

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量


连续随机变量

第二章：随机变量及其分布¹

Xiao Yuan¹

¹Center on Frontiers of Computing Studies, Peking University, Beijing
100871, China
xiaoyuan@pku.edu.cn

February 27, 2024

¹讲义内容基于三本参考教材以及网络素材 

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

1 随机变量

2 离散随机变量

- 0-1 分布
- 二项分布
- 泊松分布
- 几何分布和负二项分布

3 连续型随机变量

- 分布函数
- 概率密度
- 均匀分布
- 指数分布
- 伽马分布
- 不同分布的关系
- 正态分布

4 随机变量的函数分布

- 离散随机变量
- 连续随机变量

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

1 随机变量

2 离散随机变量

- 0-1 分布
- 二项分布
- 泊松分布
- 几何分布和负二项分布

3 连续型随机变量

- 分布函数
- 概率密度
- 均匀分布
- 指数分布
- 伽马分布
- 不同分布的关系
- 正态分布

4 随机变量的函数分布

- 离散随机变量
- 连续随机变量

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

- 随机变量为定量化和数量化描述随机现象。例如：降雨量的大小、硬币正反面、产品品质等

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

- 随机变量为定量化和数量化描述随机现象。例如：降雨量的大小、硬币正反面、产品品质等
- 随机变量一般用大写字母 X 、 Y 、 Z 表示

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

- 随机变量为定量化和数量化描述随机现象。例如：降雨量的大小、硬币正反面、产品品质等
- 随机变量一般用大写字母 X 、 Y 、 Z 表示
- 随机变量和事件的关系：对于取值 x , 定义事件 $A = \{e: X(e) = x\}$, 则有 $P(A) = P(X = x)$.

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

- 随机变量为定量化和数量化描述随机现象。例如：降雨量的大小、硬币正反面、产品品质等
- 随机变量一般用大写字母 X 、 Y 、 Z 表示
- 随机变量和事件的关系：对于取值 x , 定义事件 $A = \{e: X(e) = x\}$, 则有 $P(A) = P(X = x)$.
- 相反, 对于任意事件 A , 是否总是可以对随机变量取特定的值 \mathcal{I} 使得 $P(A) = P(X \in \mathcal{I})$?

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

- 随机变量为定量化和数量化描述随机现象。例如：降雨量的大小、硬币正反面、产品品质等
- 随机变量一般用大写字母 X 、 Y 、 Z 表示
- 随机变量和事件的关系：对于取值 x , 定义事件 $A = \{e: X(e) = x\}$, 则有 $P(A) = P(X = x)$.
- 相反, 对于任意事件 A , 是否总是可以对随机变量取特定的值 \mathcal{I} 使得 $P(A) = P(X \in \mathcal{I})$?
- 随机变量和样本空间划分的关系?

随机变量

对于样本空间 $S = \{e\}$ 和样本点 e , 随机变量 $X = X(e)$ 为定义在 S 上的实、单值函数

- 随机变量为定量化和数量化描述随机现象。例如：降雨量的大小、硬币正反面、产品品质等
- 随机变量一般用大写字母 X 、 Y 、 Z 表示
- 随机变量和事件的关系：对于取值 x , 定义事件 $A = \{e: X(e) = x\}$, 则有 $P(A) = P(X = x)$.
- 相反, 对于任意事件 A , 是否总是可以对随机变量取特定的值 \mathcal{I} 使得 $P(A) = P(X \in \mathcal{I})$?
- 随机变量和样本空间划分的关系?
- 随机变量分为离散和连续随机变量

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

1 随机变量

2 离散随机变量

- 0-1 分布
- 二项分布
- 泊松分布
- 几何分布和负二项分布

3 连续型随机变量

- 分布函数
- 概率密度
- 均匀分布
- 指数分布
- 伽马分布
- 不同分布的关系
- 正态分布

4 随机变量的函数分布

- 离散随机变量
- 连续随机变量

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

离散随机变量

取值至多可数的随机变量为离散型的随机变量

离散随机变量

取值至多可数的随机变量为离散型的随机变量

- 给定随机变量取值 $X \in \{x_1, x_2, \dots\}$, $p_i = P(X = x_i)$ 为随机变量 X 的概率分布, $\{p_k\}$ 为概率分布列 (或概率分布律), 简称分布列 (或分布律)

离散随机变量

取值至多可数的随机变量为离散型的随机变量

- 给定随机变量取值 $X \in \{x_1, x_2, \dots\}$, $p_i = P(X = x_i)$ 为随机变量 X 的概率分布, $\{p_k\}$ 为概率分布列 (或概率分布律), 简称分布列 (或分布律)
- $\{p_k\}$ 满足 $p_i \geq 0$ 和 $\sum_i p_i = 1$

离散随机变量

取值至多可数的随机变量为离散型的随机变量

- 给定随机变量取值 $X \in \{x_1, x_2, \dots\}$, $p_i = P(X = x_i)$ 为随机变量 X 的概率分布, $\{p_k\}$ 为概率分布列 (或概率分布律), 简称分布列 (或分布律)
- $\{p_k\}$ 满足 $p_i \geq 0$ 和 $\sum_i p_i = 1$
- 例: 某人骑自行车从学校到火车站, 一路上要经过 3 个独立的交通灯, 设各灯工作独立, 且设各灯为红灯的概率为 p , $0 < p < 1$, 以 X 表示首次停车时所通过的交通灯数, 求 X 的概率分布列

离散随机变量

取值至多可数的随机变量为离散型的随机变量

- 给定随机变量取值 $X \in \{x_1, x_2, \dots\}$, $p_i = P(X = x_i)$ 为随机变量 X 的概率分布, $\{p_k\}$ 为概率分布列 (或概率分布律), 简称分布列 (或分布律)
- $\{p_k\}$ 满足 $p_i \geq 0$ 和 $\sum_i p_i = 1$
- 例: 某人骑自行车从学校到火车站, 一路上要经过 3 个独立的交通灯, 设各灯工作独立, 且设各灯为红灯的概率为 p , $0 < p < 1$, 以 X 表示首次停车时所通过的交通灯数, 求 X 的概率分布列

$$P(X=0) = p, P(X=1) = (1-p)p,$$

$$P(X=2) = (1-p)^2 p, P(X=3) = (1-p)^3 p.$$

离散随机变量

取值至多可数的随机变量为离散型的随机变量

- 给定随机变量取值 $X \in \{x_1, x_2, \dots\}$, $p_i = P(X = x_i)$ 为随机变量 X 的概率分布, $\{p_k\}$ 为概率分布列 (或概率分布律), 简称分布列 (或分布律)
- $\{p_k\}$ 满足 $p_i \geq 0$ 和 $\sum_i p_i = 1$
- 例: 某人骑自行车从学校到火车站, 一路上要经过 3 个独立的交通灯, 设各灯工作独立, 且设各灯为红灯的概率为 p , $0 < p < 1$, 以 X 表示首次停车时所通过的交通灯数, 求 X 的概率分布列

$$P(X=0) = p, P(X=1) = (1-p)p,$$

$$P(X=2) = (1-p)^2 p, P(X=3) = (1-p)^3 p.$$
- 给定 $\lambda > 0$ 和概率分布 $P(X=k) = c\lambda^k/k!$, $k=1, 2, \dots$, 求 c

