VARIANCE, ENTROPY AND UNCERTAINTY MEASURE

Yuan Wei
Dept. of Statistics, People's Univ. of China

The unity of opposites of the mean and variance is one of manifestation of the law of the unity of opposites about certainty and uncertainty of materialist dialectics. Under the mass observations, the mean reflects a quantitative characteristics of certainty of a random variable after cancelling out the quantitative differences of uncertainty. The variance is a measure of differences of a random variable. Since the difference is the quantitative manifestation of inner uncertainty causality, the variance can be considered as a measure of uncertainty of a random variable.

As everyone knows that the mean measures the central tendency and the variance measures the dispersion of probability distribution. The central tendency and the dispersion are the unity of opposites. We do not know the degree of the central tendency without the dispersion and there is no the dispersion without the central tendency. It is the mean and variance that, as the two basic statistical tools of the unity of opposites, can control uncertainty and seek certainty. As a basic conception of information theory, the entropy is defined by professor Shannon as a measure of uncertainty. So, there should be certain scientific relationship between the variance and the entropy in measuring uncertainty.

1. The variance satisfies conditionally the propositions and properties of the entropy

(1). Both $H(p(A_i))$ and $Var(a)$ are continuous functions of $p(A_i)$.

(2). Let us consider a probabilistic experiment having n possible results (or outcomes) $a_1 a_2 \dots a_n$ with the same probabilities $p_1 = p_2 = \dots p_n = \frac{1}{n}$. Then $H_n(\frac{1}{n}\, \frac{1}{n} \dots \frac{1}{n})$ is the monotone function of n.

The variance of the continuous experiment with the same probabilities (i.e. the uniform distribution) is $\frac{(b-a)^2}{12}$, which is the monotone function of interval (b-a). For the discrete experiment, if we assume that the values of a random variable form an arithmetic sequence, Var(a) is also the monotone function of n. For example, suppose the values x_i of a random variable X are o,1, 2,3,..., experiments a_n of n=1,2,3,4,... are:

$$a_1 : \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad a : \begin{pmatrix} 0, & 1 \\ \frac{1}{2}, & \frac{1}{2} \end{pmatrix} \quad a : \begin{pmatrix} 0, & 1, & 2 \\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3} \end{pmatrix} \dots$$

Then

$$Var(a_n) = \frac{\sum_{k=2 \text{ or } 3}^{n} (k-1)^2}{2n}, \quad \begin{cases} k = 2,4,6,\dots & (\text{even}) \\ \text{or } k = 3,5,7,\dots & (\text{odd}) \end{cases}$$

So, $Var(a_n)$ is also the monotone function of n in discrete experiment.

(3). We have

$$H(p_1 \dots p_n) = H(p_1 + \dots + p_r, p_{r+1} + \dots + p_n) + \sum_{i=1}^{r} p_i \cdot H\left(\frac{p_1}{\sum_{i=1}^{r} p_i}, \dots \frac{p_r}{\sum_{i=1}^{r} p_i}\right) + $$

$$+ \sum_{i=r+1}^{n} p_i \cdot H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{n} p_i}, \dots \frac{p_n}{\sum_{i=r+1}^{n} p_i}\right)$$

If We divide the values of a random variable into two parts, represented as $x_{11}, x_{12}, x_{13}, \dots x_{1n_1}$ and $x_{21}, x_{22}, x_{23}, \dots x_{2n_2}$ the variance before dividing is the weighted sum of two variances divided.

$$Var(x) = (\bar{x}_1 - \bar{x})^2 \cdot \sum_{j=1}^{n_1} p(x_{1j}) + (\bar{x}_2 - \bar{x})^2 \cdot \sum_{j=1}^{n_2} p(x_{2j}) +$$

$$+ \sum_{k=1}^{n_1} p(x_{1k}) \cdot \left[\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 \frac{p(x_{1j})}{\sum_{k=1}^{n_1} p(x_{1k})} \right] + \sum_{k=1}^{n_2} p(x_{2k}) \left[\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 \frac{p(x_{2j})}{\sum_{k=1}^{n_2} p(x_{2k})} \right]$$

Proof:

$$\begin{aligned} Var(x) &= \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot p(x_i) \\ &= \sum_{i=1}^{2} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \cdot p(x_{ij}) + \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 \cdot p(x_{ij}) \\ &= \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 \cdot p(x_{1j}) + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 \cdot p(x_{2j}) + \\ &\quad + \sum_{i=1}^{2} (\bar{x}_i - \bar{x})^2 \cdot \sum_{j=1}^{n_i} p(x_{ij}) \\ &= (\bar{x}_1 - \bar{x})^2 \cdot \sum_{j=1}^{n_1} p(x_{1j}) + (\bar{x}_2 - \bar{x})^2 \cdot \sum_{j=1}^{n_2} p(x_{2j}) + \\ &\quad + \sum_{k=1}^{n_1} p(x_{1k}) \left[\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 \frac{p(x_{1j})}{\sum_{k=1}^{n_1} p(x_{1k})} \right] + \\ &\quad + \sum_{k=1}^{n_2} p(x_{2k}) \cdot \left[\sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 \frac{p(x_{2j})}{\sum_{k=1}^{n_2} p(x_{2k})} \right] \end{aligned}$$

where $\bar{x}_1 = \sum_{j=1}^{n_1} x_{1j} \cdot \frac{p(x_{1j})}{\sum_{j=1}^{n_1} p(x_{1j})}$, $\bar{x}_2 = \sum_{j=1}^{n_2} x_{2j} \cdot \frac{p(x_{2j})}{\sum_{j=1}^{n_2} p(x_{2j})}$

they are two means of the experiment 1 and the experiment 2.

