SCNUSoS

《概率论与数理统计》讲义

EVAN GAO

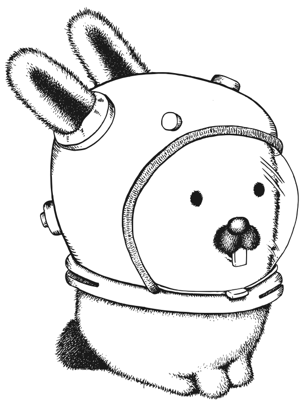
大二上

🚒🔥

Made with XeLaTeX

And Love

evan@notch1p.xyz



Compiled on **D:20231113**

Contents

Chapter 1

随机事件及其概率

1.1 随机事件

事件是样本空间的子集

Definition 1.1.1: 样本空间

考虑样本空间集合 S, 我们有 $S := \{所有样本点\}$.

由定义, 我们可以得到几种特殊的样本空间:

- $\varnothing$ 事件: 不可能发生的事件.
- $S - \varnothing$ 发生的事件.
- 基本事件 ω: $|\omega| = 1$ i.e. 基本事件只含有一个样本点.

Theorem 1.1.1 De Morgan 律

由于 $\varnothing$ 事件和 $S - \varnothing$ 事件是互为对偶的, 我们可以得到

$$\begin{aligned} \overline{A \cup B} &= \overline{A} \cap \overline{B} \\ \overline{A \cap B} &= \overline{A} \cup \overline{B} \end{aligned} \tag{1.1}$$

我们定义**和, 积事件为**

$$\begin{aligned} \bigcup_{i=1}^{n} A_i \text{ happen} &\iff \exists i \in [1, n] \text{ s.t. } A_i \text{ happens} \\ \bigcap_{i=1}^{n} A_i \text{ happen} &\iff \forall i \in [1, n] \text{ s.t. } A_i \text{ happens} \end{aligned} \tag{1.2}$$

1.2 频率

Definition 1.2.1: 频率

考虑事件 A, 其发生的**频率**是

$$f_n(A) \stackrel{\text{def}}{=} \frac{r_n(A)}{n} \in [0, 1]$$

其中频数 $\frac{r_n(A)}{n} \in [0, n]$. 显然有 $f_n(S) = 1$.

Corollary 1.2.1 有限可加性

若 $A_i \cap A_j = \varnothing$ 且 $i \neq j,\ i,j \in [1,k]$ i.e. 互斥事件, 则

$$f_n\left(\bigcap_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} f_n\left(A_i\right)$$

1.3 概率

Definition 1.3.1: 概率

(Kolmogorov 公理化定义) 设有随机试验 E 且与之对应的样本空间 S, 考虑事件 A

$$\text{for } \forall A \in E, \text{if}$$

① $0 \leqslant P(A) \leqslant 1$

② $P(S) = 1$

③ $\Pr\{\bigcup_{i=1}^{\infty} A_i\} = \sum_{i=1}^{\infty} \Pr\{A_i\}$ i.e. **可列可加性**

则称 P 为 S 上的**概率**.

1.3.1 概率的性质

由上面的定义我们能够得到概率的性质:

1. $\Pr\{\varnothing\} = 0$
2. $\Pr\{\bigcup_{i=1}^{n} A_i\} = \sum_{i=1}^{n} \Pr\{A_i\} \iff \forall i,j\ (i \neq j \to\ A_iA_j = \varnothing)$
3. $\Pr\left\{\overline{A}\right\} + \Pr\{A\} = 1$
4. $\Pr\{A - B\} = \Pr\{A\} - \Pr\{AB\} \Rightarrow\ (A - B) \cap B = \varnothing$
5. (单调性)$B \subseteq A \Rightarrow \Pr\{B\} \leqslant \Pr\{A\}$
6. 若满足 5, 由 4 可得 $\Pr\{A - B\} = \Pr\{A\} - \Pr\{B\}$
7. (容斥原理) $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}$

1.4 条件概率

Definition 1.4.1: 条件概率

设 A B 是两个事件, 且 $P(A) \neq 0$, 则称

$$\Pr\{A \mid B\} \overset{\text{def}}{=} \frac{\Pr\{AB\}}{\Pr\{B\}} \tag{1.3}$$

为在事件 A 发生的条件下, 事件 B 的**条件概率**.

Corollary 1.4.1 条件概率之性质

概率满足的性质条件概率都满足.

Theorem 1.4.1
设 $A_1, A_2, \cdots, A_n$ 是 n 个互斥事件, 则有

$$\Pr\left\{\bigcup_{i=1}^{n} A_i | A\right\} = \sum_{i=1}^{n} \Pr\{A_i | A\} \tag{1.4}$$

Question 1: 证明

$$\Pr\left\{\overline{B} | A\right\} = 1 - \Pr\{B | A\} \tag{1.5}$$

Proof:

$$\begin{aligned}
\Pr\left\{\overline{B} | A\right\} &= \frac{\Pr\left\{\overline{B} A\right\}}{\Pr\{A\}} \\
&= \frac{\Pr\{A\} - \Pr\{BA\}}{\Pr\{A\}} \\
&= 1 - \frac{\Pr\{BA\}}{\Pr\{A\}} \\
&= 1 - \Pr\{B | A\}
\end{aligned}$$

☺

1.4.1 乘法公式

由条件概率的定义, 我们可以得到**乘法公式**:

Theorem 1.4.2 乘法公式

$$\begin{aligned}
\Pr\{A_1 A_2 \cdots A_n\} &= \Pr\{A_1\} \Pr\{A_2 | A_1\} \Pr\{A_3 | A_1 A_2\} \cdots \Pr\{A_n | A_1 A_2 \cdots A_{n-1}\} \\
&= \prod_{i=1}^{n} \Pr\left\{A_i | \bigcup_{i=1}^{n-1} A_i\right\}
\end{aligned} \tag{1.6}$$

Example 1.4.1 (“彩票”)
第一次买中的概率为 $\dfrac{1}{2}$, 第二次买中而第一次未中的概率是 $\dfrac{7}{10}$, 第三次买中而前两次未中的概率是 $\dfrac{9}{10}$, 求三次都未中的概率.

Solution: 以 $A_i (i = 1, 2, 3)$ 表示事件 “第 i 次买中”, 以 B 表示事件 “三次都未中”, 那么由(1.6)

$$\begin{aligned}
\because B &= \overline{A_1 A_2 A_3} \\
\therefore \Pr\{B\} &= \Pr\left\{\overline{A_1 A_2 A_3}\right\} \\
&= \Pr\left\{\overline{A_1}\right\} \Pr\left\{\overline{A_2} | \overline{A_1}\right\} \Pr\left\{\overline{A_3} | \overline{A_1 A_2}\right\} \\
&= \left(1 - \frac{1}{2}\right)\left(1 - \frac{7}{10}\right)\left(1 - \frac{9}{10}\right) \\
&= \frac{3}{200}
\end{aligned}$$

Question 2: $P_{28}19$

袋中装有 a 个红球,b 个白球, 每次自袋中有放回地任取一球, 并同时再放入 m 个与之相同的求, 连续如此进行 $2n$ 次, 求前 n 次为红球, 后 $n-1$ 次为白球, 第 $2n$ 次为红球的概率.

Solution: 设事件 A_i 表示第 i 次取出的球为红球,B_i 表示白球, 那么可知所求概率为

$$\Pr\{A_1A_2A_3\cdots A_nB_{n+1}B_{n+2}\cdots B_{2n-1}A_{2n}\}$$

利用(1.6)展开上式, 我们立即有

$$\Pr\{A_1\}\Pr\{A_2|A_1\}\cdots\Pr\{A_n|A_1A_2\cdots A_{n-1}\}\times\Pr\{B_{n+1}|A_1A_2\cdots A_n\}\Pr\{B_{n+2}|\cdots B_{n+1}\}\\ \cdots\Pr\{A_{2n}|A_1\cdots A_nB_{n+1}B_{n+2}\cdots B_{2n-1}\}$$

此式太长, 不妨将其分成三部分:

① $\Pr\{A_1\}\Pr\{A_2|A_1\}\cdots\Pr\{A_n|A_1A_2\cdots A_{n-1}\}$ (前 n 个红球)

② $\Pr\{B_{n+1}|A_1A_2\cdots A_n\}\Pr\{B_{n+2}|\cdots B_{n+1}\}\cdots\Pr\{B_{2n-1}|\cdots B_{2n-2}\}$ (后 n-1 个白球)

③ $\Pr\{A_{2n}|A_1\cdots A_nB_{n+1}B_{n+2}\cdots B_{2n-1}\}$ (最后的红球)

三式相乘, 有

$$\prod_{i=1}^{n}\frac{a+(i-1)m}{a+b+(i-1)m}\prod_{i=n+1}^{2n-1}\frac{b+(i-n-1)m}{a+b+(i-1)m}\times\frac{a+(n-1)m}{a+b+(2n-1)m}$$

即为所求.