0-1 (两点) 分布

随机变量 X 的取值只有两个, 为 0 或 1, 记 $p = P(X=1)$, 则称 X 服从参数为 p 的 **0-1 分布**, 或**两点分布**, 或**伯努利 (Bernoulli) 分布**, 记为 $X \sim B(1, p)$

0-1 (两点) 分布

随机变量 X 的取值只有两个, 为 0 或 1, 记 $p = P(X=1)$, 则称 X 服从参数为 p 的 **0-1 分布**, 或两点分布, 或伯努利 (Bernoulli) 分布, 记为 $X \sim B(1, p)$

- 例: 检查产品的质量是否合格, 对新生婴儿的性别进行登记, 检验种子是否发芽以及前面多次讨论过的“抛硬币”试验都可以用 (0-1) 分布的随机变量来描述

0-1 (两点) 分布

随机变量 X 的取值只有两个, 为 0 或 1, 记 $p = P(X=1)$, 则称 X 服从参数为 p 的 **0-1 分布**, 或两点分布, 或伯努利 (**Bernoulli**) 分布, 记为 $X \sim B(1, p)$

- 例: 检查产品的质量是否合格, 对新生婴儿的性别进行登记, 检验种子是否发芽以及前面多次讨论过的“抛硬币”试验都可以用 (0-1) 分布的随机变量来描述
- 一个随机试验, 设 A 是一随机事件, 且 $P(A) = p$, ($0 < p < 1$)。若仅考虑事件 A 发生与否, 定义一个服从参数为 p 的 0-1 分布的随机变量:

$$X = \begin{cases} 1 & A \text{ 发生} \\ 0 & A \text{ 不发生} \end{cases}$$

- 只有两个可能结果的试验, 称为伯努利试验

二项分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

n 重伯努利试验

设试验只有两个可能的结果： A 和 \bar{A} ，且 $P(A) = p$ ， $0 < p < 1$ ，将试验独立地重复进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验

二项分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

n 重伯努利试验

设试验只有两个可能的结果： A 和 \bar{A} ，且 $P(A) = p$ ， $0 < p < 1$ ，将试验独立地重复进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验

二项分布

记随机变量 X 为 n 重伯努利试验中 A 发生的次数，则 $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ ， $k = 0, 1, 2, \dots, n$ ，称 X 为服从参数为 n, p 的二项分布，记作 $X \sim B(n, p)$

二项分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

n 重伯努利试验

设试验只有两个可能的结果： A 和 \bar{A} ，且 $P(A) = p$ ， $0 < p < 1$ ，将试验独立地重复进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验

二项分布

记随机变量 X 为 n 重伯努利试验中 A 发生的次数，则 $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ ， $k = 0, 1, 2, \dots, n$ ，称 X 为服从参数为 n, p 的二项分布，记作 $X \sim B(n, p)$

■ 容易验证 $\sum_k P(X = k) = 1$

n 重伯努利试验

设试验只有两个可能的结果： A 和 \bar{A} ，且 $P(A) = p$ ， $0 < p < 1$ ，将试验独立地重复进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验

二项分布

记随机变量 X 为 n 重伯努利试验中 A 发生的次数，则 $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ ， $k = 0, 1, 2, \dots, n$ ，称 X 为服从参数为 n, p 的二项分布，记作 $X \sim B(n, p)$

- 容易验证 $\sum_k P(X = k) = 1$
- 已知 $X \sim B(n, p)$ ， $P(X = k)$ 的最大值取值为 $X = [(n + 1)p]$

泊松分布 (Poisson 分布)

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

泊松分布

随机变量 X 取值为非负整数, 当 $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \dots$, $\lambda > 0$ 时, 称 X 服从参数为 λ 的泊松分布, 记 $X \sim \pi(\lambda)$

泊松分布 (Poisson 分布)

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

泊松分布

随机变量 X 取值为非负整数, 当 $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \dots$, $\lambda > 0$ 时, 称 X 服从参数为 λ 的泊松分布, 记 $X \sim \pi(\lambda)$

- 描述了单位时间、单位空间等 (均匀可能性) 事件发生的次数概率
- 例: 假设我们每时每刻看到流星的概率是均匀随机, 那么我们在一个晚上看到的流星数量服从泊松分布

泊松分布 (Poisson 分布)

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

泊松分布

随机变量 X 取值为非负整数, 当 $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \dots$, $\lambda > 0$ 时, 称 X 服从参数为 λ 的泊松分布, 记 $X \sim \pi(\lambda)$

- 描述了单位时间、单位空间等 (均匀可能性) 事件发生的次数概率
- 例: 假设我们每时每刻看到流星的概率是均匀随机, 那么我们在一个晚上看到的流星数量服从泊松分布
- 例: 假设每个人每时每刻生病的概率是均匀随机, 那么我们在一年生病的次数服从泊松分布

泊松分布 (Poisson 分布)

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

泊松分布

随机变量 X 取值为非负整数, 当 $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \dots$, $\lambda > 0$ 时, 称 X 服从参数为 λ 的泊松分布, 记 $X \sim \pi(\lambda)$

- 描述了单位时间、单位空间等 (均匀可能性) 事件发生的次数概率
- 例: 假设我们每时每刻看到流星的概率是均匀随机, 那么我们在一个晚上看到的流星数量服从泊松分布
- 例: 假设每个人每时每刻生病的概率是均匀随机, 那么我们在一年生病的次数服从泊松分布
- $\pi(\lambda)$ 概率最大为 $X = [\lambda]$

