概率统计 B
第四 + 章 期中复习

原著：陈家鼎、刘婉如、汪仁官
制作：李东风，邓明华

2025 年春季

本章目录

本节目录

样本空间与事件

- **样本空间**：随机试验的可能结果组成的集合。比如掷骰子的结果是 $\Omega=(\omega_1,\omega_2,\cdots,\omega_6)$; 而观察某个路口一段时间内走过的车辆个数的所有可能的结果就是全部非负整数 $\Omega=\{1,2,\cdots,n\cdots,\}$.
- **事件**：样本空间的子集合。
- 事件关系与运算：集合论描述
  1. 包含关系: $A\subset B$, 事件 B 发生则事件 A 一定发生。
  2. 运算：交，并，差。韦恩图表示
  3. 对立事件：$\bar{A}=\Omega-A$.
  4. 两个事件互补相容：$A\cap B=\phi$.

常用事件运算公式

- 集合分解

$$A \cup B = A + \overline{A}B$$

$$A = AB + A\overline{B}$$

其中 + 一般表示两个互不相容事件的并。

- De-Morgan 律

$$\overline{\cup_{i=1}^{n} A_j} = \cap_{i=1}^{n} \overline{A_j}$$

$$\overline{\cap_{i=1}^{n} A_j} = \cup_{i=1}^{n} \overline{A_j}$$

等概完备事件组

- **定义** 称一个事件组 $A_1, A_2, \ldots, A_n$ 为一个 **等概完备事件组**，如果它具有下列三条性质：
  (1). $A_1, A_2, \ldots, A_n$ 发生的机会相同（等可能性）；
  (2). 在任一次试验中，$A_1, A_2, \ldots, A_n$ 至少有一个发生（也就是所谓“除此之外，不可能有别的结果”）（完备性）；
  (3). 在任一次试验中，$A_1, A_2, \ldots, A_n$ 至多有一个发生（也就是所谓“它们是互相排斥的”）（互不相容性）。
- 等概完备事件组也称“**等概基本事件组**”，其中任一事件 $A_i(i = 1, 2, \ldots, n)$ 称为**基本事件**。

古典概型

- 若 $A_1, A_2, \ldots, A_n$ 是一个等概基本事件组，事件 B 由其中的 m 个基本事件所构成，则

$$P(B) = \frac{m}{n}.$$

- **古典概型**就是用等概基本事件组和上述公式来计算事件概率的模型。
- 从而概率的计算转化为排列组合问题。

σ 代数

- 设 Ω 为一个非空集合，叫做**基本事件空间**。
- Ω 的一些子集组成的集合 $\mathscr{F}$ 叫做 σ **代数**，如果

$$\begin{aligned}
&(1).\Omega \in \mathscr{F} \\
&(2).\text{若} A \in \mathscr{F}, \text{则} A^c = \Omega - A \in \mathscr{F} \\
&(3).\text{若} A_n \in \mathscr{F}(n = 1, 2, \dots), \text{则} \cup_{n=1}^{\infty} A_n \in \mathscr{F}
\end{aligned}$$

- 事件是 $\mathcal{F}$ 中的集合。

概率空间定义

- $\mathscr{F}$ 上有定义的函数 $P=P(\cdot)$ 叫做**概率测度** (简称**概率**), 若

  (1). $P(A)\geq 0 \qquad (\forall A\in\mathscr{F})$.

  (2). $P(\Omega)=1$.

  (3). 若 $A_n\in\mathscr{F}(n=1,2,\dots)$ 且两两不相交, 则

$$P\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P(A_n)$$

  (可列可加性)

- $(\Omega,\mathscr{F},P)$ 称之为**概率空间**

可列样本空间概率确定

- **定理：** 若样本空间 Ω 可列，不妨记为 $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n, \cdots\}$，只要找到非负数列 p_i, 满足

$$\sum_{i=1}^{+\infty} p_i = 1$$

使得基本事件的概率 $P(\omega_i) = p_i$, 那么 $(\Omega, \mathscr{F}, P)$ 定义了一个概率空间。

- 特别地：如果 Ω 有限，$|\Omega| = n$, n 个和为 1 的数 $p_1, \cdots, p_n$ 定义了一个概率空间。

概率的性质

1. $P(\emptyset) = 0$
2. 若 $A \in \mathscr{F}$ 则 $P(A^c) = 1 - P(A)$
3. 若 $A_1, \ldots, A_n$ 都属于 $\mathscr{F}$ 且两两不相交，则

