WHY IS THE VARIANCE THE SMALLEST SECOND MOMENT ? A GEOMETRICAL VIEW POINT.

Louis H. Y. Chen
University of Singapore

It has been said that geometrical intution provides the clearest, if not the shortest, path to solutions of problems. In this note I shall give a geometrical meaning to a very simple and well known problem in elementary probability. Before I set out to define the problem, let me add that although the geometrical approach will appear a little too sophisticated for this problem, this way of thinking has influenced to a large extent the development of certain areas in probability and statistics - for example, time series and linear statistical inference.

In almost every elementary textbook on probability or statistics, it is stated that the second moment of a random variable X about any real number c is minimum when $c = EX$, that is

$$\operatorname{var}(X) = \min_{c} E(X-c)^2 ,$$

There are two proofs which we would normally find in these books. One is by calculus : $E(X-c)^2$ is expanded into $EX^2 - 2cEX + c^2$ whose derivative with respect to c is then equated to zero to yield the solution $c = EX$. The proof is completed by showing that the second derivative is positive. The other proof is algebraic which depends on the completion of a square :

$$(1) \qquad E(X-c)^2 = EX^2 - 2cEX + c^2$$
$$= EX^2 - (EX)^2 + (EX-c)^2$$
$$= \operatorname{var}(X) + (EX-c)^2$$

Clearly the minimum value of $E(X-c)^2$ is var(X) which corresponds to $c = EX$. Sometimes the algebraic proof is presented in a slightly different version as follows :

$$
\begin{aligned}
(2) \qquad E(X-c)^2 &= E[(X-EX) + (EX-c)]^2 \\
&= E(X-EX)^2 + 2E(X-EX)(EX-c) + (EX-c)^2 \\
&= \operatorname{var}(X) + (EX-c)^2
\end{aligned}
$$

where the cross product term vanishes.

We shall now take a geometrical approach to this problem and give a geometrical interpretation of the algebraic proof. We shall see that the geometrical approach will provide a deeper insight into the problem. Let us begin with the simplest case where the random variable X takes on the values $x_1, x_2, \ldots, x_n$ with equal probabilities $1/n$. Then $E(X-c)^2 = n^{-1}\sum_{i=1}^{n}(x_i-c)^2$. With a little imagination, we see that the expression $\sum_{i=1}^{n}(x_i-c)^2$ is nothing but the square of the distance between the the points $P \equiv (x_1, \ldots, x_n)$ and $Q \equiv (c, \ldots, c)$ in n-dimensional Euclidean space; that is the set of all n-tuples of real numbers with the operations of addition and scalar multiplication defined as follows :

$$(y_1, \ldots, y_n) + (z_1, \ldots, z_n) = (y_1+z_1, \ldots, y_n z_n),$$

$$a(y_1, \ldots, y_n) = (ay_1, \ldots, ay_n) \text{ for every real number } a,$$

and the distance between two points $(y_1, \ldots, y_n)$ and $(z_1, \ldots, z_n)$ defined to be $[(y_1-z_1)^2 + \ldots + (y_n-z_n)^2]^{\frac{1}{2}}$. If we vary the values of c, we see that the point Q moves along the line L joining the origin O and the point $T \equiv (1, \ldots, 1)$; that is the set of $\{\xi : \xi = c(1, \ldots, 1),\ c \text{ is a real number}\}$. It is clear that if the geometry of 3-dimensional space can be extended to that of n-dimensional space, then $E(X-c)^2$ is minimum when Q is the "foot of the perpendicular" from the point P to the line L (see Figure 1). (The reader is recommended to picture the problem in 3-dimensional

Euclidean space.)

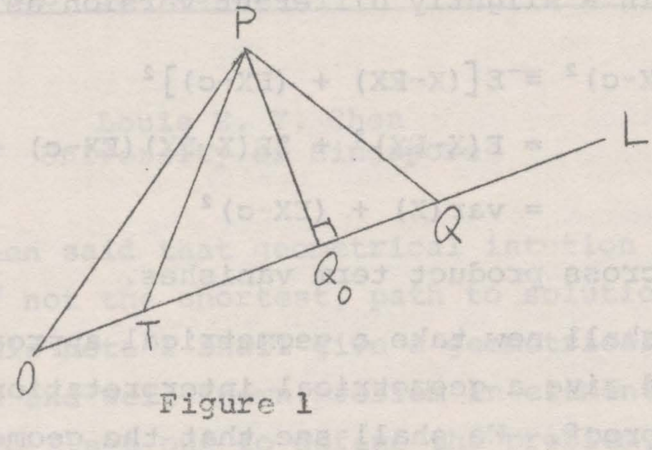

Figure 1

So we set out to define the notion of an angle between two vectors in n-dimensional Euclidean space. Where ambiguity does not arise we shall for simplicity use the same notation for a point and the vector from the origin to the point. For example, let R and S be two points whose coordinates are $y_1, \ldots, y_n$ and $z_1, \ldots, z_n$ respectively, and let $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$. Both the point R and the vector $\overline{OR}$ will be denoted by $(y_1, \ldots, y_n)$ or more simply by y. Likewise the vector from the origin, which is parallel and equal in magnitude to the localized vector $\overline{SR}$, will be denoted by $(y_1 - z_1, \ldots, y_n - z_n)$ or more simply by $y - z$. We first define the inner product (y,z) of two vectors $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_n)$ as the sum $y_1 z_1 + \ldots + y_n z_n$. Note that this is an n-dimensional analogue of the dot product in 3-dimensional vector analysis (see for example Cunningham [1], pp. 29-38). It can easily be verified that the inner product so defined satisfies the following properties :