二项分布与泊松分布的关系



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

二项分布与泊松分布的关系

对于二项分布 $P_{B,n}(X=k) \sim B(n, p_n)$, 考虑 $n \rightarrow \infty$ 且 $\lim_{n \rightarrow \infty} np_n = \lambda$, 则对于任意 $k=0, 1, \dots$, 有

$$\lim_{n \rightarrow \infty} P_{B,n}(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}. \quad (1)$$

Proof: 已知 $P_B(X=k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$ 和 $\lambda_n = p_n n$, 我们有

$$\begin{aligned} P_B(X=k) &= \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k}, \\ &= \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right), \end{aligned} \quad (2)$$

注意到 $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} = e^{-\lambda}$ 和 $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) = 1$, 得到结论。

二项分布与泊松分布的关系



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：某种疾病的发病率为 $1/1000$ ，某社区有 5000 个成年人，求发病人数为 k 的概率

二项分布与泊松分布的关系

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：某种疾病的发病率为 $1/1000$ ，某社区有 5000 个成年人，求发病人数为 k 的概率
- 记发病人数为随机变量 $X \sim B(5000, 1/1000)$ ，则
- $$P(X = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} \approx e^{-5} \frac{5^k}{k!}.$$

二项分布与泊松分布的关系

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：某种疾病的发病率为 $1/1000$ ，某社区有 5000 个成年人，求发病人数为 k 的概率

记发病人数为随机变量 $X \sim B(5000, 1/1000)$ ，则

$$P(X = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} \approx e^{-5} \frac{5^k}{k!}.$$

考虑时间 t 分为 n 段，假设每一段时间事件 A 发生概率为 p_n ，则 t 时间事件 A 发生的次数服从二项分布 $B(n, p_n)$ 。注意到当 n 很大时，事件 A 发生的概率将变小，假设 $p_n \propto \lambda/n$ ，那么我们有

$$\lim_{n \rightarrow \infty} B(n, p_n) = \pi(\lambda) \quad (3)$$

也即是，泊松分布可以理解为均匀可能性发生的事件，在单位时间内发生次数的分布。

几何分布和负二项分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

几何分布

重复伯努利试验直到事件 A 发生, 记随机变量 X 为事件不发生次数, 则 $P(X = k) = p(1 - p)^k$, $k = 0, 1, \dots$, 称 X 服从参数 p 的几何分布, 记作 $X \sim G(p)$

几何分布和负二项分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

几何分布

重复伯努利试验直到事件 A 发生, 记随机变量 X 为事件不发生次数, 则 $P(X = k) = p(1 - p)^k$, $k = 0, 1, \dots$, 称 X 服从参数 p 的几何分布, 记作 $X \sim G(p)$

几何分布的无记忆性: $P(X > m + n | X > n) = P(X > m)$

几何分布和负二项分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

几何分布

重复伯努利试验直到事件 A 发生, 记随机变量 X 为事件不发生次数, 则 $P(X = k) = p(1 - p)^k$, $k = 0, 1, \dots$, 称 X 服从参数 p 的几何分布, 记作 $X \sim G(p)$

几何分布的无记忆性: $P(X > m + n | X > n) = P(X > m)$

负二项分布 (帕斯卡分布)

重复伯努利试验直到事件 A 发生 r 次, 记随机变量 X 为事件不发生次数, 则 $P(X = k) = \binom{k+r-1}{r-1} p^r (1 - p)^k$, $k = 0, 1, \dots$, 称 X 服从参数 r, p 的负二项分布, 记作 $X \sim NB(r, p)$ 。注意到 $G(p) = NB(1, p)$

几何分布和负二项分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

几何分布

重复伯努利试验直到事件 A 发生, 记随机变量 X 为事件不发生次数, 则 $P(X = k) = p(1 - p)^k$, $k = 0, 1, \dots$, 称 X 服从参数 p 的几何分布, 记作 $X \sim G(p)$

几何分布的无记忆性: $P(X > m + n | X > n) = P(X > m)$

负二项分布 (帕斯卡分布)

重复伯努利试验直到事件 A 发生 r 次, 记随机变量 X 为事件不发生次数, 则 $P(X = k) = \binom{k+r-1}{r-1} p^r (1 - p)^k$, $k = 0, 1, \dots$, 称 X 服从参数 r, p 的负二项分布, 记作 $X \sim NB(r, p)$ 。注意到 $G(p) = NB(1, p)$

例: 从生产线上随机抽产品进行检测, 设产品的次品率为 p , $0 < p < 1$, 若查到一只次品就得停机检修, 则已检测产品数 X 服从 $G(p)$ 。若查到 r 次停机, 则已检测产品中良品数 Y 服从 $NB(r, p)$

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

1 随机变量

2 离散随机变量

- 0-1 分布
- 二项分布
- 泊松分布
- 几何分布和负二项分布

3 连续型随机变量

- 分布函数
- 概率密度
- 均匀分布
- 指数分布
- 伽马分布
- 不同分布的关系
- 正态分布

4 随机变量的函数分布

- 离散随机变量
- 连续随机变量

随机变量的分布函数



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

随机变量的分布函数

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

分布函数的性质

- 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

分布函数的性质

- 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$
- 单调性: $F(x)$ 单调不减

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

分布函数的性质

- 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$
- 单调性: $F(x)$ 单调不减
- 右连续: $\lim_{\varepsilon \rightarrow 0+} F(x + \varepsilon) = F(x)$

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

分布函数的性质

- 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$
- 单调性: $F(x)$ 单调不减
- 右连续: $\lim_{\varepsilon \rightarrow 0+} F(x + \varepsilon) = F(x)$
- 是否左连续?

