$. It can be shown that $\{X_n\}$ can obey the CLT without validity of the Feller condition, or without the uniform asymptotic negligibility condition. Further, it can happen that $\{X_n\}$ satisfies the integral CLT but not the local CLT (for densities).

17. For stochastic processes $X = (X_t, t \geq 0)$ we consider several important properties such as measurability, separability, continuity, integrability. There are interesting counterexamples illustrating these properties. In particular, let $X = (X_t, 0 \leq t \leq 1)$ be a process with covariance function $\Gamma(s,t) = E[X_s,X_t]$, s, $t \in [0,1]$. Then there is a classical result: The integral $\int_0^1 X_t\, dt$ exists in L^2–sense if and only if the Riemann integral

$$\int_0^1 \int_0^1 \Gamma(s,t)dsdt$$

exists. However, this classical result is not correct. There is a process X whose covariance function Γ is not Riemann integrable but nevertheless the integral $(L^2)\int_0^1 X_t\, dt$ exists.

18. If X is a Markov process, then its transition probabilities satisfies the Chapman–Kolmogorov equation. But not conversely. There are random sequences $(X_n, n = 0,1,2,\cdots)$ whose transition probabilities satisfy the Chapman–Kolmogorov equation, and despite this fact, these sequences do not form Markov processes.

19. Every strictly stationary $\mathscr{L}^2$ process is also weakly stationary. However a process can be weakly stationary without being strictly stationary.

20. If X is a point process such that the number of the points in any interval has a Poisson distribution and any two (or N) increments are independent, it does not imply that X is a Poisson process.

Detailed Description of a Few Counterexamples

Now we shall describe in details some specific examples. A part of them is given specially in a form which is suitable even for students of the secondary schools.

Example A (S. Bernstein, 1928). Suppose a box contains four tickets labelled by 112, 121, 211 and 222. Let us choose one ticket at random and consider the random events: $A_1 = \{1 \text{ occurs at the first place}\}$, $A_2 = \{1 \text{ occurs at the second place}\}$ and $A_3 = \{1 \text{ occurs at the third place}\}$. Obviously we have: $P(A_1) = \frac{1}{2}$, $P(A_2) = \frac{1}{2}$, $P(A_3) = \frac{1}{2}$. Further, since $A_1A_2 = \{112\}$, $A_1A_3 = \{121\}$ and $A_2A_3 = \{211\}$, we find easily that $P(A_1A_2) = \frac{1}{4}$, $P(A_1A_3) = \frac{1}{4}$, $P(A_2A_3) = \frac{1}{4}$. So we conclude that the three events A_1, A_2, A_3 are pairwise independent. However $A_1A_2A_3 = \phi$ the empty set, and hence $P(A_1A_2A_3) = 0 \neq (\frac{1}{2})^3 = P(A_1)P(A_2)P(A_3)$. Therefore these three events are not mutually independent. Thus, the general conclusion is that the pairwise independence of a given set of random events does not imply that these events are mutually independent.

Note that the Bernstein's example can be generalized. namely, we can describe an experiment and define n random events $A_1, A_2, \cdots, A_n$ such that any $n-1$ of them are mutually independent but all n are not.

Example B. Suppose A_1, A_2, A_3 are random events such that

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$$

The question is whether these events are mutually independent. The answer will follow from a simple and instructive example. Indeed, let us consider the following experiment. Toss two different standard dice, white and black, and let the space Ω of the outcomes consist of all ordered pairs ij, $i, j = 1, \cdots, 6$ and each point of Ω has probability $1/36$. Consider the events:

$A_1 = \{\text{first die} = 1, 2 \text{ or } 3\}$, $A_2 = \{\text{first die} = 3, 4 \text{ or } 5\}$, $A_3 = \{\text{sum of faces is } 9\}$.

Clearly we have

$A_1A_2 = \{31, 32, 33, 34, 35, 36\}$, $A_1A_3 = \{36\}$, $A_2A_3 = \{36, 45, 54\}$, $A_1A_2A_3 = \{36\}$.

