

# 关于信息度量问题的思考<sup>\*</sup>

王乃信, 张慧凤

(西北农林科技大学 生命科学学院, 陕西 杨凌 712100)

〔摘 要〕 为克服 Shannon 熵的局限性, 修改了 Shannon 熵的公理, 用公理化的方法, 推导出具有有限分布列的离散型随机变量的信息量系。并且将它推广到任意具有可数无限分布列的离散型随机变量和随机向量。证明了信息量系的最大信息量定理, 可加性定理, 广义可加性定理、互信息定理、独立性定理等性质。对信息量系中参数值的选取提出建议。

〔关键词〕 Shannon 熵; 信息度量; 信息量系

〔中图分类号〕 O 236

〔文献标识码〕 A

〔文章编号〕 1671-9387(2004)01-0108-05

1948 年, Shannon 提出熵的概念, 并以之作为信息的度量, 宣告了信息论作为一门科学学科的诞生<sup>[1, 2]</sup>。Shannon 熵在信息论的发展和应用中起着关键的作用。但是, 众所周知, Shannon 熵还存在许多局限性<sup>[3, 4]</sup>, 尤其是 Shannon 熵的无界性, 使许多常见的具有可数无限分布列的离散型随机变量不具有 Shannon 熵, 从而使 Shannon 熵的应用受到限制。这引发了作者的思考: 能否对 Shannon 熵进行改进, 能否引入新的更加合理的信息度量标准以克服这些局限性, 须知, 在信息论的公理化体系建立以后, Shannon 熵就是公理化方法求解时的惟一存在的解<sup>[5]</sup>, 因而在该公理化体系下, Shannon 熵是不可能改进的。因此, 如果要真正对 Shannon 熵进行改进, 就必须修改其公理化体系, 建立起新的公理化体系, 并且也用公理化的方法进行求解, 只有这样, 才可望使信息的度量得到实质上的改进。

使 Shannon 熵成为惟一存在解的一个公理化体系包括连续性公理, 单调性公理, 可加性公理。简述如下<sup>[3]</sup>:

定义 对于离散型随机变量的有限分布列  $(p_1, p_2, \dots, p_n)$ , 该分布列的 Shannon 熵记为  $H(p_1, p_2, \dots, p_n)$ 。特别地, 当  $n$  为正整数时, 记  $H_n = H(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ 。Shannon 熵满足如下公理:

I 连续性公理:  $H(p_1, p_2, \dots, p_n)$  是每一个变量的连续函数;

II 单调性公理:  $H_n$  是  $n$  的单调递增函数;

III 可加性公理: 设  $n = \sum_{i=1}^k m_i$ , 其中  $m_i$  为正整数, 则有

$$H_n = H(\frac{m_1}{n}, \frac{m_2}{n}, \dots, \frac{m_k}{n}) + \sum_{i=1}^k \frac{m_i}{n} H_{m_i}$$

已经严格证明, 满足上述公理的 Shannon 熵的表达式必为

$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log p_i,$$

并且, 这是满足上述公理的惟一存在解。

## 1 关于具有有限分布列的一维离散型随机变量的信息量系

在 Shannon 熵的公理中, 连续性公理是自然的, 应予保留; 单调性公理不具有独立性, 可以删除, 并且可以考虑代之以有界性公理; 惟一需要修改的是可加性公理。

定义 对于离散型随机变量的有限分布列  $(p_1, p_2, \dots, p_n)$ , 该分布列的  $\lambda$  级信息量记为  $H^{(\lambda)}(p_1, p_2, \dots, p_n)$ , 特别地, 当  $n$  为正整数时, 记  $H_n^{(\lambda)} = H^{(\lambda)}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ 。 $\lambda$  级信息量满足如下公理:

I 连续性公理:  $H^{(\lambda)}(p_1, p_2, \dots, p_n)$  是每一个变量的连续函数;

II 有界性公理:  $H^{(\lambda)}(p_1, p_2, \dots, p_n)$  是有界函数;

III 可加性公理: 设  $n = \sum_{i=1}^k m_i$ , 其中  $m_i$  为正整

\* 〔收稿日期〕 2002-12-30

〔作者简介〕 王乃信(1942-), 男, 陕西户县人, 教授, 博士生导师, 主要从事数学和计算机应用研究。

数, 则有

$$H^{(\lambda)}_n = H^{(\lambda)}\left(\frac{m_1}{n}, \frac{m_2}{n}, \dots, \frac{m_k}{n}\right) + \sum_{i=1}^k \left(\frac{m_i}{n}\right)^{1+\lambda} H^{(\lambda)}_{m_i}$$

