The Philippine Statistician
1991, Volume 40, Nos. 1-2 & 3-4
Pages 47-56

PROBABILITY DISTRIBUTION IN MATHEMATICS

by

E. E. Escultura
Institute of Mathematical Sciences and Physics,
College of Arts and Sciences,
University of the Philippines at Los Baños
Los Baños, Laguna

Abstract

Counterexamples to two major theorems in mathematical analysis are presented and they are attributed to present inadequacy of the notions function, derivative and integral. The inadequacy is rectified with the introduction of set-valued function, generalized derivative and generalized integral and appropriate probability distributions.

1. Counterexamples

This paper serves as a general introduction to a series of papers on the applications of probability distributions.

Counterexamples to two major theorems in analysis have been discovered lately. The theorems are as follows:

I. If a function is absolutely continuous then it is differentiable (Royden, 3rd. ed. [4]; this is really meant differentiable, a.e.).

II. (Lebesgue Theorem) A bounded function is Riemann integrable if and only if its set of discontinuity has measure zero (given as an exercise in Royden, end. ed. [4]).

The counterexample to I was a paradox to Lebesgue which was explained by Young [6]. It goes as follows:

We start with the triangle ABD in the figure whose base AB joins the points A(0,0) and B(1,0) and the slopes of sides AD and DB are, respectiviely, +1 and -1. We form a sequence

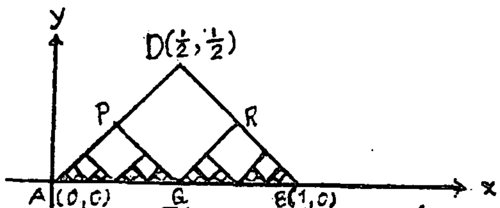

Figure 1

of saw-tooth functions $C_n : y_n = y_n(x)$, $0 \le x \le 1$, $n = 1,2,\ldots,$ whose first term has the graph which forms the sides AD and DB of the given triangle. For the second term, join the midpoint P of AD to the midpoint Q of AB and also the point Q to the midpoint R of DB to form the polygonal line APQRB from A to B. We denote this function by $C_2 : y_2 = y_2(x)$, $0 \le x \le 1$. By the geometry of the figure the slope of C_2 at any point either is +1 or -1 (except of corner points where the derivative does not exist in the ordinary sense in this set of measure 0). Also, the length $|C_2|$ of C_2 = length $|C_1|$ of $C_1 = \sqrt{2}$. Continuing the same construction we obtain a polygonal line C_m in the sequence C_n, $n = 1,2,\ldots$ which has the following properties:

$$\text{(a)} \qquad |C_m| = \int_{\{x|y_m=1\}} \sqrt{1+(y_m)^2}\,dx + \int_{\{x|y_m=-1\}} \sqrt{1+(y_m)^2}\,dx = \sqrt{2} = |C_1|$$

(the set at which y_m is simultaneously +1 and -1 has measure 0).

(b) The sequence C_n, $n = 1,2,\ldots,$ is uniformly convergent and since each C_n is continuous the limit which we denote by $C_o : y_o = y_o(x)$, $0 \le x \le 1$, is continuous. In fact, y_o coincides with the segment [0,1] whose equation is $y = 0$ which is absolutely continuous. Hence, y_o is also absolutely continuous.

(c) What about the derivative of y_o? Does it exist? If it does, what is it? We cannot have $y_o = 0$ as claimed by a visiting mathematician here. For, if that were so, it would violate the dominated convergence theorem applied to (a) since this would imply

$$\int_0^1 \sqrt{1+(y_o)^2}dx = 1 \neq \sqrt{2} \lim_{n\to\infty} |C_n| = \lim_{n\to\infty} \int_0^1 \sqrt{1+(y_n)^2}dx$$

In fact, the derivative of y_o does not exist in the ordinary sense because the sequence y_n does not converge to a single point. But the limit points of the sequence y_n, n = 1,2,..., form the set {1, 1}, i.e., y' is set-valued. (To be precise, the set-values of y_o are the limit points of the sequence <1, -1, 1, ...> or <-1, 1, -1, ...> or limit set of the sequence of sets <{1,-1}>.

