STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 87
(S 348)

Undelining random variables

J. Hemelrijk

1966

MC

Underlining random variables*)

by J. Hemelrijk**)

UDC 519.211

Summary

This paper gives an exposition of the convention, introduced by D. VAN DANTZIG *in 1947, of indicating randomness by underlining the relevant symbols. This method of distinguishing between random variables (or elements) and non-random ones serves a multiple purpose. It saves symbols and may be applied to symbols which have been introduced first in a non-random quality. It clearly marks the line between measure theory and probability theory and, in statistics, between descriptive and stochastic models. This underlining notation has proved to be very useful during its nearly twenty years of use in Holland and many Dutch authors use it in their publications. The present description of the rules of this convention is meant to facilitate the reading of these papers and to deepen the understanding of the merits of this notation.*

1. Introduction

Many Dutch statisticians and probabilists have contracted the habit of denoting random variables by underlined symbols: $\underline{x}$, $\underline{y}$, etc. This notation was introduced by D. VAN DANTZIG (1947) as a variant of the other notations which distinguish between random and non-random variables. Two of these are: upper case symbols for random variables and lower case for their realizations: e.g. $P(X = x)$, and Greek letters for random variables, Latin letters for non-random ones; e.g. $P(\xi = x)$. The step to underlining, thus writing $P(\underline{x} = x)$, seems only a small one and one would not expect that any misunderstanding about the meaning of this notation would arise. Thus VAN DANTZIG never deemed it necessary to publish a full exposition of the convention he introduced. He was content to be able to save a complete alphabet in this way – he was always short of symbols as mathematicians often are – and to have a means to distinguish the random elements emphatically and easily from the non-random ones. Another evident advantage of his notations is that expressions with an already established notation, like $S^2 = \Sigma (x_i - \bar{x})^2$ and $s^2 = S^2/(n - 1)$, lead to conflicts in the other notations mentioned, but not in his system. If one wants to use upper and lower case symbols S^2 would have to be random but s^2 not, in VAN DANTZIG'S notation $\underline{S}^2$ and $\underline{s}^2$ are random but S^2 and s^2 are not. The Greek and Latin symbols lead to complications in cases like χ^2, where VAN DANTZIG

*) Report SP 87 of the Statistical Dept. of the Mathematical Centre, Amsterdam.

**) University of Amsterdam.

could simply write $P(\underline{\chi}^2 \leqslant \chi^2)$, without getting into difficulties. Also probabilities themselves, and expected values, may be random; in VAN DANTZIG's notation the universally used symbols P and E may then be maintained: $\underline{P}$ and $\underline{E}$. We return to this point later. We will also show that the underlining-convention distinguishes neatly between probability theory proper and measure theory.

The notation has now been in use in Holland for nearly twenty years and it has developed somewhat. In some cases there have been differences of opinion about how it should be used. It therefore seems useful to give a description of some completeness of its rules and use, and to point out its character. For, although it does not involve any new theory and does not change the mathematical model essentially, it still goes a little deeper than was thought originally.

Section 2 contains the simple statistical case of sampling from a finite population; the terminology is the naive statistical idiom. In section 3 the more sophisticated and more general propabilistic model is treated; the terminology there is more purely mathematical.

2. The simplest case: a finite population

Let Π be a finite population consisting of N elements:

$$e_1, e_2, \cdots, e_N. \tag{1}$$

Let further x be a variable which has a well-defined real value for each element of Π. Then, if we sort out the population according to x, we obtain a frequency table of the form

$$\begin{cases} \text{value} & x_1 < x_2 < \cdots < x_k \ (k \leqslant N) \\ \text{relative freq.} & f_1 \quad f_2 \quad \cdots \quad f_k \ (\Sigma f_i = 1) \end{cases} \tag{2}$$

and this is a complete statistical description of Π as far as the distribution of x is concerned.