现在推导  $\lambda$  级信息量的表达式:

显然, 按 Shannon 熵的要求, 应有  $\lambda = 0$ 。为了从实质上对 Shannon 熵进行改进, 必须冲破  $\lambda = 0$  的限制。

任取正整数  $m, k$ , 令  $n = mk$ , 由可加性公理知

$$H^{(\lambda)}_n = H^{(\lambda)}_k + \frac{1}{k^\lambda} H^{(\lambda)}_m, H^{(\lambda)}_n = H^{(\lambda)}_m + \frac{1}{m^\lambda} H^{(\lambda)}_k。$$

如果  $\lambda = 0$ , 上式为  $H^{(0)}_{mk} = H^{(0)}_m + H^{(0)}_k$ , 由此知,  $H^{(0)}_n$  或者恒为零而无意义, 或者具有无界性, 故应否定  $\lambda = 0$ 。

以下设  $\lambda \neq 0$ 。

当  $m = k = 1$  时, 上式即  $H^{(\lambda)}_1 = H^{(\lambda)}_1 + H^{(\lambda)}_1$ , 由此知  $H^{(\lambda)}_1 = 0$ 。对于任意的  $m > 1, k > 1$ , 恒有

$$\begin{aligned} (1 - \frac{1}{m^\lambda})H^{(\lambda)}_k - (1 - \frac{1}{k^\lambda})H^{(\lambda)}_m &= 0, \\ \frac{H^{(\lambda)}_m}{1 - \frac{1}{m^\lambda}} &= \frac{H^{(\lambda)}_k}{1 - \frac{1}{k^\lambda}} = C。 \end{aligned}$$

由此知, 必有  $H^{(\lambda)}_n = C(1 - \frac{1}{n^\lambda})$ 。由有界性公理知, 应有  $\lambda > 0$ 。不失一般性, 可设  $C = 1$ , 则有  $H^{(\lambda)}_n = 1 - \frac{1}{n^\lambda}$ 。

设  $n = \sum_{i=1}^k m_i$ , 由可加性公理知

$$\begin{aligned} H^{(\lambda)}\left(\frac{m_1}{n}, \frac{m_2}{n}, \dots, \frac{m_k}{n}\right) &= \\ H^{(\lambda)}_n - \sum_{i=1}^k \left(\frac{m_i}{n}\right)^{1+\lambda} H^{(\lambda)}_{m_i} &= \\ 1 - \frac{1}{n^\lambda} - \sum_{i=1}^k \left(\frac{m_i}{n}\right)^{1+\lambda} \left(1 - \frac{1}{m_i^\lambda}\right) &= \\ 1 - \frac{1}{n^\lambda} - \sum_{i=1}^k \left(\frac{m_i}{n}\right)^{1+\lambda} + \sum_{i=1}^k \frac{m_i}{n^{1+\lambda}} &= \\ 1 - \sum_{i=1}^k \left(\frac{m_i}{n}\right)^{1+\lambda}。 \end{aligned}$$

这就是说: 当分布列  $(p_1, p_2, \dots, p_n)$  中的概率  $p_i$  为有理数时, 均有

$$H^{(\lambda)}(p_1, p_2, \dots, p_n) = 1 - \sum_{i=1}^n p_i^{1+\lambda}。$$

根据有理数的稠密性, 由连续性公理知, 对任意的分布列  $(p_1, p_2, \dots, p_n)$ , 恒有

$$H^{(\lambda)}(p_1, p_2, \dots, p_n) = 1 - \sum_{i=1}^n p_i^{1+\lambda}。$$

反之, 也容易验证, 当  $\lambda > 0$  时, 这样获得的  $\lambda$  级信息量  $H^{(\lambda)}(p_1, p_2, \dots, p_n)$  满足前述的连续性公

理, 有界性公理, 可加性公理, 并且具有规范性:  $0 \leq H^{(\lambda)}(p_1, p_2, \dots, p_n) \leq 1$ 。

这样, 就构造了以  $\lambda$  为参数的信息量系, 它们不仅保留了 Shannon 熵的许多重要性质, 而且还克服了 Shannon 熵的一些不合理性和局限性。