The question here is quite categorical: Is $C_o: y_o = y_o(x)$, $0 \leq x \leq 1$, a counterexample to the above theorem? The answer is also a categorical <u>yes</u> and this counterexample is neither hair-splitting nor ambiguous. It does raise the following important points which are the source of this particular contradiction:

(a) Inadequacy of the present notion of <u>function</u>; this was pointed out by Young exactly 55 years ago!

(b) Inadequacy of the notion <u>derivative</u>; that the derivative of a function cannot be adequately expressed by its values and that there is a need to expand the notion to include set-valued derivative.

(c) Existence of a new kind of function called infinitesimal zigzag (introduced by Young in 1969).

(d) The fallacy in proving existence by approximation or convergence

For the Lebesgue theorem on the Riemann integral we present, as counterexample, the topologist sine curve given by f(x) = sin1/x. Known proofs of supposed Riemann integrability of sin1/x involves standard approximation techniques by a sequence of Riemann integrals or Riemann sums. One of the most common involves the construction of its Riemann integral outside the neighborhood $(-\epsilon, \epsilon)$ of x = 0, which certainly exists, and taking a sequence of such integrals by letting $\epsilon \to 0$, which certainly converges. But, necessarily, because the limit of a convergent sequence does not necessarily inherit a property (in this case that of being a Riemann integral) of the elements of the sequence. In this case, the limit is not the Riemann integral of sin1/x over an interval containing x = 0 because this function cannot be approximated by a step function from below, the closest it could get to the function near 0 is at y = -1 and as an approximation from above, at y = 1. Thus, it is impossible to form a Riemann sum for this function. The figure, hopefully, illustrates the situation for an oscillating function such as this one.

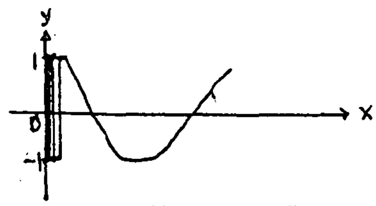

Figure 2

best we can say here is that we can construct a convergent sequence of Riemann integrals with some relation to sin1/x but the limit is not its integral.

This may sound like hair-splitting but it does raise some important points:

(a) The behavior (including non-Riemann integrability) of a function that tends towards a set, in this case the limit set or segment [-1,1]. (This is the case with oscillatory functions).

(b) Inadequacy or limitation of the present notions of function, limit and continuity. (In other words, there is a need for a calculus of set-valued functions including more general notions of integral and derivative).

(c) Inadequacy of the Riemann integral.

(d) This is another counter example on the fallacy of proving existence by approximation or convergence.

These counterexamples have far-reaching significance for mathematics especially that part of analysis that spring off real analysis. In general, contradictions play a very positive role in mathematics. The Russell paradoxes [1], the Perron Paradox [6] and the Lebesgue paradox [6], [7], to mention just a few, have led to monumental works in mathematics.

The two counterexamples together give us an idea of how a set-valued function may arise as well as set-valued derivative. The function sin1/x is set-valued at x = 0 if we define its value there as the set or vertical line segment [-1,1]. We denote this segment by sin1/0. In the interval $[-\epsilon,\epsilon]$ we can construct a set-valued function defined by

$$\{f(x)\} = \sin 1/0_x,$$

where $\sin 1/0_x$ is the vertical segment [-1,1] at the point x.

2. Probability distribution

A set-valued function would not have much use unless the set has some structure. A useful structure is a probability distribution. It will allow us to introduce both the generalized integral and generalized derivative which have applications to differential equations and the probabilistic motion of subatomic particles. We first formally develop the notion of probability distribution.