Now consider the *procedure* of selecting one element from Π at random. The practical side of this consists of constructing an unbiased randomizer (a "true" lottery). This may be a mechanical or an electronic device, a table of random numbers or there may be a "natural" randomizer at work. This is not the place to go into that matter, it is enough for our purpose to point out that a *realization* of the random procedure is an element of Π, pointed out by the randomizer and that the usual mathematical model will be

$$P(e_j) = N^{-1} \quad (j = 1, \cdots, N); \ P(x_i) = f_i \quad (i = 1, \cdots, k). \tag{3}$$

This model does not contain an explicit notation for the procedure of random selection itself, but only for the possible results and the corresponding probabilities. In VAN DANTZIG's notation a new symbol, $\underline{e}$, is introduced and called a

random element on Π. This then represents the procedure itself; the notation for the *event* that the procedure selects a given element e_j is

$$\underline{e} = e_j, \tag{4}$$

and, if A is any subset of n elements of Π, the event that an element of A is selected, is written as

$$\underline{e} \in A. \tag{5}$$

The probability of these events is, of course, the same as in (3) and is denoted by

$$Pr(\underline{e} = e_j) = N^{-1} \quad (j = 1, \cdots, N); \; Pr(\underline{e} \in A) = n/N, \tag{6}$$

where "*Pr*" stands for "probability".

As a matter of fact (6) can – but need not – be taken as a definition and then it provides a very elementary way to introduce probability on a non-axiomatic basis; many introductions to statistics follow this method, which seems fairly adequate if enough is said about the character of unbiased randomizers.

Each realization of the procedure $\underline{e}$ also points out a value of the variable x and this is expressed by the introduction of the *random variable* $\underline{x}$, which can now be defined as the x-value of (attached to) $\underline{e}$. The probability of an *event* of the form

$$\underline{x} = x_i \tag{7}$$

is then, according to (6):

$$Pr(\underline{x} = x_i) = f_i \quad (i = 1, \cdots, k) \tag{8}$$

and, without ambiguity, the r can be dropped from Pr.

Extension to more complicated events or more variables is simple and we will therefore skip any intermediaries and proceed to the completely general case.

3. The general case: a probability space

The general probabilistic model for an experiment or observation with uncertain outcome consists of the well-known triple

$$(\Omega, \mathscr{A}, P), \tag{9}$$

where Ω is a set with elements ω, $\mathscr{A}$ a sigma-field of subsets A of Ω, and P a normed measure on $\mathscr{A}$. Sets A are called events and $P(A)$ their probability. A measurable function $x(\cdot)$ on Ω is called a random variable.

The build-up of the model according to the underlining convention also starts from (9), but from then onwards it is a little different in conception, notation and terminology.

The main new element in this build-up is the introduction of $\underline{\omega}$, called a

random element on Ω. This random element is used to form *events*, or more completely *events for* $\underline{\omega}$, as follows

$$A \in \mathscr{A} \text{ then } \underline{\omega} \in A \text{ is an event for } \underline{\omega}. \tag{10}$$

These events are then assigned probabilities $Pr\,(\underline{\omega} \in A)$ by means of the relation

$$Pr\,(\underline{\omega} \in A) = P(A). \tag{11}$$

This model as a whole is called a *probability space*. The transition from measure theory (where $\underline{\omega}$ has no place) to probability theory is clearly indicated by the appearance of underlinings. This also means that the model stands for a situation where some kind of randomizer, often a "natural" one, is at work. The randomizer is, in general, not unbiased (as it was in section 2): its effect is represented by means of (11).

Here, but also in the sequel, the term "event" will only be used if a probability has been assigned to this event.

Now let $\mathscr{X}$ be an arbitrary set, with elements x, let $\mathscr{B}$ be a sigma-field of subsets B of $\mathscr{X}$ and let $x\,(\cdot)$ be a measurable mapping of Ω into $\mathscr{X}$. The measurability means that

$$B \in \mathscr{B} \rightarrow \{\,\omega \mid x(\omega) \in B\} \in \mathscr{A}. \tag{12}$$

The probability space on Ω and the mapping $x\,(\cdot)$ induce a probability space on $\mathscr{X}$ in the usual way; the measure P on Ω is carried over, by means of $x\,(\cdot)$, to $\mathscr{X}$. The description of this probability space on $\mathscr{X}$ in our notation runs as follows.

