分布族

续本达

复习

分布族

指数分布族

分布族

续本达

清华大学 工程物理系

2024-10-28 清华

分布族

续本达

复习

分布族

指数分布族

复习

使用 **随机数** 来解决 **计算问题** 方法的统称。

- ↓ 已知的信息越来越少，适用范围越来越广

	方法	已知的分布信息	维度
✓	逆函数法	完全已知	低
✓	舍选法	与标准分布的比值	低
✓	马尔可夫链法	与自身的比值	高
	对抗生成网络法*	无解析形式，无法计算比值	高

蒙特卡罗方法的普适性

- 蒙特卡罗方法在物理之外的几乎所有学科中都有应用。

分布族

续本达

复习

分布族

指数分布族

分布族

离散型

- ① 二项
- ② 泊松
- ③ 几何
- ④ 负二项

连续型

- ① 正态
- ② 指数
- ③ 伽马
- ④ 贝塔

整理分类

这么多种分布，它们有什么共同性质？

- 当改变随机变量分布的参数时，分布函数变化。
 - 说明“分布”本身是可变的。
 - “分布”与“分布族”的概念划分是人为的。
- 相互关联的分布，组成 **分布族**
 - 位置分布族：如高斯分布的 μ 参数变化
 - 尺度分布族：如高斯分布的 σ^2 参数变化
 - 指数分布族 → 本节重点

分类是数学的核心任务之一。

- 概率分布的分类
- 李群的分类

实现由繁及简的提升。

学术会议讲座例子

作为 2024 年数理逻辑年会的公众报告环节，我们有幸邀请到南开大学讲席教授高速老师在下周周五（10 月 25 日）下午 3 点至 5 点为我们带来题为《数学中分类问题的复杂性》的报告。

物理学、化学中分类也是核心任务

元素分类，核素分类，作用力分类等等。

分布族

续本达

复习

分布族

指数分布族

指数分布族

指数分布族

一族的分布，它们的概率密度函数或分布律记为 $f(x|\vec{\theta})$ ，如果有

$$f(x|\vec{\theta}) = \exp \left[\sum_{i=1}^s \eta_i(\vec{\theta}) T_i(x) - B(\vec{\theta}) \right] h(x)$$

的形式，则称为 s 维指数分布族。

如果把 $\eta_i(\vec{\theta})$ 当作自变量反解 θ ，那么形式有进一步的化简。

指数分布族标准形式

$$f(x|\vec{\eta}) = \exp \left[\sum_{i=1}^s \eta_i T_i(x) - A(\vec{\eta}) \right] h(x)$$

其中 $\eta_i = \eta_i(\vec{\theta})$, $A[\vec{\eta}(\vec{\theta})] = B(\vec{\theta})$ 。

- 二项分布是否属于指数分布族？

$$P(k|p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned} P(k|p, n) &= \binom{n}{k} (1-p)^n \left(\frac{p}{1-p}\right)^k \\ &= \binom{n}{k} \exp \left[k \log \left(\frac{p}{1-p}\right) + n \log(1-p) \right] \\ &= h(k) \exp [T(k)\eta(p) - B(p)] \end{aligned}$$

其中， $h(k) = \binom{n}{k}$, $T(k) = k$, $\eta(p) = \log \left(\frac{p}{1-p}\right)$, $B(p) = -n \log(1-p)$ 。
当把 n 视为不变时，二项分布属于指数分布族。

- 思考：若 n 也视为参数，二项分布是否属于指数分布族？

- 二项分布是否属于指数分布族？

$$P(k|p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned} P(k|p, n) &= \binom{n}{k} (1-p)^n \left(\frac{p}{1-p}\right)^k \\ &= \binom{n}{k} \exp \left[k \log \left(\frac{p}{1-p}\right) + n \log(1-p) \right] \\ &= h(k) \exp [T(k)\eta(p) - B(p)] \end{aligned}$$

其中， $h(k) = \binom{n}{k}$, $T(k) = k$, $\eta(p) = \log \left(\frac{p}{1-p}\right)$, $B(p) = -n \log(1-p)$ 。
当把 n 视为不变时，二项分布属于指数分布族。

- 思考：若 n 也视为参数，二项分布是否属于指数分布族？

- 二项分布是否属于指数分布族？

$$P(k|p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned} P(k|p, n) &= \binom{n}{k} (1-p)^n \left(\frac{p}{1-p}\right)^k \\ &= \binom{n}{k} \exp \left[k \log \left(\frac{p}{1-p}\right) + n \log(1-p) \right] \\ &= h(k) \exp [T(k)\eta(p) - B(p)] \end{aligned}$$

其中， $h(k) = \binom{n}{k}$, $T(k) = k$, $\eta(p) = \log \left(\frac{p}{1-p}\right)$, $B(p) = -n \log(1-p)$ 。
当把 n 视为不变时，二项分布属于指数分布族。

- 思考：若 n 也视为参数，二项分布是否属于指数分布族？

令

$$\eta = \log \left(\frac{p}{1-p} \right)$$

那么

$$(1-p) = 1/(1+e^\eta) \implies \log(1-p) = -\log(1+e^\eta)$$

可得 $A(\eta)$

$$A(\eta) = B[p(\eta)] = -n \log(1-p) = n \log(1+e^\eta)$$

- 期望： $E(X) = A'(\eta)$
- 方差： $\text{Var}(X) = A''(\eta)$

- 正态分布