第九讲：随机变量定义及相关性质

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

- 赌徒输光：甲和乙初始资金分别为 $i, a-i$ 元，每一局甲赢的概率为 p
- 关注的问题
  - 甲最终获胜的概率
  - 甲乙两人在任意时刻的剩余资产:k 轮赌博后恰好剩下 j 元
  - k 轮赌博后甲乙两人资产的差额 Z
  - 赌博持续时间 R
- 表示方法:
  - $E :=$ {甲最终获胜}, $Q_i := P(\mathrm{E})$
  - $A_{jk} :=$ {甲在 k 轮赌博后恰好剩下 j 元}
  - $B_{jk} :=$ {乙在 k 轮赌博后恰好剩下 j 元}
  - k 轮赌博后甲乙两人资产的差额如何表示?
  - 赌博持续时间 R 如何表示?
  - 很难用事件来表示或者表示很复杂

随机变量的引入

- $X_k :=$ 甲在 k 轮赌博后的资产
  - 乙在 k 轮赌博后的资产 $Y_k = a - X_k$
  - 资产差额:$Z = X_k - Y_k = 2X_k - a$
  - 持续时间:$R = \min\{n : X_n = 0, 或 Y_n = 0\}$
- X_k 取值的特点
  - 依赖于前面 k 次赌博这一“随机试验”的结果
  - 在“随机试验”完成之前，X_k 取值不确定，因此具有不确定性
  - k 次赌博“随机试验”一旦完成，X_k 的值必然确定
  - 记 k 次赌博“随机试验”样本空间为 Ω, 则给定 $\omega \in \Omega$, 则 X_k 值确定
- 以 $k = 2$ 为例，看一下 X_2 的取值情况，设 $i \geq 2$
  - $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, 其中
    $\omega_1 = (胜，胜), \omega_2 = (胜，败), \omega_3 = (败，胜), \omega_4 = (败，败)$
  - $X_2(\omega_1) = i + 2,\ X_2(\omega_2) = X_2(\omega_3) = i,\ X_2(\omega_4) = i - 2$
- X_k 可以看作定义在样本空间 Ω 上的函数，即
  $X_k : \Omega \to \{0, 1, \cdots, a\}$
- 一般的，随机变量可以看作从样本空间到实数的映射:$X : \Omega \to R$

定义 1.1 (直观定义) 称 X 为随机变量，如果 X 是从样本空间 Ω 到实数的映射，即 $X:\Omega\to R$.

一个随机变量的例子

例 1.1 考虑硬币抛掷两次的随机试验，其样本空间为

$$\Omega = \{HH, HT, TH, TT\}.$$

- 令 X 表示正面朝上的次数，则 X 是一个随机变量，相应的映射如下

$$X(TT) = 0, X(HT) = X(TH) = 1, X(HH) = 2.$$

- 令 Y 表示反面朝上的次数，则 $Y = 2 - X$，对于任意的 $\omega \in \Omega$，有 $Y(\omega) = 2 - X(\omega)$.
- 设 I 是由第一次掷硬币的结果决定的随机变量：若第一次硬币正面朝上则 $I = 1$，反之 $I = 0$，即

$$I(HH) = I(HT) = 1, \quad I(TH) = I(TT) = 0.$$

- 若正面记 1，反面记 0，此时 $\Omega = \{(1,1),(1,0),(0,1),(0,0)\}$. 上述的 X, Y 和 I 可表示为：

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2, Y(\omega_1, \omega_2) = 2 - \omega_1 - \omega_2, I(\omega_1, \omega_2) = \omega_1.$$

与函数对比分析

- 数学分析中的函数 $f(x)$
  - $f(x): D \to R$, 其中 $D \subset R$
  - R 上可定义距离 $d(x,y) = |x-y|$
  - 可根据距离 d 定义函数的连续性
- 概率论中的随机变量 X
  - $X(\omega): \Omega \to R$
  - 定义域 Ω 没有距离的定义，但有事件域 $\mathcal{F}$ 及定义其上的概率 P, 即具有结构 $(\Omega, \mathcal{F}, P)$
  - 值域 R 有距离，但因定义域无距离，故无法考虑随机变量的连续性
  - 值域 R 还有 σ- 代数 $\mathcal{B} := \mathcal{B}(R)$, 有可测空间结构 $(R, \mathcal{B})$
- 是否需要对 $X(\omega): \Omega \to R$ 做一些额外的限定，以便更好的研究 X?
- 从赌徒输光问题可以看出，对随机变量 X, 我们会关注
  - $X(\omega) = x$ 的概率，$X(\omega) \le x, X(\omega) \in [b, c]$ 的概率
  - 更一般的，$\forall B \in \mathcal{B}(R), X(\omega) \in B$ 的概率
- 概率的定义域是 $\mathcal{F}$, 要想计算 $X(\omega) \in B$ 的概率，当且仅当

$$\forall B \in \mathcal{B}(R),\ X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

随机变量的严格数学定义

定义 1.2 (可测映射) 设 $(\Omega, \mathcal{F})$ 和 $(E, \mathcal{E})$ 为两个可测空间并令 X 为从样本空间 Ω 到 E 的映射，即 $X(\omega): \Omega \to E$. 若对任意的 $B \in \mathcal{E}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

则称 X 为 $(\Omega, \mathcal{F})$ 到 $(E, \mathcal{E})$ 的可测映射.

若将上述定义中的可测空间 $(E, \mathcal{E})$ 更换为 $(R, \mathcal{B})$，则

定义 1.3 (可测函数或随机变量) 设 $(\Omega, \mathcal{F})$ 是可测空间，X 为从样本空间 Ω 到实数集 R 的映射，即 $X(\omega): \Omega \to R$. 如果对 $\forall B \in \mathcal{B}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

则称 X 为可测空间 $(\Omega, \mathcal{F})$ 上的可测函数或随机变量.

定义 1.4 (随机变量的另一定义) 设 $(\Omega, \mathcal{F})$ 是可测空间，X 为从样本空间 Ω 到实数集 R 的映射，即 $X(\omega): \Omega \to R$. 如果对任意的 $x \in R$ 均有

$$\{\omega : X(\omega) \le x\} \in \mathcal{F} \text{ 或} \{\omega : X(\omega) < x\} \in \mathcal{F}$$

则称 $X(\omega)$ 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，简称随机变量.

随机变量两种定义的等价性

由定义 1.4 推定义 1.3: 仅需说明若定义 1.4 成立，则对任意 $B \in \mathcal{B}$ 均有

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}.$$

即只需说明以下集合包含关系成立即可

$$\mathcal{A} := \{A : X^{-1}(A) \in \mathcal{F}\} \supset \mathcal{B}$$

欲证上面包含关系成立，我们只需说明以下两点即可:

1. $\mathcal{A}$ 是 σ 代数;
2. $O_1 := \{(-\infty, x] : x \in R\} \subset \mathcal{A}$.

再由 $\mathcal{B} := \sigma(O_1)$ 知 $\mathcal{B} \subset \mathcal{A}$.

$\mathcal{A} := \{A : X^{-1}(A) \in \mathcal{F}\}$ 为 σ 代数

- $X^{-1}(R) = \{\omega : X(\omega) \in R\} = \Omega \in \mathcal{F}$, 故 $R \in \mathcal{A}$;
- 若 $A \in \mathcal{A}$, 即 $X^{-1}(A) \in \mathcal{F}$, 则

$$\begin{aligned} X^{-1}(\overline{A}) &= \{\omega : X(\omega) \in \overline{A}\} = \{\omega : X(\omega) \notin A\} \\ &= \overline{\{\omega : X(\omega) \in A\}} = \overline{X^{-1}(A)} \\ &\in \mathcal{F} \end{aligned}$$

- 对于 $A_j \in \mathcal{A}, j = 1, 2, \cdots$, 有 $X^{-1}(A_j) \in \mathcal{F}, j = 1, 2, \cdots$. 从而

$$\begin{aligned} X^{-1}(\cup_{j=1}^{\infty} A_j) &= \{\omega : X(\omega) \in \cup_{j=1}^{\infty} A_j\} = \cup_{j=1}^{\infty} \{\omega : X(\omega) \in A_j\} \\ &= \cup_{j=1}^{\infty} X^{-1}(A_j) \\ &\in \mathcal{F} \end{aligned}$$

随机变量的分类以及两个注记

- 随机变量 X 是从样本空间 Ω 到实数 R 的映射，故根据其值域集合可粗略的分为两大类
  - 离散型随机变量：其值域集合是有限点集或可数点集即

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\} = \{a_n\}_{n\geq 1}$$

  - 非离散型随机变量：其值域集合不是有限点集或可数点集

注 1.1 随机变量的两点注记:

- 首先, 随机变量是确定性函数, 自身并没有随机性. 给定样本空间上的样本点, 有唯一确定的实数值与之相对应, 这种对应关系并没有不确定性. 所有的不确定性都体现在样本点是否在试验结果中出现, 和随机变量本身没有关系。随机变量的引入, 更多地是为了数学处理上的方便.
- 其次, 随机变量并不是概率论中独有的概念. 若将前面可测映射中的 $(\Omega, \mathcal{F}), (E, \mathcal{E})$ 均取为 $(R, \mathcal{B}(\mathbb{R}))$, 则可测映射的定义便退化为实分析中的 "可测函数". 随机变量是一特殊的可测函数.

例 1.2 设 Ω 是某随机试验的样本空间，$\mathcal{F}$ 为其事件域 (σ 代数)，则对于任意的 $A\in\mathcal{F}$，示性函数 $I_A(\omega):=\begin{cases}0, & \omega\notin A\\ 1, & \omega\in A\end{cases}$ 是随机变量.

解：由示性函数的定义知：

$$\{\omega: I_A(\omega)\le x\}=\begin{cases}\emptyset, & x<0,\\ \overline{A}, & x\in[0,1),\\ \Omega, & x\ge 1.\end{cases}$$

显然，无论 x 取何值，均有 $\{\omega: I_A(\omega)\le x\}\in\mathcal{F}$

定理 1.1 若 $X, Y, \{X_n, n \geq 1\}$ 都为概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量，则

1. $|X|$，$aX+bY, (a, b \in R)$ 均为随机变量;
2. $X^+ := X \vee 0, X^- := (-X) \vee 0$ 均为随机变量;
3. XY 为随机变量;
4. 若 X/Y 处处有意义，则 X/Y 为随机变量;
5. $\inf_n X_n, \sup_n X_n, \liminf_{n\to\infty} X_n, \limsup_{n\to\infty} X_n$ 均为随机变量.

证明:

1.
- $\{\omega : |X| < x\} = \{\omega : -x < X < x\} = \{\omega : X < x\} \cap \overline{\{\omega : X \leq -x\}} \in \mathcal{F}$;
- $\{\omega : aX < x\} = \{\omega : X < \frac{x}{a}\} \in \mathcal{F}$, (当 $a > 0$ 时);
- 设 Q 为有理数集，则

$$\begin{aligned}
\{\omega : X+Y < x\} &= \{\omega : X < x - Y\} = \cup_{r \in Q} \{\omega : X < r < x - Y\} \\
&= \cup_{r \in Q} \{\omega : X < r, Y < x - r\} \\
&= \cup_{r \in Q} (\{\omega : X < r\} \cap \{\omega : Y < x - r\}) \\
&\in \mathcal{F}
\end{aligned}$$

2. 注意到 $X^+ = \dfrac{|X|+X}{2}, X^- = \dfrac{|X|-X}{2}$，易得 X^+, X^- 均为随机变量;

3. 首先假定 X, Y 非负，则对任意的 $x > 0$ 有

$$\begin{aligned}\{XY < x\} &= \{X=0\} \cup \{Y=0\} \cup \left(\cup_{r\in Q_+}\left[\{X<r\}\cap\{Y<\frac{x}{r}\}\right]\right)\\ &\in \mathcal{F}.\end{aligned}$$

对一般的 X, Y, 由 X^+, X^-, Y^+, Y^- 为随机变量，可得

$$XY = (X^+ - X^-)(Y^+ - Y^-) = (X^+Y^+ + X^-Y^-) - (X^+Y^- + X^-Y^+)$$

为随机变量.

4. 设 $|Y| > 0$ 处处成立，易证 $\dfrac{1}{Y}$ 是随机变量，故 $\dfrac{X}{Y} = X \cdot \dfrac{1}{Y}$ 为随机变量.

5. 对任意的 $x \in R$, 我们有

$$\{\inf_n X_n < x\} = \cup_n\{X_n < x\}, \quad \{\sup_n X_n \le x\} = \cap_n\{X_n \le x\}$$

例 1.3 设 $(\Omega, \mathcal{F}, P)$ 为一概率空间，$A_i \in \mathcal{F}, i = 1, \cdots, n$ 为 Ω 的一个分割，$a_i, i = 1, \cdots, n$ 为 n 个不同的实数，则

$$X(\omega) := \sum_{i=1}^{n} a_i I_{A_i}(\omega) \tag{1}$$

作为 n 个示性随机变量的线性组合，仍为随机变量。我们称形如 (1) 的 $X(\omega)$ 为简单随机变量.

定理 1.2 设 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量，$g(x)$ 为 $(R, \mathcal{B}) \to (R, \mathcal{B})$ 上的 Borel 可测函数，证 $Y := g(X)$ 为 $(\Omega, \mathcal{F})$ 上的随机变量.

证明: 注意到，对任意的 $B \in \mathcal{B}$, $g^{-1}(B) \in \mathcal{B}$, 故

$$\begin{aligned}
Y^{-1}(B) &= \{\omega : Y(\omega) \in B\} \\
&= \{\omega : g(X(\omega)) \in B\} \\
&= \{\omega : X(\omega) \in g^{-1}(B)\} \\
&= X^{-1}(g^{-1}(B)) \\
&\in \mathcal{F}
\end{aligned}$$

对于 $(\Omega, \mathcal{F}, P)$ 上的任意非负随机变量 X 及自然数 n，我们可将 Ω 按 X 的取值进行分割。即令

$$\begin{aligned} A_k(\omega) &:= \{\omega : \frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n}\}, k = 0, 1, \cdots, n2^n - 1 \\ A_{n2^n}(\omega) &:= \{\omega : X(\omega) \ge n\} \end{aligned}$$

则

$$X_n(\omega) := \sum_{k=0}^{n2^n} \frac{k}{2^n} I_{A_k}(\omega)$$

为简单随机变量且随机变量序列 $\{X_n, n \ge 1\}$ 满足

$$0 \le X_1(\omega) \le X_2(\omega) \le \cdots \le X_n(\omega) \to X(\omega)$$

注意到对任意的 $k=0,1,\cdots,n2^n-1$,

$$\begin{aligned}
A_k(\omega) &= \{\omega:\frac{k}{2^n}\le X(\omega)<\frac{k+1}{2^n}\}=\{\omega:\frac{2k}{2^{n+1}}\le X(\omega)<\frac{2(k+1)}{2^{n+1}}\}\\
&= \{\omega:\frac{2k}{2^{n+1}}\le X(\omega)<\frac{2k+1}{2^{n+1}}\}\cup\{\omega:\frac{2k+1}{2^{n+1}}\le X(\omega)<\frac{2k+2)}{2^{n+1}}\}\\
&= A_k^1(\omega)\cup A_k^2(\omega)
\end{aligned}$$

故在集合 $A_k(\omega),k=0,1,\cdots,n2^n-1$ 上,

$$X_{n+1}(\omega)=\begin{cases}\frac{2k}{2^{n+1}}, & \omega\in A_k^1(\omega)\\ \frac{2k+1}{2^{n+1}}, & \omega\in A_k^2(\omega)\end{cases}\quad (\ge\frac{k}{2^n}=X_n(\omega))$$

而在 $A_{n2^n}(\omega):=\{\omega:X(\omega)\ge n\}=\{\omega:X(\omega)\ge\frac{n2^{n+1}}{2^{n+1}}\}$ 上，显然有

$$X_{n+1}(\omega)\ge n=X_n(\omega).$$

注意到，对任意的 ω, 必定存在 k 使得

$$\frac{k}{2^n} \le X(\omega) < \frac{k+1}{2^n},$$

从而

$$0 \le X(\omega) - X_n(\omega) \le \frac{1}{2^n}$$

显然有

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

定理 1.3 对 $(\Omega, \mathcal{F}, P)$ 上的实值变量 $X(\omega)$ 为随机变量的充要条件是：存在简单随机变量序列 $\{X_n(\omega), n \geq 1\}$ 使得

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega$$

而且当 X 非负时，还可选取 $\{X_n(\omega), n \geq 1\}$ 为非负单调不减的简单随机变量序列.