Then $P(A_1) = 1/2$, $P(A_2) = 1/2$, $P(A_3) = 1/9$ and

$$P(A_1A_2A_3) = \frac{1}{36} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{9} = P(A_1)P(A_2)P(A_3).$$

Nevertheless the events A_1, A_2, A_3 are not mutually independent. This follows from the relations:

$$P(A_1A_2) = \frac{1}{6} \neq \frac{1}{4} = P(A_1)P(A_2), \qquad P(A_1A_3) = \frac{1}{36} \neq \frac{1}{18} = P(A_1)P(A_3),$$

$$P(A_2A_3) = \frac{1}{12} \neq \frac{1}{18} = P(A_2)P(A_3).$$

Example C. Let X and Y be r.v.s whose joint distribution is absolutely continuous with density

$$f(x,y) = \frac{1}{2\pi\sqrt{3}}\left[\exp\left[-\tfrac{2}{3}(x^2+xy+y^2)\right]+\exp\left[-\tfrac{2}{3}(x^2-xy+y^2)\right]\right], \ (x,y) \in R^2.$$

Direct calculation shows that each of X and Y has a standard normal distribution, $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$. Further we easily find that $E[X\cdot Y] = 0$ which implies that the variables X and Y are uncorrelated. So, we have two r.v.s which are normally distributed and are uncorrelated. Question: Are X and Y independent?

Note that this question is followed sometimes by a wrong answer. The correct answer is negative, i.e. X and Y are not independent. Being uncorrelated they would be independent only if their joint density f(x,y) is a bivariate normal density function. The explicit form of f(x,y) shows that this is not the case.

Note additionally that if X and Y are independent r.v.'s (with finite variances), then they are uncorrelated. The above example shows that in general the converse statement is not true (even if X and Y are normally distributed).

Example D. Let as usual Φ and φ denote the standard normal d.f. and its density, respectively. Consider another function

$$H(x,y) = \Phi(x)\Phi(y)[1 + \epsilon(1-\Phi(x))(1-\Phi(y))], \qquad (x,y) \in R^2$$

where ϵ is any number in the interval $[-1,1]$. It is easy to verify that H(x,y) is a two–dimensional d.f. with marginal distributions $\Phi(x)$ and $\Phi(y)$, respectively. Clearly, if $\epsilon \neq 0$ then H is non–normal. So, we have obtained infinitely many non–normal distributions such that all their marginals coincide with the standard normal distribution. If we take $h(x,y) = \varphi(x)\varphi(y)[1 + \epsilon(2\Phi(x)-1)(2\Phi(y)-1)]$ we get infinitely many non–normal densities (for $\epsilon \neq 0$) such that all marginal densities coincide with the standard normal density function.

Example E. Let X be a r.v. with distribution $\mathcal{N}(a, \sigma^2)$ and let f be its density. Take n copies of X, say $X_1,\cdots,X_n$, $n \geq 3$ and define their joint density g_n as follows:

$$g_n(x_1,\ldots,x_n) = \left[\prod_{i=1}^{n} f(x_i)\right]\left[1 + \prod_{j=1}^{n} (x_j-a)f(x_j)\right].$$

Obviously, g_n is not a n–dimensional normal density. Let us choose k of the given n variables where $k = 2, 3,\cdots,n-1$. If our choice is $X_{i_1}, \cdots, X_{i_k}$ and g_k denotes the density of $(X_{i_1},\cdots,X_{i_k})$ we find easily that $g_k(x_1,\cdots,x_k) = f(x_1) \ldots f(x_k)$. Therefore the variables X_{i_j} are jointly normally distributed and moreover they are mutually independent. Recall that this holds for $2 \leq k \leq n-1$ despite the fact that the random vector $(X_1,\cdots,X_n)$ is non–normally distributed.

Concluding Remarks

a) In the literature we can find many other interesting counterexamples. Those 20 given above are chosen from my much bigger collection of more than 500 counterexamples. The best ones are prepared as a book which according to the plan will be published in 1987 by "John Wiley & Sons" (Chichester). The variety of the counterexamples makes then suitable for a wide circle of readers: students in the high schools and their teachers, but mainly for university students and even for professional stochasticians.

b) Clearly there are a lot of problems in statistics and stochastics and they must be solved jointly by the statisticians and the stochasticians. The ICOTS II (Victoria, August 1986) is an example, not a counterexample, of good collaboration between representatives of many countries.

References

Feller, W. (1968, 1971). An Introduction to Probability Theory and Its Applications, Vol 1, Vol 2, John Wiley & Sons, New York.

Gnedenko, B. V. & Kolmogorov, A. N. (1954). Limit Distributions for Sums of Independent Random Variables, Addison Wesley, Cambridge (MA).

Shiryaev, A. N. (1984). Probability, Springer–Verlag, New York–Berlin.

Stoyanov, J. M. (1987). Counterexamples in Probability, John Wiley & Sons, Chichester.