A random element $\underline{x}$ on $\mathscr{X}$ is introduced, with events of the form

$$\underline{x} \in B \quad (B \in \mathscr{B}) \tag{13}$$

and the event $\underline{x} \in B$ is assigned the probability of the equivalent event on Ω

$$\underline{\omega} \in \{\omega \mid x(\omega) \in B\}. \tag{14}$$

In order to indicate clearly how the random element $\underline{x}$ arises, the notation

$$\underline{x} = x(\underline{\omega}) \tag{15}$$

is introduced and the event (13) is written as

$$x(\underline{\omega}) \in B. \tag{16}$$

The process of carrying over probability from Ω into $\mathscr{X}$ is then given by

$$Pr\,(\underline{x} \in B) = Pr\,(x(\underline{\omega}) \in B) \quad (B \in \mathscr{B}), \tag{17}$$

which indeed expresses the process clearly and concisely.

There is no need to keep on writing "*Pr*" if one does not want to. Usually the "*r*" is dropped: $P(\underline{x} \in B)$. The fact that P means probability follows from the underlining occurring between the brackets.

If $\mathscr{X}_i$ $(i = 1, \ldots, n)$, with elements x_i, sigma-fields B_i and measurable mappings $x_i(\cdot)$ from Ω into $\mathscr{X}_i$ are given, then n random elements $\underline{x}_i$ arise in the way indicated above. The n mappings considered

$$x_1(\cdot), \cdots, x_n(\cdot) \tag{18}$$

induce a new measurable mapping $x(\cdot)$ from Ω into the product set

$$\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_n. \tag{19}$$

On this product set $\underline{x} = x(\underline{\omega})$ is a random element which, according to the underlining convention, can be written as

$$(\underline{x}_1, \cdots, \underline{x}_n) \text{ or } (x_1(\underline{\omega}), \cdots, x_n(\underline{\omega})); \tag{20}$$

$\underline{x}_1, \ldots, \underline{x}_n$ are called the components of the composite random element $\underline{x}$. If, in a probabilistic model more than one underlined symbol occurs, every finite set of these symbols represents the components of a composite random element.

A random element $\underline{x}$ is called a *random variable* or, following M. G. KENDALL, a *variate* if $\mathscr{X} = R$, the real axis, and $\mathscr{B}$ is the sigma-field of Borel sets. This also holds for $\underline{\omega}$ itself, if $\Omega = R$ and $\mathscr{A}$ is the Borel sigma-field.

If in (19) $\mathscr{X}_i = R$ $(i = 1, \ldots, n)$, thus $\mathscr{X} = R_n$, the n-dimensional Euclidean space, and if the $\mathscr{B}_i$ are Borel sigma-fields, then $\underline{x}$ is called an *n-dimensional variate* or a *random vector* or *random point* on R_n and $\underline{x}_1, \ldots, \underline{x}_n$, the (random) coordinates of $\underline{x}$, have a joint probability distribution. This always holds for any finite set of variates occurring in a probability model.

If $\mathscr{X}$ is a set of real functions $x(\cdot)$ with real argument t, a measurable mapping of Ω into $\mathscr{X}$ must be indicated a little more explicitly, e.g. by

$$\omega \to x(\cdot\,; \omega), \tag{21}$$

where $x(\cdot; \omega)$ is the element of $\mathscr{X}$ corresponding to ω. The random element induced on $\mathscr{X}$ is then written as $\underline{x}(\cdot)$ or $x(\cdot; \underline{\omega})$ and it is a *random function*. The insertion of a fixed value t, giving $\underline{x}(t)$, is now, according to the underlining convention, only allowed if a random element is obtained on R. For the values $x(t)$ $(x \in \mathscr{X})$ are real numbers, and the underlining is only permissable to indicate a random element. Thus the notation $\underline{x}(t)$ presupposes that the mapping $x(t; \omega)$, for fixed t, of Ω into R is measurable. If this mapping is Borel-measurable for every $t \in T \subset R$ then $\underline{x}(t)$ is, for $t \in T$, a random variable and $\{\underline{x}(t);\ t \in T\}$ is a *stochastic process*.

In this situation it may be desirable to substitute a random variable $\underline{t}$ (on T)

into $\underline{x}(t)$, thus obtaining $\underline{x}(\underline{t})$. Again according to the convention, this can only be done if $\underline{t}$ and $\underline{x}(\cdot)$ together form a composite random element $(\underline{x}(\cdot), \underline{t})$ on the product set $\mathscr{X} \times T$.

