

信息的量化:

连续情形

10

本讲介绍研究连续随机变量信息量时的一些基本事实。

在物理世界中，连续随机变量很常见，其中的信源和信道通常由连续的物理定律支配。

例如，信源可能是一个对环境进行模拟（即连续）测量的传感器，而信道可能包含最好被建模为连续随机变量（如高斯变量）的噪声。

由于对一般随机变量（包括连续、连续-离散混合，甚至更抽象的类型）的信息量进行数学上严谨的处理需要高深的数学工具，我们将满足于一种很大程度上是启发式的处理方法，并提供一些未经证明的进阶结果。

10.1 从熵到微分熵. 111

10.2 最大熵密度. 115

10.3 熵功率不等式 117

10.4 互信息. 118

10.1 从熵到微分熵

回想一下，对于一个离散随机变量  $X \sim P_X(x)$ ,  $x \in \mathcal{X}$ ，其熵在定义 2.10 中定义为

$$H(X) = -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x).$$

(10.1)

现在，考虑实数线上的一个连续随机变量  $X$ 。

我们假设  $X$  有一个概率密度函数  $f_X(x)$ ,  $x \in S \subseteq \mathbb{R}$ 。

这里  $S$  表示  $f_X(x)$  不为零的集合，称为  $X$  的支撑集。

如何将熵的概念扩展到这样一个连续的  $X$ ？

一个看似合理的想法是将  $X$  离散化为一个“量化”的离散变量，然后让离散化越来越精细。

鉴于此，固定一个离散化单元大小  $\delta$ ，并考虑一个具有概率质量函数（pmf）的离散随机变量  $X \sim \delta$

$$P_{X \sim \delta}(i) = \int_{i\delta}^{(i+1)\delta} f_X(x) dx,$$

(10.2)

对于  $i = \dots, -1, 0, 1, \dots$ 。也就是说，我们将实数线划分为大小为  $\delta$  的区间，称为“单元”，并通过为每个单元分配  $X$  在该单元中出现的概率，将连续随机变量  $X$  离散化到这些单元中。

112

10 信息的量化：连续情形

在对  $f_X$  做温和假设的情况下，根据中值定理，对每个  $i$ ，存在某个  $x_i \in [i\delta, (i+1)\delta]$ ，使得

$$P_{X \sim \delta}(i) = f_X(x_i) \delta.$$

(10.3)

据此，计算  $X \sim \delta$  的熵，我们得到

**Error: Missing  $\end{matrix}$**

$$= -\sum_i f_X(x_i) \delta \log(f_X(x_i) \delta)$$

$$= -\sum_i f_X(x_i) \delta \log f_X(x_i) - \sum_i f_X(x_i) \delta \log \delta, (10.4)$$

其中第一项收敛于

$$-\int \delta f_X(x) \log f_X(x) dx,$$

(10.5)

如果这个积分存在，当  $\delta \rightarrow 0$  时，第二项恰好是  $\log \delta$ 。

因此，我们看到

$$\lim_{\delta \rightarrow 0} [H(X \sim \delta) - \log \delta] = -\int \delta f_X(x) \log f_X(x) dx$$

(10.6)

也就是说，如果我们试图将连续随机变量的熵定义为其在精细离散化极限下的离散化熵，这通常会导致无穷大的熵，因为当  $\delta \rightarrow 0$  时， $\log \delta$  发散。

然而，在去除了这个  $\log \delta$  主导项之后，剩余的偏差项  $-\int \delta f_X(x) \log f_X(x) dx$  提供了一个衡量  $X$  熵的相对度量。