随机变量的生成 σ-代数

定义 1.5 (单个随机变量的生成 σ-代数) 设 X 为定义在 $(\Omega, \mathcal{F})$ 上的随机变量并令 $\mathcal{F}_X := \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$. 我们称 $\sigma(X) := \sigma(\mathcal{F}_X)$ 为随机变量 X 的生成 σ-代数.

定义 1.6 (有限个随机变量的生成 σ-代数) 设 $X_k, k = 1, 2, \cdots, n$ 为定义在 $(\Omega, \mathcal{F})$ 上的随机变量并令 $\mathcal{F}_{X_k} := \{X_k^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$. 我们称

$$\sigma(X_1, X_2, \cdots, X_n) := \sigma\left(\cup_{k=1}^n \mathcal{F}_{X_k}\right)$$

为随机变量 $X_1, X_2, \cdots, X_n$ 的生成 σ-代数.

- $\mathcal{F}_X, \mathcal{F}_{X_k}$ 均为 σ-代数, 但 $\cup_{k=1}^n \mathcal{F}_{X_k}$ 不一定是 σ-代数, 因此

$$\sigma(X) = \mathcal{F}_X, \quad \text{但} \quad \sigma(X_1, X_2, \cdots, X_n) \neq \cup_{k=1}^n \mathcal{F}_{X_k}.$$

- 若 X 是 $(\Omega, \mathcal{F})$ 上的随机变量, 则必有 $\sigma(X) := \mathcal{F}_X \subset \mathcal{F}$, 即 $\sigma(X)$ 是 Ω 上使得 X 成为随变量所需要的最小 σ-代数.
- 类似的, $\sigma(X_1, X_2, \cdots, X_n)$ 是 Ω 上使得 $X_1, X_2, \cdots, X_n$ 为随机变量所需要的最小 σ-代数.
- 随机变量 X 生成的 $\sigma-$ 代数 $\sigma(X)$ 集中体现了 X 的取值信息.

定义 1.7 考虑两个随机变量 X, Y，如果对任意的 $B \in \mathcal{B}(\mathbb{R})$ 均有

$$Y^{-1}(B) \in \sigma(X), \text{ 等价的 } \sigma(Y) \subset \sigma(X).$$

则称 Y 适应 (adaptive to) X.

定理 1.4 考虑可测空间 $(\Omega, \mathcal{F})$, X 和 Y 为定义其上的随机变量. 若 Y 适应 X, 则存在可测函数 $g: \mathbb{R} \to \mathbb{R}$, 使得

$$Y(\omega) = g(X(\omega)), \ \forall \omega \in \Omega.$$

注 1.2

- 随机变量的生成 σ-代数研究随机变量间关系起着重要作用.
- 直观地看,Y 适应 X, 意味着 Y 包含的信息被 X 包含的信息所涵盖. 换句话说,Y 和 X 间存在导出关系.

随机向量：如何定义 $\mathcal{B}(\mathbb{R}^n)$

- 直观上来讲，随机向量就是取值于 $\mathbb{R}^n$ 的随机变量.
- 如何定义 $\mathcal{B}(\mathbb{R}^n)$
  - 注意到
$$\mathbb{R}^n := \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} = \Pi_{k=1}^n \mathbb{R}$$
  - 我们希望
$$\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R}) = \Pi_{k=1}^n \mathcal{B}(\mathbb{R})$$
  - 但 $\Pi_{k=1}^n \mathcal{B}(\mathbb{R})$ 不是 σ-代数.
  - 因此，定义
$$\mathcal{B}(\mathbb{R}^n) := \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})) = \sigma(\Pi_{k=1}^n \mathcal{B}(\mathbb{R}))$$
  - 我们也可以类比 $\mathcal{B}(\mathbb{R})$ 的定义方法，先定义 $\mathbb{R}^n$ 上的立方体
$$I = I_1 \times I_2 \times \cdots \times I_n = \Pi_{k=1}^n I_k, \ \ I_k = (a_k, b_k].$$
然后定义 $\mathcal{B}(\mathbb{R}^n)$ 为:
$$\mathcal{B}(\mathbb{R}^n) := \sigma(\{I : I\text{为}\mathbb{R}^n\text{上的立方体}\}).$$

定义 1.8 (n 维随机向量) 考虑可测空间 $(\Omega, \mathcal{F})$, 若映射 $X:\Omega\to\mathbb{R}^n$ 使得对任意的 $B\in\mathcal{B}(\mathbb{R}^n)$ 均有

$$X^{-1}(B)\in\mathcal{F}$$

则称 X 为 n 维随机向量.

注 1.3 若记 $X=(X_1,X_2,\cdots,X_n)$, 则 X 为 n 维随机向量当且仅当 $X_k, k=1,\cdots,n$ 为随机变量.

定义 1.9 (复随机变量) 考虑可测空间 $(\Omega, \mathcal{F})$ 以及映射 Z

$$Z(\omega)=X(\omega)+\mathrm{i}Y(\omega)\in\mathbb{C}$$

其中 $\mathrm{i}=\sqrt{-1}$ 是虚数单位. 如果 $(X(\omega),Y(\omega))$ 构成二维随机向量, 则称 Z 为复随机变量.

第十讲：随机变量的分布与性质

张鑫

Email: x.zhang.seu@foxmail.com

东南大学　　数学学院

定理 1.5 设 X 为 $(\Omega, \mathcal{F}, P)$ 上的随机变量，对于 Borel 集 B, 定义集函数 $\mathbf{F}(B)$ 如下:

$$\mathbf{F}(B) := P(X^{-1}(B)) = P \circ X^{-1}(B) = P(X \in B) \tag{2}$$

则 $\mathbf{F}(\cdot)$ 为 $(R, \mathcal{B})$ 上的概率，称之为随机变量 X 的诱导概率测度.

定义 1.10 称由 (2) 式定义在 $(R, \mathcal{B})$ 上的概率测度 $\mathbf{F}(\cdot)$ 为随机变量 X 的概率分布，简称分布.

- 给定概率空间 $(\Omega, \mathcal{F}, P)$，任给随机变量均可在 $(R, \mathcal{B})$ 上诱导出一个概率测度。由此可见，在同一个可测空间上可以定义不同的概率测度.
- 对于任意的 $B \in \mathcal{B}$，随机变量 X 落入 B 中的概率可通过 B 的概率测度 $\mathbf{F}(B)$ 得出。这也就是说，概率分布 $\mathbf{F}(\cdot)$ 完全刻画了随机变量 X 取值的概率规律.

如果我们将 $(R, \mathcal{B})$ 上的测度仅局限于集类 $\mathcal{P} := \{(-\infty, x], x \in R\}$ 上，由于 $\mathcal{P}$ 中的每条半直线被它的右端点 x 所决定，于是集函数 F 就化为 R 上的点函数.

定义 1.11 对于随机变量 X 而言，称 x 的函数

$$F(x) := \mathbf{F}((-\infty, x]) = P(X \leq x)$$

为 X 的概率分布函数或累积分布函数，简称分布函数并记作 $X \sim F(x)$，有时也以 $F_X(x)$ 表明是 X 的分布函数.

注 1.4 也有一些教材按如下方式定义分布函数:

$$F(x) := \mathbf{F}((-\infty, x)) = P(X < x)$$

定理 1.6 任一分布函数 $F(x)$ 都具有以下三条基本性质

1. 单调性非降性：$F(x)$ 是单调非减函数即对任意的 $x_1 < x_2$, 有 $F(x_1) \le F(x_2)$;
2. 右连续性：$F(x)$ 是 x 的右连续函数，即

$$F(x_0) = F(x_0+) := \lim_{x \to x_0+} F(x)$$

3. 规范性：对任意的 x 有，$0 \le F(x) \le 1$ 且

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$
$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1$$

注 1.5 在函数论中, 我们一般不从随机变量出发来定义分布函数, 而是直接定义满足上面三条性质的函数为分布函数.

分布函数性质的证明

1. 对任意的 $x < y$, $F(y) - F(x) = P(x < X \le y) \ge 0$;
2. 因 $F(x)$ 是单调有界非降函数，所以其任一点 x_0 的右极限 $F(x_0+)$ 必存在，为证其连续性，只需证对单调上下降且收敛至 x_0 的数列 $\{x_n\}$ 有 $\lim_{n\to\infty} F(x_n) = F(x_0)$ 即可。注意到

$$\begin{aligned}\lim_{n\to\infty} F(x_n) &= \lim_{n\to\infty} P(X \le x_n) = P(\cap_{n=1}^{\infty}\{X \le x_n\}) = P(X \le x_0)\\ &= F(x_0)\end{aligned}$$

3. 由 F 的单调性及概率的连续性可知

$$\begin{aligned}F(+\infty) &= \lim_{n\to+\infty} F(n) = \lim_{n\to+\infty} P(X \le n)\\ &= P(\cup_{n=1}^{\infty}\{X \le n\}) = P(X < \infty) = 1\end{aligned}$$

同理可证 $F(-\infty) = 0$.

事件概率的分布函数表示

- $P(X > b) = 1 - F(b)$;
- $P(a < X \le b) = F(b) - F(a)$;
- $P(X < a) = F(a-)$;
- $P(X = a) = F(a) - F(a-)$;
- $P(X \ge b) = 1 - F(b-)$;
- $P(a \le X < b) = F(b-) - F(a-)$;
- $P(a \le X \le b) = F(b) - F(a-)$;
- $P(a < X < b) = F(b-) - F(a)$;
- 对于不交区间并: $P(X \in \cup_{i=1}^{n}(a_i, b_i]) = \sum_{i=1}^{n}[F(b_i) - F(a_i)]$;
- 一般的，$P(X \in B) = \int_B dF(x)$.

随机变量与其分布 (函数) 之间的关系

- 分布函数和随机变量密不可分. 每个随机变量都有自己的分布函数; 反过来, 每个分布函数都能找到与之相对应的随机变量.
- 分布函数是随机变量最基本的性质，也是对随机变量概率特性最全面的描述。为清晰起见，通常记随机变量 X 的分布函数为 $F_X(x)$.
- 尽管随机变量与其分布函数密切关联，但是两者间并没有一一对应关系。一个随机变量只能有唯一的分布函数, 不同的随机变量却可能具有相同的分布函数.

例 1.4 (不同的随机变量有相同的分布) 考虑随机变量 X, 其分布满足

$$P(X=1)=P(X=-1)=\frac{1}{2}.$$

设随机变量 $Y=-X$, 那么 Y 的分布为

$$P(Y=1)=P(Y=-1)=\frac{1}{2}$$

显然, X 和 Y 的分布函数完全相同, 但明显是两个不同的随机变量.

定理 1.7 任给一个定义在 Borel 域 $\mathcal{B}(\mathbb{R})$ 上的概率 $\bar{P}$, 都存在如下定义的分布函数 F 与之相对应：$F(x)=\bar{P}((-\infty,x])$. 反过来, 任给一个定义于实数轴上的分布函数 F, 都存在定义在 Borel 域 $\mathcal{B}(\mathbb{R})$ 上的概率 $\bar{P}$ 与之相对应.

证明: 第一个结论可直接验证. 由概率的单调性, 若 $x_1 \leqslant x_2$, 则

$$\bar{P}\left((-\infty,x_1]\right) \leqslant \bar{P}\left((-\infty,x_2]\right) \Rightarrow F\left(x_1\right) \leqslant F\left(x_2\right),$$

从而得到 F 的单调不减. 再由概率的连续性,

$$\begin{aligned}\lim_{x_n\downarrow x}(-\infty,x_n]=(-\infty,x] \quad &\Rightarrow \quad \lim_{x_n\downarrow x}\bar{P}\left((-\infty,x_n]\right)=\bar{P}((-\infty,x])\\ &\Rightarrow \quad \lim_{x_n\downarrow x}F\left(x_n\right)=F(x)\end{aligned}$$

定理的第二个结论本质上是概率 “扩张”. 根据测度扩张定理, 由分布函数所确定的定义在 $\mathcal{P}$ 上的集函数 $\mathbf{F}((-\infty,x]):=F(x)$ 可以唯一的扩张到 $\mathcal{B}:=\sigma(\mathcal{P})$ 上, 成为 $\mathcal{B}$ 上的概率测度. 扩张后的概率测度称之为分布函数 $F(x)$ 所引出的勒贝格 - 斯蒂尔吉斯测度。实际上这个 $\mathbf{F}$ 正好是我们前面引进的概率分布.

定义 1.12 (离散型随机变量) 如果随机变量 X 只取有限个值 $x_1, x_2, \cdots, x_n$ 或可列个值 $x_1, x_2, \cdots$, 就称 X 为离散型随机变量，简称离散随机变量，其分布函数称之为离散型的.

定义 1.13 (离散型随机变量的分布列或概率质量函数) 对于离散型随机变量 X, 称 X 取值 x_k 的概率

$$p_k := p(x_k) = P(X = x_k), k = 1, 2, \cdots,$$

为 X 的概率分布列或简称为分布列，记 $X \sim \{p_k\}$. 分布列也常用下面的矩阵来表示

$$\begin{pmatrix} x_1, & x_2, & \cdots, & x_k, & \cdots \\ p_1, & p_2, & \cdots, & p_k, & \cdots \end{pmatrix}$$

容易验证，分布列有以下性质

1. 非负性：$p_k \geq 0, k = 1, 2, \cdots$;
2. 正则性：$\sum_k p_k = 1$

- 由概率分布的定义，对任意的 $B \in \mathcal{B}$，我们有

$$\begin{aligned} \mathbf{F}(B) &= P(X \in B) = P(\cup_{k:x_k \in B}\{X = x_k\}) \\ &= \sum_{k:x_k \in B} P(X = x_k) = \sum_{k:x_k \in B} p_k \end{aligned}$$

- 由分布函数的定义知，

$$\begin{aligned} F(x) &= P(X \le x) = P(\cup_{k:x_k \le x}\{X = x_k\}) \\ &= \sum_{k:x_k \le x} P(X = x_k) = \sum_{k:x_k \le x} p_k \\ &= \sum_{k=1}^{\infty} p_k U(x - x_k) \end{aligned}$$

此处，$U(x) = \begin{cases} 1, & x \ge 0, \\ 0, & x < 0, \end{cases}$ 为阶跃函数.

- 易见离散型随机变量 X 的分布函数是一个纯跳跃函数：在 X 的每个可能取值 x_k 上有跃度 p_k，在每个不含 x_k 的区间上恒取常值.

退化随机变量

例 1.5 常数 c 可看作仅取一个值的随机变量 X, 即

$$P(X=c)=1$$

这个分布常称为 单点分布 或 退化分布，其分布函数为

$$F(x)=\begin{cases}0, & x<c\\ 1, & x\geq c\end{cases}$$

例 1.6 计算例 1.1 中的所有随机变量的分布列或概率质量函数.