Up to this point $\underline{x}(\cdot)$ is a random function and $\underline{t}$ and $\underline{x}(t)$ are random variables. If $x(\cdot)$ is a Borel-measurable function then $x(\underline{t})$ is also a variate. Under further conditions of regularity, which we will not go into, $\underline{x}(\underline{t})$ itself will also be a variate.

Going still further one may examine expressions like $\underline{x} \in \underline{B}$ and $\underline{P}(\underline{x} \in B)$. The former one would be called a random event: $\underline{B}$ would be a random element on $\mathscr{B}$ and $\underline{x}$ on $\mathscr{X}$ and they would, together, form a composite random element $(\underline{x}, \underline{B})$ on $\mathscr{X} \times \mathscr{B}$. Regularity conditions, guaranteeing $P(\underline{x} \in \underline{B})$ to be defined, would have to be fulfilled. In practice it might easily occur that first a subset B of $\mathscr{X}$ is selected at random and then a random element of $\mathscr{X}$ is taken. Random probabilities like $\underline{P}(\underline{x} \in B)$, or even $\underline{P}(\underline{x} \in \underline{B})$ may occur in a similar way; one meets them in the theory of conditional probabilities. We omit the details.

4. Neutralization of underlinings

There are a number of operations which neutralize underlinings, i.e. which take away the random character. One of these we have met already: P, which "changes" an event into a number. Similarly the operator E (for expected value), σ^2 or *var* (for variance), and other symbols for moments work upon variates and result in numbers (under condition of convergence). The underlining convention rules that only specifically probabilistic operators can have this neutralizing property, not those, like integrals and differential operators, which are in general use in mathematics.

Thus one has

$$E\underline{x} = \int_{\Omega} x(\omega)\, P(d\omega), \tag{22}$$

but one should not write for the right hand side

$$\int_{\Omega} \underline{x}(\omega)\, P(d\omega).$$

Following the convention the latter expression would be a random expectation because of the non-neutralized underlining.

Random expectations occur in the theory of conditional probability and the underlining notation is, in this case too, clarifying and versatile. Let $(\underline{x}, \underline{y})$ have a two-dimensional distribution, then

$$E(\underline{x} \mid \underline{y} = y) \tag{23}$$

is an ordinary conditional expectation, depending on y; we might write $\varphi(y)$ for it. Then, in general, $\varphi(\underline{y})$ would again be a random variable, but this can simply be written

$$E(\underline{x} \mid \underline{y}) \tag{24}$$

and this is in fact a random expectation, for the E-symbol only neutralizes the underlinings which occur before the vertical bar. The same holds for P: $P(\underline{x} \leqslant x \mid \underline{y})$ is a random probability and, in those cases where one wishes to indicate the condition behind the vertical bar measure-theoretically (i.e. without underlining) the random character should be indicated by underlining the E or P. So one may meet symbols like $\underline{E}(\underline{x} \mid \mathscr{B})$, where $\mathscr{B}$ is a sigmafield; this is a random variable: the bar under $\underline{x}$ is neutralized by the E, but the bar under E remains.

5. Final remarks

It has been pointed out as a disadvantage of the underlining convention, that the symbol x, if used to indicate realizations of the variate $\underline{x}$, cannot be used any more to indicate the function on Ω which generates the variate; one has to use the complete symbol $x(\cdot)$ to indicate this function. This may, however, also be considered an advantage.

Those who use the underlining convention often apply it to better understand papers of authors who do not use a notation which clearly distinguishes between random variables and others. Reading such papers they put in the underlinings themselves. This makes the reading much easier and if one does not succeed in placing the underlinings consistently one knows that either one does not understand the author of the paper or the author does not work consistently with his probabilistic model. Cf. e.g. D. VAN DANTZIG (1957).

A c k n o w l e d g e m e n t s. The author wants to thank Dr FABIUS for his persistance in criticizing the convention; this was the immediate cause for the writing of this paper. Discussions with him and with Prof. RUNNENBURG, who also read a first draft, clarified the subject considerably.

References

D. VAN DANTZIG (1947), Kadercursus Statistiek, Mathematisch Centrum, Amsterdam.
D. VAN DANTZIG (1957), Statistical Priesthood II, Statistica Neerlandica **11**, 185-200.