鉴于前面的讨论，我们引入微分熵的定义。

定义 10.1 对于一个在实数线上定义的连续随机变量  $X$ ，其概率密度函数为  $f_X(x)$ ， $x \in S \subseteq \mathbb{R}$ ，其微分熵定义为

$$h(X) = h(f_X) = -\int_S f_X(x) \log f_X(x) dx. (10.7)$$

如果该积分存在。

备注 10.1 前面微分熵的“推导”本质上是启发式的。

特别是，它严重依赖于使用统一大小的单元对实数线进行离散化。

很容易验证，如果应用非均匀离散化，那么离散化随机变量的熵的渐近行为将不会产生如 (10.6) 所示的简单形式。

总之，一个谨慎的观点是

--- PAGE 3 ---

## 10.1 从熵到微分熵 113

将微分熵视为一个独立的数学实体，而不是熵的扩展。

我们可以将微分熵的概念扩展到多个随机变量。

定义 10.2 对于一个在  $\mathbb{R}^k$  上定义的连续随机向量  $X$ ，其联合概率密度函数为  $f_X(x)$ ， $x \in \delta \subseteq \mathbb{R}^k$ ，其微分熵定义为

$$h(X) = h(f_X) = -\int_{\delta} f_X(x) \log f_X(x) dx, (10.8)$$

如果该积分存在。

我们也可以为一对随机变量定义条件微分熵。

定义 10.3 对于一对随机变量  $X$  和  $Y$ ，其联合概率密度函数为  $f_{X,Y}(x,y)$ ， $(x,y) \in \delta \subseteq \mathbb{R}^k \times \mathbb{R}^l$ ，条件微分熵定义为

$$h(Y|X) = -\int \int f_{X,Y}(x,y) \log f_{Y|X}(y|x) dx dy, \quad (10.9)$$

如果该积分存在。

当  $X$  是一个离散随机变量时，只要对于每个  $X=x$ ,  $f_{Y|X}$  存在，我们仍然可以定义条件微分熵  $h(Y|X)$ , 如下：

$$h(Y|X) = -\sum_{x \in \mathcal{X}} P_X(x) \int S_x f_{Y|X}(y|x) \log f_{Y|X}(y|x) dy. \quad (10.10)$$

如果该积分存在，其中  $S_x$  是  $f_{Y|X}(y|x)$  的支撑集。

示例 10.1 设  $X$  为均匀分布， $f_X(x) = \frac{1}{b-a}$ ,  $x \in [a,b]$ 。

$X$  的微分熵为

$$h(X) = \int_a^b \frac{1}{b-a} \log(b-a) dx = \log(b-a) \text{ 奈特}. \quad (10.11)$$

注意当  $b-a < 1$ ,  $h(X) < 0$ . 所以微分熵不必是非负的。

示例 10.2 设  $X$  为指数分布， $f_X(x) = be^{-bx}$ ,  $x \geq 0$ 。  $X$  的微分熵为

$$h(X) = \int_0^\infty f_X(x) (bx - \ln b) dx$$

$$= bE[X] - \ln b$$

$$= b \frac{1}{b} - \ln b = 1 - \ln b \text{ 奈特}.$$

(10.12)

--- PAGE 4 ---

114

10 信息的量化：连续情形

示例 10.3 令  $X$  为高斯分布，其  $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ ,  $x \in \mathbb{R}$ .  $X$  的微分熵为

$$h(X) = \int_{-\infty}^{\infty} f_X(x) [2\sigma^2(x-\mu)^2 + 21\ln(2\pi\sigma^2)] dx$$

$$= 2\sigma^2 \text{var}[X] + 21\ln(2\pi\sigma^2)$$

$$= 21\ln(2\pi\sigma^2) \text{ 奈特}.$$

(10.13)

示例 10.4 设  $X$  为一个  $k$  维高斯随机向量，其

$$f_X(x) = \frac{1}{(2\pi)^{k/2} |K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)}. \quad (10.14)$$