1.4.2 全概率公式

Lemma 1.4.1 完备事件组

设 $A_1,A_2,\cdots,A_n$ 是有限或可数个事件, 若其满足 1. 两两互斥, 2. $\bigcup_{i=1}^{n}A_i = S$, 则称 $A_1,A_2,\cdots,A_n$ 是一个**完备事件组**.

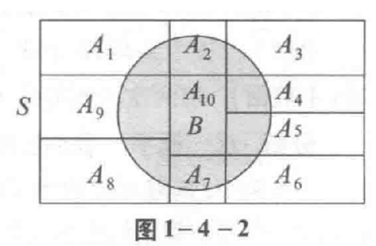


Figure 1.1: 全概公式示意

(1.7)指出在复杂情况下不易计算 $\Pr\{B\}$ 时应该根据具体情况构造完备事件组 $\{A_i\}$, 使得事件 B 发生的概率是各事件 A_i 发生的条件下造成事件 B 发生概率的和.

Theorem 1.4.3 全概公式

设 $A_1,A_2,\cdots,A_n$ 是一个完备事件组, 且
$\forall i(i\in[1,n]\to\Pr\{A_i\}>0)$, 则对任一事件 B, 有

$$\Pr\{B\}=\sum_{i=1}^{n}\Pr\{A_i\}\Pr\{B|A_i\} \tag{1.7}$$

Question 3: $P_{14}10$

从 1 到 9 的整数中有放回地依次随机抽取 3 次, 求取出的 3 个数之积能被 10 整除的概率.

显然这三个数中必有两数为 5 和 $\{2,4,6,8\}$, 因此

法一: 分情况讨论

① A = {三个数里有两个 5 和一个偶数}

② B = {三个数里有一个 5 和两个偶数}

③ C = {三个数里有一个 5 一个偶数和一个其他的奇数}

那么有 $\Pr\{A\} = \dfrac{\binom{4}{1}\cdot 3}{9^3}$ $\Pr\{B\} = \dfrac{\binom{4}{1}\cdot 3+\binom{4}{2}A_3^3}{9^3}$ $\Pr\{C\} = \dfrac{\binom{4}{1}\binom{4}{1}A_3^3}{9^3}$ 相加得 $\dfrac{156}{729}$. ☺

法二: 对立事件

不妨设三个数中出现 5 的事件为 A, 出现偶数的事件为 B, 那么

$$\begin{aligned}\Pr\{AB\} &= 1-\Pr\left\{\overline{AB}\right\}\\ &= 1-\left(\Pr\left\{\overline{A}\right\}+\Pr\left\{\overline{B}\right\}-\Pr\left\{\overline{A}\,\overline{B}\right\}\right)\\ &= 1-\left(\frac{8^3}{9^3}+\frac{5^3}{9^3}-\frac{4^3}{9^3}\right)\\ &= \frac{156}{729}\end{aligned}$$

☺

法三: 另一种分情况讨论 (笔者的分法)

① A = {三个数中有两个相同的}

此项须要分成下述两项, 否则会重

② B = {三个数全不同}

(a) B_1 = {三个数全不同且有两个偶数}

(b) B_2 = {三个数全不同且有两个奇数}

那么有 $\Pr\{A\} = \dfrac{\binom{4}{1}\binom{3}{1}+\binom{4}{1}\binom{3}{1}}{9^3}$ $\Pr\{B_1\} = \dfrac{\binom{4}{1}\binom{4}{1}\binom{3}{1}}{9^3}$ $\Pr\{B_2\} = \dfrac{\binom{4}{1}\binom{4}{1}A_3^3}{9^3}$ 相加得 $\dfrac{156}{729}$. ☺

Note

讨论各种情况的概率进而求得所求事件的概率的方法实际上是**全概率公式**的一种体现.

1.4.3 Beyes 公式

全概公式是通过计算某一事件会发生的所有原因和情况的可能性大小来计算该事件发生的概率, 而**Beyes**公式则与之相反, 考察一件已经发生的事情的各种原因或或情况的可能性大小.

$\Pr\{A_i\}$,$\Pr\{A_i|B\}$ 分别称为原因的**先验概率**和**后验概率**

Theorem 1.4.4 Beyes 公式

设 $A_1, A_2, \cdots, A_n$ 是一个完备事件组, 且 $\forall i(i\in[1,n]\to\Pr\{A_i\}>0)$, 则对任一可能发生的事件 B, 有

$$\Pr\{A_i|B\} = \frac{\Pr\{A_iB\}}{\Pr\{B\}} = \frac{\Pr\{A_i\}\Pr\{B|A_i\}}{\sum_{j=1}^{n}\Pr\left\{A_j\right\}\Pr\left\{B|A_j\right\}} \tag{1.8}$$

特别地, 当 $n=2$ 时,Beyes 公式也可以写成

$$\Pr\{A|B\} = \frac{\Pr\{A\}\Pr\{B|A\}}{\Pr\{B\}} \tag{1.9}$$

1.5 事件的独立性

独立与互斥是两种不同的概念

Definition 1.5.1: 两事件独立性

设 A, B 是两个事件, 若

$$\Pr\{AB\} = \Pr\{A\}\Pr\{B\}$$

则称 A, B**相互独立**.

Theorem 1.5.1

设 A,B 两事件相互独立且 $\Pr\{B\} > 0$, 则有

$$\Pr\{A|B\} = \Pr\{A\}$$

Theorem 1.5.2

若 A,B 两事件相互独立, 则它们对立事件和其本身 (不同事件间) 的组合也相互独立.

根据三个事件 (略) 独立性的定义, 可以类推到 n 个事件的独立性: 设 $A_1, A_2, \cdots, A_n$ 是 n 个事件, 若对于其中任意 $k(2 \leqslant k \leqslant n)$ 个事件 $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$, 有

$$\Pr\left\{A_{i_1}A_{i_2}\cdots A_{i_k}\right\} = \Pr\left\{A_{i_1}\right\}\Pr\left\{A_{i_2}\right\}\cdots\Pr\left\{A_{i_k}\right\} \tag{1.10}$$

则称 $A_1, A_2, \cdots, A_n$**相互独立**.

Definition 1.5.2

设 $A_1, A_2, \cdots, A_n$ $(n > 2)$ 是 n 个事件, 若其中任意两个事件相互独立, 则称这 n 个事件**两两独立**.

由此可以得到多个独立事件所具备的性质

Corollary 1.5.1

设 $A_1, A_2, \cdots, A_n$ $(n > 2)$ 相互独立, 则其中任意 $k(2 \leqslant k \leqslant n)$ 个 (它们的对立) 事件也相互独立.

1.6 Bernoulli 试验

即两点分布.

这相当于在实验结果序列中任取 k 次发生 A 事件

Theorem 1.6.1 Bernoulli 定理

一次试验中, 事件 A 发生的概率为 p, 进行这样的试验 n 次, 事件 A 发生 k 次的概率为

$$b(k; n; p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Note

n 次 Bernoulli 试验的概率分布就是二项分布. 其概率密度函数为

$$f(k, n, p) = \Pr(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Glossary: 两点分布 =Bernoulli 试验; 二项分布 = 多次 Bernoulli 试验 =Bernoulli 概型.

辨析: $\Pr\{A\}$ 和 $\Pr(X)$ 使用的括号不同. 大括号代表是事件, 即一系列样本点集合 (故而用大括号), 小括号代表是随机变量 (因为此时是概率密度函数). ☺

Chapter 2

随机变量及其分布

2.1 随机变量

Definition 2.1.1: 随机变量

设随机试验的样本空间为 S, 称定义在 S 上的实值单值函数 $X = X(\omega)$ 为**随机变量**.
亦即

$$X : \omega \mapsto x_i \in \mathbb{R} \text{ i.e. } S \to \mathbb{R}$$

实际上,$\Pr(X = a)$ 是 $A = \{\omega | X(\omega) = a\}$ 的简记.

2.2 离散型随机变量及其分布

设 X 是一个随机变量, 如果 X 的所有可能取值为有限个或可列无限多个, 则称 X 为**离散型随机变量**.

Definition 2.2.1: 离散型随机变量分布律

设 X 是一个离散型随机变量, 如果

$$\Pr(X = x_k) = p_k, k = 1, 2, \cdots$$

则称 p_k 为 X 的**分布律**或**概率分布**, 亦称**概率质量函数**(PMF). 记为 $f_X(x)$.