- X 表示正面朝上的次数，其概率质量函数 p_X 为:

$$p_X(0) = P(X=0) = 1/4, \quad p_X(1) = P(X=1) = 1/2,$$
$$p_X(2) = P(X=2) = 1/4, \quad p_X(x) = P(X=x) = 0, x \neq 0, 1, 2.$$

- $Y = 2 - X$, 表示反面朝上的次数。注意到

$$P(Y=y) = P(2-X=y) = P(X=2-y) = p_X(2-y),$$

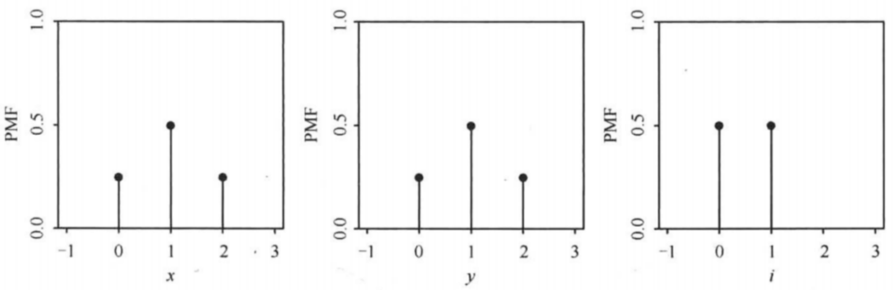
因此，随机变量 Y 的概率质量函数为

$$p_Y(0) = P(Y=0) = 1/4, \quad p_Y(1) = P(Y=1) = 1/2,$$
$$p_Y(2) = P(Y=2) = 1/4, \quad p_Y(y) = P(Y=y) = 0, y \neq 0, 1, 2.$$

- I 表示第一次是否正面朝上的示性随机变量.

$$p_I(0) = P(I=0) = 1/2, \quad p_I(1) = P(I=1) = 1/2,$$
$$p_I(i) = P(I=i) = 0, i \neq 0, 1.$$

X, Y 和 I 的概率质量函数图



例 1.7 掷两颗骰子，其样本空间 Ω 含有 36 个等可能的样本点

$$\Omega=\{(x,y):x,y=1,2,\cdots,6\}$$

令 X 和 Y 表示每个骰子分别出现的点数。试求下面随机变量的分布列:

1. $T_1:=X+Y=$ 骰子点数之和;
2. $T_2:=14-(X+Y)$;
3. $T_3:=$ 点数为 6 点的骰子的个数;
4. $T_4:=\max\{X,Y\}=$ 骰子的最大点数

- T_1, T_2 的概率分布列为

$$\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \frac{1}{36} & \frac{2}{36} & \frac{3}{36} & \frac{4}{36} & \frac{5}{36} & \frac{6}{36} & \frac{5}{36} & \frac{4}{36} & \frac{3}{36} & \frac{2}{36} & \frac{1}{36} \end{pmatrix}$$

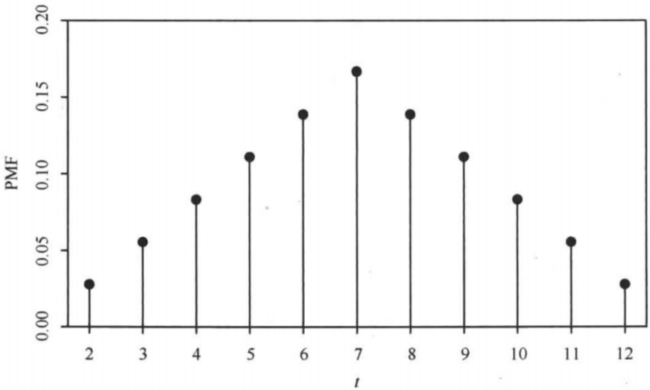


图 3.4　两个骰子的点数之和的概率质量函数。

- T_3 的概率分布列为

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{25}{36} & \frac{10}{36} & \frac{1}{36} \end{pmatrix}$$

- T_4 的概率分布列为

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{36} & \frac{3}{36} & \frac{5}{36} & \frac{7}{36} & \frac{9}{36} & \frac{11}{36} \end{pmatrix}$$

定义 1.14 (连续型随机变量) 设 X 为一随机变量，$F(x)$ 为随机变量 X 的分布函数，如果存在非负可积函数 $p(x)$ 使得

$$F(x) = \int_{-\infty}^{x} p(y)dy \tag{3}$$

则称 X 为连续型随机变量，其分布函数称之为连续型分布函数，函数 $p(x)$ 称为 X 的概率密度函数，简称密度函数或密度.

注 1.6

- 能够表为 (3) 式变上限积分的函数 $F(x)$ 在分析中称为绝对连续函数。绝对连续函数必为连续函数.
- 在若干个点上或零测集上改变密度函数 $p(x)$ 的值并不影响其积分的值，从而不影响分布函数 $F(x)$ 的值，这意味着连续分布的密度函数不唯一.

密度函数的性质

容易验证，随机变量 X 的密度函数有以下性质

1. 非负性：$p(x) \geq 0$;
2. 正则性：$\int_{-\infty}^{\infty} p(x)dx = 1$.

连续型随机变量的一些常见性质

- $p(x) = F'(x)$;
- $P(a < X \leq b) = F(b) - F(a) = \int_a^b p(x)dx$;
- $0 \leq P(X = a) \leq P(X \in (a - \epsilon, a)) = \int_{a-\epsilon}^a p(x)dx \xrightarrow{\epsilon \to 0} 0$, 故 $P(X = a) = 0$，即连续型随机变量取值单点的概率为 0;
- $P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X < b)$;
- 对任意的 Borel 集 B,

$$P(X \in B) = \int_B p(x)dx$$

- $P(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} p(y)dy = p(\xi)\Delta x \approx p(x)\Delta x$

例 1.8 定义函数 $F(x)$ 如下

$$F(x)=\begin{cases}0, & x<0,\\ \dfrac{1+x}{2}, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

试说明

1. $F(x)$ 为分布函数;
2. $F(x)$ 既非离散型也非连续型分布;
3. $F(x)$ 可分解为

$$F(x)=\frac{1}{2}F_1(x)+\frac{1}{2}F_2(x)$$

其中

$$F_1(x)=\begin{cases}0, & x<0,\\ 1, & x\ge 0.\end{cases}\qquad F_2(x)=\begin{cases}0, & x<0,\\ x, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}$$

定理 1.8 (勒贝格分解) 对任一分布函数 $F(x)$ 有如下分解

$$F(x)=c_1F_1(x)+c_2F_2(x)+c_3F_3(x),$$

其中常数 $c_1,c_2,c_3\geq 0, c_1+c_2+c_3=1$, 而 $F_1(x),F_2(x),F_3(x)$ 都是分布函数，$F_1(x)$ 为纯跳跃离散型函数，$F_2(x)$ 为连续型分布函数，$F_3(x)$ 为奇异型分布函数 (见教材 3.6 节奇异连续型分布构造).

- 上述定理中奇异函数的含义及定理的证明可参见一般的实变函数论教科书，这里我们不再详述，仅指出几种特殊情况:
  - 在分解式中取 $c_1=1,c_2=c_3=0$ 便得到我们所讨论的离散型分布函数;
  - 在分解式中取 $c_2=1,c_1=c_3=0$ 便得到连续型分布函数;
  - 若取 $c_3=0,c_1\neq 0,c_2\neq 0,c_1+c_2=1$ 便得到离散与连续混合分布
- 从上面分析可看出，随机变量除了离散型与连续型外还有很多其他类型.

第十一讲：常见的离散型随机变量

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

伯努利实验

- 只有两种可能结果的试验称为伯努利试验；例如抽检产品，可能是合格品，也可能是次品；掷两颗骰子，可能得到同点，也可能得到不同点，等等都是伯努利试验.
- 伯努利试验的样本空间 Ω 并不一定只含有两个样本点，有时只是把我们所关心的一部分样本点归结为一种结果 A, 同时把其余的样本点的集合看作另一种结果 $\bar{A}$;
- 在上述掷骰子的试验中，样本空间 Ω 共含有 36 个样本点，如果我们只关心同点是否发生，就可以把其中的 6 个样本点组成的事件 $A := \{(i,i) : i = 1, \cdots, 6\}$ 视为一种结果，而其余的 30 个样本点组成另一结果 $\bar{A} := \{$不同点$\}$;
- 此外我们不再关心由 Ω 的其他非空子集组成的事件，于是对于伯努利试验而言，事件 σ 代数应取为 $\mathcal{F} = \{\emptyset, A, \bar{A}, \Omega\}$;
- 通常把结果 A 称作" 成功", 而把 $\bar{A}$ 称作" 失败";
- 再取定成功失败的概率 $p = P(A), q = P(\bar{A})$ ($p > 0, q > 0$ 且 $p + q = 1$), 则建立了一次伯努利试验的概率空间 $(\Omega, \mathcal{F}, P)$.

- 在概率论的理论与应用中，经常以一系列独立重复的伯努利试验作为概率模型；
- 所谓重复，粗略的说即各次试验的概率空间都是上述的 $(\Omega, \mathcal{F}, P)$;
- 而 n 个试验的独立性则是指各次试验的结果互不影响，即对于第 i 次试验的任何结果 $E_i(i = 1, \cdots, n)$ 均有

$$P(E_1E_2\cdots E_n) = P(E_1)P(E_2)\cdots P(E_n)$$

- 将一次伯努利试验独立重复 n 次，称作 n 重伯努利试验;
- 将一次伯努利试验独立地重复下去所得到的一系列试验，称为可列重伯努利试验.

二项分布:n 重伯努利试验中成功的次数 X

- X:n 重伯努利试验中成功 (事件 A 发生) 的次数;
- X 的所有可能取值为: $0, 1, \cdots, n$;
- 下面我们考虑 X 的分布列
  - n 重伯努利试验的样本空间:
    $\Omega := \{\omega = (\omega_1, \cdots, \omega_n) : \omega_i \text{或者为} A \text{或者为} \bar{A}\}$
  - 样本空间样本点的个数为 2^n 个;
  - $\{X = k\} = \{\omega = (\omega_1, \cdots, \omega_n) : \omega_1, \cdots, \omega_n \text{中有} k \text{个} A\}$, 共包含 C_n^k 个样本点;
  - 若任给样本点 $\omega = (\omega_1, \cdots, \omega_n) \in \{X = k\}$, 则意味着 $\omega_1, \cdots, \omega_n$ 中有 k 个 A, $n-k$ 个 $\bar{A}$, 故由独立性可知

$$P(\omega) = p^k(1-p)^{n-k}$$

  - 而事件 $\{X = k\}$ 中共有 C_n^k 个类似的 ω, 故

$$P(X = k) = C_n^k p^k (1-p)^{n-k} := b(k; n, p), k = 0, 1, \cdots, n$$

    这个分布常称为二项分布，记为 $X \sim B(n, p)$.
  - 容易验证，$\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$

两点分布

- $n=1$ 时的二项分布 $B(1,p)$ 称为二点分布，或 $0-1$ 分布，或称伯努利分布，其分布列为

$$P(X=k)=p^k(1-p)^{1-k}, k=0,1$$

或记为

$$\begin{pmatrix} 0, & 1 \\ 1-p, & p \end{pmatrix}$$

- $B(1,p)$ 主要用于描述一次伯努利试验中成功的次数 (0 或 1);
- 若记 X_i 表示第 i 次伯努利试验中成功的次数，则 X_i 相互独立且有

$$X=X_1+X_2+\cdots X_n$$

即二项分布随机变量可写为 n 个独立同为两点分布随机变量的和.

二项分布的性质

- 对 $k \geq 1$, 考虑比值

$$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{C_n^k p^k (1-p)^{n-k}}{C_n^{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)}$$
$$= \frac{k(1-p)+(n+1)p-k}{k(1-p)} = 1 + \frac{(n+1)p-k}{k(1-p)}$$

- 当 $(n+1)p > k$ 时，$b(k;n,p) > b(k-1;n,p)$;
- 当 $(n+1)p < k$ 时，$b(k;n,p) < b(k-1;n,p)$;
- 从而，对于固定的 $n, p, \{X=k\}$ 的概率 $b(k;n,p)$ 先随 k 增大而增大，再随 k 增大而减小，故必有最大值:
  - 如果 $m := (n+1)p$ 为整数，则 $b(m;n,p) = b(m-1;n,p)$ 同为 $b(k;n,p)$ 的最大值
  - 如果 $(n+1)p$ 不为整数，则 $b(k;n,p)$ 在 $m = [(n+1)p]$ 处取到最大值 (此处 $[a]$ 表示不超过 a 的最大整数)
- 我们称使得 $b(k;n,p)$ 取得最大值的 m 为二项分布随机变量的最可能值，或称为 n 重伯努利试验中最可能的成功次数.

定理 1.9 设 $X \sim B(n,p)$, 且 $q = 1-p$(通常用 q 表示伯努利试验失败的概率), 则有 $n-X \sim B(n,q)$.

- 借用二项分布的直观定义：将 X 为 n 次独立伯努利试验成功的次数，则 $n-X$ 为这些试验中失败的次数.
- 互相交换成功与失败的角色，可知 $n-X \sim B(n,q)$.
- 也可从分布列 (概率质量函数) 的角度出发得到 $n-X \sim B(n,q)$.
- 令 $Y = n-X$, 则 Y 的分布列 (概率质量函数) 为

$$\begin{aligned} P(Y=k) =& P(n-X=k) = P(X=n-k) \\ =& \binom{n}{n-k} p^{n-k} q^k = \binom{n}{k} q^k p^{n-k}, \end{aligned}$$

定理 1.10 设 $X \sim B(n,p)$, 其中 n 为偶数，$p = 1/2$, 则 X 的分布关于 $n/2$ 对称，即对任意的非负整数 j, 均有

$$P(X = \frac{n}{2} + j) = P(X = \frac{n}{2} - j).$$

解: 由定理 1.9 可知，$n - X$ 同样服从 $B(n, 1/2)$. 因此对任意非负整数 k 均有

$$P(X = k) = P(n - X = k) = P(X = n - k).$$

令 $k = n/2 + j$, 即可得证.

例 1.9 设每台自动机床在运行过程中需要维修的概率均为 $p=0.01$, 并且各机床是否需要维修相互独立。如果:

1. 每名维修工人负责看管 20 台机床;
2. 3 名维修工人负责看管 80 台机床;

求机床不能及维修的概率.

解: 1. 这是 $n=20$ 重伯努利试验，参数 $p=0.01$, 故需要维修的机床数 X 服从 $B(20,0.01)$ 分布。故不能及时维修的概率为

$$\begin{aligned} P(X>1) &= 1-P(X\leq 1)=1-P(X=0)-P(X=1) \\ &= 1-C_{20}^{0}0.01^{0}(1-0.01)^{20}-C_{20}^{1}0.01(1-0.01)^{20-1}\approx 0.0169 \end{aligned}$$

2. 此时需要维修的机床数 X 服从 $B(80,0.01)$ 分布，类似可得不能及时维修的概率为

$$P(X>3)=1-\sum_{k=0}^{3}b(k;80,0.01)\approx 0.0087$$

小概率事件终将发生

例 1.10 在可列重伯努利试验中，求事件 $E := \{$试验终将成功$\}$ 的概率.

解: 考虑所求概率事件的反面即 $\overline{E} := \{$试验永不成功$\}$.若我们记

$$F_n := \{\text{前}n\text{次试验均失败}\},$$

则易知，$\{F_n\}$ 为单调下降事件序列，且

$$\lim_{n\to\infty} F_n = \cap_{n=1}^{\infty} F_n = \overline{E}$$

从而

$$P(\overline{E}) = P(\lim_{n\to\infty} F_n) = \lim_{n\to\infty} P(F_n) = \lim_{n\to\infty} C_n^0 p^0 (1-p)^n = 0$$

故

$$P(E) = 1 - P(\overline{E}) = 1 - 0 = 1$$

无论成功的概率有多小，但是试验最终成功的概率为 1, 也就是说小概率事件终将发生的概率为 1.

- 记 X 为可列重伯努利试验中首次成功的等待时间即首次成功所需要试验的次数;
- $\{X=k\}=\{\underbrace{\overline{A}\cdots\overline{A}}_{k-1\text{个}}A\}$, 故

$$P(X=k)=(1-p)^{k-1}p:=g(k;p),k=1,\cdots,$$

这个分布常称为几何分布，记为 $X\sim G(p)$.