其中协方差矩阵  $K$  是正定的；

即  $X \sim N(\mu, K)$ 。  $X$  的微分熵为

$$h(X) = \int \frac{1}{|K|^{1/2}} f_X(x) [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$= \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \int \frac{1}{|K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)} [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$= \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \int \frac{1}{|K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)} [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$= \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \int \frac{1}{|K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)} [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$= \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \int \frac{1}{|K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)} [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$= \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \int \frac{1}{|K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)} [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$= \frac{1}{(2\pi)^{k/2} |K|^{1/2}} \int \frac{1}{|K|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T K^{-1}(x-\mu)} [21(x-\mu)^T K^{-1}(x-\mu) + 2k\ln(2\pi) + 21\ln |K|] dx$$

$$=21\text{EN}(0, I_k \times k)[\text{XTX}] + 2k\ln(2\pi) + 21\ln |K|$$

$$= (b) 21\text{trEN}(0, I_k \times k)[X \sim X \sim T] + 2k\ln(2\pi) + 21\ln |K|$$

$$= 21\text{tr} I_k \times k + 2k\ln(2\pi) + 21\ln |K|$$

$$= 2k\ln(2\pi e) + 21\ln |K| \text{ nats,}$$

其中在(a)中我们进行了变量替换  $x \sim = K^{-1/2}(x - \mu)$ ，在(b)中我们使用了迹的一个恒等式  $\text{tr}(AB) = \text{tr}(BA)$ 。

(10.15)

熵的一些重要性质在微分熵中也有类似的形式。

命题 10.1 微分熵满足链式法则，即

$$h(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h(X_i | X_{i-1}, \dots, X_1), \quad (10.16)$$

其中  $X_0$  理解为退化的。

命题 10.2 条件作用减少微分熵，即

$$h(Y | X) \leq h(Y),$$

(10.17)

--- PAGE 5 ---

## 10.2 最大熵密度 115

当且仅当  $X$  和  $Y$  独立时等号成立。

回想一下，对于熵， $H(X) = H(f(X))$  当且仅当  $f$  是一个双射（推论 3.6）。

一般来说，微分熵在变换下不再是不变的，即使变换是双射的。

命题 10.3 考虑一个具有概率密度函数  $f_X(x)$ ,  $x \in S \subseteq \mathbb{R}^k$  的连续随机向量  $X$ 。我们有

► 对于任何给定的向量  $b$ ,  $h(X+b) = h(X)$ 。

► 对于任何给定的非奇异方阵  $A$ ,  $h(AX) = h(X) + \log |A|$ 。

对于两个概率密度函数  $f_X$  和  $\delta_X$ ，我们可以以类似于离散情况的方式定义它们的相对熵。

定义 10.4 对于两个概率密度函数  $f_X$  和  $g_X$ ，它们的相对熵定义为

$$D(f_X || g_X) = \int_S f_X(x) \log \frac{f_X(x)}{g_X(x)} dx$$

(10.18)

其中  $S$  是  $f_X$  或  $g_X$  不为零的支撑集。

注意，如果存在一个区间，其中  $g_X(x)$  为零但  $f_X(x)$  严格为正，则  $D(f_X || g_X) = \infty$ 。

可以用类似于离散情况的方式证明相对熵总是非负的。

，等号成立

命题 10.4 对于任何  $f_X$  和  $g_X$ ，当且仅当在  $S$  上几乎处处有  $f_X(x) = g_X(x)$  时，等号成立。

## 10.2 最大熵密度

考虑所有满足以下条件的概率密度函数：

► 约束 [0]:  $\int S f_X(x) dx = 1$  支撑集为  $\mathcal{S}$ ;

► 约束 [1] 到 [m]:  $\int S f_X(x) r_i(x) dx = a_i, i=1, \dots, m$

定义 10.5 在约束 [0] 到 [m] 下, 最大熵概率密度函数定义为

$$f_X^*(x) = e^{\lambda_0 + \sum_{i=1}^m \lambda_i r_i(x)} \quad x \in \mathcal{S},$$

(10.19)

其中  $\lambda_0, \lambda_1, \dots, \lambda_m$  的选择使得  $f_X^*(x)$  同时满足约束 [0] 到 [m]。