就是高中所学之分布列.

2.2.1 常用离散分布

两点分布

Definition 2.2.2: 两点分布

设 X 的分布律为

$$\Pr(X = k) = \begin{cases} p, & k = x_1 \\ 1 - p, & k = x_2 \end{cases} \tag{2.1}$$

其中 $0 < p < 1$, 则称 X 服从以 p 为参数的**两点分布**.

若 X 服从 $x_1 = 1, x_2 = 0$ 处参数为 p 的两点分布, 则称其服从参数为 p 的**0-1 分布**.

二项分布

Definition 2.2.3: 二项分布

当 $n=1$ 时, 二项分布退化为 0-1 分布.

设 X 的分布律为

$$\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\cdots,n \tag{2.2}$$

其中 n 为正整数,$0<p<1$, 则称 X 服从参数为 n,p 的**二项分布**, 记为 $X\sim B(n,p)$.

$\sim$ 记号就是**服从**的意思

二项概率总存在一个最大值 M s.t. $(n+1)p-1\leqslant M<(n+1)p$.

Poisson 分布

Definition 2.2.4: Poisson 分布

设 X 的分布律为

$$\Pr(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,\cdots \tag{2.3}$$

其中 $\lambda>0$, 则称 X 服从参数为 λ 的**Poisson 分布**, 记为 $X\sim P(\lambda)$.

因为 $e^x=1+\sum_{n=1}^{\infty}\frac{x^n}{n!}$, 将 x 换为 λ 可得泊松分布的概率和为 1.

2.3 累积分布函数

2.3.1 随机变量的分布函数

Definition 2.3.1: 随机变量 CDF

设 X 是一个随机变量, 对任意实数 x, 定义

$$F(x)\stackrel{\text{def}}{=}\Pr(X\leqslant x)\in[0,1]$$

称 $F(x)$ 为 X 的**累积分布函数**(CDF). 有时记作 $F_X(x)$ 或 $X\sim F(x)$.

分布函数 $F(x)$ 的值就表示 X 落在区间 $(-\infty,x]$ 的概率.

性质

1. 单调非减性: $x_1<x_2\Rightarrow F(x_1)\leqslant F(x_2)$
2. $F(-\infty)=\lim_{x\to-\infty}F(x)=0,\ F(+\infty)=\lim_{x\to+\infty}F(x)=1.$
3. 右连续性: $\lim_{x\to x_0^+}F(x)=F(x_0)$

具有上述性质的函数一定是某个随机变量的 CDF.

2.3.2 离散型随机变量的分布函数

在上述基础上, 对于离散型随机变量 X 有 CDF:

$$F(x) = \Pr(X \leqslant X) = \sum_{x_i \leqslant x} \Pr(X = x_i)$$

$F(x)$ 是一个**阶梯型函数**, 在点 x_i 处的跃度为 $\Pr(X = x)$. 可以看出阶梯型函数符合 CDF 之性质: 总是右连续的; 而作为一个分段函数来说, 其各段区间总是 (习惯上) 左闭右开的.

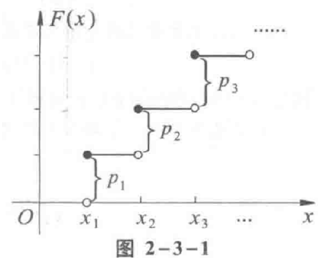


Figure 2.1: 阶梯型 CDF

> **Note**
> 考题中由 CDF 反推分布列可以根据每一个左侧端点得到对应的随机变量 x_i.
> 计算分布列时, 有公式 $\Pr\{a \leqslant X < b\} = F(b) - F(a)$.

随机变量 X 的 CDF 为阶梯型函数 $\Leftrightarrow$ X 是离散型随机变量.

2.4 连续型随机变量及其概率密度

Definition 2.4.1: PDF

设 X 是一个随机变量, 如果存在非负可积函数 $f(x)$, 使对任意实数 x 有

$$F(x) = \Pr\{X \leqslant x\} = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \tag{2.4}$$

则称 X 为**连续型随机变量**,$f(x)$ 为 X 的**概率密度函数**(PDF).

立即可以得到 PDF 的性质

1. $f(x) \geqslant 0$;
2. $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$.
3. $F'(x) = f(x)$. 若 $f(x)$ 在 x 处连续

> **Note**
> 若某个随机变量的 CDF 可以写成 PDF 积分函数的形式, 则称其为连续型随机变量.

证明 X 取任一特值的概率为 0:

$$\text{考虑极限} \Pr\{X = a\} = \lim_{\Delta x \to 0^+} \Pr\{a - \Delta x \leqslant X \leqslant a\}$$

对 RHS 积分, 有

$$\begin{aligned}\Pr\{X = a\} &= \lim_{\Delta x \to 0^+} \int_{a-\Delta x}^{a} f(x)\,\mathrm{d}x \\ &= 0.\end{aligned}$$

☺

由上面的证明我们可以知道以下的性质

$$\Pr\{a < X < b\} = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x\,. \tag{2.5}$$

这个等式无论 X 两侧端点开闭都成立.

2.4.1 常用连续型分布

均匀分布

Definition 2.4.2: 均匀分布

设 X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & otherwise\end{cases} \tag{2.6}$$

U for Uniform

其中 $a<b$, 则称 X 服从参数为 a,b 的**均匀分布**, 记为 $X\sim U(a,b)$.

设 $[a,b]$ 之间的任意子区间为 $[c,d]$, 由上式可得 $\Pr\{c<X<d\}=\int_c^d\frac{1}{b-a}\,\mathrm{d}x=\frac{d-c}{b-a}$. 因而我们可以得到

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\leqslant x<b\\ 1, & x\geqslant b\end{cases} \tag{2.7}$$

指数分布

Definition 2.4.3: 指数分布

若随机变量 X 的概率密度为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\leqslant 0\end{cases},\lambda>0 \tag{2.8}$$

则称 X 服从参数为 λ 的**指数分布**, 记为 $X\sim E(\lambda)$.

积分得

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\leqslant 0\end{cases} \tag{2.9}$$

正态分布

Definition 2.4.4: 正态分布

若随机变量 X 的概率密度为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ (-\infty<x<+\infty) \tag{2.10}$$

其中 $\mu,\sigma(\sigma>0)$ 为常数, 则称 X 服从参数为 μ,σ^2 的**正态分布**, 记为 $X\sim N(\mu,\sigma^2)$.

Question 4: Poisson 积分

不使用 Wallis 公式, 求 Poisson 积分 $\displaystyle\int_{-\infty}^{+\infty}e^{-t^2}\,\mathrm{d}t$.

Solution: 转化为求累次积分

要求原式, 不妨先求其平方, 亦即

$$\left(\int_{-\infty}^{+\infty}e^{-t^2}\,\mathrm{d}t\right)^2=\int_{-\infty}^{+\infty}e^{-x^2}\,\mathrm{d}x\int_{-\infty}^{+\infty}e^{-y^2}\,\mathrm{d}y$$

化为累次积分有

$$=\iint_{\mathbb{R}^2}e^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y\,.$$

$\mathbb{R}^2$ 即 xOy 面, 化为极坐标有

$$\iint_{\mathbb{R}^2} e^{-\rho^2}\rho\,\mathrm{d}\rho\,\mathrm{d}\theta = \int_0^{2\pi}\mathrm{d}\theta\int_0^{+\infty} e^{-\rho^2}\rho\,\mathrm{d}\rho$$
$$= \pi$$
$$\therefore \int_{-\infty}^{+\infty} e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}. \tag{2.11}$$

通过计算 Poisson 积分, 我们可以得到正态分布的 PDF 重要性质:

