第一章随机变量及其信息度量

第三节随机事件的自信息

陈兴同

中国矿业大学 数学学院

2021 年 8 月

内容提要

1. 自信息

内容提要

1. 自信息
2. 自信息量单位

内容提要

1. 自信息
2. 自信息量单位
3. 联合自信息

内容提要

内容提要

随机事件与信息量

现在来考虑随机事件的信息度量问题。随机事件 A 发生与否具有不确定性，概率 $p=P(A)$ 是度量这种不确定性的一种数值。因为概率小的事件不容易发生，要推测它何时发生比较困难，通常需要更多的信息；另一方面，概率小的事件一旦发生，一般会造成更大的轰动，产生强大新闻效果，它提供的信息更多，因此可以说小概率事件包含的信息量大。大概率事件因为较容易发生，要推测它何时发生需要的信息量少，它发生时产生的轰动效应也小，故可以说大概率事件包含的信息量少。
常用 $I(A)$或$I(p)$ 来表示事件 A 包含的信息量大小，称为事件 A 的**自信息量**。

度量信息的函数

它应当具有什么表达式呢？根据上面分析，它应当是概率的函数 $I(A)=f(p)$，并且应当满足

（1） 是概率 p 的单调递减函数；

（2） 具有可加性。即当两事件 A, B 独立时，$I(AB)=I(A)+I(B)$；

（3） 是非负函数；

（4） 当概率 $P(A)\to 0$ 时，$I(A)\to +\infty$；

（5） 当 $P(A)=1$ 时 $I(A)=0$。

续：可以度量信息的函数

1.3.1

满足这些性质的函数 $I(A)$ 可以由下面引理确定。

如果非负函数 $f(x), x \geq 1$ 满足下面两个条件：

（1） 当 $1 \leq x < y$ 时 $f(x) < f(y)$，

（2） 当 $x, y \geq 1$ 时 $f(xy) = f(x) + f(y)$，

则这个函数必有表达式 $f(x) = C\log x$，其中 $C > 0$。

证明：

Step1: 对任何 $x \geq 1$ 和正整数 k 有：

$$f(x^k) = f(x \cdot x^{k-1}) = f(x) + f(x^{k-1}) = \cdots = kf(x),$$

于是

$$f(1) = kf(1) \Rightarrow f(1) = 0.$$

再由单调性条件得：

$$f(x) > 0, \text{当} x > 1,$$

因此只有当 $x = 1$ 时才会 $f(x) = 0$。

证明（续）：

Step2: 对任意自然数 $k>1$ 及实数 $x>1, y>1$ 有自然数 n 使

$$y^n \le x^k < y^{n+1}. \tag*{$\cancel{(1.1)}$}$$

对式 (1.1) 取对数得：

$$n\log y \le k\log x < (n+1)\log y,$$

$$\frac{n}{k} \le \frac{\log x}{\log y} < \frac{n+1}{k}. \tag*{$\cancel{(1.2)}$}$$

另外从式 (1.1) 还可得：

$$\frac{n}{k} \le \frac{f(x)}{f(y)} < \frac{n+1}{k}. \tag*{$\cancel{(1.3)}$}$$

证明（续）：

Step3：根据式 (1.2) 与 (1.3) 可得：

$$\left|\frac{f(x)}{f(y)}-\frac{\log x}{\log y}\right|<\frac{1}{k},$$

由于 k 的任意性可知：

$$\frac{f(x)}{f(y)}=\frac{\log x}{\log y},$$

即比值

$$\frac{f(x)}{\log x}=C \quad \text{常数.}$$

定义 1.3.1：自信息量定义

随机事件 A 的自信息量 $I(A)$ 定义为

$$I(A) = \log \frac{1}{P(A)}.$$

基本事件 $\{X = x\}$ 自信息量也记为 $I(x)$:

$$I(x) = \log \frac{1}{p(x)} = -\log p(x), \qquad \overset{1.14}{\cancel{(1.4)}}$$

其中 $p(x) = P\{X = x\} > 0$。显然 $I(x), x \in \mathcal{X}$ 是一个离散函数，故可以确定随机变量 X 的函数 $I(X) = -\log p(X)$。