定理 1.11 取值自然数的随机变量 X 为几何分布当且仅当 X 有无记忆性:

$$P(X>m+n|X>m)=P(X>n),\text{对任意的}m,n\geq 1. \tag{4}$$

证明: 若 X 为几何分布，则

$$P(X>m+n|X>m)=\frac{P(X>m+n,X>m)}{P(X>m)}=\frac{P(X>m+n)}{P(X>m)}$$

而

$$P(X>n)=\sum_{k=n+1}^{\infty}(1-p)^{k-1}p=\frac{(1-p)^n p}{1-(1-p)}=(1-p)^n$$

故

$$P(X>m+n|X>m)=\frac{(1-p)^{m+n}}{(1-p)^m}=(1-p)^n=P(X>n)$$

若 X 具有无记忆性，则由 ~ (4) 知

$$Q_n := P(X > n) > 0, \quad \text{对任意的} n \geq 1$$

并且有

$$Q_{m+n} = P(X > m+n) = P(X > m)P(X > m+n | X > m) = Q_m Q_n$$

从而

$$Q_m = Q_1^m$$

注意到 $Q_1 \in (0,1)$, 事实上，

- $Q_1 = P(X > 1) > 0$ 显然；
- 若 $Q_1 = 1$, 则对一切的 m 均有 $Q_m = P(X > m) = 1$, 这与 X 取自然数矛盾，故 $Q_1 \in (0,1)$.

故取 $p = 1 - Q_1 \in (0,1)$, 且对任意的 $k \geq 1$ 有

$$\begin{aligned} P(X = k) &= P(X > k-1) - P(X > k) = Q_{k-1} - Q_k \\ &= (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1} p \end{aligned}$$

无记忆性

$$P(X > m+n | X > m) = P(X > n)$$

- 上述的无记忆性表明：已知试验了 m 次未获得成功，再加做 n 次试验仍不成功的概率，等于从开始算起做 n 次试验都不成成功的概率.
- 换句话放，已做过的 m 次失败的试验被忘记了；
- 产生几何分布的这种无记忆性的根本原因在于，我们进行的是独立重复试验，这是不学习，不总结经验的试验，已经做过的试验当然不会留下记忆.

例 1.11 10 把外形相同的钥匙中只有一把能打开门。现一一试开，试对每次试毕放回与不放回两种情形，分别求事件
$E :=$ {至多试3次能打开门} 的概率.

解: 1. 放回情形是独立重复试验，属伯努利概型。以 X 表示首次打开门的等待时间，则 X 服从几何分布 $G(0.1)$. 故所求概率为

$$P(E) = P(X \le 3) = \sum_{k=1}^{3} (1-0.1)^{k-1} 0.1 = 0.271$$

2. 不放回情形不再是独立重复试验，适用于古典概型。样本点总数 $n(\Omega) = C_{10}^3$, 而 $n(\overline{E}) = C_9^3$. 故

$$P(E) = 1 - P(\overline{E}) = 1 - \frac{C_9^3}{C_{10}^3} = 0.3$$

或令 $A_i :=$ {第i次取到能开门的钥匙}, 则 A_1, A_2, A_3 互不相容，由可加性及抽签的公平性可得

$$P(E) = P(\cup_{i=1}^3 A_i) = \sum_{i=1}^{3} P(A_i) = 0.3$$

- 记 X_r 为可列重伯努利试验中第 r 次成功的等待时间即第 r 次成功所需要试验的次数;
- 易见 X 的所有可能取值为 $k=r,r+1,\cdots$，并且有

$$\begin{aligned}
\{X_r=k\} &= \{\text{前}k-1\text{次试验恰有}r-1\text{次成功且第}k\text{次成功}\} \\
P(X_r=k) &= P(\text{前}k-1\text{次试验恰有}r-1\text{次成功})P(\text{第}k\text{次成功}) \\
&= C_{k-1}^{r-1}p^{r-1}(1-p)^{k-1-(r-1)}p \\
&= C_{k-1}^{r-1}p^{r}(1-p)^{k-r} := f(k;r,p), \quad k=r,r+1,\cdots
\end{aligned}$$

  这个分布常称为帕斯卡 (Pascal) 分布或负二项分布.
- $f(k;r,p), k=r,r+1,\cdots$，可以成为离散型分布的密度，事实上: $f(k;r,p)>0$ 显然，其和

$$\begin{aligned}
\sum_{k=r}^{\infty} f(k;r,p) &= \sum_{k=r}^{\infty} C_{k-1}^{r-1}p^{r}(1-p)^{k-r} \xlongequal[q=1-p]{k-r=i} \sum_{i=0}^{\infty} C_{r+i-1}^{r-1}p^{r}q^{r+i-r} \\
&= \sum_{i=0}^{\infty} C_{r+i-1}^{i}p^{r}q^{i} = \sum_{i=0}^{\infty} C_{-r}^{i}p^{r}(-q)^{i} = p^{r}(1-q)^{-r} = 1
\end{aligned}$$

帕斯卡分布的性质

- 若帕斯卡分布中的 $r=1$, 则此时的帕斯卡分布即为几何分布;
- 如果记

$$\tau_1 = X_1, \quad \tau_n = X_n - X_{n-1}, \quad n > 1,$$

则随机变量 τ_n 是第 $n-1$ 次成功到第 n 次成功的间隔时间。显然有

$$X_r = \tau_1 + \tau_2 + \cdots + \tau_r$$

- 以后我们会看到: $\tau_1, \cdots, \tau_r$ 是 r 个相互独立的随机变量且每个 τ_k 均服从几何分布.

例 1.12 某人口袋中有两盒火柴，开始时每盒各装 n 根。每次他从口袋中任取一盒使用其中的一根火柴。求此人掏出一盒发现已空，而另一盒尚余 r 根的概率.

解: 记

$$E = \{\text{掏出甲盒已空而乙盒尚余}r\text{根}\}$$

则由对称性可知所求概率为 $2P(E)$. 若我们以取出甲盒为" 成功", 这便是一个成功率 $p = 1/2$ 的独立重复伯努利试验. 而

$$E = \{\text{第}n+1\text{次成功发生在第}2n-r+1\text{次试验}\}$$

故所求概率为

$$\begin{aligned} 2P(E) &= 2f(2n-r+1; n+1, 1/2) = 2C_{2n-r}^{n}\left(\frac{1}{2}\right)^{n+1}\left(\frac{1}{2}\right)^{2n-r-n} \\ &= C_{2n-r}^{n}2^{r-2n} \end{aligned}$$

例 1.13 1654 年，当时的职业赌徒 DeMere 爵士向法国的大数学家 Pascal 提出如下问题：甲乙两人各下赌注 m 元，商定先胜三局者取得全部赌金。假定在每一局中二人获胜的机会相等，且各局胜负相互独立。如果当甲胜一局而乙胜零局时赌博被迫中止，问赌注如何分？

- 为解决这个问题，Pascal 与当时声望很高的数学家 Fermat 建立了通信联系。他们进行了卓有成效的讨论，不仅完满的回答了分赌注问题，而且为解决其他概率问题建立起了框架，极大的促进了概率论的建立与发展；
- Pascal 令人信服的指出，赌金的分法应当取决于若赌博能继续进行下去甲乙各自获胜的概率，这个概率即为在 $p = 0.5$ 的可列重伯努利试验中 2 次成功发生在 3 次失败之前的概率；
- 更一般的，下面我们求一下 n 次成功发生在 m 次失败之前的概率.

例 1.14 在可列重伯努利试验中，求下面事件的概率:

$$E = \{n\text{次成功发生在}m\text{次失败之前}\}$$

解: 记 $F_k = \{\text{第}n\text{次成功发生在第}k\text{次试验}\}$, 则

$$E = \cup_{k=n}^{n+m-1} F_k$$

从而由 F_k 的互不相容性可得

$$P(E) = \sum_{k=n}^{n+m-1} P(F_k) = \sum_{k=n}^{n+m-1} C_{k-1}^{n-1} p^n (1-p)^{k-n}$$

利用上面的公式可计算 $n=2, m=3, p=1/2$ 时，其相应的概率为

$$P(\text{甲胜}) = P(E) \xlongequal{n=2,m=3,p=1/2} \frac{11}{16}$$

故赌注应以 $11:5$ 的比例分配给甲乙两人.

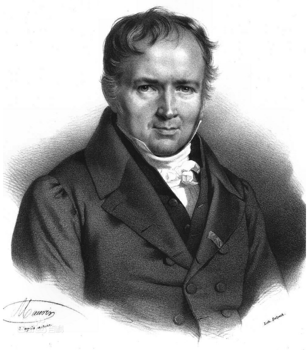


图 1: 泊松

- 西莫恩 德尼 泊松：法国数学家、几何学家和物理学家；
- 泊松的科学生涯开始于研究微分方程及其在摆的运动和声学理论中的应用；
- 对积分理论、行星运动理论、热物理、电磁理论、位势理论和概率论都有重要贡献；
- 19 世纪概率统计领域里的卓越人物，改进了概率论的运用方法，特别是用于统计方面的方法，建立了描述随机现象的一种概率分布 泊松分布；
- 推广了“大数定律”，并导出了在概率论与数理方程中有重要应用的泊松积分。

泊松 (Poisson) 分布

- Poisson 分布是概率论中一种重要的离散型分布，它在理论与实践中都有广泛的应用，常与单位时间 (面积) 内上的计数过程相联系;
    - 一天内到达某商场的顾客数;
    - 单位时间内，电路受外界电磁波的冲击次数;
    - 一定时期内，某放射性物质放射出来的粒子数等
- 在二项分布中，当参数 n 较大时，计算二项概率 $b(k;n,p)$ 会非常麻烦.

定理 1.12 设有一列二项分布 $\{b(k;n,p_n)\}$, 若其参数列 p_n 满足

$$\lim_{n\to\infty} np_n = \lambda > 0,$$

则对任何非负整数 k 有

$$\lim_{n\to\infty} b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

证明: 记 $\lambda_n := np_n$, 则

$$b(k;n,p_n) = C_n^k p_n^k(1-p_n)^{n-k} = \frac{n!}{k!(n-k)!}(\frac{\lambda_n}{n})^k(1-\frac{\lambda_n}{n})^{n-k}$$

$$= \frac{\lambda_n^k}{k!}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})(1-\frac{\lambda_n}{n})^{n-k}$$

注意到

$$\lim_{n\to\infty}\lambda_n^k = \lambda^k,$$

$$\lim_{n\to\infty}(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n}) = 1$$

$$\lim_{n\to\infty}(1-\frac{\lambda_n}{n})^{n-k} = \lim_{n\to\infty} e^{(n-k)\ln(1-\frac{\lambda_n}{n})}$$

$$= e^{\lim_{n\to\infty}(n-k)\ln(1-\frac{\lambda_n}{n})} = e^{\lim_{n\to\infty}(n-k)(-\frac{\lambda_n}{n}+o(\frac{1}{n}))}$$

$$= e^{\lim_{n\to\infty}(-\lambda_n+\frac{k\lambda_n}{n}+(n-k)o(\frac{1}{n}))} = e^{-\lambda}$$

从而

$$\lim_{n\to\infty} b(k;n,p_n) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0,1,2,\cdots$$

有了上述定理，当 n 很大而 p 很小时，可以用近似公式计算二项概率

$$b(k;n,p) \approx \frac{(np)^k}{k!}e^{-np}$$

这里我们要求 p 很小，为保证 n 很大时，乘积 np 有适度的大小.

定义 1.15 (泊松分布) 对参数 $\lambda > 0$, 记

$$p(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, \cdots .$$

易见 $p(k;\lambda) > 0$ 且

$$\sum_{k=0}^{\infty} p(k;\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda}e^{-\lambda} = 1$$

即 $\{p(k;\lambda)\}$ 可以看成离散型分布的密度，我们就把它称为以 λ 为参数的泊松分布，记作 $P(\lambda)$.

泊松分布的性质

- 泊松分布列 $p(k;\lambda)$ 随 k 变化情况与二项分布相似，事实上，考虑比值

$$\frac{p(k;\lambda)}{p(k-1;\lambda)} = \frac{\lambda^k (k-1)! e^{-\lambda}}{\lambda^{k-1} k! e^{-\lambda}} = \frac{\lambda}{k}, k \geq 1$$

  - 当 $k < \lambda$ 时有，$p(k;\lambda) > p(k-1;\lambda)$;
  - 当 $k > \lambda$ 时有，$p(k;\lambda) < p(k-1;\lambda)$;
  - 因此 $p(k;\lambda)$ 随 k 先升后降，在 $m = [\lambda]$ 处达到最大值，而 λ 为整数时，$p(k;\lambda)$ 在 $m = \lambda, \lambda - 1$ 处同时取到最大值.
- 泊松分布在随机选择下的不变性：假设某块放射性物质在单位时间内发射出的粒子数 X 服从 $P(\lambda)$ 分布。而每个粒子被记录下来的概率为 p 即粒子有 $1-p$ 的概率被计数器遗漏，如果各粒子是否被记录相互独立，试求记录下的粒子数 Y 的分布.

$$
\begin{aligned}
P(Y=k) &= \sum_{n=0}^{\infty} P(X=n, Y=k) = \sum_{n=0}^{\infty} P(X=n)P(Y=k|X=n) \\
&= \sum_{n=k}^{\infty} P(X=n)P(Y=k|X=n) = \sum_{n=k}^{\infty} p(n;\lambda)B(k;n,p) \\
&= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} C_n^k p^k (1-p)^{n-k} = \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} \frac{1}{k!} e^{-\lambda} (\lambda p)^k \\
&= \frac{1}{k!} e^{-\lambda p} (\lambda p)^k
\end{aligned}
$$

超几何分布

假设某个罐子中有 w 个白球和 b 个黑球，考虑如下两种取球方式

- 有放回的抓取 n 个球
  - $X := n$ 个球中白球的数量;
  - $X \sim B(n, \frac{w}{w+b})$
- 不放回的抓取 n 个球
  - $X := n$ 个球中白球的数量;
  - X 服从超几何分布，记为 $X \sim H(w, b, n)$;
  - 易知

$$P(X=k) = \frac{C_w^k C_b^{n-k}}{C_{w+b}^n} := h(k, w, b, n), k = 0, 1, \cdots, r,$$

    其中 $r := \min\{w, n\}$.
  - 若要验证以上给出的确实为一个概率分布列，只需注意到下面的组合等式成立

$$\sum_{k=0}^{r} C_w^k C_b^{n-k} = C_{w+b}^n$$

超几何分布：两个例子

- 不合格品抽检问题
  - 设有 N 件产品，其中有 M 件不合格品，从中不放回的抽取 n 件;
  - 令 $X := n$ 件抽取的产品中不合格品的件数;
  - $X \sim H(M, N-M, n)$
- 麋鹿的捕获 - 再捕获问题
  - 森林中共有 N 头麋鹿。某天，捕获了 M 头麋鹿，标记后将这 M 头麋鹿再放回野外.
  - 几天后，又重新随机地捕获 n 头麋鹿。假设重新捕获的麋鹿也同样可能是之前捕获的麋鹿.
  - 令 $X :=$ 再次被捕获的麋鹿的数量;
  - $X \sim H(M, N-M, n)$;
  - 第一次捕获的 M 头麋鹿相当于前面例子中白球的总数，第一次未捕获的 $N-M$ 头麋鹿相当于前面例子中黑球的总数，再次被捕获的 n 头麋鹿相当于前面例子中抽样的数量.

超几何分布的适用基础

- 除上面例子外，超几何分布还可出现在许多情况下;
- 超几何分布的适用基础是总体根据两套标签进行分类:
  - 在罐子的示例中，每个球不是白色就是黑色 (第一套标签);
  - 每个球要么是样本要么不是样本 (第二套标签);
- 两套标签中至少有一个是被完全随机分配的 (在罐子的例子中，球是随机抽样的);
- X 代表被两套标签都标记的数量：在罐子的例子中，关注的是既被抽样又是白色的球。
- X 服从超几何分布.