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = 1 \tag{2.12}$$

图形特征

σ 越小, 方差越大, 曲线中峰越陡

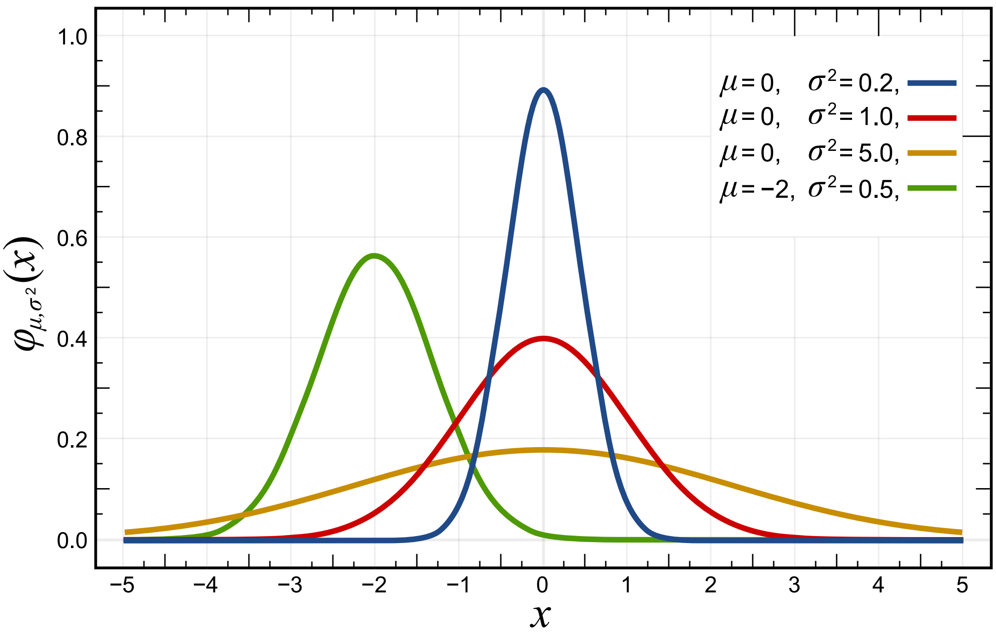


Figure 2.2: PDF

1. μ 确定曲线位置, σ 确定曲线中峰的陡峭程度
2. 密度曲线关于 $x=\mu$ 对称
3. 密度曲线在 $x=\mu$ 处有最大值 $\frac{1}{\sqrt{2\pi}\sigma}$
4. 密度曲线在 $x=\mu\pm\sigma$ 处有拐点且以 x 轴为渐近线

当 $\mu=0,\sigma=1$ 时, 称为**标准正态分布**, 记为 $X\sim N(0,1)$,PDF 用 $\varphi(x)$ 表示,CDF 用 $\Phi(x)$ 表示:

显然 $\varphi(x)$ 为偶函数.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \qquad\qquad \Phi(x) = \int_{-\infty}^{x}\varphi(t)\,\mathrm{d}t \tag{2.13}$$

Note

设 $X\sim N(\mu,\sigma^2)$, 则 $Y=\frac{X-\mu}{\sigma}\sim N(0,1)$.
计算 $\Pr\{Y\leqslant x\}$ 时, 可转化为计算 $\Pr\{X\leqslant \mu+\sigma x\}$, 计算此时的 $\Phi(x)$ 即得证.

这说明任意一般的正态分布都可以通过**线性变换**转化为标准正态分布. 因此 X 的分布函数也可以写成

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\Pr\{a<X<b\} = \Pr\left\{\frac{a-\mu}{\sigma}<Y<\frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

2.5 随机变量函数的分布

前面提到计算 $\Pr\{Y \leqslant x\}$ 时, 可转化为计算 $\Pr\{X \leqslant \mu + \sigma x\}$, 其中的 $Y = \frac{X-\mu}{\sigma}$ 是随机变量 X 的函数, 因此 Y 也是随机变量.

Definition 2.5.1: 随机变量函数

若存在一个函数 $g(x)$ s.t. 随机变量 X, Y 满足

$$Y = g(X)$$

称 Y 为 X 的一个**随机变量函数**.

2.5.1 离散型随机变量函数的分布

根据上述定义, 显然 X 的函数 Y 也是一个随机变量.

导出 Y 的分布律

首先确定 Y 的所有取值, 通过 Y 的每一个值 $y_i\ (i = 1, 2, \cdots)$ 确定相应的
(注: 我们用 C_i 来表示 X 的取值为 x_i 时, 与对应 Y 匹配的 X 的取值的集合)

$$C_i = \{x_i \,|\, g(x_j) = y_i\},$$

$$\{Y = y_i\} = \{X \in C_i\},$$

那么有

$$\Pr\{Y = y_i\} = \Pr\{X \in C_i\} = \sum_{x_j \in C_i} \Pr\left\{X = x_j\right\} \tag{2.14}$$

为 Y 的分布律. 这表明:**Y 的分布律完全由 X 的分布律确定**.

2.5.2 连续型随机变量函数的分布

分布律: 离散变量
PDF: 连续变量

通过上述定义, 不难发现连续随机变量的函数不一定是连续型随机变量.

Note

考虑符号函数作为映射法则:

$$Y = \operatorname{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1 & x < 0 \end{cases}$$

显然 $Y = 0, \pm 1$, 是一个离散型随机变量. 因此, 在连续型的情况下, Y 的连续性由其自身决定.

导出 Y 的 PDF

已知 X 的 CDF $F_X(x)$ 或 PDF $f_X(x)$, 则 $Y = g(x)$ 的 CDF 可以通过以下公式求得:

$$F_Y(y) = \Pr\{Y \leqslant y\} = \Pr\{g(X) \leqslant y\} = \Pr\left\{X \in C_y\right\}, \text{其中 } C_y = \{x \,|\, g(x) \leqslant y\}. \tag{2.15}$$

之后通过积分 $\int_{C_y} f_X(x)\,\mathrm{d}x$ 求得 CDF.

Exercise 2.5.1 离散变量函数导出 PMF

给出随机变量 X 的分布律如下, 试求 $Y = X^2$ 的分布律:

X	-2	-1	0	1	3
p_i	$1/5$	$1/6$	$1/5$	$1/15$	$11/30$

Solution: 由题意得 $Y = 0, 1, 4, 9,$, 因此利用(2.14)计算得

$$\Pr\{Y=0\} = \Pr\{X=0\} = \frac{1}{5} \qquad \Pr\{Y=1\} = \Pr\{X=\pm 1\} = \frac{7}{30}$$
$$\Pr\{Y=4\} = \Pr\{X=-2\} = \frac{1}{5} \qquad \Pr\{Y=9\} = \Pr\{X=3\} = \frac{11}{30}$$

$\therefore$

Y	0	1	4	9
p_i	1/5	7/30	1/5	11/30

Exercise 2.5.2 连续变量函数导出 PDF

设 X 服从标准正态分布, 求 $Y = 2X^2 + 1$ 的 PDF

Solution: 要求 PDF, 不妨先求 Y 的 CDF $F_Y(y)$ i.e. $\Pr\{Y \leqslant y\}$, 代入有

$$F_Y(y) = \Pr\left\{2X^2 + 1 \leqslant y\right\} = \Pr\left\{\sqrt{-\frac{y-1}{2}} \leqslant X \leqslant \sqrt{\frac{y-1}{2}}\right\}$$
$$= \begin{cases} \int_{-\sqrt{\frac{y-1}{2}}}^{\sqrt{\frac{y-1}{2}}} f_X(x)\,\mathrm{d}x & y > 1 \\ 0 & \text{otherwise} \end{cases}$$
$$= \begin{cases} F_X\left(\sqrt{\frac{y-1}{2}}\right) - F_X\left(-\sqrt{\frac{y-1}{2}}\right) & y > 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore f_Y(y) = F_Y'(y) = \begin{cases} \sqrt{2}\,\frac{f_X\left(\sqrt{\frac{y-1}{2}}\right)}{4\sqrt{y-1}} + \sqrt{2}\,\frac{f_X\left(-\sqrt{\frac{y-1}{2}}\right)}{4\sqrt{y-1}} & y > 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \ (2.14)$$ 注意到 $\varphi(x)$ 为偶函数, 合并上述两项

$$\Longrightarrow f_Y(y) = \begin{cases} \frac{1}{2\sqrt{\pi(y-1)}} e^{-\frac{y-1}{4}} & y > 1 \\ 0 & \text{otherwise} \end{cases}$$

Chapter 3

多维随机变量及其分布

3.1 二维随机变量及其分布

3.1.1 二维随机变量

Definition 3.1.1: 二维随机变量

设随机试验的样本空间为 S, 定义在 S 上的实值单值函数 $X = X(\omega)$ $Y = Y(\omega)$ 为两个随机变量. 称向量 (X, Y) 为定义在 S 上的**二维随机变量**

$$\text{i.e.}\quad \begin{aligned} X &: S \to \mathbb{R} \\ Y &: S \to \mathbb{R} \\ (X,Y) &: S \to \mathbb{R}^2 \text{ i.e. } \omega \mapsto (X(\omega), Y(\omega)) \end{aligned}$$

类似地, 我们可以得到 n 维随机变量的定义.

3.1.2 二维随机变量的分布函数

Definition 3.1.2: 二维随机变量的分布函数

设 (X, Y) 是二维随机变量, 对于任意实数 x, y, 定义二元函数

$$F(x, y) = \Pr\{X \leqslant x \cap Y \leqslant y\} \overset{\text{def}}{=} \Pr\{X \leqslant x, Y \leqslant y\} \tag{3.1}$$

称为二维随机变量 (X, Y) 的**分布函数**, 或**联合分布函数 (Joint CDF)**.