自信息量单位

自信息量的单位与对数底的选择有关，对不同的底用不同的单位。

（1） 用以 2 为底的对数时，$I(A)$ 的单位为**比特**，符号为：bit，常常用于工程上。

（2） 用以 e 为底的对数时，$I(A)$ 的单位为**奈特**，符号为：nat，常常用于理论推导上。

（3） 用以 10 为底的对数时，$I(A)$ 的单位为**哈特**，符号为：hat。

（4） 用其它大于 1 的正整数 d 为底的对数时，$I(A)$ 的单位为“d 进制信息单位”。

单位换算

常见单位之间的换算：

（1） $1\text{nat} = \log_2 e \approx 1.442695\text{bits}$；

（2） $1\text{hat} = \log_2 10 \approx 3.321928\text{bits}$；

（3） $1\text{hat} = \ln 10 \approx 2.302585\text{nats}$。

定义联合自信息

设两个随机事件 $\{X = x\}$ 与 $\{Y = y\}$ 同时发生的概率（不妨称为联合概率）$p(x, y) > 0$，则这两个事件的**联合自信息量**定义为

$$I(x, y) = \log \frac{1}{p(x, y)} = -\log p(x, y). \qquad \cancel{(3.1)}\ 1.15$$

多个随机事件联合自信息

可以推广到多个随机事件的情况，称

$$I(x_1,x_2,\cdots,x_n)=\log\frac{1}{p(x_1,x_2,\cdots,x_n)},x_i\in\mathcal{X}_i,i=1,2,\cdots,n, \qquad \overset{1.16}{\cancel{(3.2)}}$$

为随机事件 $\{X_1=x_1\},\{X_2=x_2\},\cdots,\{X_n=x_n\}$ 的联合自信息量，其中 $p(x_1,x_2,\cdots,x_n)$ 为这 n 随机事件的联合概率。

显然联合自信息量 1.15 (3.1) 定义了二元离散函数 $I(x,y),(x,y)\in\mathcal{X}\times\mathcal{Y}$，故可以确定随机变量 X,Y 的函数 $I(X,Y)$，类似地也有函数 $I(X_1,X_2,\cdots,X_n)$。

定义条件自信息

如果以某个事件发生为先决条件，还可以定义条件自信息量。设在事件 $\{X=x\}$ 发生条件下事件 $\{Y=y\}$ 条件概率 $p(y|x)>0$，则称

$$I(y|x)=\log\frac{1}{p(y|x)}=-\log p(y|x) \tag{$\cancel{4.1}$ 1.17}$$

为在事件 $\{X=x\}$ 发生的条件下事件 $\{Y=y\}$ 的**条件自信息量**。

多个条件的条件自信息

可以推广到多个条件情况，称

$$I(y|x_1x_2\cdots x_n)=\log\frac{1}{p(y|x_1x_2\cdots x_n)}, \qquad \cancel{(4.2)}$$

1.18

为在事件 $\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$ 发生的条件下事件 $\{Y=y\}$ 的条件自信息量。更一般地，称

$$I(y_1y_2\cdots y_m|x_1x_2\cdots x_n)=\log\frac{1}{p(y_1y_2\cdots y_m|x_1x_2\cdots x_n)}$$

为在事件 $\{X_1=x_1,\cdots,X_n=x_n\}$ 发生的条件下事件 $\{Y_1=y_1,\cdots,Y_m=y_m\}$ 的条件自信息量。

条件自信息作用

条件自信息用于刻画条件事件 $\{X=x\}$ 的发生对事件 $\{Y=y\}$ 包含信息量的影响，若事件 $\{X=x\}$ 发生对事件 $\{Y=y\}$ 的发生有利，则条件自信息量 $I(y|x)$ 会比 $I(y)$ 减少，否则会增大。

条件自信息量一定比自信息量小吗?