--- PAGE 6 ---

116

## 10 信息的量化: 连续情形

定理 10.1 最大熵概率密度函数  $f_X^*(x)$  在所有满足约束条件 [0] 至 [m] 的概率密度函数中, 使微分熵最大化。

证明: 思路是利用相对熵的非负性。

具体来说, 考虑一个满足约束条件 [0] 至 [m] 的任意概率密度函数  $f_X(x)$ , 并计算  $D(f_X \parallel f_X^*)$ :

$$\begin{aligned} D(f_X \parallel f_X^*) &= \int S f_X(x) \ln f_X^*(x) f_X(x) dx \\ &= -h(f_X) - \int S f_X(x) [\lambda_0 + \sum_{i=1}^m \lambda_i r_i(x)] dx \\ &= -h(f_X) - \lambda_0 \int S f_X(x) dx - \sum_{i=1}^m \lambda_i \int S f_X(x) r_i(x) dx \\ &= -h(f_X) - \lambda_0 - \sum_{i=1}^m \lambda_i a_i \\ &= -h(f_X) - \lambda_0 \int S f_X^*(x) dx - \sum_{i=1}^m \lambda_i \int S f_X^*(x) r_i(x) dx \\ &= -h(f_X) - \int S f_X^*(x) [\lambda_0 + \sum_{i=1}^m \lambda_i r_i(x)] dx \\ &= -h(f_X) + h(f_X^*). \end{aligned}$$

(10.20)

由于  $D(f_X \parallel f_X^*) \geq 0$ , 当且仅当在  $\mathcal{S}$  上几乎处处  $f_X(x) = f_X^*(x)$  时等号成立, 我们得出结论  $h(f_X^*) \geq h(f_X)$ 。

□

示例 10.5 考虑在  $\mathcal{S} = [a, b]$  上的所有概率密度函数, 没有任何进一步的约束。

所以只有约束 [0] 存在, 最大熵概率密度函数是

$$f_X^*(x) = e^{\lambda_0}, \quad x \in [a, b].$$

(10.21)

这意味着  $f_X^*(x)$  在  $\mathcal{S}$  上是一个常数, 立即表明  $f_X^*(x)$  应该是一个均匀概率密度函数, 即

$$f_X^*(x) = \frac{1}{b-a} \quad x \in [a, b].$$

(10.22)

示例 10.6 考虑在  $\mathcal{S} = \mathbb{R}$  上的所有概率密度函数, 满足  $E[X] = 0$  和  $E[X^2] = \sigma^2$ 。因此, 最大熵概率密度函数是

$$f_X^*(x) = e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2},$$

$x \in \mathbb{R}$ .

(10.23)

该概率密度函数的形式与二次型的指数成正比，因此它应该是高斯分布。

均值为零、方差为  $\sigma^2$  的高斯概率密度函数为

--- PAGE 7 ---

函数是

10.3 熵功率不等式 117

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

(10.24)

这个最大熵概率密度函数在下一讲计算高斯信道的容量-成本函数时将非常有用。

例 10.7 考虑在  $S=[0, \infty)$  上的所有概率密度函数，满足  $E[X]=\mu$ 。那么最大熵概率密度函数是

$$f_X(x) = e^{\lambda_0 + \lambda_1 x} \quad x \in [0, \infty).$$

(10.25)

这样的形式立即意味着  $f_X(x)$  应该遵循指数概率密度函数，即

$$f_X(x) = \mu e^{-\mu x}, \quad x \geq 0$$

(10.26)

在这个例子中，如果我们将支撑集  $S$  从  $[0, \infty)$  改为  $\mathbb{R}$ ，那么很容易验证不存在最大熵概率密度。

事实上，令  $X$  是一个均值为、方差为  $\sigma^2$  的高斯随机变量，当  $\sigma^2 \rightarrow 0$  时，得到的微分熵是发散的。

10.3 熵功率不等式

对于一个在  $\mathbb{R}$  上的连续随机变量  $X$ ，其微分熵为  $h(X)$ ，我们称

$$N(X) = \frac{1}{2\pi e} e^{2h(X)}$$

$$h(X) = \frac{1}{2} \ln(2\pi e N(X)),$$

为  $X$  的熵功率。重写

(10.27)

(10.28)

我们立即观察到，当  $X$  是高斯分布时， $N(X)$  正是  $X$  的方差。所以总的来说， $N(X)$  可以被看作是  $X$  的方差（即“功率”）的“推广”。