Suppose we have a set of real numbers x with a σ-algebra Σ of measurable sets (e.g., the Borel sets). We define a measure μ as a mapping from Σ into the real numbers which is countably additive on pair-wise disjoint elements of Σ. If $\mu(X) = k$ is finite we can normalize μ by dividing it by k. Thus, we have a measure p such that $p(X) = 1$. We call p a unit measure or probability measure. Imagine this unit to be distributed over the whole set X so that the integral with respect to this measure satisfies

$$(1) \qquad \int_X 1dp = 1$$

then we call this distribution a unit measure distribution or probability distribution p. If we introduce a dummy variable w that ranges over the set X we denote this probability distribution function by p(w) and the integral in (1) becomes

$$(2) \qquad \int_X p(w)dw = 1.$$

Probability distribution is a special case of measure distribution (when the measure of X is not necessarily 1) and the latter is a more general notion than density of an object with unit mass. The density may be function of the points (dummy variable) which ranges over the object.

A measure is said to have compact support K if K is compact and every measurable set in the complement of K has measure zero. In a Hausdorff space where every singleton is compact, a probability measure is said to be concentrated at a point x if x is its compact support. This means that $p(\{x\}) = 1$ and $p(X-\{x\}) = 0$.

A set-valued function is a mapping from the real numbers into the power set of a given set. In this paper the image sets are subsets of a vector space, in particular, subsets of the Euclidean plane. They are called set-valued functions. To each image set we attach a probability distribution $p_x(.)$. This notation is standard in functional analysis where the symbol (.) is reserved for the dummy variable which ranges over the particular image set.

Suppose we have a set A with unit measure p so that $p(A) = 1$. Then the value of the integral

$$\int_A wpdw, \tag{3}$$

where the dummy variable w ranges over the set A, is called the expectation point of A. If the measure of A were a unit mass this would represent the centroid of A.

If we have a set-valued function $\{f(x)\}$ so that the image of x is the set $\{f(x)\}$ and we attach a unit measure p_x on the image set $\{f(x)\}$ so that $p_x(\{f(x)\}) = 1$, then we map the set $\{f(x)\}$ by means of the integral

$$\int_{\{f(x)\}} (.)dp_x(.) \tag{4}$$

into its expectation point E(x). (It is explained in VIII(2) of [2] why $dp(w)=p(w)dw$). This defines a function called the expectation function which maps x into E(x) and we consider only cases where this mapping is measurable. In the applications considered here E(x) is also bounded so that it is Lebesgue integrable. Thus we set up the double integral

$$\int \left(\int_{\{f(x)\}} (1)dp_x(.) \right) dx, \tag{5}$$

which we call a generalized integral, where the inner integral maps the set {f(x)} at each x into its expectation point E(x) so that the outer integral becomes an ordinary indefinite Lebesgue integral

(6) $$\int E(x)dx.$$

The generalized integral reduces to the ordinary Lebesgue integral for ordinary integrable function f(x) if we define the probability distribution at each point x to be concentrated at the image point f(x) (a singleton) for this ordinary function.

Set-valued function and the generalized integral are useful in solving a differential equation of the type

(7) $$y' = \sin^n 1/x,$$

where the right side is set-valued at x=0, its set-value there being the segment [-1,1] or the segment [0,1] depending on whether the positive integer n is odd or even. The function on the right side of (7) is called an oscillation. It can be shown that the product of an oscillation, such as $\sin^n 1/x$ $\cos^m 1/x$, is an oscillation called compound oscillation. This method enables us to solve a differential equation with compound oscillation of the type

$$y' = \sin^n 1/x \cos^m 1/x$$

(The solution will be discussed in a sequel to this paper).

A set-valued function can also approximate a wave packet in quantum mechanics more accurately as it can accommodate other requirements such as symmetry and indeterminacy and upgrade path integration there with the use of the generalized integral. The present representation of a wave packet as unsymmetric superimposed wave function on another or wave function with an envelope and with continuous derivative violates the quantization principle, the uncertainty principle and symmetry. The use of set-valued function with probability distribution resolves the problem (details are given in the sequel).

For a set-valued vector function we introduce probability distributions on the component sets and use the fact that the components of the expectation point of a vector set are the expectation points of its component sets.