定义互自信息

设两个随机事件 $\{X=x\}$ 与 $\{Y=y\}$ 发生的概率 $p_X(x)>0, p_Y(y)>0$，并且联合概率 $p(x,y)>0$，则这两个随机事件的**互自信息量**定义为

1.19

$$I(x;y)=\log\frac{p(x,y)}{p_X(x)p_Y(y)}. \tag{~~5.1~~}$$

互自信息的意义

注意到互自信息的表达式 (5.1) 还可以写成

$$I(x;y) = \log p(x|y) - \log p_X(x) = I(x) - I(x|y),$$

故互自信息表示在已知事件 $\{Y = y\}$ 发生的条件下事件 $\{X = x\}$ 的自信息量的减少量；另外 (5.1) 也可以写成

$$I(x;y) = \log p(y|x) - \log p_Y(y) = I(y) - I(y|x),$$

故互自信息也表示在已知事件 $\{X = x\}$ 发生的条件下事件 $\{Y = y\}$ 的自信息量的减少量。这两个量是相等的，从而可以认为互自信息量 $I(x,y)$ 是两随机事件包含的公共信息量。

多个随机事件的互自信息定义

可以推广到多个随机事件的情况，称

$$I(x_1,\cdots,x_n;y_1,\cdots,y_m)=\log\frac{p(x_1,\cdots,x_n,y_1,\cdots,y_m)}{p_X(x_1,\cdots,x_n)p_Y(y_1,\cdots,y_m)},$$

为 $\{X_1=x_1,\cdots,X_n=x_n\}$ 与 $\{Y_1=y_1,\cdots,Y_m=y_m\}$ 的互自信息量，其中 $p_X(x_1,x_2,\cdots,x_n)$ 为随机事件 $\{X_1=x_1,\cdots,X_n=x_n\}$ 的联合概率，$p_Y(y_1,y_2,\cdots,y_m)$ 为随机事件 $\{Y_1=y_1,\cdots,Y_m=y_m\}$ 联合概率，$p(x_1,x_2,\cdots,x_n,y_1,y_2,\cdots,y_n)$ 为随机事件 $\{X_1=x_1,\cdots,X_n=x_n,Y_1=y_1,\cdots,Y_m=y_m\}$ 的联合概率。

（续）多个随机事件的互自信息定义

根据条件概率

$$p(x,y|z)=\frac{p(x,y,x)}{p_Z(z)},p(x|z)=\frac{p(x,z)}{p_Z(z)},q(y|z)=\frac{p(y,z)}{p_Z(z)},$$

也可以定义**条件互自信息量**

$$I(x;y|z)=\log\frac{p(x,y|z)}{p(x|z)q(y|z)} \qquad \overset{1.20}{\cancel{(5.2)}}$$

例题 1.3.1

如果随机变量 (X, Y) 具有分布律 $p(x, y)$:

$Y \setminus X$	1	2	3	4
1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{16}$
2	0	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{16}$
3	0	0	$\frac{1}{12}$	$\frac{1}{16}$
4	0	0	0	$\frac{1}{16}$

1. 试求随机事件 $\{X = 2, Y = 1\}$ 的联合自信息量。
2. 求条件自信息量 $I(\{Y = 1\} \mid \{X = 2\})$。
3. 求互自信息量 $I(\{Y = 1\}; \{X = 2\})$。

解:

设边缘分布分别为 $p_X(x), p_Y(y)$.

$$
\begin{aligned}
I(\{X=2\},\{Y=1\}) &= -\ln p(2,1) = \ln 8 = 2.0794 \text{ nats},\\
I(\{Y=1\}\,|\,\{X=2\}) &= -\ln p(1|2) = \ln \frac{p_X(2)}{p(2,1)} = \ln 2 = 0.6931 \text{ nats}\\
I(\{Y=1\}) &= -\ln p_Y(1) = \ln \frac{48}{25} = 0.6523 \text{ nats.}
\end{aligned}
$$

这说明事件 $\{X=2\}$ 发生不利于事件 $\{Y=1\}$ 的发生，故条件自信息量增大了。

$$
I(\{Y=1\};\{X=2\}) = \ln \frac{p(2,1)}{p_X(2)p_Y(1)} = \ln \frac{1/8}{1/4*25/48} = \ln \frac{24}{25} = -0.
$$

这说明互自信息量可能为负值，因为条件事件可能是不利事件！

问题：

互自信息量一定非负吗？