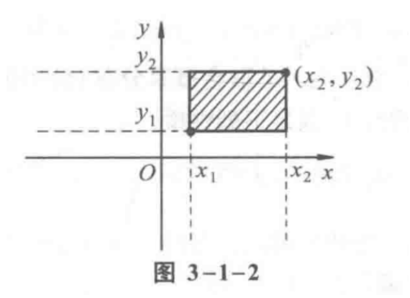


Figure 3.1: CDF

将这个向量视作平面内随机点的坐标, 分布函数就是随机点 (X, Y) 落入矩形域的概率. 通过容斥原理计算得到

$$\Pr\{x_1 \leqslant X \leqslant x_2, y_1 \leqslant Y \leqslant y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \tag{3.2}$$

联合 CDF 的性质

1. $0 \leqslant F(x,y) \leqslant 1$

右侧几个等式可通过矩形域的面积来理解.

   (a) 固定 y,$F(-\infty, y) = 0$;
   (b) 固定 x,$F(x, -\infty) = 0$;
   (c) $\lim_{\substack{x\to+\infty\\y\to+\infty}} F(x,y) = 1, \lim_{\substack{x\to-\infty\\y\to-\infty}} F(x,y) = 0$

2. (单调性)$F(x,y)$ 关于 x 和 y 都是单调不减的.
   (a) 固定 y,$F(x,y)$ 关于 x 单调不减;
   (b) 固定 x,$F(x,y)$ 关于 y 单调不减.
3. (连续性)$F(x,y)$ 关于 x 和 y 都是右连续的.

若已知 (X,Y) 的分布函数 $F(X,Y)$, 则可由之导出各个参数 (在固定另一个参数的情况下) 各自的分布函数:

$$F_X(X) = \Pr\{X \leqslant x\} = \Pr\{X \leqslant x\} = \lim_{y\to+\infty} F(x,y), \tag{3.3}$$

$$F_Y(y) = \Pr\{Y \leqslant y\} = \Pr\{Y \leqslant y\} = \lim_{x\to+\infty} F(x,y), \tag{3.4}$$

如此导出的 CDF 称为**边缘分布函数**.

3.1.3 二维离散型随机变量及其分布律

Definition 3.1.3: 联合概率分布律

设 (X,Y) 是二维离散型随机变量, 若存在非负函数 p_{ij}, 使得

$$\Pr\left\{X = x_i, Y = y_j\right\} = p_{ij}, \quad i,j = 1,2,\cdots \tag{3.5}$$

则称 (X,Y) 为二维离散型随机变量, 并称 p_{ij} 为 (X,Y) 的**联合概率分布律**.

显然,p_{ij} 满足性质: 1. $p_{ij} \geqslant 0$; 2. $\sum_i \sum_j p_{ij} = 1$. 利用其分布律, 易见 (X,Y) 在 D 上的概率为

$$\Pr\{(X,Y) \in D\} = \sum_{(x_i,y_j)\in D} p_{ij} \tag{3.6}$$

而其联合 CDF 为

$$F(x,y) = \Pr\{X \leqslant x, Y \leqslant y\} = \sum_{x_i\leqslant x,\, y_j\leqslant y} p_{ij} \tag{3.7}$$

边缘分布律

通过联合概率分布, 我们可以得到 X, Y 各自的概率分布:

即行和, 列和

$$p_{i\cdot} = \Pr\{X = x_i\} = \sum_j p_{ij} \tag{3.8}$$

$$p_{\cdot j} = \Pr\left\{Y = y_j\right\} = \sum_i p_{ij} \tag{3.9}$$

这称为 (X,Y) 的**边缘分布**.

Question 5: 例 1

设随机整数变量 $X = 1,2,3,4$,$Y = 1\cdots X$ 等可能地取值, 试求 (X,Y) 的分布律.

Solution:

$$\text{由题意可得}\Pr\{X=i,Y=j\}=\Pr\{Y=j|X=i\}\Pr\{X=i\}(1.6)$$
$$=\frac{1}{i}\cdot\frac{1}{4}, i=1,2,3,4, j\leqslant i.$$

X,Y	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

3.1.4 二维连续型随机变量及其概率密度

Definition 3.1.4: 二维连续型随机变量

设 (X,Y) 是二维随机变量,$F(X,Y)$ 为其分布函数, 若存在非负可积函数 $f(x,y)$ s.t.

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\,\mathrm{d}u\,\mathrm{d}v$$

则称 (X,Y) 为**二维连续型随机变量**, 称 $f(x,y)$ 为其**概率密度**

显然,$f(x,y)$ 满足性质: ①$f(x,y)\geqslant 0$;②$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=F(+\infty,+\infty)=1$.

重要性质 设 D 是 xOy 平面上的区域,(X,Y) 落入其中的概率为

$$\Pr\{(X,Y)\in D\}=\iint_D f(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{3.10}$$

若 $f(x,y)$ 在 (x,y) 处连续, 则有

$$\frac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y). \tag{3.11}$$

边缘概率密度

我们知道此处的 X 和 Y 都是都是连续型随机变量, 可得它们的**边缘密度**为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}y\,; \qquad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}x\,; \tag{3.12}$$

结合之前的性质, 我们可以知道对 $F(x,y)$ 求一个变量的偏导可以得到另一个变量的边缘密度.

Exercise 3.1.1 习题 3-1-6 改

设随机变量 (X,Y) 的概率密度为

通常都是分段函数, 且一段为 0 (不然很难算)

$$f(x,y)=\begin{cases}k(6-x-y), & 0<x<2, 2<y<4\\ 0, & \text{otherwise}\end{cases}$$

求① 常数 k; ② $\Pr\{X+Y\leqslant 4\}$; ③ X,Y 的边缘密度.

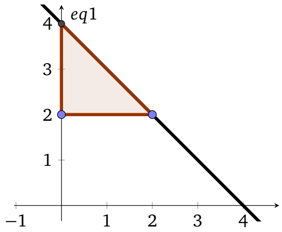
Solution: (1) 因为

$$\iint_{\mathbb{R}^2}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=\begin{cases}\iint_D k(6-x-y)\,\mathrm{d}x\,\mathrm{d}y\,, & x\in D=\{(x,y)|x\in(0,2)y\in(2,4)\}\\ 0, & \text{otherwise}\end{cases}=1$$

$$\text{i.e.}\iint_D k(6-x-y)\,\mathrm{d}x\,\mathrm{d}y=\int_0^2\int_2^4 k(6-x-y)\,\mathrm{d}x\,\mathrm{d}y=1,\text{计算得}k=\frac{1}{8}.$$

(2) 由题意, 所有符合的 (X, Y) 至少落在直线 $x + y = 4$ 的左侧. 由 (1) 知点只可能落入 D 中, 故令 $G = D \cap \{x + y \leqslant 4\}$, 即下图棕色区域

w/ Geogebra



那么, 我们有

$$k \cdot \iint_G 6 - x - y \,\mathrm{d}x\,\mathrm{d}y = \Pr\{X + Y \leqslant 4\} = \frac{2}{3} \quad \text{即为所求.}$$

(3) 根据3.1.4, 可以得到

$$f_X(x) = \begin{cases} \dfrac{1}{8} \cdot \displaystyle\int_2^{4-x} (6 - x - y)\,\mathrm{d}y = \frac{1}{8}(6 - 4x + \frac{1}{2}x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

相当于固定一个变量, 对另一个变量积分

同理可得

$$f_Y(y) = \begin{cases} \dfrac{1}{8} \cdot \displaystyle\int_0^{4-y} (6 - x - y)\,\mathrm{d}x = (\ldots \text{略}), & 2 < y < 4 \\ 0, & \text{otherwise} \end{cases}$$

3.1.5 二维均匀分布

Definition 3.1.5: 二维均匀分布

设 G 是平面上的有界区域, 其面积为 A, 若二维随机变量 (X, Y) 具有概率密度函数

$$f(x, y) = \begin{cases} \frac{1}{A}, & (x, y) \in G \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

则称 (X, Y) 在 G 上服从**二维均匀分布**.

若 (X, Y) 在 G 上服从均匀分布, 则其概率密度函数反映在几何上为定义在 xOy 平面内区域 G 上的空间的一块平面. 考虑向平面 G 上投掷一质点, 若点落在区域 $B \subseteq G$ 内的概率与 B 的面积成正比且与 B 的位置无关, 则称 G 为均匀分布的区域, 坐标 (X, Y) 在 G 上服从均匀分布.