为了尽可能多地保留 Shannon 熵的性质, 还可以对  $\lambda$  级信息量作如下变换:

$$S^{(\lambda)}(p_1, p_2, \dots, p_n) = (1 - H^{(\lambda)}(p_1, p_2, \dots, p_n))^{-\frac{1}{\lambda}}。$$

## 2 关于具有可数无限分布列的一维离散型随机变量的信息量系

上述信息量系, 很容易推广到具有可数无限分布列的一维离散型随机变量的情形。

定义 设一维离散型随机变量的可数无限分布列为  $(p_1, p_2, \dots, p_n, \dots)$ , 则该分布列的  $\lambda$  级信息量为

$$H^{(\lambda)}(p_1, p_2, \dots, p_n, \dots) = 1 - \sum_{i=1}^{\infty} p_i^{1+\lambda},$$

其中  $\lambda > 0$ 。

由于以  $p_i$  为一般项的级数收敛:  $\sum_{i=1}^{\infty} p_i = 1$ , 且  $0 \leq p_i^{1+\lambda} \leq p_i$ , 因而,

$$H^{(\lambda)}(p_1, p_2, \dots, p_n, \dots) = 1 - \sum_{i=1}^{\infty} p_i^{1+\lambda}$$

恒收敛。

由此可见, 这里定义的  $\lambda$  级信息量与 Shannon 熵不同, 对于任何具有可数无限分布列的一维离散型随机变量,  $\lambda$  级信息量恒有意义, 这就克服了 Shannon 熵在这方面的局限性。

## 3 关于高维离散型随机向量的信息量系

仿照 Shannon 熵的推广方式, 也可以定义高维随机向量的联合信息量、边缘信息量、条件信息量、互信息量等。

定义 已知  $n$  维离散型随机向量  $(X_1, X_2, \dots, X_n)$  的分布阵为  $(p_{i_1, i_2, \dots, i_n})$ ,  $(X_1, X_2, \dots, X_m)$  的边缘分布阵为  $(p_{i_1, i_2, \dots, i_m})$ , 其中  $p_{i_1, i_2, \dots, i_m} = \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_1, i_2, \dots, i_n}$ , 则

I  $(X_1, X_2, \dots, X_n)$  的  $\lambda$  级联合信息量为

$$H^{(\lambda)}(X_1, X_2, \dots, X_n) = 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_n} p_{i_1, i_2, \dots, i_n}^{1+\lambda},$$

II  $(X_1, X_2, \dots, X_m)$  的  $\lambda$  级边缘信息量为

$$H^{(\lambda)}(X_1, X_2, \dots, X_m) = 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} p_{i_1, i_2, \dots, i_m}^{1+\lambda};$$

III 在  $(X_{m+1}, X_{m+2}, \dots, X_n)$  条件下的  $(X_1, X_2, \dots, X_m)$  的  $\lambda$  级条件信息量为

$$H^{(\lambda)}((X_1, X_2, \dots, X_m) | (X_{m+1}, X_{m+2}, \dots, X_n)) = \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda} (1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} (\frac{p_{i_1, i_2, \dots, i_m}}{p_{i_{m+1}, i_{m+2}, \dots, i_n}})^{1+\lambda});$$

IV  $(X_1, X_2, \dots, X_m)$  与  $(X_{m+1}, X_{m+2}, \dots, X_n)$  的  $\lambda$  级互信息量为

$$I^{(\lambda)}((X_1, X_2, \dots, X_m), (X_{m+1}, X_{m+2}, \dots, X_n)) = H^{(\lambda)}(X_1, X_2, \dots, X_m) + H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n) - H^{(\lambda)}(X_1, X_2, \dots, X_n).$$

类似地, 这里定义的概念均可推广到具有可数无限分布阵的高维随机向量的情形。

这里定义的联合信息量、边缘信息量、条件信息量、互信息量等, 不仅保留了 Shannon 熵的许多性质, 而且还引出一些新的性质。

## 4 关于信息量系的性质

定理 (最大信息量定理) 对于一维离散型随机变量的有限分布列  $(p_1, p_2, \dots, p_n)$ , 恒有

$$H^{(\lambda)}(p_1, p_2, \dots, p_n) \leq H^{(\lambda)}_n = H^{(\lambda)}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}).$$