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

分布函数的性质

- 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$
- 单调性: $F(x)$ 单调不减
- 右连续: $\lim_{\varepsilon \rightarrow 0+} F(x + \varepsilon) = F(x)$
- 是否左连续?
- 离散随机变量的分布函数一般是跳跃的

分布函数

考虑随机变量 X 和实变量 x , 函数 $F(x) := P(X \leq x)$ 为 x 的概率分布函数

分布函数的性质

- 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$
- 单调性: $F(x)$ 单调不减
- 右连续: $\lim_{\varepsilon \rightarrow 0+} F(x + \varepsilon) = F(x)$
- 是否左连续?
- 离散随机变量的分布函数一般是跳跃的

写出上述离散随机变量的分布函数

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

概率密度

对于随机变量 X ，若存在 $f(x)$ 使得 $F(x) = \int_{-\infty}^x f(t)dt$ ，则 X 为连续型随机变量， $F(x)$ 为分布函数， $f(x)$ 为概率密度函数

概率密度

对于随机变量 X ，若存在 $f(x)$ 使得 $F(x) = \int_{-\infty}^x f(t)dt$ ，则 X 为连续型随机变量， $F(x)$ 为分布函数， $f(x)$ 为概率密度函数

概率密度的性质

- 非负性： $f(x) \geq 0$
- 归一性： $\int_{-\infty}^{\infty} f(x)dx = 1$

概率密度

对于随机变量 X ，若存在 $f(x)$ 使得 $F(x) = \int_{-\infty}^x f(t)dt$ ，则 X 为连续型随机变量， $F(x)$ 为分布函数， $f(x)$ 为概率密度函数

概率密度的性质

- 非负性： $f(x) \geq 0$
- 归一性： $\int_{-\infty}^{\infty} f(x)dx = 1$

任意满足以上性质的函数 $f(x)$ 都可以用来定义概率密度和分布函数。

概率密度

对于随机变量 X , 若存在 $f(x)$ 使得 $F(x) = \int_{-\infty}^x f(t)dt$, 则 X 为连续型随机变量, $F(x)$ 为分布函数, $f(x)$ 为概率密度函数

概率密度的性质

- 非负性: $f(x) \geq 0$
- 归一性: $\int_{-\infty}^{\infty} f(x)dx = 1$

任意满足以上性质的函数 $f(x)$ 都可以用来定义概率密度和分布函数。

概率密度和概率的关系为 $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$ 其中 $x_2 \geq x_1$, 因此, $P(X = a) = 0$ 。

同时, 在 $f(x)$ 的连续点我们有 $F'(x) = f(x)$ 。

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 概率密度刻画了在无穷小时间内事件发生的概率大小，与物体的密度类似。

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 概率密度刻画了在无穷小时间内事件发生的概率大小，与物体的密度类似。
- 当分布函数 $F(x)$ 不连续时，不对应连续性随机变量

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 概率密度刻画了在无穷小时间内事件发生的概率大小，与物体的密度类似。
- 当分布函数 $F(x)$ 不连续时，不对应连续性随机变量
- 我们可以随意改变概率密度函数在有限个点的值而不影响对应的分布函数

- 概率密度刻画了在无穷小时间内事件发生的概率大小，与物体的密度类似。
- 当分布函数 $F(x)$ 不连续时，不对应连续性随机变量
- 我们可以随意改变概率密度函数在有限个点的值而不影响对应的分布函数
- 例：一银行服务需要等待，设等待时间 X （分钟）的概率密度为 $f(x) = \begin{cases} e^{-x/10}/10 & x > 0 \\ 0 & x < 0 \end{cases}$ 某人到银行办理业务，写出等待时间的随机变量的分布函数。若他等待超过 30 分钟后就离开，写出对应的分布函数，判断是否是离散型还是连续型随机变量。

均匀分布

假设随机变量 X 取值范围为 (a, b) , 概率密度为

$$f(x) = \begin{cases} 1/(b-a) & x \in (a, b) \\ 0 & \text{else} \end{cases} \quad \text{则称 } X \text{ 为区间 } (a, b) \text{ 上的均匀分布, 记为 } X \sim U(a, b)$$

均匀分布

假设随机变量 X 取值范围为 (a, b) , 概率密度为

$$f(x) = \begin{cases} 1/(b-a) & x \in (a, b) \\ 0 & \text{else} \end{cases} \quad \text{则称 } X \text{ 为区间 } (a, b) \text{ 上的均匀分布, 记为 } X \sim U(a, b)$$

- 例: (1) 在区间 $(-1, 2)$ 上随机取一数 X , 试写出 X 的概率密度。并求 $P(X > 0)$ 的值; (2) 若在该区间上随机取 10 个数, 求 10 个数中恰有两个数大于 0 的概率。

均匀分布

假设随机变量 X 取值范围为 (a, b) , 概率密度为

$$f(x) = \begin{cases} 1/(b-a) & x \in (a, b) \\ 0 & \text{else} \end{cases} \quad \text{则称 } X \text{ 为区间 } (a, b) \text{ 上的均匀分布, 记为 } X \sim U(a, b)$$

- 例: (1) 在区间 $(-1, 2)$ 上随机取一数 X , 试写出 X 的概率密度。并求 $P(X > 0)$ 的值; (2) 若在该区间上随机取 10 个数, 求 10 个数中恰有两个数大于 0 的概率。

$$(1) P(X > 0) = 2/3 \quad (2) B(10, P(X > 0))$$

均匀分布

假设随机变量 X 取值范围为 (a, b) , 概率密度为

$$f(x) = \begin{cases} 1/(b-a) & x \in (a, b) \\ 0 & \text{else} \end{cases} \quad \text{则称 } X \text{ 为区间 } (a, b) \text{ 上的均匀分布, 记为 } X \sim U(a, b)$$

- 例: (1) 在区间 $(-1, 2)$ 上随机取一数 X , 试写出 X 的概率密度。并求 $P(X > 0)$ 的值; (2) 若在该区间上随机取 10 个数, 求 10 个数中恰有两个数大于 0 的概率。

$$(1) P(X > 0) = 2/3 \quad (2) B(10, P(X > 0))$$

- 例: 某车站每天从早上 6 点开始, 每隔 30 分钟有一班车开往上海。王先生在早上 6:20 过 X 分钟到达车站, 设 X 服从 $(0, 50)$ 上的均匀分布, (1) 求王先生候车时间不超过 15 分钟的概率; (2) 如果王先生一月中有两次按此方式独立地去候车, 求他一次候车不超过 15 分钟, 另一次候车大于 10 分钟的概率。

均匀分布

假设随机变量 X 取值范围为 (a, b) , 概率密度为

$$f(x) = \begin{cases} 1/(b-a) & x \in (a, b) \\ 0 & \text{else} \end{cases} \quad \text{则称 } X \text{ 为区间 } (a, b) \text{ 上的均匀分布, 记为 } X \sim U(a, b)$$