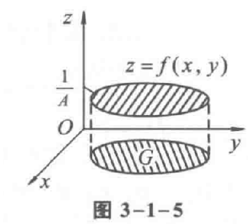


Figure 3.2: 均匀分布

Corollary 3.1.1

均匀分布 (X, Y) 的两个边缘分布仍为均匀分布且分别为

$$\begin{aligned} f_X(x) &= \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \\ f_Y(y) &= \begin{cases} \frac{1}{d-c}, & c < y < d \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{3.14}$$

对于**矩形域** G 该公式成立, 其他形状则不一定.

Question 6: 例 5

设 (X, Y) 服从**单位圆域**上的均匀分布, 求对应的两个边缘分布.

Solution: 由题意得, 其 PDF 为

$$f(x,y)=\begin{cases}\frac{1}{\pi}, & x^2+y^2\leqslant 1\\ 0, & \text{otherwise}\end{cases}$$

根据(3.12), 只需考虑点在圆域内的情况, 易得

$$f_X(x)=\frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\mathrm{d}y=\frac{2}{\pi}\sqrt{1-x^2}.$$
$$\therefore f_X(x)=\begin{cases}\frac{2}{\pi}\sqrt{1-x^2}, & -1\leqslant x\leqslant 1\\ 0, & \text{otherwise}\end{cases}$$

因为 $x^2+y^2\leqslant 1$ 是轮换式, 将 x 换成 y 即可得到 $f_Y(y)$.

3.2 随机变量的独立性

Definition 3.2.1: 独立性

设 (X,Y) 是二维随机变量, 若对于任意 x,y,

$$\begin{aligned}\Pr\{X\leqslant x,Y\leqslant y\}&=\Pr\{X\leqslant x\}\Pr\{Y\leqslant y\}\\ \text{i.e. } F(x,y)&=F_X(x)F_Y(y)\text{充要条件}\end{aligned}\tag{3.15}$$

则称 X 和 Y 是**相互独立**的.

Theorem 3.2.1
随机变量 X,Y 相互独立的充要条件是对 X 和 Y 所生成的任何事件相互独立, 即

$$\Pr\{X\in A,Y\in B\}=\Pr\{X\in A\}\Pr\{Y\in B\}\tag{3.16}$$

Theorem 3.2.2
如果 X,Y 相互独立, 则对任意函数 $g_1(x),g_2(y)$ 均有两者相互独立.

3.2.1 离散型随机变量的独立性

若 X,Y 是离散型随机变量, 可以由此判断其独立性:
定义 2 若对 (X,Y) 的所有可能取值 (X_i,Y_i) 有

$$\begin{aligned}\Pr\left\{X=X_i,Y=Y_j\right\}&=\Pr\{X=X_i\}\Pr\left\{Y=Y_j\right\},\\ \text{i.e. } p_{ij}&=p_{i\cdot}p_{\cdot j}\end{aligned}\tag{3.17}$$

则称 X,Y 相互独立.

Exercise 3.2.1 $P_{77}11$
设 X,Y 相互独立, 完成下面的联合分布:

Y,X	x_1	x_2	x_3	$p_{i\cdot}$
y_1	a	$1/9$	c	
y_2	$1/9$	b	$1/3$	
$p_{\cdot j}$				1

Solution: 首先求 a,b,c:

$$\because p_{21}=\frac{1}{9}=p_{2\cdot}p_{\cdot 1}$$
$$p_{23}=\frac{1}{3}=p_{2\cdot}p_{\cdot 3}$$

联立上两式可得 $c=3a$. 又

$$\sum_i\sum_j p_{ij}=1$$

得 $a+b+c=\frac{4}{9}$. 又

$$p_{22}=b=p_{2\cdot}p_{\cdot 2}$$

解得 $b=\frac{2}{9}$.

注意计算准确

综上可以得到 $a=\dfrac{1}{18},b=\dfrac{2}{9},c=\dfrac{1}{6}$. 即

Y,X	x_1	x_2	x_3	$p_{i\cdot}$
y_1	1/18	1/9	1/6	1/3
y_2	1/9	2/9	1/3	2/3
$p_{\cdot j}$	1/6	1/3	1/2	1

3.2.2 连续型随机变量的独立性

若 X,Y 是连续型随机变量, 可以由此判断其独立性:

严谨的说法是: 几乎处处成立. 亦即不包括平面上面积为 0 的集合.

$$f(x,y)=f_X(x)f_Y(y)\quad \forall x,y \tag{3.18}$$

则称 X,Y 相互独立.

Proposition 3.2.1
(3.18)和(3.15)是等价命题.

证明: (充分性) 对(3.18)两边积分有

$$\begin{aligned}\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=F(x,y)&=\int_{-\infty}^{x}\int_{-\infty}^{y}f_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y\\&=\int_{-\infty}^{x}f_X(x)\,\mathrm{d}x\int_{-\infty}^{y}f_Y(y)\,\mathrm{d}y\ (\text{分离变量})\\&=F_X(x)F_Y(y).\end{aligned}$$

(必要性) 对(3.15)两边求两次偏导有

$$\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)=\frac{\partial F_X}{\partial x}\cdot\frac{\partial F_Y}{\partial y}=f_X(x)f_Y(y)$$

☺

同时, 我们通过上述证明, 得到了重要的推论:

Corollary 3.2.1 快速判断独立性
若 $f(x,y)$ 是可分离变量的, 且 x,y 的取值范围独立 $\iff$ X,Y 独立.

Exercise 3.2.2 $P_{78}14$

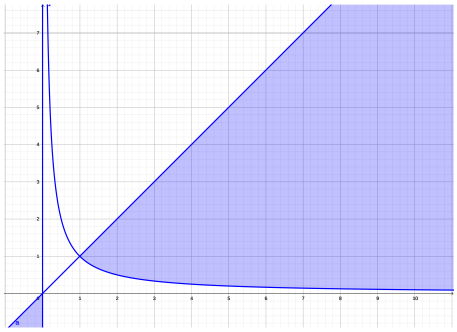
$$f(x,y)=\begin{cases}\dfrac{1}{2x^2y}, & 1\leqslant x\leqslant+\infty,\dfrac{1}{x}\leqslant y\leqslant x,\\ 0, & \text{otherwise}\end{cases}$$

判断 X,Y 是否独立.

Solution: 要判断其独立性, 先求两个边缘密度:

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}y=\begin{cases}\displaystyle\int_{1/x}^{x}\frac{1}{2x^2y}=\frac{\ln x}{2x^2}, & 1\leqslant x\leqslant+\infty,\\ 0, & \text{otherwise}\end{cases}$$

对于 Y, 在下图第一象限阴影区域内对 x 积分



此处的 x 的范围并不是 $[1,+\infty]$. 应将 D 投影到 y 轴上来讨论 x 的范围 (想累次积分定义)

设积分区域 (阴影部分) 为 D, 则将 D 投影到 y 轴上.(复习)

$$D=D_1+D_2=\{x\geqslant y,1\leqslant y\}+\{x\geqslant 1/y,0<y\leqslant 1\}$$

$$\text{i.e.}\quad f_Y(y)=\int_{D_1}f(x,y)\,\mathrm{d}x+\int_{D_2}f(x,y)\,\mathrm{d}x$$

$$=\begin{cases}\displaystyle\int_{1/y}^{+\infty}\frac{1}{2x^2y}\,\mathrm{d}x, & 0<y\leqslant 1,\\ \displaystyle\int_{y}^{+\infty}\frac{1}{2x^2y}\,\mathrm{d}x, & 1\leqslant y,\\ 0, & \text{otherwise}\end{cases}=(\cdots)$$

要写出乘积. 若判断独立性先枚举特值找反例.

两者相乘, 显然不与 $f(x,y)$ 相等. $\therefore\ XY$ 不独立.