证明 设  $p_i(t) = \frac{1}{n} + t(p_i - \frac{1}{n})$ , 则  $p_i(0) = \frac{1}{n}$ ,  $p_i(1) = p_i$ , 令

$$H^{(\lambda)}(p_1(t), p_2(t), \dots, p_n(t)) = 1 - \sum_{i=1}^n p_i^{1+\lambda}(t) = \varphi(t),$$

则有  $H^{(\lambda)}_n = H^{(\lambda)}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) = \varphi(0)$ ,  $H^{(\lambda)}(p_1, p_2, \dots, p_n) = \varphi(1)$ 。注意到

$$\varphi(t) = - (1 + \lambda) \sum_{i=1}^n (p_i - \frac{1}{n}) \times (\frac{1}{n} + t(p_i - \frac{1}{n}))^\lambda,$$

$$\varphi(0) = - (1 + \lambda) \sum_{i=1}^n (p_i - \frac{1}{n}) \frac{1}{n} = 0,$$

$$\varphi(t) = - \lambda(1 + \lambda) \sum_{i=1}^n (p_i - \frac{1}{n})^2 \times (\frac{1}{n} + t(p_i - \frac{1}{n}))^{\lambda-1},$$

则当  $0 \leq t \leq 1$  时, 恒有  $\varphi(t) \leq 0$ , 故知当  $0 \leq t \leq 1$  时,  $\varphi(t) \leq 0$ ,  $\varphi(t)$  为减函数, 从而

$$H^{(\lambda)}(p_1, p_2, \dots, p_n) \leq H^{(\lambda)}_n =$$

$$H^{(\lambda)}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}).$$

定理 (可加性定理) 对于一维离散型随机变

量的有限分布列  $(p_1, p_2, \dots, p_n)$ , 已知  $p_i = \sum_{j=1}^m q_{ij}$ , 其中  $q_{ij} \geq 0$ , 则有

$$H^{(\lambda)}(q_{11}, q_{12}, \dots, q_{1m}, q_{21}, q_{22}, \dots, q_{2m}, \dots, q_{n1}, q_{n2}, \dots, q_{nm}) = H^{(\lambda)}(p_1, p_2, \dots, p_n) + \sum_{i=1}^n p_i^{1+\lambda} H^{(\lambda)}(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \dots, \frac{q_{im}}{p_i}).$$

证明 根据  $\lambda$  级信息量的表达式,

$$\begin{aligned} \sum_{i=1}^n p_i^{1+\lambda} H^{(\lambda)}(\frac{q_{i1}}{p_i}, \frac{q_{i2}}{p_i}, \dots, \frac{q_{im}}{p_i}) &= \sum_{i=1}^n p_i^{1+\lambda} \times (1 - \sum_{j=1}^m (\frac{q_{ij}}{p_i})^{1+\lambda}) = \\ &= \sum_{i=1}^n p_i^{1+\lambda} - \sum_{i=1}^n \sum_{j=1}^m q_{ij}^{1+\lambda} = \\ &= (1 - \sum_{i=1}^n p_i^{1+\lambda}) + (1 - \sum_{i=1}^n \sum_{j=1}^m q_{ij}^{1+\lambda}) = \\ &= H^{(\lambda)}(q_{11}, q_{12}, \dots, q_{1m}, q_{21}, q_{22}, \dots, q_{2m}, \dots, q_{n1}, q_{n2}, \dots, q_{nm}) - H^{(\lambda)}(p_1, p_2, \dots, p_n). \end{aligned}$$

定理 (广义可加性定理) 对于  $n$  维离散型随机向量  $(X_1, X_2, \dots, X_n)$ , 恒有

$$H^{(\lambda)}(X_1, X_2, \dots, X_n) = H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n) + H^{(\lambda)}((X_1, X_2, \dots, X_m) | (X_{m+1}, X_{m+2}, \dots, X_n)).$$

证明 根据定义, 在  $(X_{m+1}, X_{m+2}, \dots, X_n)$  条件下的  $(X_1, X_2, \dots, X_m)$  的条件信息量为

$$H^{(\lambda)}((X_1, X_2, \dots, X_m) | (X_{m+1}, X_{m+2}, \dots, X_n)) = \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda} (1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} (\frac{p_{i_1, i_2, \dots, i_m}}{p_{i_{m+1}, i_{m+2}, \dots, i_n}})^{1+\lambda}) =$$

$$1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} p_{i_1, i_2, \dots, i_m}^{1+\lambda} - (1 - \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda}) =$$

$$H^{(\lambda)}(X_1, X_2, \dots, X_n) - H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n).$$