The properties of the variance are also similar to that of the entropy.

(1). If and only if one of $p(a_i)$ is 1, then $H = 0$. It is all the same to the variance.

(2). If the experiment α and experiment β are independent, we get

$$H(\alpha\beta) = H(\alpha) + H(\beta)$$

The variance has the similar property

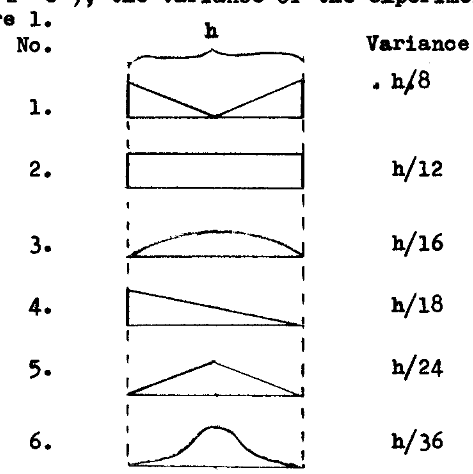
$$Var(\alpha + \beta) = Var(\alpha) + Var(\beta)$$

(3). We have

$$H(p_1, p_2, \dots p_n) \leqslant H\left(\frac{1}{n}, \frac{1}{n}, \dots \frac{1}{n}\right)$$

The variance corresponds conditionally to this property. If the distribution has only one central tendency or no central tendency(in Fig. 1, No. 2--6), the variance of the experiment

Figure 1.



with the same probabilities(i.e. the uniform distribution) is maximum. If the distribution has two or more central tendencies(in Fig. 1, No. 1), the variance does not correspond to this property because the entropy does not depend on the values of a random variable and only depends on the probabilities.

2. Conditional variance and information

For the experiments which are not independent, the conditional variance and the conditional entropy are also similar.

(1). $H(\alpha\beta) = H(\alpha) + H_{\alpha}(\beta)$

The variance has

$$Var[\alpha + (\beta \mid \alpha = x)] = Var(\alpha) + Var(\beta)(1-R^2)$$

where R is the linear coefficient of correlation.

(2). The amount of information contained in α about β is

$$I(\alpha, \beta) = H(\beta) - H_{\alpha}(\beta)$$

In the variance, the amount of information contained in α about β is

$$I(\alpha, \beta) = Var(\beta) - Var(\beta) \cdot (1-R^2)$$

$$= R^2 \cdot Var(\beta)$$

It tells us that through the linear prediction of β given α, the uncertainty of the experiment β can be reduced.

3. The entropies and variance of the important distributions of random variable

In the Table 1, we list the entropies and variances of reveral important discrete and continuous distributions of random variable.

We find that the entropies of these distributions are all the increasing functions of the variances. Thus, we have proved that the variance is equivalent to the entropy in the important random variable distributions in uncertainty measure.

Table 1.

Distribution	Entropy	Variance
Bernoulli	$-p.\log p - q.\log q$	$p.q$
binomial	$-\sum_{k=0}^{n} C_n^k p^k q^{n-k} \log C_n^k - np.\log p - nq.\log q$	npq
geometrical	$-\log p - \frac{q}{p}\log q$	$\frac{q}{p^2}$
Poisson $\lambda=1$	$\log e + \frac{1}{e}\sum_{k=0}^{\infty}\frac{\log k!}{k!}$	λ
$\lambda \gg 1$	$\log\sqrt{2\pi e\lambda}$	
$\lambda \ll 1$	$\lambda e^{-\lambda}.(1-\log\lambda+\lambda)$	
uniform	$\log(b-a)$	$\frac{(b-a)^2}{12}$
normal	$\log\sqrt{2e\pi}\cdot\sigma$	σ^2
exponential	$\log ea$	a^2

4. Comments on the variance and entropy

We have discussed the relationship and difference between the variance and entropy in uncertainty measure. During the practice of the study and application of the information theory, we should select a suitable statistic between the variance and entropy according to the different data available, which are listed below

Table 2.

Practical data available	Statistic
1.values(few or difficult to set up the frequency distribution)	variance
2.values and probabilities	variance or entropy
3.probabilities(easy to apply the dummy variable)	entropy or variance
4.probabilities(difficult to apply the dummy variable)	entropy

5. Acknowledgement

The author is greatly indebted to Prof. Dai Shiguang for helpful suggestions and comments on the manuscript. Remaining errors are my own.

REFERENCES

Ash, R. (1965). Information Theory. John Wiley & Sons

Cederlund, J. (1979). Information Content of Poisson Images. Uppsala

Suiasu, S. (1977). Information Theory with Applications. McGraw-Hill. Inc.

Young, J.F. (1971). Information Theory. Butterworth & Co. Ltd.

戴世光 (Dai Shiguang) (1985) 马克思主义哲学是统计学应用的理论基础．中国人民大学出版社 (Dialectical Materialsm, The Guiding Ideology of Applied Statistics)

复旦大学 (Fudan Univ.) (1979) 概率论基础．高等教育出版社．(The Foundations of Probability)