§5 熵

熵是用来描述不确定性的量，对于离散型随机变量，记它的分布列为 $p_1, p_2, \cdots, p_n$，熵是 $p_1, p_2, \cdots, p_n$ 的函数，满足以下几个条件：

(1) 熵是 $p_1, p_2, \cdots, p_n$ 的连续函数；

(2) 在等概情况下，熵是 n 的连续函数；

(3) $H(p_1, p_2, p_3) = H(p_1, p_2 + p_3) + (o_2 + p_3)H(\frac{p_2}{p_2 + p_3}, \frac{p_3}{p_2 + p_3})$。

在这些条件下，可以求得

$$H(p_1, \cdots, p_n) = -c \sum_{i=1}^{n} p_i \ln p_i, \quad c > 0$$

由 $f(x) = -\ln x$ 是上凸函数知 $H(p_1, \cdots, p_n) \geqslant -c\ln \sum_{i=1}^{n} p_i^2$.

熵还有以下性质：

(1) $H(p_1, \cdots, p_n) \geqslant 0$，且 $H(p_1, \cdots, p_n) = 0 \Longleftrightarrow P(X = c) = 1$。

(2) 等概事件的熵最大。

(3) 若试验 α, β 独立，则 $H(\alpha, \beta) = H(\alpha) + H(\beta)$。

对连续型随机变量，定义 $H(\alpha) = \int_{-\infty}^{\infty} p(x) \ln p(x) \mathrm{d}x$。

对连续型随机变量的熵，有下面的结论：

(1) 若 $p(x)$ 集中在 $[a, b]$ 上，以均匀分布的熵最大。

(2) 若 $p(x)$ 集中在一位直线上，方差为 σ，以正态分布的熵最大。

(3) 若 $p(x)$ 集中在 $[0, \infty)$ 上，均值为 α，以指数分布的熵最大。

第五章　母函数与特征函数及极限定理

§1　母函数

定义　设 X 为非负整值随机变量，它的概率分布列为

$$P(X=k)=p_k,\quad k=0,1,2,\cdots,$$

则它的母函数为

$$g(z)=\sum_{k=0}^{\infty}p_k z^k.$$

由于 $p_k\geqslant 0$，且 $\sum\limits_{k=0}^{\infty}p_k=1$，故 $g(z)$ 的收敛半径 $R\geqslant 1$。

下面给出几个常见分布的母函数。

(1) 二项分布

设 $X\sim B(n,p)$，则它的母函数为

$$G(z)=\sum_{k=0}^{\infty}C_n^k p^k q^{n-k}z^k=(pz+q)^n.$$

(2) $Poisson$ 分布

设 $X\sim P(\lambda)$，则它的母函数为

$$g(z)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}z^k=\mathrm{e}^{\lambda(z-1)}.$$

(3) 几何分布

设 X 服从参数为 p 的几何分布，则它的母函数为

$$g(z)=\sum_{k=0}^{\infty}pq^{k-1}z^k=\frac{pz}{1-qz}.$$

母函数可以表示成期望的形式，根据随机变量函数的期望公式有

$$g(z)=\sum_{k=0}^{\infty}p_k z^k=E(z^X).$$

母函数有以下性质：

(1) 母函数唯一确定了 X 的概率分布，$p_k = \dfrac{g^{(k)}(0)}{k!}$。

(2) $EX = g^{'}(1)$, $EX^2 = g^{''}(1) + g^{'}(1)$。

(3) X 与 Y 相互独立，则

$$g_{X+Y}(z) = g_X(z)g_Y(z).$$

(4) 设 $X_1, X_2, \cdots, X_n \cdots$ 独立同分布，Y 与 $X_1, X_2, \cdots$ 独立，$Z = \sum_{i=1}^{Y} X_i$，则 Z 的母函数为

$$g_Z(s) = g_Y(g_X(s)).$$

(1)(2) 将 $g(z)$ 展开为 $Taylor$ 级数即可知。下面先证明 (4)。

由母函数的期望表示形式，以及 $X_1, X_2, \cdots$ 相互独立且都和 Y 独立，有

$$\begin{aligned} g_Z(s) &= Es^Z \\ &= Es^{x_1+\cdots+X_Y} \\ &= E[s^{X_1}s^{X_2}\cdots s^{X_Y}] \\ &= E[Es^{X_1}\cdots Es^{X_Y}] \\ &= Eg_X(s)^Y \\ &= g_Y(g_X(s)). \end{aligned}$$

(3) 只需在 (4) 中取 $P(Y=2)=1$ 即可。