(a) $(y,z) = (z,y)$;

(b) $(y+z,w) = (y,w) + (z,w)$;

(c) $(ay,z) = a(y,z)$ for every real number a ;

(d) $(y,y) \geq 0$;

(e) $(y,y) = 0$ if and only if $y = (0,\ldots,0)$.

The norm or length of the vector y, denoted by $\|y\|$, is defined to be $(y,y)^{\frac{1}{2}}$. Thus the distance between the two points $(y_1,\ldots,y_n)$ and $(z_1,\ldots,z_n)$ is the norm of the vector $y-z$ where $y = (y_1,\ldots,y_n)$ and $z = (z_1,\ldots,z_n)$.

We now use the properties (a) and (d) to prove the Cauchy-Schwarz inequality :

$$|(y,z)| \leqslant \|y\|.\|z\| \quad .$$

Let $A = \|y\|^2$, $B = (y,z)$ and $C = \|z\|^2$. For every real number r we then have

$$
\begin{aligned}
(3) \quad \|y-rz\|^2 &= (y-rz, y-rz) \\
&= (y,y) - 2r(y,z) + r^2(z,z) \\
&= \|y\|^2 - 2r(y,z) + r^2\|z\|^2 \\
&= A - 2Br + Cr^2 \ .
\end{aligned}
$$

Since the expression on the left of (3) is nonnegative, we must have $A - 2Br + Cr^2 \geqslant 0$ for every real number r. Now $C \geqslant 0$. If $C = 0$, then B must be zero and the inequality $B^2 \leqslant AC$ is trivially true; otherwise by choosing $r = (A+1)/B$ we get the absurd statement that $-A-2 \geqslant 0$. On the other hand if $C > 0$, then by choosing $r = B/C$ we get $B^2 \leqslant AC$. This proves the Cauchy-Schwarz inequality.

Note that there is a great similarity between (1) and (3). However while the Cauchy-Schwarz inequality may be proved by completion of squares, we have seen that the proof need not be so. On the other hand, it does not seem possible to avoid completion of squares in minimising the quadratic expression in c in (1) by algebraic method.

The Cauchy-Schwarz inequality enables us to define, as in 3-dimensional case, the angle θ $(0 \leqslant \theta \leqslant \pi)$ between the two nonzero vectors y and z by

$$\cos\theta = \frac{(y,z)}{\|y\|\cdot\|z\|} \quad .$$

Hence two vectors y and z are said to be orthogonal or perpendicular to each other if and only if $(y,z) = 0$. Next we see that Pythagoras' Theorem holds. Indeed, let y and z be two orthogonal vectors. By (a) and (b),

$$(4)\qquad \|y+z\|^2 = (y+z,y+z) = (y,y) + 2(y,z) + (z,z)$$
$$= \|y\|^2 + \|z\|^2 ,$$

(see Figure 2).

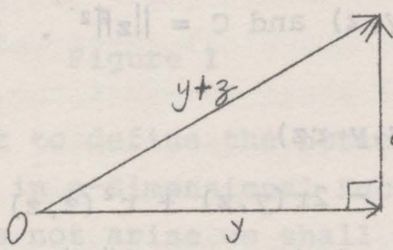

Figure 2

To find the value of c for which $E(X-c)^2 = n^{-1}\|x-\xi\|^2$ is minimum where $x = (x_1,\dots,x_n)$ and $\xi = (c,\dots,c) = c(1,\dots,1)$ amounts now to finding the value of c for which $\overline{QP}$ is orthogonal to $\overline{OT}$ (See Figure 1); that is

$$(5)\qquad (x-\xi,e) = 0$$

where $e = (1,\dots,1)$. Now by (b) and (c),

$$(x-\xi,e) = (x,e) - (\xi,e) = (x,e) - c(e,e)$$
$$= \sum_{i=1}^{n} x_i - nc = n(EX-c)$$

and so (5) yields $c = EX$. Note that (5) is equivalent to the equation obtained by equating the derivative of $EX^2 - 2cEX + c^2$ to zero, and we obtained it without differentiation. Let $\xi_o = (EX,\dots,EX)$ and Q_o be the point with coordinates $EX,\dots,EX$. The point Q_o is defined to be the foot of the perpendicular from P to L and the vector ξ_o $(\overline{OQ_o})$ is defined to be the orthogonal projection of the vector $x(\overline{OP})$ onto the line L. To see that $\|x-\xi_o\|^2$ must be the unique minimum value we use Pythagoras' theorem. First we note that by (c),