超几何分布之间的对称性

定理 1.13 $X \sim H(w,b,n), Y \sim H(n,w+b-n,w)$, 则 X 和 Y 是同分布的.

证明:

- 考虑一个由 w 个白球和 b 个黑球充满的罐子，现在随机不放回地从罐子里抓取 n 个球;
- 白球或者黑球看作第一套标签，有没有被抽取看作第二套标签;
- X 表示抽取的样本球中白球的数量 $\sim H(w,b,n)$;
- 若将是否被抽取看作第一套标签，白球或者黑球看作第二套标签;
- Y 表示所有白球中被抽中的数量 $\sim H(n,w+b-n,w)$;
- 显然 X 和 Y 都是表示被抽取的白球数量，所以它们有相同的分布;
- 也可以用代数的方法来检查 X 和 Y 是否具有相同的概率质量函数，即验证 $P(X=k)=P(Y=k)$.

定理 1.14 如果 $X \sim B(n,p), Y \sim B(m,p)$, 且 X 和 Y 相互独立，则当给定条件 $X+Y=r$ 时，X 的条件分布为超几何分布 $H(n,m,r)$.

定理 1.15 如果 $X \sim H(M,b,n)$, 且当 $N=M+b \to \infty$ 时 $p=\dfrac{M}{N}$ 保持不变，则 X 的分布列 (概率质量函数) 收敛到 $B(n,p)$ 的分布列.

注意到

$$h(k;M,b,n)=\frac{C_M^k C_b^{n-k}}{C_{M+b}^n}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

$$=\frac{\dfrac{M!}{k!(M-k)!}\dfrac{(N-M)!}{(n-k)!(N-M-(n-k))!}}{\dfrac{N!}{n!(N-n)!}}$$

$$=\frac{n!}{k!(n-k)!}\frac{M!}{(M-k)!}\frac{(N-M)!}{(N-M-(n-k))!}\frac{(N-n)!}{N!}$$

$$=\frac{n!}{k!(n-k)!}\frac{M(M-1)\cdots(M-(k-1))}{N(N-1)\cdots(N-(k-1))}\frac{(N-M)\cdots(N-M-(n-k)+1)}{(N-k)\cdots(N-n+1)}$$

$$=\frac{n!}{k!(n-k)!}\frac{\frac{M}{N}(\frac{M}{N}-\frac{1}{N})\cdots(\frac{M}{N}-\frac{(k-1)}{N})}{1(1-\frac{1}{N})\cdots(1-\frac{(k-1)}{N})}\frac{(1-\frac{M}{N})\cdots(1-\frac{M}{N}-\frac{(n-k)-1}{N})}{(1-\frac{k}{N})\cdots(1-\frac{n-1}{N})}$$

$$\to C_n^k\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}=b(k;n,p)$$

第十二讲：常见的连续型分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

均匀分布

定义 1.16 任给参数 $a<b$，函数

$$p(x)=\frac{1}{b-a},\quad a<x<b$$

满足密度函数的两个性质即:$p(x)\geq 0$ 且 $\int_{-\infty}^{\infty}p(x)dx=1$. 我们称以上式中的 $p(x)$ 为密度的连续型分布为区间 (a,b) 上的均匀分布，记作 $U(a,b)$.

- 易见，均匀分布的分布函数为

$$F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\leq x<b\\ 1. & x\geq b\end{cases}$$

- 如果 X 服从 $U(a,b)$ 分布，则对任何 $a\leq x<y\leq b$ 有

$$P(x<X\leq y)=\int_x^y\frac{1}{b-a}dt=\frac{y-x}{b-a}$$


GU5672972S2
DEUTSCHE BUNDESBANK
Banknote
10
ZEHN DEUTSCHE MARK
1777–1855 Carl Friedr. Gauß
Deutsche Bundesbank
Frankfurt am Main
1. September 1999
GU5672972S2
ZEHN DEUTSCHE MARK
10
Banknote
Zehn
Deutsche Mark
10
DEUTSCHE BUNDESBANK

图 2: 高斯

- 卡尔。弗里德里希。高斯 1777 年出生于布伦瑞克;
- 11 岁时就发现了二项式定理;
- 19 岁发现正十七边形的尺规作图法，解决了困扰数学家们两千多年的难题;
- 1809 年提出“最小二乘法”并在此基础上建立的正态分布方程，是概率统计中一个非常重要的工具，广泛应用于数学、物理学等领域;
- 一生发表了 155 篇论文，对数论、代数学、非欧几何、复变函数和微分几何等领域都做出了开创性的贡献，被誉为数学王子.

正态分布 (高斯分布)

考虑下述函数

$$p_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

下面我们验证 $p_{\mu,\sigma}(x)$ 满足密度函数的两个性质即非负性与正则性：非负性显然；正则性则需要计算积分 $I := \int_{-\infty}^{\infty} p_{\mu,\sigma}(x)dx$.

由于 $p_{\mu,\sigma}(x)$ 的原函数不是初等函数，我们无法通过微积分基本公式计算 I. 我们考虑计算 I^2:

$$\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy \\
&= \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdrd\theta \\
&= 1
\end{aligned}$$

正态分布的定义

定义 1.17 若随机变量 X 的密度函数为

$$p_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

则称 X 服从参数为 (μ, σ^2) 的正态分布，记作 $X \sim N(\mu, \sigma^2)$. 特别的称 $N(0, 1)$ 分布为标准正态分布，并简写 $p_{0,1}(x)$ 为 $\varphi(x)$. 正态分布又称高斯分布.

正态分布的性质

- 分布密度关于参数 μ 对称，即有

$$p_{\mu,\sigma}(\mu - x) = p_{\mu,\sigma}(\mu + x), \quad \forall x \in R$$

特别的，$N(0,1)$ 分布的密度函数为偶函数.

- $p_{\mu,\sigma}(x)$ 在 $x=\mu$ 处取得最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$. 注意到密度曲线下的面积应保持等于 1, 并且密度函数在 x 处的值反映了此分布取值 x 附近的概率大小。故 σ^2 越小，密度的曲线越尖陡，分布取值越集中；σ^2 越大，密度曲线越平缓，分布取值越分散.
- 正态分布的分布函数为

$$\Phi_{\mu,\sigma}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

特别的记标准正态分布函数为 $\Phi(x)$. 其与 $\Phi_{\mu,\sigma}(x)$ 有以下关系:

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

例 1.15 (3σ 原则) 设随机变量 X 服从 $N(\mu, \sigma^2)$ 分布，求 $P(|X-\mu| < 3\sigma)$.

解: 注意到

$$
\begin{aligned}
P(|X-\mu| < 3\sigma) &= P(\mu - 3\sigma < X < \mu + 3\sigma) \\
&= \Phi_{\mu,\sigma}(\mu + 3\sigma) - \Phi_{\mu,\sigma}(\mu - 3\sigma) \\
&= \Phi(3) - \Phi(-3) = \Phi(3) - (1 - \Phi(3)) \\
&= 2\Phi(3) - 1 = 0.9973
\end{aligned}
$$

这表明，尽管正态分布的取值范围是 $(-\infty, +\infty)$, 但是其取值落在区间 $(\mu - 3\sigma, \mu + 3\sigma)$ 的概率高达 99.73%

例 1.16 若 $X \sim N(\mu, \sigma^2)$, 则 $Y := \dfrac{X-\mu}{\sigma} \sim N(0,1)$; 反之，若 $X \sim N(0,1)$, 则 $Y := \sigma X + \mu \sim N(\mu, \sigma^2)$.

解: 我们仅给出第一种情况的证明。事实上，我们只需要说明 Y 的分布密度为

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

即可. 事实上，考虑 Y 的分布函数

$$\begin{aligned} F_Y(x) &= P(Y \le x) = P\left(\frac{X-\mu}{\sigma} \le x\right) = P(X \le \sigma x + \mu) \\ &= \int_{-\infty}^{\sigma x+\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \end{aligned}$$

从而 Y 的分布密度为

$$p_Y(x) = F_Y'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x)$$

正态分布函数的简单性质

记 $\Phi(x)$ 为标准正态分布的分布函数，$U \sim N(0,1)$，$\Phi_{\mu,\sigma}(x)$ 为 $N(\mu,\sigma^2)$ 分布的分布函数，$X \sim N(\mu,\sigma^2)$，则

- $\Phi(0) = \dfrac{1}{2}$;
- $\Phi_{\mu,\sigma}(\mu) = \dfrac{1}{2}$;
- $\Phi(-x) = 1 - \Phi(x)$;
- $P(|U| < x) = P(-x < U < x) = P(U < x) - P(U \le -x)$
  $= \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1$;
- $\Phi_{\mu,\sigma}(x) = P(X \le x) = P(\dfrac{X-\mu}{\sigma} \le \dfrac{x-\mu}{\sigma}) = \Phi(\dfrac{x-\mu}{\sigma})$
- $P(a < X \le b) = \Phi(\dfrac{b-\mu}{\sigma}) - \Phi(\dfrac{a-\mu}{\sigma})$.

定义 1.18 称以下函数

$$\Gamma(\alpha) := \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \alpha > 0,$$

为 Gamma 函数. Gamma 函数有以下性质:

- $\Gamma(1) = \int_0^{\infty} x^{1-1} e^{-x} dx = 1;$
- $\Gamma(\frac{1}{2}) = \int_0^{\infty} x^{\frac{1}{2}-1} e^{-x} dx = 2\sqrt{\pi} \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x} d\sqrt{2} x^{\frac{1}{2}}$
  $= 2\sqrt{\pi} \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \sqrt{\pi};$
- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, 故 $\Gamma(n+1) = n\Gamma(n) = \cdots = n!$; 事实上,

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^{\infty} x^{\alpha} e^{-x} dx = \int_0^{\infty} -x^{\alpha} de^{-x} \\ &= -x^{\alpha} e^{-x}|_0^{\infty} + \int_0^{\infty} \alpha x^{\alpha-1} e^{-x} dx = \alpha\Gamma(\alpha)\end{aligned}$$

定义 1.19 (Gamma 分布) 若随机变量 X 的密度函数为

$$p(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称 X 服从参数为 (α, λ) 的 Gamma 分布，记作 $X \sim \Gamma(\alpha, \lambda)$.

定义 1.20 称 $\Gamma(1,\lambda)$ 分布为指数分布，其密度函数与分布函数分别为

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

定义 1.21 称 $\Gamma(\frac{n}{2},\frac{1}{2})$ 分布为 χ^2 分布，记作 $\chi^2(n)$, 其密度函数为

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

例 1.17 假定有一个于随机时刻陆续到来的质点流，我们若以 $N(t)$ 表示在 $[0,t]$ 时间内到来的质点的个数，并假设其服从参数为 λt 的泊松分布。试证明第 n 个质点到达的时间 S_n 服从 $\Gamma(n,\lambda)$ 分布.

证明: 首先我们考虑一下 S_n 的分布函数

$$\begin{aligned} P(S_n \le t) &= P(N(t) \ge n) = \sum_{k=n}^{\infty} P(N(t)=k) \\ &= \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

故其密度 $p(t)$ 为

$$\begin{aligned} p(t) = F'(t) &= \sum_{k=n}^{\infty} \left[\frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} - \frac{\lambda(\lambda t)^k}{k!} e^{-\lambda t} \right] \\ &= \frac{\lambda(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \end{aligned}$$

第十三讲：多维概率分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

November 4, 2024

定义 1.22 (n 维随机向量或随机变量) 如果 $X_1, \cdots, X_n$ 是定义在同一个概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 个随机变量，即

$$\{\omega : X_i(\omega) \in B_i\} \in \mathcal{F}, \quad 任意的B_i \in \mathcal{B},$$

则称 $X(\omega) := (X_1(\omega), \cdots, X_n(\omega))$ 为概率空间 $(\Omega, \mathcal{F}, P)$ 上的 n 维随机向量或 n 维随机变量.

注 1.7

- 如果 $(X_1, \cdots, X_n)$ 是一个 n 维随机向量，则对任何不超过 n 的正整数 k 和 $1 \leqslant j_1 < \cdots < j_k \leqslant n$, $\left(X_{j_1}, \cdots, X_{j_k}\right)$ 都是一个 k 维随机向量;
- 特别地，当 $k = 1$ 时就是随机变量.

定理 1.16 $X(\omega)=(X_1(\omega),\cdots,X_n(\omega))$ 是定义在某个概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 维随机向量当且仅当对任意的 $(x_1,\cdots,x_n)\in\mathbb{R}^n$ 均有

$$\{\omega:X_1(\omega)\le x_1,\cdots,X_n(\omega)\le x_n\}=\{X_1\le x_1,\cdots,X_n\le x_n\}\in\mathcal{F},$$

或等价的有对任意的 $B\in\mathcal{B}(\mathbb{R}^n)$

$$\{\omega:X(\omega)\in B\}=\{\omega:(X_1(\omega),\cdots,X_n(\omega))\in B\}\in\mathcal{F}.$$

定义 1.23 (多维分布) 设 $X=(X_1,\cdots,X_n)$ 是定义在某个概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 维随机向量，则对任意的 $B\in\mathcal{B}(\mathbb{R}^n)$，称

$$\mathbf{F}(B):=P(X\in B)=P((X_1,\cdots,X_n)\in B)$$

为 n 维随机向量 $X=(X_1,\cdots,X_n)$ 的分布.

定义 1.24 (分布函数或联合分布函数) 设 $X=(X_1,\cdots,X_n)$ 是定义在某个概率空间 $(\Omega,\mathcal{F},P)$ 上的 n 维随机向量，称

$$F\left(x_1,\cdots,x_n\right)=P\left(X_1\leqslant x_1,\cdots,X_n\leqslant x_n\right),\quad\forall\left(x_1,\cdots,x_n\right)\in\mathbb{R}^n$$

为随机向量 X 的分布函数，也称为随机变量 $X_1,\cdots,X_n$ 的联合分布函数.

注 1.8

- n 维随机向量的分布函数 $F(x_1,\cdots,x_n)$ 是定义在 $\mathbb{R}^n$ 上的 n 元函数;
- 称一个 n 元函数 $F(x_1,\cdots,x_n)$ 为 n 维分布函数，如果存在某个随机向量以它作为分布函数.

定理 1.17 n 维随机向量 $X=(X_1,\cdots,X_n)$ 的分布函数 $F(x_1,\cdots,x_n)$ 具有下述性质:

1. $F(x_1,\cdots,x_n)$ 对每个变元非降;
2. $F(x_1,\cdots,x_n)$ 对每个变元右连续;
3. 对任意的 $1\leq j\leq n$,
$$\lim_{x_j\to-\infty} F\left(x_1,\cdots,x_j,\cdots,x_n\right)=0,$$
$$\lim_{x_1\to\infty,\cdots,x_n\to\infty} F(x_1,\cdots,x_n)=1$$
4. $F(x_1,\cdots,x_n)$ 具有增量非负性: 对任意的 $1\leq j\leq n$ 及 $a_j\leqslant b_j$ 均有
$$\Delta_{(a_1,\cdots,a_n)}^{(b_1,\cdots,b_n)}F=\sum_{\boldsymbol{x}=(x_1,\cdots,x_n)\in\mathcal{O}}\operatorname{sgn}(\boldsymbol{x})F(\boldsymbol{x})\geqslant 0,$$
其中
$$\mathcal{O}:=\{(x_1,\cdots,x_n):x_j=a_j\text{或}b_j,1\leq j\leq n\},$$
$$\operatorname{sgn}(\boldsymbol{x})=\begin{cases}1, & |\{j:x_j=a_j\}|\text{为偶数},\\ -1, & |\{j:x_j=a_j\}|\text{为奇数}.\end{cases}$$

特别的，当 $n=2$ 或 3 时，$\Delta_{(a_1,\cdots,a_n)}^{(b_1,\cdots,b_n)}F$ 有以下具体形式:

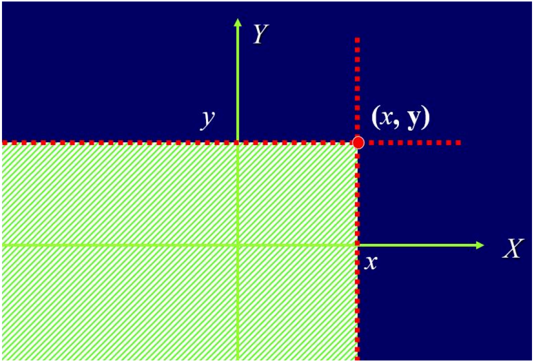
$$\Delta_{(a_1,a_2)}^{(b_1,b_2)}F = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2)$$

$$\begin{aligned}\Delta_{(a_1,a_2,a_3)}^{(b_1,b_2,b_3)}F = {} & F(b_1,b_2,b_3) - F(a_1,b_2,b_3) - F(b_1,a_2,b_3) \\ & - F(b_1,b_2,a_3) + F(a_1,a_2,b_3) + F(a_1,b_2,a_3) \\ & + F(b_1,a_2,a_3) - F(a_1,a_2,a_3)\end{aligned}$$

定义 1.25 设 (X, Y) 为 $(\Omega, \mathcal{F}, P)$ 上的随机向量，称

$$F(x,y) := P(X \leq x, Y \leq y)$$

为 (X, Y) 的二维 (联合) 分布函数.