Chapter 4

随机变量的数字特征

4.1 数学期望

4.1.1 随机变量的数学期望

Definition 4.1.1

设离散型随机变量 X 的概率分布为

$$\Pr\{X = x_i\} = p_i,\ i = 1, 2, \cdots$$

复习: 绝对收敛: $\sum_{i=1}^{+\infty} |x_i p_i| < \infty$

若级数 $\sum_{i=1}^{\infty} x_i p_i$ 绝对收敛, 则定义 X 的**数学期望**为

$\mathbb{E}[X]$, $\mathrm{E}(X)$, EX 都是等价的写法.

$$\mathbb{E}[X] = \sum_i x_i p_i. \tag{4.1}$$

上述定义限于离散型随机变量. 考虑连续型随机变量 X, 其 PDF 为 $f(x)$. 在数轴上取一段长度为 $\Delta x \to 0$ 的区间 $[x, x + \Delta x]$, 则 X 落入该区间的概率为

$$\Pr\{x \leqslant X \leqslant x + \Delta x\} = \int_x^{x+\Delta x} f(x)\,\mathrm{d}x \approx f(x)\Delta x.$$

这称为离散化

在数轴上取无限多段这样的小区间并对概率求和, 我们有

$$\sum_i x_i f(x_i)\Delta x_i \approx \int_{\mathbb{R}} x f(x)\,\mathrm{d}x\,.$$

并非所有随机变量都有数学期望

因此可以得到定义:

Definition 4.1.2

设 X 是连续型随机变量,PDF 为 $f(x)$, 若上述积分绝对收敛, 定义 X 的**数学期望**为

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x)\,\mathrm{d}x\,. \tag{4.2}$$

4.1.2 随机变量函数的数学期望

对于随机变量 X 的函数 $Y = g(X)$, 我们可以通过定义求出 X 的分布而求出 Y 的分布, 进而由定义得 Y 的数学期望 $\mathbb{E}[g(X)]$, 这做法较繁.

Theorem 4.1.1
设 X 是一个随机变量,$Y=g(X)$, 且 $E(Y)$ 存在, 于是
(1) 若 X 为离散型随机变量:

$$\mathbb{E}[Y]=\mathbb{E}[g(X)]=\sum_i g(x_i)p_i, \tag{4.3}$$

(2) 若 X 为连续型随机变量:

$$\mathbb{E}[Y]=\mathbb{E}[g(X)]=\int_{\mathbb{R}} g(x)f(x)\,\mathrm{d}x\,. \tag{4.4}$$

推广至二维: 设 (X,Y) 是二维随机变量,$Z=g(X,Y)$, 若 $\mathbb{E}[Z]$ 存在, 则
(1) 若为离散型:

$$\mathbb{E}[Z]=\mathbb{E}[g(X,Y)]=\sum_i\sum_j g(x_i,y_j)p_{ij}, \tag{4.5}$$

(2) 若为连续型:

$$\mathbb{E}[Z]=\mathbb{E}[g(X,Y)]=\iint_{\mathbb{R}^2} g(x,y)f(x,y)\,\mathrm{d}\sigma\,. \tag{4.6}$$

注意 实际做题中, 积分区域 $D\subseteq\mathbb{R}$, 根据所要求期望的随机变量来决定积分次序. E.g. $\mathbb{E}[X]$—先对 y 积分, 再对 x 积分;$\mathbb{E}[Y]$—先对 x 积分, 再对 y 积分;$\mathbb{E}[XY]$—两者谁先皆可, 但注意不满足 Fubini 定理的函数.

4.1.3 数学期望的性质

1. $\mathbb{E}[c]=c$(c 为常数, 下同);
2. $\mathbb{E}[cX]=c\mathbb{E}[X]$;
3. $\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]$;
4. 若 X,Y 相互独立, 则 $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$.

Exercise 4.1.1 P_{85} 5
设随机变量 X 的分布律为

X	-2	0	2
p_i	0.4	0.3	0.3

求 $\mathbb{E}[X],\mathbb{E}\left[X^2\right],\mathbb{E}\left[3X^2+5\right]$

Solution: 由题意得,$\mathbb{E}[X]=-0.2$. 设 $Y=X^2$,由(4.5)得 $\mathbb{E}\left[X^2\right]=2.8$. 由性质123得 $\mathbb{E}\left[3X^2+5\right]=3\mathbb{E}\left[X^2\right]+\mathbb{E}[5]=13.4$.

Exercise 4.1.2 P_{85} 10
设 (X,Y) 的概率密度为

$$f(x,y)=\begin{cases}12y^2, & 0\leqslant y\leqslant x\leqslant 1,\\ 0, & \text{otherwise}\end{cases}$$

求 $\mathbb{E}[X],\mathbb{E}[Y],\mathbb{E}[XY],\mathbb{E}\left[X^2+Y^2\right]$.

Solution: 由题意得, 实际的积分区域为 x 轴,$x=1$ 和 $y=x$ 围成的三角形区域 D.
(1) 求 $\mathbb{E}[X]$, 由定义可得 $\mathbb{E}[X]=\int_{-\infty}^{+\infty}xf_X(x)\,\mathrm{d}x$, 展开有

$$\mathbb{E}[X]=\iint_{\mathbb{R}^2}xf(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$D_1 = D_2$, 分开写是强调积分次序.

注意到 $f(x,y)$ 在 $D_1 = \{0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant x\}$ 上不为 0, 计算得 $\mathbb{E}[X] = \frac{4}{5}$.

(2) 同理, 求 $\mathbb{E}[Y]$,$D_2 = \{0 \leqslant y \leqslant 1, y \leqslant x \leqslant 1\}$. 可得

$$\mathbb{E}[Y] = \underbrace{\int_0^1 \int_y^1}_{D_2} y f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{3}{5}.$$

这个次序计算简单

(3) 由于 $f_X(x) f_Y(y) \neq f(x,y)$,$X, Y$ 不具有独立性, 无法使用性质 4.
由于积分次序不影响结果, 因此先积 y 再积 x 可得

$$\mathbb{E}[XY] = \iint_{D_1} xy f(x,y)\,\mathrm{d}\sigma = \frac{1}{2}.$$

(4) 由性质 3 得 $\mathbb{E}\left[X^2 + Y^2\right] = \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right]$. 由 (1)(2) 可得

$$\mathbb{E}\left[X^2 + Y^2\right] = \iint_{D_1} x^2 f(x,y)\,\mathrm{d}\sigma + \iint_{D_2} y^2 f(x,y)\,\mathrm{d}\sigma = \frac{16}{15}.$$

4.2 方差

4.2.1 方差的定义

Definition 4.2.1: 方差

设 X 是一个随机变量, 若 $\mathbb{E}^2[X-\mu]$ 存在, 则称其为 X 的**方差**, 记为 $\mathbb{V}[X]$, 其中 $\mu = \mathbb{E}[X]$. 方差亦记作 $\operatorname{Var} X, D(X), DX$.
方差的算数平方根 $\sqrt{\mathbb{V}[X]}$ 称为 X 的**标准差**, 记为 σ_X.

常见分布的期望和方差见书 P_{215}

方差刻画了随机变量 X 的取值与其期望 μ 的偏离程度, 方差越大, 随机变量的取值越分散. 而且, 若 $\mathbb{V}[X] = 0$, 则随机变量 X 一定取常数值, 此时 X 不是随机变量.

4.2.2 方差的计算

若 X 是**离散型**随机变量, 其分布律为 $\Pr\{X = x_i\} = p_i,\ i = 1, 2, \cdots$, 则

$$\mathbb{V}[X] = \mathbb{E}^2[X-\mu] = \sum_i (x_i - \mu)^2 p_i. \tag{4.7}$$

若 X 是**连续型**随机变量, 其概率密度为 $f(x)$, 则

$$\mathbb{V}[X] = \mathbb{E}^2[X-\mu] = \int_{\mathbb{R}} (x-\mu)^2 f(x)\,\mathrm{d}x\,. \tag{4.8}$$

由数学期望的性质可得方差计算的重要公式:

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]\,. \tag{4.9}$$

证明:

$$\begin{aligned}
\mathbb{V}[X] = \mathbb{E}^2[X-\mu] &= \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mu\mathbb{E}[X] + \mu^2 \\
&= \mathbb{E}\left[X^2\right] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]\,.
\end{aligned}$$

☺

Question 7: Poisson 分布

设 $X \sim P(\lambda)$, 求 $\mathbb{E}[X]$, $\mathbb{V}[X]$.

Solution: 随机变量 X 的分布律为

$$\Pr\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,\cdots;\ \lambda>0.$$

$$\begin{aligned}
\text{由定义得 } \mathbb{E}[X] &= \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\cdot e^{\lambda} = \lambda, \\
\mathbb{E}\left[X^2\right] &= \mathbb{E}[X(X-1)+X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] \\
&= \sum_{k=0} \infty k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} + \lambda = \lambda^2 e^{-\lambda} \underbrace{\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}}_{\text{Poisson 分布}=1} + \lambda \\
&= \lambda^2 e^{-\lambda}\cdot e^{\lambda} + \lambda = \lambda^2 + \lambda,
\end{aligned}$$

$$\text{故方差 } \mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X] = \lambda. \tag{4.10}$$

Question 8: 均匀分布

设 $X \sim U(a,b)$, 求 $\mathbb{E}[X]$, $\mathbb{V}[X]$.