x and ay are orthogonal if x and y are, where a is any real number. Therefore the vector $x-\xi$ ($\overline{Q_oP}$) is orthogonal to the vector $\xi-\xi_o$ ($\overline{Q_oQ}$) since $\xi-\xi_o = (c-EX)e$. Hence

$$(6) \qquad \|x-\xi\|^2 = \|(x-\xi_o) + (\xi_o-\xi)\|^2$$

$$= \|x-\xi_o\|^2 + \|\xi_o-\xi\|^2 .$$

If $\xi \neq \xi_o$, then by (e), $\|\xi_o-\xi\|^2 > 0$. This shows that $\|x-\xi\|^2$ (and hence $E(X-c)^2$ achieves minimum value at the unique point $\xi = \xi_o$ (namely for $c = EX$).

Now we can give a geometrical interpretation of the algebraic proof which is nothing but an application of Pythagoras' Theorem. This fact is particularly apparent when we note the similarity between (4) and (6).

Suppose X takes on the values $x_1,\ldots,x_n$ with probabilities $p_1,\ldots,p_n$. Then $E(X-c)^2 = \sum_{i=1}^{n} p_i(x_i-c)^2$ which is not the square of the distance between the points P and Q in the ordinary sense. However, if we examine our geometrical proof carefully we see that all that is needed is the definition of an inner product which must satisfy the properties (a) to (e) and which is such that $E(X-c)^2$ is proportional to the norm square of $x-\xi$, where $x = (x_1,\ldots,x_n)$, $\xi = (c,\ldots,c)$, and the norm or length of a vector is defined to be the square root of its inner product with itself. The rest of the proof is just a consequence of these properties. So we define a new inner product by

$$(y,z) = \sum_{i=1}^{n} p_i y_i z_i$$

where $y = (y_1,\ldots,y_n)$ and $z = (z_1,\ldots,z_n)$. Then $E(X-c)^2 = \sum_{i=1}^{n} p_i(x_i-c)^2 = \|x-\xi\|^2$. Note that the norm or length $\|y\|$ of a vector y no longer has the ordinary meaning. I leave it to the reader to verify that the inner product just defined satisfies the properties (a) to (e). This being so, the value of c for which $E(X-c)^2$ is minimum can be determined in exactly the same way as before.

What if X takes on an infinite (but countable) number of values $x_1, x_2, \ldots$ with respective probabilities $p_1, p_2, \ldots$? In this case we have to consider an appropriate "infinite-dimensional" space which is the set of all infinite sequences of real numbers $(y_1, y_2, \ldots)$ such that $\sum_{i=1}^{\infty} y_i^2$ converges, and with the operations of addition and scalar multiplication defined by

$$(y_1, y_2, \ldots) + (z_1, z_2, \ldots) = (y_1+z_1, y_2+z_2, \ldots)$$

and

$$a(y_1, y_2, \ldots) = (ay_1, ay_2, \ldots) \text{ for every real number } a .$$

Each vector is now identified with such an infinite sequence. Then we define the inner product by

$$(y, z) = \sum_{i=1}^{\infty} p_i y_i z_i$$

and the norm $\|y\|$ of the vector y by $\|y\| = (y, y)^{1/2}$

The inner product can be shown to satisfy the properties (a) and (e). Again $E(X-c)^2 = \|x-\xi\|^2$ where $x = (x_1, x_2, \ldots)$ and $\xi = (c, c, \ldots)$. The rest of the proof is clear.

To solve the problem in its utmost generality, we need to consider the space of all real-valued functions f (strictly speaking, equivalence classes of functions) defined on the real line such that the Lebesque-Stieltjes integral $\int f^2(u)dF(u)$ is finite where F is the distribution function of the random variable X. The operations of addition and scalar multiplication, and the inner product and the norm are defined in an analogous way as in the special cases :

$$(f+g)(u) = f(u)+g(u),$$

$$(af)(u) = af(u) \text{ for every real number } a,$$

$$(f,g) = \int f(u)g(u)dF(u)$$

and

$$\|f\| = \left(\int f^2(u)dF(u)\right)^{\frac{1}{2}} .$$

Again the inner product satisfies the properties (a) to (e) and $E(X-c)^2 = \int (u-c)^2 dF(u) = \|x-\xi\|^2$ where x is now the identity function and ξ the constant function with value c. As before $c = \int u dF(u) = EX$ corresponds to the minimum value of $E(X-c)^2$.

The reader can see that, in the foregoing discussion, the general difficulty of the material increases with the level of generalization. However the geometrical concepts and properties remain the same throughout. So does the method of proof. This is the advantage of abstraction. The above geometrical approach in an example of abstraction, and in taking this approach we have touched on the basic concepts and properties of a very important object in mathematics — the Hilbert space (see Rudin [2], pp. 79-99).

References

[1] John Cunningham : Vectors, Heinemann, London, 1969.

[2] Walter Rudin : Real and complex analysis, Second edition, McGraw-Hill, New York, 1974.