- 例: (1) 在区间 $(-1, 2)$ 上随机取一数 X , 试写出 X 的概率密度。并求 $P(X > 0)$ 的值; (2) 若在该区间上随机取 10 个数, 求 10 个数中恰有两个数大于 0 的概率。

$$(1) P(X > 0) = 2/3 \quad (2) B(10, P(X > 0))$$

- 例: 某车站每天从早上 6 点开始, 每隔 30 分钟有一班车开往上海。王先生在早上 6:20 过 X 分钟到达车站, 设 X 服从 $(0, 50)$ 上的均匀分布, (1) 求王先生候车时间不超过 15 分钟的概率; (2) 如果王先生一月中有两次按此方式独立地去候车, 求他一次候车不超过 15 分钟, 另一次候车大于 10 分钟的概率。

$$(1) 0.5 \quad (2) 2 \cdot 1/2 \cdot 3/5 = 3/5$$

指数分布

对于 $\lambda > 0$, 概率密度为 $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为指数分布, 记作 $X \sim \text{Exp}(\lambda)$

指数分布

对于 $\lambda > 0$, 概率密度为 $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为指数分布, 记作 $X \sim \text{Exp}(\lambda)$

- 指数分布的特征为无记忆性 (等价), 也即是 $P(X > x_0 + x | X > x_0) = P(X > x)$ 证明只需要注意到 $P(X > x) = e^{-\lambda x}$.

指数分布

对于 $\lambda > 0$, 概率密度为 $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为指数分布, 记作 $X \sim \text{Exp}(\lambda)$

- 指数分布的特征为无记忆性 (等价), 也即是 $P(X > x_0 + x | X > x_0) = P(X > x)$ 证明只需要注意到 $P(X > x) = e^{-\lambda x}$.
- 例: 考虑从 $t=0$ 到任意时间 t 内事件 A 发生的次数服从泊松分布 $\pi(\lambda t)$, 那么从 $t=0$ 开始事件 A 发生的时间从什么分布?

指数分布

对于 $\lambda > 0$, 概率密度为 $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为指数分布, 记作 $X \sim \text{Exp}(\lambda)$

- 指数分布的特征为无记忆性 (等价), 也即是 $P(X > x_0 + x | X > x_0) = P(X > x)$ 证明只需要注意到 $P(X > x) = e^{-\lambda x}$.
- 例: 考虑从 $t = 0$ 到任意时间 t 内事件 A 发生的次数服从泊松分布 $\pi(\lambda t)$, 那么从 $t = 0$ 开始事件 A 发生的时间从什么分布?
记 A 第一次发生的时间为 X , 则 $P(X \leq x) = 1 - e^{-\lambda x}$, 也即是 $f(x) = \lambda e^{-\lambda x}$

伽马分布

对于 $\alpha, \lambda > 0$, 概率密度为 $f(x) = \begin{cases} \frac{x^{(\alpha-1)} \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为伽马分布, 记作 $X \sim \Gamma(\alpha, \lambda)$

■ 这里 $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \alpha > 0$ 且 $\Gamma(n) = (n-1)!$

伽马分布

对于 $\alpha, \lambda > 0$, 概率密度为 $f(x) = \begin{cases} \frac{x^{(\alpha-1)} \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为伽马分布, 记作 $X \sim \Gamma(\alpha, \lambda)$

- 这里 $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \alpha > 0$ 且 $\Gamma(n) = (n-1)!$
- 例: 考虑从 $t=0$ 到任意时间 t 内事件 A 发生的次数服从泊松分布 $\pi(\lambda t)$, 那么从 $t=0$ 开始事件 A 发生 n 次的时间从什么分布?

伽马分布

对于 $\alpha, \lambda > 0$, 概率密度为 $f(x) = \begin{cases} \frac{x^{(\alpha-1)}\lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{else} \end{cases}$ 的随机变量 X 为伽马分布, 记作 $X \sim \Gamma(\alpha, \lambda)$

- 这里 $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \alpha > 0$ 且 $\Gamma(n) = (n-1)!$
- 例: 考虑从 $t=0$ 到任意时间 t 内事件 A 发生的次数服从泊松分布 $\pi(\lambda t)$, 那么从 $t=0$ 开始事件 A 发生 n 次的时间从什么分布?

记 A 发生 n 次的时间为 X , 则

$$P(X \leq x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}, \text{ 也即是}$$

$$f(x) = F'(x) = \frac{x^{(n-1)}\lambda^n e^{-\lambda x}}{(n-1)!}$$

不同分布的关系



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

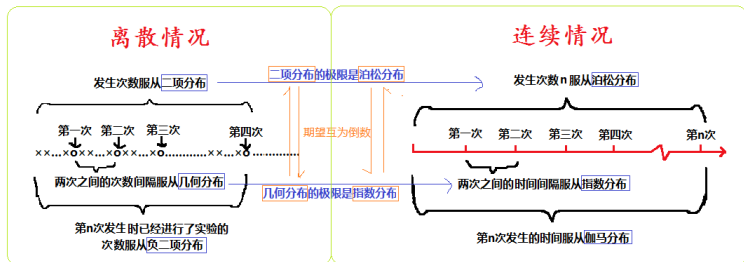
不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量



制作者：武辰 2018

Figure: 二项分布、几何分布、负二项分布、泊松分布、指数分布、伽马分布的关系 (from <https://zhuanlan.zhihu.com/p/32932782>)

正态分布

对于实数 μ 和 $\sigma > 0$, 称概率密度为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 的随机变量满足为正态 (**Guass**) 分布, 记作 $X \sim N(\mu, \sigma^2)$

正态分布

对于实数 μ 和 $\sigma > 0$, 称概率密度为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 的随机变量满足为正态 (**Guass**) 分布, 记作 $X \sim N(\mu, \sigma^2)$

- 归一性: $\int f(x) dx = 1$. 首先 $\int f(x) dx = \frac{1}{\sqrt{2\pi}} \int e^{-t^2/2}$. 记 $\alpha = \int e^{-t^2/2}$, 注意到 $\alpha^2 = 2\pi$ 得证。
- 对称性: $f(x)$ 关于 $x = \mu$ 对称
- 最大值: $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$
- $\lim_{|x| \rightarrow \infty} f(x) = 0$
- $N(0, 1)$ 为标准正态分布

正态分布的性质



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 称 μ 为位置参数 (决定对称轴位置). 当固定 σ 时, 改变 μ 只会平移正态分布的位置
- σ 为尺度参数 (决定曲线分散性). 当固定 μ 时, σ 越大, 曲线的峰越低, 落在 μ 附近的概率越小, 取值就越分散, 即 σ 是反映 X 的取值分散性的一个指标。