这个想法可以推广到更高维度。

对于在  $\mathbb{R}^k$  上的  $X$ ，其微分熵为  $h(X)$ ，我们称

$$N(X) = \frac{1}{(2\pi e)^k} e^{2h(X)}$$

(10.29)

--- PAGE 8 ---

## 10 信息的量化：连续情形

X 的熵功率。当 X 是具有协方差矩阵 K 的高斯分布时， $N(X) = |K|^{1/2}/k$ 。

以下基本不等式被称为熵功率不等式（EPI），它关系到两个独立随机向量及其和的熵功率。

定理 10.2 对于在  $R^k$  上的独立连续随机向量 X 和 Y，我们有

$$N(X+Y) \geq N(X) + N(Y),$$

(10.30)

当且仅当 X 和 Y 是具有成比例协方差矩阵的高斯分布时，等号成立。

证明 EPI 有几种方法，但似乎没有一种简单到可以在我们的讲义中讲授。

所以我们只给出结果，不提供证明。

### 10.4 互信息

我们可以通过相对熵（定义 10.4）为一对连续随机变量定义互信息，如下所示。

定义 10.6 对于一对随机变量 X 和 Y，其联合概率密度函数为  $f_{X,Y}(x,y)$ ， $(x,y) \in S \subseteq R^k \times R^l$ ，X 和 Y 之间的互信息定义为

$$I(X;Y) = D(f_{X,Y} \parallel f_X f_Y)$$

(10.31)

$$= - \int \delta f_{X,Y}(x,y) \log f_X(x) f_Y(y) f_{X,Y}(x,y) dx dy, \quad (10.32)$$

如果该积分存在。

备注 10.2 定义 (10.31) 并非一般随机变量互信息的最基本形式。

两个随机变量 X 和 Y 之间互信息最基本也是最通用的定义如下：

$$I(X;Y) = \sup_{\mathcal{Q}_X \times \mathcal{Q}_Y} I([X]_{\mathcal{Q}_X}; [Y]_{\mathcal{Q}_Y}),$$

$\mathcal{Q}_X \times \mathcal{Q}_Y$

(10.33)

其中  $\mathcal{Q}_X$  (或  $\mathcal{Q}_Y$ ) 表示 X (或 Y) 的一个划分，即 X (或 Y) 的有限或可数个子集，这些子集不相交且其并集为 X (或 Y)。

因此， $[X]_{\mathcal{Q}_X}$  (或  $[Y]_{\mathcal{Q}_Y}$ ) 表示通过将  $[X]_{\mathcal{Q}_X} = i$  (或  $[Y]_{\mathcal{Q}_Y} = j$ ) 赋给 X (或 Y) 属于由 i (或 j) 索引的子集的概率而获得的离散化随机变量，该划分依据为  $\mathcal{Q}_X$  (或  $\mathcal{Q}_Y$ )。

这个定义适用于非常一般的随机变量，

--- PAGE 9 ---

并且对于具有联合概率密度函数的连续随机变量，它简化为由 (10.31) 给出的形式。

我们也可以用类似于离散随机变量的方式为三个连续随机变量定义条件互信息。

定义 10.7 对于三个随机变量 X、Y 和 Z，其联合概率密度函数为  $f_{X,Y,Z}(x,y,z)$ ， $(x,y,z) \in \delta \subseteq R^k \times R^l \times R^m$ ，给定 Z 时 X 和 Y 之间的条件互信息定义为

$$I(X;Y \mid Z) = \int f_{X,Y,Z}(x,y,z) \log f_{X \mid Z}(x \mid z) f_{Y \mid Z}(y \mid z) f_{X,Y \mid Z}(x,y \mid z) dx dy dz. \quad (10.34)$$

