

第 5 章 累積分布関数

浅川 伸一

2006 年 05 月 11 日

補足

4.5 分散の意味

分散の幾何学的意味は、各データの平均からの距離の 2 乗の平均であった。

メーリングリストより... とありますが、二行目の意味をグッとにらんで、ジューッと考えると分散とは、期待値（平均値）からの自乗、つまり一辺が平均からの偏差の正方形の面積を足し合わせていることになります。幾何学的に説明すると、分散とは一辺が平均偏差である正方形の面積の平均なのです。従って、分散が大きいとは、正方形の面積が大きいことを意味し、平均からの距離（平均偏差）が大きい値、極端に平均から外れたデータがあると大きな値を取りやすいという性質があることが分かります。分散とは散らばり具合を正方形の大きさで表現したものなのです。正方形、面積、などごく普通の算数しか使っていません。

n 個からなる任意のデータに対して c を任意の定数としたとき

$$\frac{1}{n} \sum (x - c)^2 \quad (1)$$

を最小にする c を求めることを考える。この式は x の 2 次関数の和であるから実数の範囲では必ず正またはゼロになる。この最小値を求めるために c で微分して結果をゼロとおいてみる。

$$\frac{d}{dc} \frac{1}{n} \sum (x - c)^2 = \frac{1}{n} \cdot -2 \sum (x - c) = 0 \quad (2)$$

であるから

$$\sum (x - c) = 0 \quad (3)$$

$$\sum x - n c = 0 \quad (4)$$

$$c = \frac{1}{n} \sum x \quad (5)$$

このことから、分散とは任意の値 c を選んで、その 2 乗和が最小となる値である。

4.6 少数の法則

Amos Tversky and Daniel Kahneman,(1971), Belief in the Law of Small Numbers, Psychological Bulletin, Vol. 76, No. 2. 105-110.

原文は

<http://www.compositesconsultants.com/Other/SmallNumbers.htm>

日本語訳が

<http://d.hatena.ne.jp/lionfan/20060306>

にある。

2002 年 受賞。

Tversky と Kahneman は人間の確率判断が如何に不合理かということを繰り返し示した。「リンダ問題」(割愛)

なお、心理学者以外ではポワソン分布のことを少数の法則ということがある。

4.7 整数以外の次元

心理学においては David Marr の $2\frac{1}{2}$ 次元スケッチが有名。Marr は視覚情報処理における計算論的目標が、網膜に写る 2 次元情報から如何に 3 次元情報を再構成することであると主張した。

その他の次元については、「フラクタル次元」を検索して調べる。

4.8 \sum 記号の復習

$$\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5.$$

$$\sum_{i=1}^5 x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2, \neq \left(\sum x_i\right)^2$$

$$\sum_{i=1}^n ax_i^2 = a \cdot x_1^2 + a \cdot x_2^2 + \cdots + a \cdot x_{n-1}^2 + a_n^2 = a \sum_{i=1}^n x_i^2$$

$$\begin{aligned} \sum_{i=1}^n (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_n + y_n) \\ &= \sum_{i=1}^n x_i + \sum_{i=1}^n y_i. \end{aligned}$$

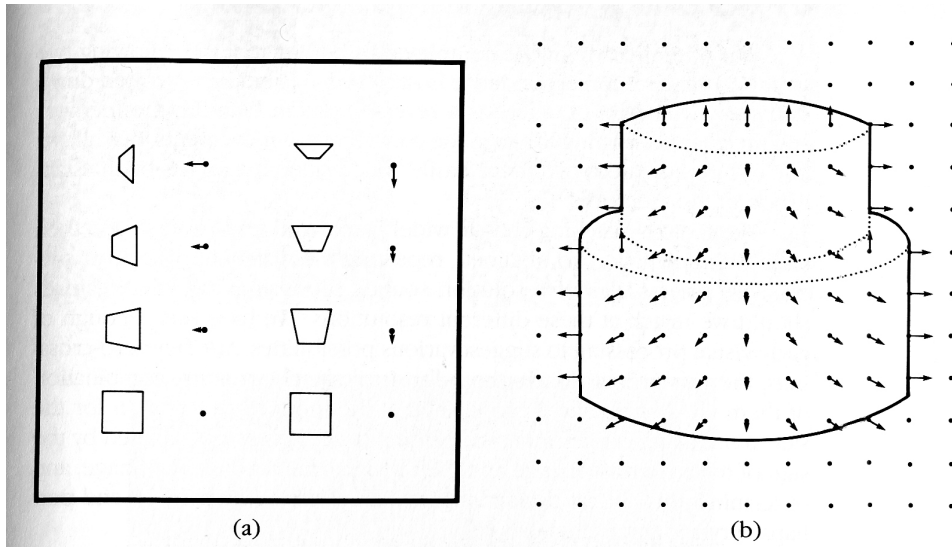


Figure 3-12. Illustration of the 2½-dimensional sketch. In (a), the perspective views of small squares placed at various orientations to the viewer are shown. The dots with arrows symbolically represent the orientations of such surfaces. In (b), this symbolic representation is used to show the surface orientations of two cylindrical surfaces in front of a background orthogonal to the viewer. The full 2½-dimensional sketch would include rough distances to the surfaces as well as their orientations; contours where surface orientations change sharply, which are shown dotted; and contours where depth is discontinuous (subjective contours), which are shown with full lines. See Chapter 4 for more details. (D. Marr and H. K. Nishihara, 1978.)

$$\begin{aligned}
 \sum_{i=1}^n (x_i + y_i)^2 &= \sum_{i=1}^n (x_i^2 + 2x_i y_i + y_i^2) \\
 &= \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \\
 &\neq \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2.
 \end{aligned}$$

5 累積密度関数

確率変数 $x_i (i = 1, 2, 3, \dots)$ の確率密度関数 $P(x_i)$ を

$$f(x_i) = P(x = x_i) \quad (6)$$

とする。このとき累積分布関数 cumulative distribution function, CDF は

$$F(x_i) = P(x \leq x_i) = f(x_1) + f(x_2) + \dots + f(x_i) \quad (7)$$

と表される。これは、確率変数を $x = 0$ から $x = i$ まで足し合わせて行ったときの累積確率である。

5.1 Excel の BINOMDIST() の第 4 引数

true のとき累積分布関数。false のとき確率密度関数になる。