定理 (互信息量定理)  $(X_1, X_2, \dots, X_m)$  与  $(X_{m+1}, X_{m+2}, \dots, X_n)$  的互信息量满足

$$\begin{aligned} I^{(\lambda)}((X_1, X_2, \dots, X_m), (X_{m+1}, X_{m+2}, \dots, X_n)) &= \\ H^{(\lambda)}(X_1, X_2, \dots, X_n) - H^{(\lambda)}((X_1, X_2, \dots, X_m) | \\ & (X_{m+1}, X_{m+2}, \dots, X_n)) = \\ & H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n) - \\ & H^{(\lambda)}((X_{m+1}, X_{m+2}, \dots, X_n) | (X_1, X_2, \dots, X_m)). \end{aligned}$$

证明 由广义可加性定理知,

$$\begin{aligned} H^{(\lambda)}(X_1, X_2, \dots, X_m) - H^{(\lambda)}((X_1, X_2, \dots, X_m) | \\ (X_{m+1}, X_{m+2}, \dots, X_n)) = \\ H^{(\lambda)}(X_1, X_2, \dots, X_m) - H^{(\lambda)}(X_1, X_2, \dots, X_n) + \\ H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n) = \\ I^{(\lambda)}((X_1, X_2, \dots, X_m), (X_{m+1}, X_{m+2}, \dots, X_n)). \end{aligned}$$

定理 (独立性定理) 如果 \$(X\_1, X\_2, \dots, X\_m)\$ 与 \$(X\_{m+1}, X\_{m+2}, \dots, X\_n)\$ 互相独立, 则

$$\begin{aligned} H^{(\lambda)}(X_1, X_2, \dots, X_n) = 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_n} p_{i_1, i_2, \dots, i_n}^{1+\lambda} \\ \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda}, \\ I^{(\lambda)}((X_1, X_2, \dots, X_m), (X_{m+1}, X_{m+2}, \dots, X_n)) = \\ H^{(\lambda)}(X_1, X_2, \dots, X_m) H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n). \end{aligned}$$

证明 如果 \$(X\_1, X\_2, \dots, X\_m)\$ 与 \$(X\_{m+1}, X\_{m+2}, \dots, X\_n)\$ 互相独立, 则

$$\begin{aligned} p_{i_1, i_2, \dots, i_n} = p_{i_1, i_2, \dots, i_m} p_{i_{m+1}, i_{m+2}, \dots, i_n}, \\ H^{(\lambda)}(X_1, X_2, \dots, X_n) = \\ 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_n} p_{i_1, i_2, \dots, i_n}^{1+\lambda} = \\ 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_n} p_{i_1, i_2, \dots, i_m}^{1+\lambda} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda} = \\ 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} p_{i_1, i_2, \dots, i_m}^{1+\lambda} \\ \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda}; \\ I^{(\lambda)}((X_1, X_2, \dots, X_m), \\ (X_{m+1}, X_{m+2}, \dots, X_n)) = \\ H^{(\lambda)}(X_1, X_2, \dots, X_m) + \\ H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n) - \\ H^{(\lambda)}(X_1, X_2, \dots, X_n) = \\ 1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} p_{i_1, i_2, \dots, i_m}^{1+\lambda} - \\ \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda} + \\ \sum_{i_1} \sum_{i_2} \dots \sum_{i_n} p_{i_1, i_2, \dots, i_n}^{1+\lambda} = \end{aligned}$$

$$(1 - \sum_{i_1} \sum_{i_2} \dots \sum_{i_m} p_{i_1, i_2, \dots, i_m}^{1+\lambda}) \cdot$$

$$(1 - \sum_{i_{m+1}} \sum_{i_{m+2}} \dots \sum_{i_n} p_{i_{m+1}, i_{m+2}, \dots, i_n}^{1+\lambda}) = \\ H^{(\lambda)}(X_1, X_2, \dots, X_m) H^{(\lambda)}(X_{m+1}, X_{m+2}, \dots, X_n).$$