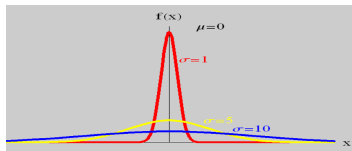
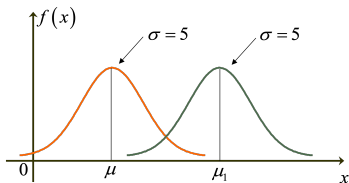


Figure: 正态分布的参数性质

标准正态分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

标准正态分布

$N(0, 1)$ 的分布为标准正态分布，且有概率密度

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ 和分布函数 } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

标准正态分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

标准正态分布

$N(0, 1)$ 的分布为标准正态分布，且有概率密度

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ 和分布函数 } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

■ 注意 $\Phi(x) + \Phi(-x) = 1$

标准正态分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

标准正态分布

$N(0, 1)$ 的分布为标准正态分布，且有概率密度

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ 和分布函数 } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

- 注意 $\Phi(x) + \Phi(-x) = 1$
- 对于随机变量 $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim N(0, 1)$

标准正态分布

$N(0, 1)$ 的分布为标准正态分布，且有概率密度

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ 和分布函数 } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

- 注意 $\Phi(x) + \Phi(-x) = 1$
- 对于随机变量 $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim N(0, 1)$
- 对于随机变量 $X \sim N(\mu, \sigma^2)$, 我们有
$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

标准正态分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

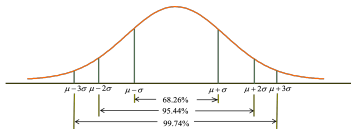
连续随机变量

标准正态分布

$N(0, 1)$ 的分布为标准正态分布，且有概率密度

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ 和分布函数 } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

- 注意 $\Phi(x) + \Phi(-x) = 1$
- 对于随机变量 $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim N(0, 1)$
- 对于随机变量 $X \sim N(\mu, \sigma^2)$, 我们有
$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$
- 一些常用的结论有: $\Phi(1) - \Phi(-1) = 0.6826$,
 $\Phi(2) - \Phi(-2) = 0.9544$, $\Phi(3) - \Phi(-3) = 0.9974$



例子

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：一批钢材长度 $X(\text{cm}) \sim N(\mu, \sigma)$ (1) 若 $\mu = 100$, $\sigma = 2$, 求这批钢材长度小于 97.8cm 的概率; (2) 若 $\mu = 100$, 要使这批钢材的长度至少有 90% 落在区间 (97, 103) 内, 问 σ 至多取何值?

- 例：一批钢材长度 $X(\text{cm}) \sim N(\mu, \sigma)$ (1) 若 $\mu = 100$, $\sigma = 2$, 求这批钢材长度小于 97.8cm 的概率; (2) 若 $\mu = 100$, 要使这批钢材的长度至少有 90% 落在区间 (97, 103) 内, 问 σ 至多取何值?
(1) $P(X < 97.8) = 0.1357$ (2) $\sigma \leq 1.8237$
- 例：设一天中经过一高速公路某一入口的重型车辆数 X 近似服从 $N(\mu, \sigma^2)$, 已知有 25% 的天数超过 400 辆, 有 33% 的天数不到 350 辆, 求 μ, σ

- 例：一批钢材长度 $X(\text{cm}) \sim N(\mu, \sigma)$ (1) 若 $\mu = 100$, $\sigma = 2$, 求这批钢材长度小于 97.8cm 的概率; (2) 若 $\mu = 100$, 要使这批钢材的长度至少有 90% 落在区间 (97, 103) 内, 问 σ 至多取何值?
(1) $P(X < 97.8) = 0.1357$ (2) $\sigma \leq 1.8237$
- 例：设一天中经过一高速公路某一入口的重型车辆数 X 近似服从 $N(\mu, \sigma^2)$, 已知有 25% 的天数超过 400 辆, 有 33% 的天数不到 350 辆, 求 μ, σ
 $\mu \approx 369.7, \sigma \approx 44.8$

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

1 随机变量

2 离散随机变量

- 0-1 分布
- 二项分布
- 泊松分布
- 几何分布和负二项分布

3 连续型随机变量

- 分布函数
- 概率密度
- 均匀分布
- 指数分布
- 伽马分布
- 不同分布的关系
- 正态分布

4 随机变量的函数分布

- 离散随机变量
- 连续随机变量

离散随机变量的函数分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

问题

问题：已知随机变量 X 的概率分布，且已知 $Y = g(X)$ ，求 Y 的概率分布。

离散随机变量的函数分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

问题

问题：已知随机变量 X 的概率分布，且已知 $Y = g(X)$ ，求 Y 的概率分布。

- 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布
 $p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = 2X$ 服从什么分布？

离散随机变量的函数分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

问题

问题：已知随机变量 X 的概率分布，且已知 $Y = g(X)$ ，求 Y 的概率分布。

■ 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布

$p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = 2X$ 服从什么分布？

$Y \in \{-2, 0, 2\}$ ， $P(Y = -2) = P(X = -1) = 0.2$ ，

$P(Y = 0) = P(X = 0) = 0.5$ ， $P(Y = 2) = P(X = 1) = 0.3$

离散随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

问题

问题：已知随机变量 X 的概率分布，且已知 $Y = g(X)$ ，求 Y 的概率分布。

■ 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布

$p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = 2X$

服从什么分布？

$Y \in \{-2, 0, 2\}$ ， $P(Y = -2) = P(X = -1) = 0.2$ ，

$P(Y = 0) = P(X = 0) = 0.5$ ， $P(Y = 2) = P(X = 1) = 0.3$

离散随机变量函数的概率分布

对于随机变量 X ，若 $Y = g(X)$ 为离散随机变量，则可以首先找出 Y 的所有可能值，再找出每个值对应的等价事件来求出概率

离散随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布
 $p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = X^2$
服从什么分布？

离散随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布
 $p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = X^2$
服从什么分布？
 $Y \in \{0, 1\}$ ， $P(Y = 0) = P(Y = 1) = 0.5$

离散随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布 $p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = X^2$ 服从什么分布？
 $Y \in \{0, 1\}, P(Y = 0) = P(Y = 1) = 0.5$
- 例：已知离散型随机变量 X 服从的泊松分布 $\pi(\lambda)$ ，求 $Y = (X - 1)^2$ 服从什么分布？