定理 1.18 $F(x,y)$ 具有性质

1. $F(x,y)$ 对每个自变量是单调不减的;
2. $F(x,y)$ 对每个自变量都是右连续的;
3. $\lim_{x\to-\infty}F(x,y)=\lim_{y\to-\infty}F(x,y)=1-\lim_{x\to+\infty,y\to+\infty}F(x,y)=0$;
4. $F(x,y)$ 在任一矩形 $(a_1,b_1]\times(a_2,b_2]$ 上具有非负增量，即有

$$\Delta F=F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2)\geq 0.$$

二维离散型分布

定义 1.26 若随机向量 (X,Y) 有至多可列对可能值 $(x_i,y_i), i,j=1,2,\cdots$,, 则称随机向量及其联合分布是离散型的，而

$$p_{ij} := P(X=x_i, Y=y_j), \quad i,j=1,2,\cdots,$$

称为 (X,Y) 的联合分布列 (律). 我们常用下表来表示 (X,Y) 的分布列

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

二维离散型分布列的基本性质

- 非负性: $p_{ij} \geq 0, i, j = 1, 2, \cdots;$
- 正则性: $\sum_i \sum_j p_{ij} = 1;$
- 与一维情形类似，对任意的 $B \in \mathcal{B}^2$,

$$P((X, Y) \in B) = \sum_{i,j:(x_i, y_j) \in B} p_{ij}.$$

计算联合分布列的方法

- 确定随机变量 (X, Y) 的所有取值数对;
- 计算取每个数值对的概率;
- 列出表格

例 1.18 从 $1, 2, 3, 4$ 中任取一数记为 X, 再从 $1, \cdots, X$ 中任取一数记作 Y, 求 (X, Y) 的联合分布列及 $P(X = Y)$.

定义 1.27 若存在二元非负可积函数 $p(x,y)$ 使得 (X,Y) 的二维 (联合) 分布函数

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv,\forall(x,y)\in\mathbb{R}^2$$

则称 (X,Y) 及其概率分布为连续型的，称 $p(x,y)$ 为其联合分布密度.

易知，联合分布密度具有以下两个基本性质:

1. 非负性: $p(x,y)\geq 0$;
2. 正则性: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)dxdy=1.$

联合分布密度与概率分布之间的关系

- $p(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$
- 对于连续型随机变量，对任意的 $B \in \mathcal{B}^2$ 有

$$P((X,Y) \in B) = \iint_B p(x,y)dxdy$$

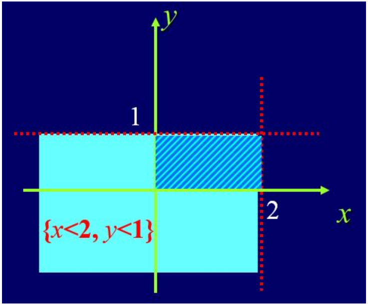
例 1.19 设 (X, Y) 的联合密度函数为

$$p(x,y)=\begin{cases} Ae^{-2x-3y}, & x>0, y>0 \\ 0, & \text{其他} \end{cases}$$

试确定常数 A, 并求概率 $P(X<2, Y<1)$ 及 $P(2X+3Y<6)$.

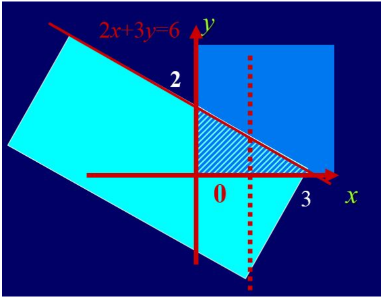
解:

- 由 $\iint p(x,y)dxdy = 1$ 可得 $A=6$;
- $P(X<2, Y<1) = \int_{-\infty}^{2}\int_{-\infty}^{1} p(x,y)dxdy = \int_0^2\int_0^1 p(x,y)dxdy = (1-e^{-4})(1-e^{-3})$;



注意到如果我们令 $D := \{(x,y) : 2x+3y<6\}$, 则

$$\begin{aligned} P(2X+3Y<6) &= P((X,Y)\in D) = \iint_D p(x,y)dxdy \\ &= \int_0^3 \int_0^{\frac{1}{3}(6-2x)} 6e^{-(2x+3y)}dxdy \\ &= 6\int_0^3 e^{-2x}(-\frac{1}{3}e^{-3y})|_0^{\frac{1}{3}(6-2x)}dx = 1-7e^{-6} \end{aligned}$$



边缘分布：随机向量的分量各自的概率分布

定义 1.28 (边缘分布) 在 (X,Y) 的联合分布函数中，令 $y\uparrow+\infty$ 可得

$$\begin{aligned}\lim_{y\to+\infty}F(x,y) &= \lim_{y\to+\infty}P(X\le x, Y\le y)\\ &= P(X\le x, Y\le +\infty)=F(x,+\infty)\\ &= P(X\le x):=F_1(x).\end{aligned}$$

称 $F_1(x)$ 为 X 的边缘分布. 类似的，Y 的边缘分布为

$$F_2(y)=P(Y\le y)=\lim_{x\to+\infty}F(x,y)=F(+\infty,y)$$

从边缘分布的定义可知，若 (X,Y) 的联合分布为 $F(x,y)$, 则

$$\begin{aligned}X&\sim F_1(x)=F(x,+\infty)\\ Y&\sim F_2(y)=F(+\infty,y)\end{aligned}$$

边缘分布的离散情形

- 边缘一词源于离散型情形。在二维离散型概率分布 $\{p_{ij}\}$ 的列表中，将各行求和写在表的最右一列，再将各列求和写在表的最下一行.
- 若 (X,Y) 的联合分布列为 $\{p_{ij}\}$, 则

$$X\text{的分布列为}:P(X=x_i)=\sum_j P(X=x_i,Y=y_j)=\sum_j p_{ij}=p_{i\cdot};$$

$$Y\text{的分布列为}:P(Y=y_j)=\sum_i P(X=x_i,Y=y_j)=\sum_i p_{ij}=p_{\cdot j}$$

i,j	1	2	$\cdots$	j	$\cdots$	$p_{i\cdot}$
1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	

- 注意到

$$
\begin{aligned}
F_1(x) &= P(X \le x) = F(x,+\infty) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} p(u,v)\,du\,dv \\
&= \int_{-\infty}^{x} \underbrace{\left[\int_{-\infty}^{+\infty} p(u,v)\,dv\right]}_{p_1(u)} du = \int_{-\infty}^{x} p_1(u)\,du
\end{aligned}
$$

类似的有，

$$
\begin{aligned}
F_2(y) &= P(Y \le y) = F(+\infty,y) = \int_{-\infty}^{y}\int_{-\infty}^{+\infty} p(u,v)\,du\,dv \\
&= \int_{-\infty}^{x} \underbrace{\left[\int_{-\infty}^{+\infty} p(u,v)\,du\right]}_{p_2(v)} dv = \int_{-\infty}^{y} p_2(v)\,dv
\end{aligned}
$$

- 对于连续型随机向量，其分量仍为连续型，相应的边缘密度分别为

$$
p_1(x) = \int_{-\infty}^{+\infty} p(x,y)\,dy, \quad p_2(y) = \int_{-\infty}^{+\infty} p(x,y)\,dx
$$

- 多项分布：进行 n 次独立重复试验，每次试验有 r 个互不相容的结果:$A_1, \cdots, A_r$ 之一发生；
- 每次试验中 A_i 发生的概率为 $p_i = P(A_i), i = 1, \cdots, r$ 且 $p_1 + \cdots + p_r = 1$;
- 记 X_i 为 n 次独立试验中 A_i 出现的次数，则 $(X_1, X_2, \cdots, X_r)$ 取值 $(n_1, n_2, \cdots, n_r)$ 的概率，即 $A_i, i = 1, \cdots, r$ 出现 n_i 次的概率为

$$P(X_1 = n_1, \cdots, X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

  其中 $n = n_1 + \cdots + n_r$;
- 这个联合分布列称为 r 项分布，又称多项分布，记作 $M(n, p_1, \cdots, p_r)$.
- 上述概率是多项式 $(p_1 + p_2 + \cdots + p_r)^n$ 的一项，故其和为 1.
- $r = 2$ 时即为二项分布.

常见的多维离散型分布：多维超几何分布

- 多维超几何分布：袋中有 N 个球，其中有 N_i 个 i 号球，$i=1,2,\cdots,r$, 且 $N_1+N_2+\cdots+N_r=N$;
- 从中任取 n 个球，若记 X_i 为取出的 n 个球中 i 号球的个数，$i=1,2,\cdots,r$, 则

$$P(X_1=n_1,X_2=n_2,\cdots,X_r=n_r)=\frac{C_{N_1}^{n_1}C_{N_2}^{n_2}\cdots C_{N_r}^{n_r}}{C_N^n}$$

  其中 $n_1+n_2+\cdots+n_r=n$;
- $r=2$ 时即为超几何分布.

多维均匀分布

定义 1.29 设 D 为 $\mathbb{R}^n$ 中的一个有界区域，其度量为 S_D, 如果多维随机变量 $(X_1, \cdots, X_n)$ 的联合密度函数为

$$p(x_1, \cdots, x_n) = \begin{cases} \dfrac{1}{S_D}, & (x_1, x_2, \cdots, x_n) \in D \\ 0, & \text{其他} \end{cases}$$

则称 $(X_1, \cdots, X_n)$ 服从 D 上的多维均匀分布，记作 $(X_1, \cdots, X_n) \sim U(D)$.

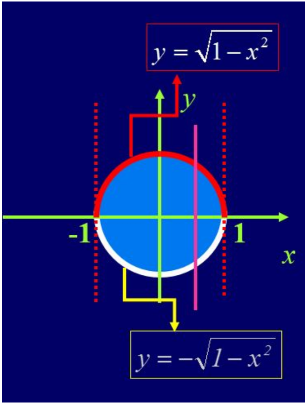
二维均匀分布所描述的随机现象就是向平面区域 D 中随机投点，如果该点的坐标 (X, Y) 落在 D 的子区域 G 中的概率只与 G 的面积有关，而与 G 的位置无关，则

$$P((X, Y) \in G) = \iint_G p(x, y)dxdy = \iint_G \frac{1}{S_D}dxdy = \frac{S_G}{S_D}$$

例 1.20 设 (X,Y) 服从单位圆 $D=\{(x,y):x^2+y^2<1\}$ 上的均匀分布，试求其边缘密度函数 (非均匀分布).

(X,Y) 的联合密度为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2<1\\ 0, & 其他\end{cases}$$



故当 $|x|\geq 1$, $p(x,y)=0$, 从而 $p_1(x)=0$; 当 $|x|<1$ 时，

$$\begin{aligned} p_1(x) &= \int_{-\infty}^{+\infty} p(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy \\ &= \frac{2}{\pi}\sqrt{1-x^2} \end{aligned}$$

类似可得 Y 的密度为

$$p_2(y)=\frac{2}{\pi}\sqrt{1-y^2}$$

定义 1.30 设参数 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 满足 $\sigma_1>0,\sigma_2>0$ 及 $-1<\rho<1$, 称以

$$
\begin{aligned}
p(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\times \\
&\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}
\end{aligned}
$$

为密度函数的连续型分布为二维正态分布，记作

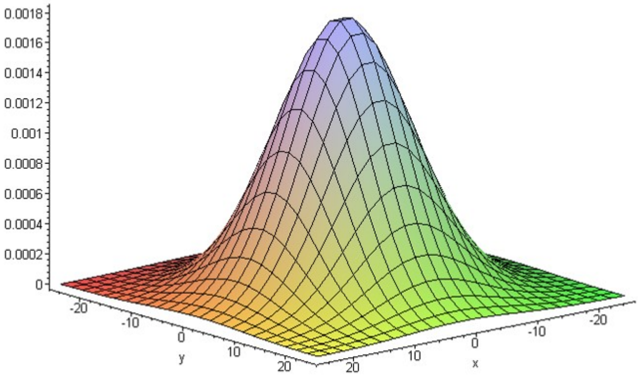
$$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho).$$

后面我们会给出

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}p(x,y)dxdy=1$$

的证明.

二维正态分布密度图像



0.0018
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
-20
-10
0
y
10
20
20
10
0
x
-10
-20

定理 1.19 若 (X,Y) 为二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

$$X\sim N(\mu_1,\sigma_1^2),\quad Y\sim N(\mu_2,\sigma_2^2)$$

证明: X 的边缘分布密度为

$$\begin{aligned}
p_1(x) &= \int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
&\quad\cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}dy \\
&= \int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right\}dv \\
&= \int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left[\left(\frac{v-\rho u}{\sqrt{1-\rho^2}}\right)^2+u^2\right]\right\}dv \\
&= \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}
\end{aligned}$$

类似的可得 $Y\sim N(\mu_2,\sigma_2^2)$.

第十四讲：条件分布与随机变量的独立性

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　　数学学院

条件分布

- $(\Omega, \mathcal{F}, P)$ 为一概率空间，任意固定 $B \in \mathcal{F}$ 且 $P(B) > 0$, $P(\cdot|B)$ 仍为 $(\Omega, \mathcal{F})$ 上的概率;
- 对于 $(\Omega, \mathcal{F})$ 上的随机变量 X, 考虑其关于条件概率 $P(\cdot|B)$ 的分布即在概率空间 $(\Omega, \mathcal{F}, P(\cdot|B))$ 上考虑 X 的分布:

$$F(x|B) := P(X \leq x|B) = \frac{P(X \leq x, B)}{P(B)}, x \in R$$

- 现假定有另一随机变量 Y, 则当 $P(Y = y) > 0$ 时，$F(x|Y = y)$ 则为已知 $Y = y$ 时 X 的条件分布函数.

定义 1.31 若离散型随机向量 (X,Y) 的联合分布为 $\{p_{ij}\}$, 则当

$$P(Y=y_j)=p_{\cdot j}>0$$

时，称

$$p_{i|j}:=P(X=x_i|Y=y_j)=\frac{p_{ij}}{p_{\cdot j}},i=1,2,\cdots,$$

为已知 $Y=y_j$ 时 X 的条件分布. 类似的，当 $P(X=x_i)=p_{i\cdot}>0$ 时，称

$$p_{j|i}:=P(Y=y_j|X=x_i)=\frac{p_{ij}}{p_{i\cdot}}$$

为已知 $X=x_i$ 时 Y 的条件分布.