## 5 关于信息量系中参数值的选取建议

世界上的事情是复杂的, 信息的度量也不必采用同一种度量标准。Shannon 熵只是一种度量标准, 不能适应不同情况的需求, 这也是 Shannon 熵的不合理性和局限性。这里定义的信息量系, 正好适应了不同情况的需求。

为了说明这一点, 将 \$\lambda\$ 级信息量做如下解释和理解:

$$H^{(\lambda)}(p_1, p_2, \dots, p_n) =$$

$$1 - \sum_{i=1}^n p_i^{1+\lambda} = \sum_{i=1}^n p_i (1 - p_i^\lambda),$$

其中, \$p\_i(1 - p\_i^\lambda)\$ 表示分布列中的概率 \$p\_i\$ 所贡献的分信息量的大小。设 \$f(p) = p(1 - p^\lambda)\$, 则 \$f'(p) = 1 - (1 + \lambda)p^\lambda\$, \$f''(p) = -\lambda(1 + \lambda)p^{\lambda-1}\$。由此可知, \$f(p) = p(1 - p^\lambda)\$ 是凸函数, 且当 \$p = p\_0 = (1 + \lambda)^{-\frac{1}{\lambda}}\$ 时, \$f(p)\$ 达到最大值。这是参数 \$\lambda\$ 所具有的重要性质。更具体地说: 当 \$0 < \lambda < 1\$ 时, \$p\_0 < 0.5\$; 当 \$\lambda = 1\$ 时, \$p\_0 = 0.5\$; 当 \$\lambda > 1\$ 时, \$p\_0 > 0.5\$。

根据参数 \$\lambda\$ 的这种性质, 在对信息进行度量时, 建议这样选取参数的值: 如果能够认定事件发生的概率较大时所贡献的分信息量较大, 则选 \$\lambda > 1\$; 如果能够认定事件发生的概率较小时所贡献的分信息量较大, 则选 \$0 < \lambda < 1\$; 如果能够认定事件发生与不发生等概率时所贡献的分信息量最大, 或者难以作出上述判断时, 则选 \$\lambda = 1\$, 此时, 所得到的 1 级信息量将呈极简单而又对称的形式, 并且必将具有更强的性质, 得到更广泛的应用:

$$H^{(1)}(p_1, p_2, \dots, p_n) = 1 - \sum_{i=1}^n p_i^2, H^{(1)}_n = 1 - \frac{1}{n}.$$

## 【参考文献】

- [1] Shannon C E. A Mathematical theory of communication[J]. Bell Syst Tech J, 1948, (27): 379- 423, 623- 656
- [2] 日本数学会. 数学百科词典[M]. 北京: 科学出版社, 1984. 1257- 1261.
- [3] 孟庆生. 信息论[M]. 西安: 西安交通大学出版社, 1986. 19- 33.
- [4] Roszak T. 信息崇拜[M]. 苗华健, 陈体仁, 译. 北京: 中国对外翻译出版公司, 1994. 8- 12.
- [5] 沈世镒, 陈鲁生. 信息论与编码理论[M]. 北京: 科学出版社, 2002. 18- 44.

## Thinking on the measurement problems of information

WANG Naixin, ZHANG Hui-feng

(College of Life Sciences, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

**Abstract:** In order to eliminate the limitation of the Shannon entropy, the axiom system for the Shannon entropy was modified, and an information measure system with the axiomatization method was deduced. That could be extended to all discrete random variable and vector with the infinite countable distribution series. The paper proved the theorem of maximum information measure, the theorem of additivity, the theorem of generalized additivity, the theorem of mutual information measurement and the theorem of independency. Suggestions to choose the value of the parameters in the information measurement system were put forward.

**Key words:** Shannon entropy; measurement of information; information measurement system

(上接第 107 页)

## Influence of data imbalance on the analysis results of experiment

HU Xi-yuan

(College of Agronomy, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

**Abstract:** This paper examined the influence of data imbalance on the analysis results for random block design, split-plot design and strip-plot design using the Monte Carlo simulation. The simulation results showed that the imbalance of data due to missing observations led to lower precision and lower exactness of variance component estimation, to underestimation of the error of fixed effects as well as to higher type I error rate for fixed effect test. With more missing observations these influences would be heavier. The imbalance of data influenced the analysis results in split-plot design and strip-plot design much heavier than in random block design.

**Key words:** imbalance of data; variance component estimate; fixed effect test