离散随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布 $p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = X^2$ 服从什么分布？
 $Y \in \{0, 1\}$, $P(Y = 0) = P(Y = 1) = 0.5$
- 例：已知离散型随机变量 X 服从的泊松分布 $\pi(\lambda)$ ，求 $Y = (X - 1)^2$ 服从什么分布？
 $X = 0, Y = 1, X = 1, Y = 0, X = 2, Y = 1, X = 3, Y = 4,$
...
因此， $P(Y = 0) = P(X = 1) = e^{-1}\lambda$,
 $P(Y = 1) = P(X = 0) + P(X = 2) = e^{-1}(1 + \lambda^2/2)$,
 $P(Y = (x - 1)^2) = P(X = x), \forall x \geq 3.$

离散随机变量的函数分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 $X \in \{-1, 0, 1\}$ 满足分布 $p(-1) = 0.2, p(0) = 0.5, p(1) = 0.3$ ，则随机变量 $Y = X^2$ 服从什么分布？
 $Y \in \{0, 1\}$, $P(Y = 0) = P(Y = 1) = 0.5$
- 例：已知离散型随机变量 X 服从的泊松分布 $\pi(\lambda)$ ，求 $Y = (X - 1)^2$ 服从什么分布？
 $X = 0, Y = 1, X = 1, Y = 0, X = 2, Y = 1, X = 3, Y = 4,$
...
因此， $P(Y = 0) = P(X = 1) = e^{-1}\lambda$,
 $P(Y = 1) = P(X = 0) + P(X = 2) = e^{-1}(1 + \lambda^2/2)$,
 $P(Y = (x - 1)^2) = P(X = x), \forall x \geq 3$.
- 例：已知 $X \sim N(0, 1)$, $Y = [|X|]$ ，求 Y 的概率分布。

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求 $Y = |X|$ 的概率密度

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

■ 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = |X|$ 的概率密度

$Y \in (0, 4)$ ，同时 $f_Y(Y = y) = f_X(X = -y) + f_X(X = y) = 1/4$ ，因此

有 $f_Y(y) = \begin{cases} 1/4 & y \in (0, 4) \\ 0 & \text{else} \end{cases}$

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = |X|$ 的概率密度

$Y \in (0, 4)$ ，同时 $f_Y(Y = y) = f_X(X = -y) + f_X(X = y) = 1/4$ ，因此

$$\text{有 } f_Y(y) = \begin{cases} 1/4 & y \in (0, 4) \\ 0 & \text{else} \end{cases}$$

考虑任意概率密度 $f_X(x)$ 的情况。

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = 2X$ 的概率密度

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = |X|$ 的概率密度

$Y \in (0, 4)$ ，同时 $f_Y(Y=y) = f_X(X=-y) + f_Y(X=y) = 1/4$ ，因此

$$\text{有 } f_Y(y) = \begin{cases} 1/4 & y \in (0, 4) \\ 0 & \text{else} \end{cases}$$

考虑任意概率密度 $f_X(x)$ 的情况。

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = 2X$ 的概率密度

同理， $Y \in (-8, 8)$ ，因此有 $f_Y(Y=y) = f_X(X=y/2) = 1/8$ ，因此有

$$f_Y(y) = \begin{cases} 1/8 & y \in (-8, 8) \\ 0 & \text{else} \end{cases} ?$$

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = |X|$ 的概率密度

$Y \in (0, 4)$ ，同时 $f_Y(Y=y) = f_X(X=-y) + f_Y(X=y) = 1/4$ ，因此

$$\text{有 } f_Y(y) = \begin{cases} 1/4 & y \in (0, 4) \\ 0 & \text{else} \end{cases}$$

考虑任意概率密度 $f_X(x)$ 的情况。

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = 2X$ 的概率密度

同理， $Y \in (-8, 8)$ ，因此有 $f_Y(Y=y) = f_X(X=y/2) = 1/8$ ，因此有

$$f_Y(y) = \begin{cases} 1/8 & y \in (-8, 8) \\ 0 & \text{else} \end{cases} ?$$

连续随机变量函数的概率密度

对于连续随机变量 X 的概率密度为 $f_X(x)$ ，若 $Y = g(X)$ 为连续型随机变量，则 Y 的概率密度为 $f_Y(y) = \sum_{g(x)=y} f_X(x)$?

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

■ 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = 2X$ 的概率密度

同理， $Y \in (-8, 8)$ ，因此有 $f_Y(Y = y) = f_X(X = y/2) = 1/8$ ，因此有

$$f_Y(y) = \begin{cases} 1/8 & y \in (-8, 8) \\ 0 & \text{else} \end{cases} ?$$

连续随机变量函数的概率密度

对于连续随机变量 X 的概率密度为 $f_X(x)$ ，若 $Y = g(X)$ 为连续型随机变量，则 Y 的概率密度为 $f_Y(y) = \sum_{g(x)=y} f_X(x)$?

为什么上述分析是错误的？

连续随机变量的函数分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

■ 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} 1/8 & x \in (-4, 4) \\ 0 & \text{else} \end{cases}$ ，求

$Y = 2X$ 的概率密度

同理， $Y \in (-8, 8)$ ，因此有 $f_Y(Y = y) = f_X(X = y/2) = 1/8$ ，因此有

$$f_Y(y) = \begin{cases} 1/8 & y \in (-8, 8) \\ 0 & \text{else} \end{cases} ?$$

连续随机变量函数的概率密度

对于连续随机变量 X 的概率密度为 $f_X(x)$ ，若 $Y = g(X)$ 为连续型随机变量，则 Y 的概率密度为 $f_Y(y) = \sum_{g(x)=y} f_X(x)$?

为什么上述分析是错误的？

连续随机变量函数的概率密度

概率可以相加，密度不能直接相加。

连续随机变量的函数分布

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} x/8 & x \in (0, 4) \\ 0 & \text{else} \end{cases}$ ，
求 $Y = X^2$ 的概率密度

连续随机变量的函数分布



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

■ 例：随机变量 X 概率密度为 $f_X(x) = \begin{cases} x/8 & x \in (0, 4) \\ 0 & \text{else} \end{cases}$,

求 $Y = X^2$ 的概率密度

我们注意到概率密度仅仅反应了连续变量在某个无穷小区间 $[x, x + \delta x]$ 的概率 $P(X \in [x, x + \delta x]) = f(x)dx$. 因此考虑随机变量 Y 在 $[y, y + \delta y]$ 区间的概率 $P_Y(Y \in y, y + \delta y) = f_Y(y)dy$, 对应于随机变量 X 在 $I = [\sqrt{y}, \sqrt{y + \delta y}]$, 对应概率

$$P_X(X \in I) = \int_{\sqrt{y}}^{\sqrt{y+\delta y}} \frac{x}{8} dx = \frac{x^2}{16} \Big|_{\sqrt{y}}^{\sqrt{y+\delta y}} = \frac{1}{16} \delta y$$