离散型条件分布函数

根据上面的定义，给定 $Y = y_j$ 条件下 X 的条件分布函数为

$$F(x|y_j) = \sum_{i:x_i \le x} P(X = x_i | Y = y_j) = \sum_{i:x_i \le x} p_{i|j}$$

类似的给定 $X = x_i$ 条件下 Y 的条件分布函数为

$$F(y|x_i) = \sum_{j:y_j \le y} P(Y = y_j | X = x_i) = \sum_{j:y_j \le y} p_{j|i}$$

连续型条件分布

- 设二维连续型随机变量 (X,Y) 的联合密度函数为 $p(x,y)$，边际密度函数分别为 $p_X(x), p_Y(y)$;
- 条件分布函数 $P(X\le x|Y=y)$：因为 $P(Y=y)=0$，故无法通过条件概率直接计算;
- $P(X\le x|Y=y) := \lim_{h\to 0} P(X\le x|Y\in[y,y+h])$;
- 根据如上定义:

$$
\begin{aligned}
F(x|y) &:= P(X\le x|Y=y) = \lim_{h\to 0}\frac{P(X\le x, Y\in[y,y+h])}{P(Y\in[y,y+h])} \\
&= \lim_{h\to 0}\frac{\int_{-\infty}^{x}\int_{y}^{y+h} p(u,v)dvdu}{\int_{y}^{y+h} p_Y(v)dv} = \lim_{h\to 0}\frac{\int_{-\infty}^{x}\{\frac{1}{h}\int_{y}^{y+h} p(u,v)dv\}du}{\frac{1}{h}\int_{y}^{y+h} p_Y(v)dv} \\
&\xlongequal[\text{积分中值定理}]{p(x,y),p_Y(y)\text{在}y\text{连续}} \int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}du
\end{aligned}
$$

- 这表明给定 $Y=y$ 时 X 的条件分布仍为连续型，其密度函数为

$$p(x|y)=p(x,y)/p_Y(y).$$

连续随机变量的条件分布函数及条件密度函数

定义 1.32 对一切使 $p_Y(y) > 0$ 的 y, 给定 $Y = y$ 条件下 X 的条件分布函数和条件密度函数分别为

$$
\begin{aligned}
F(x|y) &= \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)} du, \\
p(x|y) &= \frac{p(x,y)}{p_Y(y)}.
\end{aligned}
$$

同理，对一切使 $p_X(x) > 0$ 的 x, 给定 $X = x$ 条件下 X 的条件分布函数和条件密度函数分别为

$$
\begin{aligned}
F(y|x) &= \int_{-\infty}^{y} \frac{p(x,v)}{p_X(x)} dv, \\
p(y|x) &= \frac{p(x,y)}{p_X(x)}.
\end{aligned}
$$

- 由条件密度函数的定义可得

$$p(x,y) = p_X(x)p(y|x), \quad p(x,y) = p_Y(y)p(x|y)$$

- 对 $p(x,y)$ 求边际密度函数可得

$$\begin{aligned}
p_X(x) &= \int_{-\infty}^{+\infty} p_Y(y)p(x|y)dy \\
P(X \le x) &= \int_{-\infty}^{x} \left[\int_{-\infty}^{+\infty} p_Y(y)p(u|y)dy \right] du \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{x} p(u|y)du \right] p_Y(y)dy \\
&= \int_{-\infty}^{+\infty} P(X \le x|Y=y)dP(Y \le y)
\end{aligned}$$

- 类似推导可得

$$
\begin{aligned}
p_Y(y) &= \int_{-\infty}^{+\infty} p_X(x)p(y|x)dx \\
P(Y \le y) &= \int_{-\infty}^{+\infty} P(Y \le y|X = x)dP(X \le x)
\end{aligned}
$$

- 贝叶斯公式的密度函数形式:

$$
p(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)p(y|x)}{\int_{-\infty}^{+\infty} p_X(x)p(y|x)dx}
$$

$$
p(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{p_Y(y)p(x|y)}{\int_{-\infty}^{+\infty} p_Y(y)p(x|y)dy}
$$

例 1.21 若 (X,Y) 为二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 试求给定 $Y=y$ 时 X 的条件分布密度.

解: X 的条件分布密度为

$$
\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p_Y(y)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}} \\
&= \frac{\exp\left\{\frac{-1}{2\sigma_1^2(1-\rho^2)}\left[(x-\mu_1)^2-2\rho\frac{\sigma_1}{\sigma_2}(x-\mu_1)(y-\mu_2)+\rho^2\frac{\sigma_1^2}{\sigma_2^2}(y-\mu_2)^2\right]\right\}}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right]^2\right\} \\
&\sim N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2)\right) \xlongequal{\rho=0} N(\mu_1,\sigma_1^2)
\end{aligned}
$$

类似可求得 Y 关于 $X=x$ 的条件分布密度.

例 1.22 设 (X,Y) 服从单位圆 $D=\{(x,y):x^2+y^2<1\}$ 上的均匀分布，试求条件分布密度函数 $p(y|x)$.

解: (X,Y) 的联合密度及 X 的边缘密度分别为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2<1,\\ 0, & \text{其他}\end{cases}, \quad p_X(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2}, & |x|<1,\\ 0, & \text{其他}.\end{cases}$$

故

$$p(y|x)=\frac{p(x,y)}{p_X(x)}=\begin{cases}\dfrac{1}{2\sqrt{1-x^2}}, & |y|\leq\sqrt{1-x^2}\\ 0, & \text{其他}\end{cases}$$

- 首先我们回顾一下事件的独立性:

$$P(AB)=P(A)P(B)\Leftrightarrow P(A|B)=P(A),\text{当}P(B)>0;$$

- 二维离散型随机变量

$$p_{i|j}:=\frac{p_{ij}}{p_{\cdot j}}=P(X=x_i|Y=y_j)\xlongequal[\text{不影响概率}]{\text{若条件事件}}P(X=x_i)=p_{i\cdot};$$

- 故如果对任意的事件 $\{X=x_i\},\{Y=y_j\},i,j=1,2,\cdots,$ 均独立，我们则可称随机变量 X,Y 独立，即 X,Y 独立等价于

$$p_{ij}=p_{i\cdot}p_{\cdot j},\quad i,j=1,2,\cdots,$$

- 此时的联合分布函数 $F(x,y)$ 为

$$\begin{aligned}F(x,y)&=P(X\le x,Y\le y)=\sum_{x_i\le x}\sum_{y_j\le y}p_{ij}=\sum_{x_i\le x}p_{i\cdot}\sum_{y_j\le y}p_{\cdot j}\\&=\sum_{x_i\le x}P(X=x_i)\sum_{y_j\le y}P(Y=y_j)=F_X(x)F_Y(y)\end{aligned}$$

- 二维连续型随机变量

$$
\begin{aligned}
F(x|y) &= \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)}du = P(X \le x|Y=y) \xlongequal[\text{不影响概率}]{\text{若条件事件}} P(X \le x) \\
&= \int_{-\infty}^{x} p_X(u)du \\
&\Rightarrow \ p(x,y) = p_X(x)p_Y(y)
\end{aligned}
$$

- 此时，其分布函数

$$
\begin{aligned}
F(x,y) &= P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)dudv \\
&= \int_{-\infty}^{x} p_X(u)du \int_{-\infty}^{y} p_Y(v)dv = F_X(x)F_Y(y)
\end{aligned}
$$

随机变量独立性的定义

定义 1.33 设 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的联合分布函数为 $F(x_1,\cdots,x_n)$, $F_i(x_i)$ 为 X_i 的边缘分布函数。如果对任意的 n 个实数 $x_1,x_2,\cdots,x_n$, 有

$$F(x_1,x_2,\cdots,x_n)=F_1(x_1)F_2(x_2)\cdots F_n(x_n)$$

则称 $X_1,X_2,\cdots,X_n$ 相互独立。特别的,

1. 在离散随机变量场合,如果对任意 n 个取值 $x_1,x_2,\cdots,x_n$, 有

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=P(X_1=x_1)P(X_2=x_2)\cdots P(X_n=x_n),$$

   则称 $X_1,X_2,\cdots,X_n$ 相互独立.

2. 在连续随机变量场合,如果对任意的 n 个取值 $x_1,x_2,\cdots,x_n$, 有

$$p(x_1,x_2,\cdots,x_n)=p_1(x_1)p_2(x_2)\cdots p_n(x_n)$$

   则称 $X_1,X_2,\cdots,X_n$ 相互独立.

第十三讲：随机变量函数的分布

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

定理 1.20 设 $X \sim \{p_k\}$ 为一维离散型随机变量，若 $Z = f(X)$，则对随机变量 Z 的所有可能值 z_k 有

$$P(Z = z_k) = \sum_{i:f(x_i)=z_k} p_i$$

证明: 注意到 $\{Z = z_k\} = \{f(X) = z_k\} = \cup_{i:f(x_i)=z_k}\{X = x_i\}$

故 $P(Z = z_k) = \sum_{i:f(x_i)=z_k} P(X = x_i) = \sum_{i:f(x_i)=z_k} p_i$.

例 1.23 若 $P(X = 1) = P(X = -1) = 0.25, P(X = 0) = 0.5$，求 $Z = X^2$ 的分布.

解: $P(Z = 1) = \sum_{i:x_i^2=1} p_i = P(X = 1) + P(X = -1) = 0.5$

$$P(Z = 0) = P(X = 0) = 0.5$$

定理 1.21 设 X 是连续随机变量，其密度函数为 $p_X(x)$. $Y=f(X)$ 是另一个随机变量。若 $y=f(x)$ 严格单调，其反函数 $h(y)$ 有连续导数，则 $Y=f(X)$ 的密度函数为

$$p_Y(y)=\begin{cases} p_X(h(y))|h'(y)|, & a\le y\le b \\ 0, & \text{其他} \end{cases}$$

其中 $a=\min\{f(-\infty),f(+\infty)\},b=\max\{f(-\infty),f(+\infty)\}$.

证明: 由 $f(x)$ 为严格单调函数可知:

$$\begin{aligned} F_Y(y) &= P(Y\le y)=P(f(X)\le y) \\ &= \begin{cases} 0, & y<a \\ 1, & y>b \\ P(X\le h(y))=\int_{-\infty}^{h(y)}p_X(x)dx, & a\le y\le b\text{且}f(x)\text{严格递增} \\ P(X\ge h(y))=\int_{h(y)}^{+\infty}p_X(x)dx, & a\le y\le b\text{且}f(x)\text{严格递减} \end{cases} \end{aligned}$$

- $f(x)=ax+b$：当 X 为均匀分布，正态分布，伽玛分布时，求 $Y=f(X)$ 的分布密度;
- 当 $X\sim N(\mu,\sigma^2)$ 时，可以验证 $aX+b\sim N(a\mu+b,a^2\sigma^2)$;
- $f(x)=e^x$：当 X 为均匀分布，正态分布，伽玛分布时，求 $Y=f(X)$ 的分布密度;
- 对无单调性的函数 $f(x)$, 我们一般根据 X 的分布密度想法求出 Y 的分布函数 $F_Y(y)$,

$$\begin{aligned} F_Y(y) &= P(Y\le y)=P(f(X)\le y)=P(X\in f^{-1}((-\infty,y])) \\ &= \int_{f^{-1}((-\infty,y])} p_X(x)dx \end{aligned}$$

- 然后对 $F_Y(y)$ 关于 y 求导即可得到其分布密度 $p_Y(y)$.

定理 1.22 (离散卷积公式) 设 X 与 Y 为相互独立的非负整值随机变量，各有分布 $\{a_k\}$ 与 $\{b_k\}$. 则其和有分布

$$P(X+Y=n)=\sum_{k=0}^{n}a_k b_{n-k}, n=0,1,2,\cdots.$$

证明: 注意到 X 与 Y 相互独立，故对任何的 $n=0,1,2,\cdots$，有

$$P(X+Y=n)=\sum_{k=0}^{\infty}P(X+Y=n,X=k)=\sum_{k=0}^{n}P(X+Y=n,X=k)$$

$$=\sum_{k=0}^{n}P(X=k)P(X+Y=n|X=k)=\sum_{k=0}^{n}P(X=k)P(Y=n-k|X=k)$$

$$=\sum_{k=0}^{n}P(X=k)P(Y=n-k)=\sum_{k=0}^{n}a_k b_{n-k}$$

泊松分布的可加性

例 1.24 (泊松分布的可加性) 设随机变量 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$，且 X, Y 相互独立，则 $Z = X + Y \sim P(\lambda_1 + \lambda_2)$

解: 由题意知

$$
\begin{aligned}
P(Z=k) &= P(X+Y=k) = \sum_{i=0}^{\infty} P(X=i, X+Y=k) \\
&= \sum_{i=0}^{\infty} P(X=i, Y=k-i) = \sum_{i=0}^{k} P(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} P(X=i) P(Y=k-i) = \sum_{i=0}^{n} \left(\frac{\lambda_1^i}{i!} e^{-\lambda_1} \right) \left(\frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2} \right) \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} = \frac{(\lambda_1+\lambda_2)^k}{k!} e^{-(\lambda_1+\lambda_2)} \\
& \quad k = 0, 1, \cdots,
\end{aligned}
$$

泊松分布的可加性

- 泊松分布的上述性质可以叙述为：泊松分布的卷积仍是泊松分布，并记为

$$P(\lambda_1) * P(\lambda_2) = P(\lambda_1 + \lambda_2);$$

- 这里的卷积是指" 寻求两具独立随机变量和的分布运算";
- 上述泊松的性质可推广至有限个独立泊松随机变量之和的分布上去，即:

$$P(\lambda_1) * P(\lambda_2) * \cdots * P(\lambda_n) = P(\lambda_1 + \lambda_2 + \cdots + \lambda_n);$$

- 特别的，当 $\lambda_1 = \lambda_2 = \cdots = \lambda_n$ 时有

$$P(\lambda) * P(\lambda) * \cdots * P(\lambda) = P(n\lambda);$$

例 1.25 (二项分布的可加性) 设随机变量 $X \sim B(n,p), Y \sim B(m,p)$, 且 X, Y 独立，则 $X+Y \sim B(n+m,p)$. 特别的，若

$$X_i \sim B(1,p), i=1,2,\cdots,n, \text{则} X_1+X_2+\cdots+X_n \sim B(n,p).$$

若 $Z=f(X_1,X_2,\cdots,X_n)$, $(X_1,X_2,\cdots,X_n)$ 的密度函数为 $p(x_1,x_2,\cdots,x_n)$, 则 Y 的分布函数为

$$G(z)=P(Z\le z)=\int\cdots\int_{f(x_1,x_2,\cdots,x_n)\le z}p(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n$$

定理 1.23 若 $Z = X + Y$, 而 (X, Y) 的联合密度函数为 $p(x, y)$, 则

$$\begin{aligned} F(z) &= P(Z \le z) = \iint_{x+y\le z} p(x,y)dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} p(x,y)dydx \end{aligned}$$

特别的，若 X, Y 相互独立时有 $p(x,y) = p_1(x)p_2(y)$, 将其代入上式得

$$\begin{aligned} F(z) &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} p_1(x)p_2(y)dy \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z} p_1(x)p_2(u-x)du \\ &= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} p_1(x)p_2(u-x)dx\right] du \end{aligned}$$

故其密度函数为 $p(z) = \int_{-\infty}^{+\infty} p_1(z)p_2(z-x)dx = \int_{-\infty}^{+\infty} p_1(z-y)p_2(y)dy$

正态分布的可加性

例 1.26 设随机变量 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$，且 X, Y 独立，则 $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

证明: 显然 $Z = X+Y$ 仍在 $(-\infty, +\infty)$ 上取值。据卷积公式可得

$$
\begin{aligned}
p_z(z) &= \int_{-\infty}^{+\infty} p_1(z-y)p_2(y)dy \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[\frac{(z-y-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}dy \\
&\xlongequal[B=\frac{(z-\mu_1)}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}]{A=\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[A(y-\frac{B}{A})^2+\frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right]\right\}dy \\
&= \frac{\sqrt{2\pi}\sqrt{1/A}}{2\pi\sigma_1\sigma_2}e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sqrt{1/A}}\exp\left\{-\frac{(y-\frac{B}{A})^2}{2(\sqrt{1/A})^2}\right\}dy \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2\sqrt{A}}e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}
\end{aligned}
$$

正态分布的可加性

- 若 $X \sim N(\mu, \sigma^2)$, 则对任意非零常数 a 有，$aX \sim N(a\mu, a^2\sigma^2)$;
- 若 $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$ 为 n 个独立的正态随机变量，则其线性组合仍服从正态分布，即

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n \sim N(\mu_0, \sigma_0^2)$$

其中 $\mu_0 = \sum_{i=1}^n a_i\mu_i, \sigma_0^2 = \sum_{i=1}^n a_i^2\sigma_i^2$.