如果该积分存在。

以下性质对  $I(X;Y)$  和  $I(X;Y \mid Z)$  成立。

推论 10.1 互信息和条件互信息满足以下性质：

- ▶  $I(X;Y) \geq 0$ , 当且仅当  $X$  和  $Y$  独立时等号成立。
- ▶  $I(X;Y|Z) \geq 0$  当且仅当  $X \leftrightarrow Z \leftrightarrow Y$  时等号成立。
- ▶ (DPI) 如果  $X \leftrightarrow Y \leftrightarrow Z$ , 则  $I(X;Y) \geq I(X;Z)$ , 当且仅当  $X \leftrightarrow Z \leftrightarrow Y$  也成立时等号成立。
- ▶  $I(X;Y) = h(Y) - h(Y|X)$ 。
- ▶  $I(X;YZ) = h(YZ) - h(YX, Z)$ 。

备注 10.3 推论 10.1 的前三个性质是普适的。

后两个性质要求相关微分熵的存在。

尽管如此, 当  $X$  是离散的, 且  $Y$  在给定  $X$  (以及适用时的  $Z$ ) 的条件下具有条件概率密度函数时, 它们也成立, 因此在计算具有离散输入  $X$  和连续输出  $Y$  的信道互信息时它们非常有用, 这在实际通信系统中经常遇到。

注释

微分熵已在香农 1948 年的两部分文章 [1] 中出现, 其几个关键性质, 包括 EPI, 也出现在其中, 但没有证明。

对一般随机变量的信息量进行最基本且最普遍的处理, 归功于俄罗斯学派;

参见, 例如, [5]。

## 10.4 互信息 119

--- PAGE 10 ---

120

## 10 信息的量化: 连续情形

练习

1. 计算以下概率密度函数的微分熵。 ▶ 柯西分布:  $f(x) = \frac{1}{\pi\lambda} \frac{1}{1+x^2/\lambda^2}$ ,  $\lambda > 0$

拉普拉斯分布:  $f(x) = \frac{1}{2\lambda} e^{-|x-\theta|/\lambda}$ ,  $\lambda > 0$

- ▶ 瑞利分布:  $f(x) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$ ,  $\sigma > 0$

1. 考虑一个独立同分布 (i.i.d.) 的随机变量序列  $\{X_i, i=1,2,\dots\}$  及其样本均值  $\{S_n, n=1,2,\dots\}$ ,  $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。

- a) 当  $X_i$  是离散的, 熵为  $H(X)$  时, 计算  $nH(S_1, S_2, \dots, S_n)$ 。
- b) 当  $X_i$  是连续的, 微分熵为  $h(X)$  时, 计算  $n h(S_1, S_2, \dots, S_n)$ 。

3. 对于独立的连续随机变量  $X$  和  $Y$ , 证明  $h(X+Y) \geq h(X)$ 。

1. 考虑一个  $k$  维连续随机向量  $X$ 。

- ▶ 如果  $X$  的均值为零, 协方差矩阵为  $K$ , 那么  $X_i$  的最大微分熵是多少?
- ▶ 证明哈达玛不等式, 即  $|K| \leq \prod_{i=1}^k K_{ii}$ 。
- ▶ 证明对数行列式  $\ln |K|$  对于  $K$  是凹函数。

1. 证明最大熵原理的以下推广: 对于任意给定的概率密度函数  $g_X(x)$ ,  $x \in \mathcal{X}$ , 最小化  $D(f_X | g_X)$  同时满足  $\int_{\mathcal{X}} f_X(x) r_i(x) dx = a_i$ ,  $i=1,2,\dots,m$ , 的概率密度函数  $f_X(x)$  由以下形式给出:

$$f_X(x) = g_X(x) e^{\lambda_0 + \sum_{i=1}^m \lambda_i r_i(x)},$$



其中  $\{\lambda_i\}_{i=0,1,\dots,m}$  是参数。

6. 验证在标量情况下，EPI可以重写为以下等价形式：对于在 $\mathbb{R}$ 上的独立连续随机变量 $X$ 和 $Y$ ，令 $X\sim$ 和 $Y\sim$ 为满足 $h(X\sim)=h(X)$ 和 $h(Y\sim)=h(Y)$ 的独立高斯随机变量，则以下不等式成立：

$$h(X+Y) \geq h(X\sim+Y\sim)$$

(10.35)