也即是 $f_Y(y) = \begin{cases} 1/16 & y \in (0, 16) \\ 0 & \text{else} \end{cases}$

连续随机变量的函数分布

设 X 的概率密度为 $f_X(x)$, 函数 $g(x)$ 严格单调且其反函数 $h(y)$ 有连续导数, 则 $Y = g(X)$ 也是一个随机变量, 且概率密度为

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)| & y \in (\alpha, \beta) \\ 0 & \text{else} \end{cases} \quad (4)$$

其中 $\alpha = \min\{g(-\infty), g(+\infty)\}$ 和 $\beta = \max\{g(-\infty), g(+\infty)\}$

连续随机变量的函数分布

设 X 的概率密度为 $f_X(x)$, 函数 $g(x)$ 严格单调且其反函数 $h(y)$ 有连续导数, 则 $Y = g(X)$ 也是一个随机变量, 且概率密度为

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)| & y \in (\alpha, \beta) \\ 0 & \text{else} \end{cases} \quad (4)$$

其中 $\alpha = \min\{g(-\infty), g(+\infty)\}$ 和 $\beta = \max\{g(-\infty), g(+\infty)\}$

proof: 不妨假设 $g(x)$ 严格单调上升, 则 $g'(x), h'(y) > 0$, $\alpha = g(-\infty)$, $\beta = g(+\infty)$. 考虑 $Y \in (\alpha, \beta)$ 的分布函数为

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq h(y)) = F_X(h(y))$$

因此

$$f_Y(y) = F'_Y(y) = f_X(h(y))' \cdot h'(y)$$

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim N(\mu, \sigma^2)$, $Y = aX + b (a \neq 0)$, 求 Y 的概率密度

例子



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim N(\mu, \sigma^2)$, $Y = aX + b (a \neq 0)$, 求 Y 的概率密度

$$y = g(x) = ax + b, g'(x) = a \neq 0, x = h(y) = (y - b)/a,$$

$$\begin{aligned} f_Y(y) &= f_X(h(y)) \cdot |h'(y)| = f_X((y - b)/a) / |a| \\ &= \frac{1}{|a|\sqrt{2\pi}\sigma} e^{-\frac{((y-b)/a - \mu)^2}{2\sigma^2}} \end{aligned}$$

也即是 $Y \sim N(a\mu + b, a^2\sigma^2)$

例子



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim N(\mu, \sigma^2)$, $Y = aX + b (a \neq 0)$, 求 Y 的概率密度

$$y = g(x) = ax + b, g'(x) = a \neq 0, x = h(y) = (y - b)/a,$$

$$\begin{aligned} f_Y(y) &= f_X(h(y)) \cdot |h'(y)| = f_X((y - b)/a) / |a| \\ &= \frac{1}{|a|\sqrt{2\pi}\sigma} e^{-\frac{((y-b)/a - \mu)^2}{2\sigma^2}} \end{aligned}$$

也即是 $Y \sim N(a\mu + b, a^2\sigma^2)$

- 例：设 $X \sim U(-\pi/2, \pi/2)$, 求 $Y = \sin X$ 的概率密度

例子



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim N(\mu, \sigma^2)$, $Y = aX + b (a \neq 0)$, 求 Y 的概率密度

$$y = g(x) = ax + b, g'(x) = a \neq 0, x = h(y) = (y - b)/a,$$

$$\begin{aligned} f_Y(y) &= f_X(h(y)) \cdot |h'(y)| = f_X((y - b)/a) / |a| \\ &= \frac{1}{|a|\sqrt{2\pi}\sigma} e^{-\frac{((y-b)/a-\mu)^2}{2\sigma^2}} \end{aligned}$$

也即是 $Y \sim N(a\mu + b, a^2\sigma^2)$

- 例：设 $X \sim U(-\pi/2, \pi/2)$, 求 $Y = \sin X$ 的概率密度
- $$y = g(x) = \sin x, g'(x) = \cos x > 0, x = h(y) = \arcsin y,$$
- $$h'(y) = 1/\sqrt{1-y^2} \text{ 因此}$$
- $$f_Y(y) = \begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{1-y^2}} & y \in (-1, 1) \\ 0 & \text{else} \end{cases}$$

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim \text{Exp}(\lambda)$, $F(x)$ 为 X 的分布函数，求 $F(x)$ ，以及证明 $Y = F(X) \sim U(0, 1)$

例子



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim \text{Exp}(\lambda)$, $F(x)$ 为 X 的分布函数, 求 $F(x)$, 以及证明 $Y = F(X) \sim U(0, 1)$
- 首先我们有 $Y = F(X) = \begin{cases} 1 - e^{-\lambda X} & X > 0 \\ 0 & X \leq 0 \end{cases}$ 也即是 Y 的取值范围为 $[0, 1]$ 。当 $y \leq 0$, 我们有 $F_Y(y) = 0$, 当 $y \geq 1$ 我们有 $F_Y(y) = 1$ 。对于 $y \in (0, 1)$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(1 - e^{-\lambda X} \leq y), \\ &= P(e^{-\lambda X} \geq 1 - y), \\ &= P(X \leq -\frac{1}{\lambda} \ln(1 - y)), \\ &= 1 - e^{\lambda[-\frac{1}{\lambda} \ln(1 - y)]}, \\ &= y \end{aligned}$$

例子



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

■ 例：设 $X \sim N(0, 1)$, $Y = X^2$, 求 Y 的概率密度

例子



PEKING
UNIVERSITY

概率统计

Xiao Yuan

随机变量

离散随机变量

0-1 分布

二项分布

泊松分布

几何分布和负二项分布

连续型随机变量

分布函数

概率密度

均匀分布

指数分布

伽马分布

不同分布的关系

正态分布

随机变量的函数分布

离散随机变量

连续随机变量

- 例：设 $X \sim N(0, 1)$, $Y = X^2$, 求 Y 的概率密度
已知 $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, Y 的取值范围为 $[0, \infty)$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y), \\ &= 2P(X \in (0, \sqrt{y})), \\ &= 2(P(X \leq \sqrt{y}) - 1/2), \end{aligned} \tag{5}$$

$$\text{因此 } f_Y(y) = f_X(\sqrt{y})y^{-1/2} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$