Gamma 分布的可加性

例 1.27 若 $X \sim \Gamma(\alpha_1, \lambda), Y \sim \Gamma(\alpha_2, \lambda)$, 且 X, Y 独立，则

$$Z = X + Y \sim \Gamma(\alpha_1 + \alpha_2, \lambda).$$

例 1.28 若 $X_i \sim \exp(\lambda) = \Gamma(1, \lambda), i = 1, 2, \cdots, n$, 且相互独立独立，则

$$X_1 + X_2 + \cdots + X_n \sim \Gamma(n, \lambda).$$

例 1.29 若 $X_i \sim \chi^2(n_i) = \Gamma(\frac{n_i}{2}, \frac{1}{2}), i = 1, 2, \cdots, m$, 且相互独立独立，则

$$\begin{aligned} X_1 + X_2 + \cdots + X_m &\sim \Gamma(\frac{n_1 + n_2 + \cdots + n_m}{2}, \frac{1}{2}) \\ &= \chi^2(n_1 + n_2 + \cdots + n_m) \end{aligned}$$

二维随机变量商的分布: $Z = X/Y$

定理 1.24 (商的密度) 如果随机变量 (X, Y) 的联合密度函数为 $p(x, y)$, 则它们的商 $Z := X/Y$ 仍为连续型，且其密度为

$$p(z) = \int_{-\infty}^{+\infty} |y| p(zy, y) dy.$$

证明: 商的分布函数为

$$\begin{aligned} F(z) &= P(X/Y \le z) = \iint_{x/y \le z} p(x, y) dxdy \\ &= \iint_{x/y \le z, y>0} p(x, y) dxdy + \iint_{x/y \le z, y<0} p(x, y) dxdy \\ &= \int_0^{+\infty} dy \int_{-\infty}^{yz} p(x, y) dx + \int_{-\infty}^{0} dy \int_{yz}^{+\infty} p(x, y) dx \\ &= \int_0^{+\infty} dy \int_{-\infty}^{z} yp(yu, y) du + \int_{-\infty}^{0} dy \int_{z}^{-\infty} yp(yu, y) du \\ &= \int_{-\infty}^{z} \left[\int_{-\infty}^{+\infty} |y| p(yu, y) dy \right] du \end{aligned}$$

定理 1.25 若 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量，其各自的分布为 $F_k(x), k = 1, 2, \cdots, n$, 若记

$$\overline{X} := \max_{k=1}^{n} X_k, \quad \underline{X} = \min_{k=1}^{n} X_k$$

则

$$P(\overline{X} \le x) = \Pi_{k=1}^{n} F_k(x), \quad P(\underline{X} \le x) = 1 - \Pi_{k=1}^{n}(1 - F_k(x))$$

证明： $\overline{X}, \underline{X}$ 的分布函数为

$$\begin{aligned}
F_{\overline{X}}(x) &= P(\overline{X} \le x) = P(X_1 \le x, X_2 \le x, \cdots, X_n \le x) \\
&= \Pi_{k=1}^{n} P(X_k \le x) = \Pi_{k=1}^{n} F_k(x) \\
F_{\underline{X}}(x) &= 1 - P(\underline{X} > x) = 1 - P(X_1 > x, \cdots, X_n > x) \\
&= 1 - \Pi_{k=1}^{n} P(X_k > x) = 1 - \Pi_{k=1}^{n}(1 - F_k(x))
\end{aligned}$$

推论 1.1 若上述定理中的 F_k 均相同为 $F(x)$, 并且假设 $F(x)$ 具有分布密度 $p(x)$, 则 $\overline{X}, \underline{X}$ 分布密度分别为

$$\begin{aligned} p_{\overline{X}}(x) &= n[F(x)]^{n-1}p(x), \\ p_{\underline{X}}(x) &= n[1-F(x)]^{n-1}p(x) \end{aligned}$$

思考: 计算 $(\overline{X}, \underline{X})$ 的联合分布函数.

随机向量函数的联合分布：一般情形

求连续型随机向量的函数的分布最一般的提法是:

设已知 $(X_1, X_2, \cdots, X_n)$ 的联合密度函数 $p(x_1, x_2, \cdots, x_n)$, 而

$$\begin{cases} Y_1 = f_1(X_1, X_2, \cdots, X_n) \\ Y_2 = f_2(X_1, X_2, \cdots, X_n) \\ \cdots \\ Y_m = f_m(X_1, X_2, \cdots, X_n) \end{cases} \tag{5}$$

求随机变量 $(Y_1, Y_2, \cdots, Y_n)$ 的分布.

- 保证 $(Y_1, Y_2, \cdots, Y_n)$ 是随机向量：只需 $f_i, i = 1, 2, \cdots, m$ 是 Borel 函数即可;
- 我们需要利用 $(X_1, \cdots, X_n)$ 的分布密度计算 $(Y_1, \cdots, Y_n)$, 故我们需要从 (5) 中反解出 $(X_1, \cdots, X_n)$ 即要求方程组 (5) 有解:
  - 若 $m > n$, 既使对线性函数 f_i 也不能保证方程组有解;
  - 若 $m < n$, 我们可以增补 $Y_j = X_j, j = m+1, \cdots, n$ 化为 $m = n$ 的情形;
  - 故我们只考虑 $m = n$ 的情形.

随机向量函数的联合分布：一般情形

定理 1.26设随机向量 $(X_1, X_2, \cdots, X_n)$ 的联合密度函数 $p(x_1, x_2, \cdots, x_n)$，而 $Y_k = f_k(X_1, \cdots, X_n), k = 1, 2, \cdots, n$. 若假设每个 f_k 均为 n 维 Borel 函数，并且对 $(Y_1, \cdots, Y_n)$ 的每一组可能值 $(y_1, \cdots, y_n)$, 方程组

$$\begin{cases} y_1 = f_1(x_1, \cdots, x_n) \\ \cdots \\ y_n = f_n(x_1, \cdots, x_n) \end{cases} \tag{6}$$

有唯一解

$$\begin{cases} x_1 = h_1(y_1, \cdots, y_n) \\ \cdots \\ x_n = h_n(y_1, \cdots, y_n) \end{cases} \tag{7}$$

并且每个 $h_k(y_1, \cdots, y_n)$ 有连续的一阶偏导数。那么 $(Y_1, \cdots, Y_n)$ 是连续型随机向量，并且有联合密度函数

$$q(y_1, \cdots, y_n) = \begin{cases} p\big(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\big)|J| \\ 0, \text{若}(y_1, \cdots, y_n)\text{使(6)无解} \end{cases}$$

其中 J 为变换 (7) 的 Jacobi 行列式.

证明:　$(Y_1, \cdots, Y_n)$ 的联合分布函数为

$$F(u_1, \cdots, u_n) = P(Y_1 \le u_1, \cdots, Y_n \le u_n)$$
$$= \int_{f_1(x_1,\cdots,x_n)\le u_1} \cdots \int_{f_n(x_1,\cdots,x_n)\le u_n} p(x_1, \cdots, x_n) dx_1 \cdots dx_n$$
$$= \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_n} p\big(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\big) |J| dy_1 \cdots dy_n$$

故 $(Y_1, \cdots, Y_n)$ 是连续型随机向量，并且有联合密度函数

$$q(y_1, \cdots, y_n) = \begin{cases} p\big(h_1(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n)\big) |J| \\ 0, \text{若}(y_1, \cdots, y_n)\text{使}(6)\text{无解} \end{cases}$$

注: 若 Jacobi 行列式 J 不易计算时，可计算 J^{-1} 即变换 (6) 的 Jacobi 行列式.

例 1.30 设随机向量 (X,Y) 的联合密度函数为 $p(x,y)$. 则 $U=XY$ 的密度函数为

$$p_U(u)=\int_{-\infty}^{+\infty}p(\frac{u}{v},v)\frac{1}{|v|}dv$$

证明: 增补变量 $V=Y$, 则 $\begin{cases} u=xy, \\ v=y, \end{cases}$ 的反函数为 $\begin{cases} x=\frac{u}{v}, \\ y=v, \end{cases}$. 其 Jacobi 行列式为

$$J=\begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix}=\frac{1}{v},$$

故 (U,V) 的联合密度函数为 $q(u,v)=p(\frac{u}{v},v)\frac{1}{|v|}$. 对 $q(u,v)$ 关于 v 积分便可得 $U=XY$ 的密度函数. (随机变量商的密度也可类似求出)

第十四讲：随机变量的存在性

张鑫

Email: xzhangseu@seu.edu.cn

东南大学　数学学院

分布函数与随机变量

$$\left.\begin{array}{c}(\Omega,\mathcal{F},P)\\ \\ X(\omega)\end{array}\right\}\Rightarrow P(X\le x)=:F(x)\Rightarrow\left\{\begin{array}{l}\text{单调非降性}\\ \text{右连续性}\\ \text{规范性}\end{array}\right.$$

$$\left.\begin{array}{r}\text{单调非降性}\\ \text{右连续性}\\ \text{规范性}\end{array}\right\}\text{的}F(x)\text{给定}\Rightarrow\text{存在}\left\{\begin{array}{l}(\Omega,\mathcal{F},P)\\ \\ X(\omega)\end{array}\right.\quad\text{使得}F(x)=P(X\le x)?$$

定理 1.27 若 $F(x)$ 是右连续、单调非降函数，且 $F(-\infty)=0, F(+\infty)=1$, 则存在一个概率空间 $(\Omega,\mathcal{F},P)$ 及其上的随机变量 $X(\omega)$, 使 $X(\omega)$ 的分布函数恰好是 $F(x)$.

分析:

- 对 $[0,1]$ 上的均匀分布随机变量 $\theta(\omega)=\omega$, 其分布函数为

$$P(\theta(\omega)\le x)=P(\omega\in[0,x])=x,\ \ \forall x\in[0,1];$$

- $F(x)\in[0.1]$, 若将上式中的 x 替换为 $F(x)$, 则有

$$P(\theta(\omega)\le F(x))=F(x);$$

- 若分布函数 $F(x)$ 可逆，则

$$P(F^{-1}(\theta(\omega))\le x)=P(\theta(\omega)\le F(x))=F(x);$$

- 考虑 $X(\omega):=F^{-1}(\theta(\omega))$, 则 $F_X(x)=P(F^{-1}(\theta(\omega))\le x)=F(x)$;

- $F(x)$不一定可逆, 能否定义一个函数 G 使得

$$\{\theta(\omega) \leq F(x)\} = \{G(\theta(\omega)) \leq x\}?$$

- 上述等式也即寻求如下等价性

$$G(\theta) \leq x \Leftrightarrow F(x) \geq \theta$$

- 因此，若函数 G 存在，则对任意给定的 θ 必须满足

$$G(\theta) \leq x, \quad \forall x \in \{x : F(x) \geq \theta\}$$

- 故所寻找的函数 G 需有以下性质

$$G(\theta) \leq \inf\{x : F(x) \geq \theta\}$$

- 上述不等式右侧的下确界也是 G 的一种选择，可以证明选此下确界作为函数 G 的定义的确具有我们所要求的性质.

单调逆 (一般逆) 的定义

定义 1.34 设 $F(x)$ 是右连续、单调非降函数，且 $F(-\infty)=0, F(+\infty)=1$. 对任意的 $p\in(0,1)$, 我们称

$$F^{-1}(p) := \inf\{x : F(x) \geq p\},$$

为函数 $F(x)$ 的单调逆或一般逆.

注 1.9 $F^{-1}(p)$ 在概率论中也称为函数 $F(x)$ 的 p 分位数函数或与其相对应的随机变量 X 的 p 分位数，通常用 x_p 或 ξ_p 来表示.

定理 1.28 设 $F(x), F^{-1}(p)$ 定义如上，则

$$F^{-1}(p) \leq x \Leftrightarrow F(x) \geq p$$

证明: 令 $A := \{y : F(y) \geq p\}$, 则

- $\Leftarrow$: $F(x) \geq p$ 蕴含 $F^{-1}(p) := \inf A \leq x$
- $\Rightarrow$: $F^{-1}(p) \leq x$ 即 $x \geq F^{-1}(p) := \inf A =: x_0$
  - 若 $x_0 \in A$, 则显然 $F(x) \geq F(x_0) \geq p$
  - 若 $x_0 \notin A$, 则存在 $\{x_n\}_{n\geq 1} \subset A$ 使得 $x_n > x_0$ 且 $\lim_{n\to\infty} x_n = x_0$,从而

$$F(x) \geq F(x_0) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) \geq p;$$

定理 1.27 的证明:

- 取 $\Omega=[0,1]$, $\mathcal{F}$ 为 $[0,1]$ 上的 Borel 集全体，取 P 为直线上的 Lebesgue 测度 (是长度概念的推广，但对一切 Borel 集有定义);
- 定义: $\theta(\omega)=\omega$, 则 $\theta(\omega)$ 是 $(\Omega,\mathcal{F},P)$ 上的随机变量，并且，对一切的 $0\le x\le 1$,

$$P(\theta(\omega)\le x)=P(\omega\in[0,x])=x;$$

- 考虑 $X(\omega):=F^{-1}(\theta(\omega))$, 则

$$F_X(x)=P(X\le x)=P(F^{-1}(\theta(\omega))\le x)\xlongequal[p\le F(x)]{F^{-1}(p)\le x}P(\theta(\omega)\le F(x))=F(x)$$

单调逆 (一般逆) 的性质

- $F^{-1}(F(x)) \le x, \forall x \in R$：设 $x_0 \in R$ 任意给定且 $F(x_0) = p_0$，则

$$\begin{aligned} F^{-1}(F(x_0)) = F^{-1}(p_0) &= \inf\{x : F(x) \ge p_0\} \\ &= \inf\{x : F(x) \ge F(x_0)\} \le x_0 \end{aligned}$$

- $F(F^{-1}(p)) \ge p, \forall p \in (0,1)$:设 $p_0 \in (0,1)$ 任意给定且

$$F^{-1}(p_0) = x_0 = \inf\{x : F(x) \ge p_0\} =: \inf A,$$

  - 若 $x_0 \in A$，则显然有 $F(F^{-1}(p_0)) = F(x_0) \ge p_0$;
  - 若 $x_0 \notin A$，则 存在 $\{x_n\}_{n\ge1} \subset A$ 使得 $x_n > x_0$ 且 $\lim_{n\to\infty} x_n = x_0$,从而由函数 $F(x)$ 的右连续性可知

$$F(F^{-1}(p_0)) = F(x_0) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) \ge p_0;$$

- $F^{-1}(p)$ 关于 p 单调非降：集合 $\{x : F(x) \ge p\}$ 关于 p 单调不增，故其下确界也关于 p 单调非降

单调逆 (一般逆) 的性质

下面的几条性质

- $F^{-1}(p)$ 关于 p 是左连续的;
- $F^{-1}(p) := \inf\{x : F(x) \geq p\} = \sup\{x : F(x) < p\}$
- $x < F^{-1}(p) \Leftrightarrow F(x) < p$

思考

若 $F(x)$ 是左连续、单调非降函数，且 $F(-\infty)=0, F(+\infty)=1$，是否存在一个概率空间 $(\Omega, \mathcal{F}, P)$ 及其上的随机变量 $X(\omega)$，使得

$$F(x)=P(X<x).$$