Solution: X 的概率密度为

$$f(x) = \begin{cases} \dfrac{1}{b-a}, a<x<b, \\ 0, \text{otherwise}. \end{cases}$$

$$\text{而 } \mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X] = \int_a^b x^2\frac{1}{b-a}\,\mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}. \tag{4.11}$$

4.2.3 方差的性质

1. $\mathbb{V}[c] = 0$;
2. $\mathbb{V}[X] \geqslant 0$;
3. $\mathbb{V}[cX] = c^2\mathbb{V}[X]$;
4. $\mathbb{V}[c+X] = \mathbb{V}[X]$;
5. $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\,\mathrm{Cov}(X,Y)$;
6. 若 X, Y 相互独立, 则 $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

第六个性质可以推广到 n 维, 即若 $X_1, X_2, \cdots, X_n$ 相互独立, 则

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i].$$

Exercise 4.2.1 $\mathrm{P}_{90}8$

设 $X \sim N(1,2)$, $Y \sim E(3)$, 且 X, Y 相互独立, 求 $\mathbb{V}[XY]$.

Solution: (分析法) 要求 $\mathbb{V}[XY]$, 即求 $\mathbb{E}\left[XY^2\right] - \mathbb{E}^2[XY]$. 因为 X, Y 独立, 则

$$\mathbb{V}[XY] = \mathbb{E}\left[XY^2\right] - \mathbb{E}^2[X]\,\mathbb{E}^2[Y].$$

由定理 (3.22), 因为 X, Y 独立, 则 X^2, Y^2 也独立. 有

$$= \mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right] - \mathbb{E}^2[X]\,\mathbb{E}^2[Y].$$

X^2, Y^2 这俩玩意积不出来, 继续展开有

$$= \left(\mathbb{E}^2[X] + \mathbb{V}[X]\right)\left(\mathbb{E}^2[Y] + \mathbb{V}[Y]\right) - \mathbb{E}^2[X]\,\mathbb{E}^2[Y].$$

已知 $\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = 3$, 得 $= 27$.

若设 $f(x) = \mathbb{E}^2[X - x]$, $x \in \mathbb{R}$, 求导可证明 $f(x)$ 在 $x = \mu$ 处取得最小值, 即 $\mathbb{V}[X]$ 在 μ 处取得最小值, 这说明随机变量的取值对其数学期望的偏离程度比对其他任何值的偏离程度都要小.

Theorem 4.2.1
若随机变量序列 $\{X_n\} \sim N$, 且相互独立, 则其和常数的线性组合仍然服从正态分布.

$$\text{i.e. } A\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \sim N, \text{ where } A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}. \tag{4.12}$$

Example 4.2.1 ($\mathrm{P}_{91}4$)
若 $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \cdots, n$, 且相互独立, 求 $Y = \sum_{i=1}^n (a_i X_i + b_i)$ 服从什么分布.

Solution: 由题意, 根据定理 (4.21),Y 是 X 的一个线性组合, 所以 Y 服从正态分布.

$$\mathbb{E}[Y] = \sum_{i=1}^n a_i \mathbb{E}[X_i] + \sum_{i=1}^n b_i,$$
$$\mathbb{V}[Y] = \sum_{i=1}^n a_i^2 \mathbb{V}[X_i].$$

$$\therefore\ Y \sim N\left(\sum_{i=1}^n (a_i\mu_i + b_i), \sum_{i=1}^n a_i^2\sigma_i^2\right)$$

4.3 大数定律与中心极限定理

4.3.1 Chebyshev 不等式

Theorem 4.3.1 Chebyshev 不等式
设随机变量 X 的期望为 μ, 方差为 σ^2, 则对任意 $\varepsilon > 0$, 有

$$\Pr\{|X - \mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}. \tag{4.13}$$

这个不等式从直观上表示事件 X 的取值都落在 μ 的一个邻域 $(\mu-\varepsilon,\mu+\varepsilon)$ 之外 (阴影部分) 的概率. 它同时指出随机变量 X 的方差越小, 事件 $|X-\mu|<\varepsilon$ 发生的概率越大, 即 X 的取值基本上集中在 μ 的邻域内. 因此称方差刻画了随机变量的离散程度.

只证连续型的情况. 设 X 的概率密度为 $f(x)$, 则

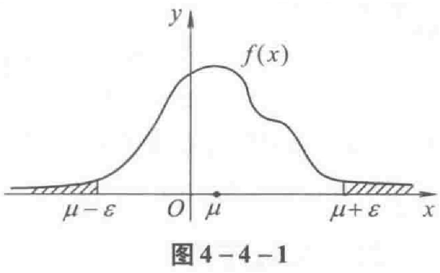


Figure 4.1: $f(x)$ 示意

证明:

$$\Pr\{|X-\mu|\geqslant\varepsilon\}=\int_{|x-\mu|\geqslant\varepsilon}f(x)\,\mathrm{d}x\,.$$

放缩, 取 $k=\dfrac{|x-\mu|^2}{\varepsilon^2}$.(因为 X 都落在领域外, 则 $k>1$.)

再次放缩, 变大区间

$$\leqslant\int_{|x-\mu|\geqslant\varepsilon}kf(x)\,\mathrm{d}x \qquad \leqslant\frac{1}{\varepsilon^2}\underbrace{\int_{\mathbb{R}}(x-\mu)^2f(x)\,\mathrm{d}x}_{\mathbb{V}[X]}=\frac{\sigma^2}{\varepsilon^2}$$

☺

4.3.2 大数定律

Lemma 4.3.1 随机变量序列相互独立

若对于任意 $n>1$,$X_1,X_2,\cdots,X_n$ 都相互独立, 则称 $X_1,X_2,\cdots,X_n,\cdots$ 是相互独立的.

Theorem 4.3.2 大数定律

设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立, 且具有相同的期望和方差

$$\mathbb{E}\left[X_i\right]=\mu,\ \mathbb{V}\left[X_i\right]=\sigma^2,\ i=1,2,\cdots.$$

记 $Y_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$, 则对任意 $\varepsilon>0$, 有

$$\lim_{n\to\infty}\Pr\left\{\left|Y_n-\mu\right|<\varepsilon\right\}=1 \tag{4.14}$$

证明: 由 Y_n 的式子可以知道

$$\mathbb{E}\left[Y_n\right]=\frac{1}{n}\sum_{i=1}n\mathbb{E}\left[X_i\right]=\mu, \qquad \mathbb{V}\left[Y_n\right]=\frac{1}{n^2}\sum_{i=1}n\mathbb{V}\left[X_i\right]=\frac{\sigma^2}{n},$$

由 Chebyshev 不等式可得

$$\begin{aligned}\Pr\left\{\left|Y_n-\mu\right|<\varepsilon\right\}&\geqslant 1-\Pr\left\{\left|Y_n-\mu\right|\geqslant\varepsilon\right\}\\&\geqslant 1-\frac{\mathbb{V}\left[Y_n\right]}{\varepsilon^2}\\&\geqslant 1-\frac{\sigma^2}{n\varepsilon^2}.\end{aligned}\tag{$*$}$$

夹逼法,$(*)$ 式两边取极限, 且因为概率不可能大于 1,

$$1\geqslant\lim_{n\to\infty}\Pr\left\{\left|Y_n-\mu\right|<\varepsilon\right\}\geqslant\lim_{n\to\infty}1-\frac{\sigma^2}{n\varepsilon^2}=1.$$
$$\therefore\lim_{n\to\infty}\Pr\left\{\left|Y_n-\mu\right|<\varepsilon\right\}=1.$$

☺

这表明对任意 $\varepsilon > 0$, 事件 $\left|Y_n - \mu\right| < \varepsilon$ 发生的概率很大 (Y_n 的取值一定落在 μ 的邻域内), 当 n 很大时,Y_n 集中于 μ. 像这样表现的收敛性成为随机变量序列 $Y_1, Y_2, \cdots, Y_n, \cdots$**依概率收敛**于 μ, 记为

$$Y_n \xrightarrow{P} \mu. \tag{4.15}$$

这还表明, 随机变量 $X_1, X_2, \cdots, X_n$ 的算术平均值序列 Y_n 依概率收敛于 μ.

Corollary 4.3.1 Bernoulli 定理

设 n_A 是 n 重 Bernoulli 试验中事件 A 发生的次数,p 是事件 A 在每次试验中发生的概率 ($n_A \sim b(n,p)$), 则对任意的 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} \Pr\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1. \tag{4.16}$$

证明同上.

$\dfrac{n_A}{n}$ 是事件 A 发生的频率, 这个定理表明当 $n \to \infty$ 时, 这个频率依概率收敛于事件 A 发生的概率 p. 这说明频率的本质是一个随机变量.