As we have seen in the first counterexample, the sequence of derivatives of a uniformly convergent sequence of (absolutely) continuous functions need not converge to a single point. Hence, the limit points of such sequence form a set-valued derivative, that is, the uniform limit of a differentiable (hence absolutely continuous) function may have set-valued derivative, which is not a derivative in the usual sense. We, therefore, allow set-valued derivative and the structure we put on each set-value is also a probability distribution. Then, for set-valued derivative, we define derivative, in a new sense, as the expectation point or average of its set-values with respect to its probability distribution.

To resolve the problem with Theorem I, we adopt the method of Young in [5] and define a function in the parametric case as a pair $(f(t),g(t))$, where $f(t)$ is a real-valued function, function in the ordinary sense and $g(t)$ is also a real number except in a set of measure zero, independent of $f(t)$. The variable t is a real-valued parameter. Thus, $g(t)$ is well defined a.e. in t and we call it the derivative of $f(t)$. We define uniform convergence of a sequence of functions in this new sense as follows:

Let $(f_n(t),\ g_n(t))$, $n = 1,2,\ldots$, be a sequence of functions where $g_n(t)$ is the derivative of $f_n(t)$ which is well defined except in a set of measure zero. We say that the sequence converges uniformly to the function $(f_o(t),\ \{g_o(t)\})$ if, for each t, $\lim f_n(t) = f_o(t)$ in the sup norm as $n \to \infty$ and at points t where $g_n(t)$ is well-defined, $\{g_o(t)\}$ is the set of limit points of $g_n(t)$ as $n \to \infty$. At points where $\{g_n(t)\}$ is set-valued we take as $\{g_o(t)\}$ the limit set of the sequence $\{g_n(t)\}$ as $n \to \infty$. We now take the closure under uniform convergence of the space of functions in the new sense and call that space the space of functions in the generalized sense. The elements of this new space includes differentiable (therefore absolutely continuous) ordinary functions, functions in the new sense as well as functions in the generalized sense where the derivative of each is set-valued at some points or at all points. We now define the generalized derivative of a function in this new sense at each point t as the expectation point of its set-values there with respect to some probability distribution. In the case where the derivative $g_o(t)$ is well defined, we define its generalized derivative by taking a probability distribution that is concentrated at that point so that the expectation point of that singleton $g_o(t)$ coincides with it. Thus, the generalized derivative reduces to the ordinary derivative when the latter is well defined.

It is clear that in this enlarged space of functions in the new sense every absolutely continuous function has a derivative provided we take the derivative in the generalized sense. The converse of this statement is proved in [6]. Now, we can state a rectified and more general version of Theorem I:

A function f is absolutely continuous if and only if its generalized derivative exists.

It is understood that f is really a pair (f,(g)) as defined earlier. It is also clear that the generalized derivative exists if the set-values of the derivative in the new sense has finite measure.

Sequels to this paper will develop a calculus of set-valued functions, generalized theory of functions and applications to differential equations and physics. In all cases probability distribution will play a major role.

References

[1] Benacerraf, P. and Putnam, H.: **Philosophy of Mathematics.** Cambridge University Press, London, 1985.

[2] Escultura, E. E.: **Introduction to Qualitative Control Theory.** Kalikasan, Quezon City, 1991.

[3] ________________: Dead Ends and Contradictions in Classical Mathematics. Technical Report, IMSP-PCASTRD-NRCP, 1992.

[4] Royden, H. L. **Real Analysis.** MacMillan, 2nd and 3rd ed., New York, 1963 and 1983.

[5] Young, L.C.: Generalized Curves and the Existence of an Attained Absolute Minimum in the Calculus of Variations. **Compt. Rend. Soc. Sci. Lettr. Varsovie,** CI III, 30(1937), 211-234.

[6] ________________: **Lectures on the Calculus of Variations and Optimal Control Theory.** W. Saunders, Philadelphia, 1969.

[7] ________________: **Mathematicians and Their Times.** North-Holland, Amsterdam, 1980.