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \text{ (有限可加性)}$$

4. 若 $A \subset B$, $A, B \in \mathscr{F}$, 则 $P(A) \leq P(B)$ 且

$$P(B \backslash A) = P(B) - P(A)$$

概率的性质

5. 若 $A_n \subset A_{n+1}, A_n \in \mathscr{F}(n=1,2,\dots)$，则

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \text{ (单调上升事件的概率极限)}$$

6 若 $A_n \supset A_{n+1}, A_n \in \mathscr{F}(n=1,2,\dots)$，则

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \text{ (单调下降事件的概率极限)}$$

条件概率

- **定义：** 设有概率空间 $(\Omega, \mathscr{F}, P)$, 若 $A \in \mathscr{F}$ 是 Ω 下的随机事件，$P(A) > 0$，对任意 $B \in \mathscr{F}$, 定义 A 发生的前提下 B 发生的 **条件概率**, 记作 $P(B|A)$ 为
$$P(B|A) = \frac{P(AB)}{P(A)}$$
- 容易验证：条件概率定义了以 B 为样本空间的概率空间。

乘法公式

- 由条件概率的定义，有

$$P(AB) = P(A)P(B|A), P(AB) = P(B)P(A|B)$$

- 乘法公式可以推广到 n 个事件上

$$\begin{aligned}
&P(A_1A_2\cdots A_n) \\
&= P(A_n|A_1A_2\cdots A_{n-1})P(A_1A_2\cdots A_{n-1}) \\
&= P(A_n|A_1A_2\cdots A_{n-1}) \times \cdots \times P(A_3|A_1A_2) \times P(A_2|A_1)P(A_1)
\end{aligned}$$

独立性

- **定义：** 称两个随机事件 A, B 是**相互独立**的, 如果

$$P(AB) = P(A)P(B)$$

- 定义中不要求 $P(A) > 0$ 或 $P(B) > 0$。
- 在 $P(A) > 0$, 则可以考虑条件概率 $P(B|A)$，于是独立性等价于

$$P(B|A) = P(A)$$

即事件的条件概率等于事件的无条件概率，因此两个事件独立。

多个事件相互独立

- **定义：** 称 $A_1, A_2, \ldots, A_n$ 是相互独立的，如果对任意整数 $k(2 \le k \le n)$ 以及从 $1, 2, \ldots, n$ 中任意取出的 k 个 $i_1, i_2, \ldots, i_k$ 都满足

$$P(A_{i_1} A_{i_2} \ldots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \ldots P(A_{i_k})$$

- 其中一个要求是

$$P(A_1 A_2 \ldots A_n) = P(A_1) P(A_2) \ldots P(A_n)$$

两两独立但不相互独立的例子

- **例子：** 三个事件两两独立不能保证三个事件独立的例子。
- 均匀正四面体，四面涂红色、黄色、蓝色、红黄蓝混杂，投掷一次，考察底面出现的颜色。
- 设 $A=$“红色出现”，$B=$“黄色出现”，$C=$“蓝色出现”。
- 基本事件：$A_i=$“第 i 面在底面”，$i=1,2,3,4$, 构成等概基本事件组。
- $A=A_1\cup A_4$, $B=A_2\cup A_4$, $C=A_3\cup A_4$。
- $P(A)=P(B)=P(C)=\frac{1}{2}$。
- $AB=AC=BC=A_4$，$P(AB)=P(AC)=P(BC)=\frac{1}{4}$, 按定义 A,B 相互独立，A,C 相互独立，B,C 相互独立。
- 但 $ABC=A_4$, $P(ABC)=\frac{1}{4}\neq P(A)P(B)P(C)$。

全概公式

- **定理（全概公式）** 如果事件组 $A_1, A_2, \ldots, A_n$ 满足：
- (1). $A_1, A_2, \ldots, A_n$ 互不相容，且 $P(A_i) > 0(i = 1, 2, \ldots, n)$。
- (2). $A_1 \cup A_2 \cup \cdots \cup A_n = U$(完备性)，
- 则对任一事件 B 皆有

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$

- 满足条件（1）和（2）的事件组 $A_1, A_2, \ldots, A_n$ 称为**完备事件组**。
- 更一般的全概公式中的完备事件组可以包含**可数个事件**。
- 运用全概公式关键在于求完备事件组。

逆概公式

- 若有多个基本事件是完备事件组，则观测到一个结果后，可以逆推原来到底是哪一个事件。
- **定理（逆概公式）** 设 $A_1, A_2, \ldots, A_n$ 为一完备事件组，则对任一事件 $B(P(B) \neq 0)$ 有

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} \tag{6.6}$$

- 逆概公式也称为贝叶斯（Bayes）公式，是决策分析的基础。
- 逆概公式是全概公式的引申。
- 例子：疾病检测；收发报；模式识别（数据特征和类型）

本节目录

随机变量的定义

- 给定概率空间 $(\Omega, \mathscr{F}, P)$, **随机变量**是样本空间 Ω 上的函数

$$X : \Omega \to R$$

满足

$$\forall x, \{\omega : X(\omega) \leq x\} \in \mathscr{F}$$

相应的，其概率定义为

$$P(X \leq x) = P(\{\omega : X(\omega) \leq x\})$$

- 随机变量将样本空间统一对应到实数空间上。
- 随机变量将样本空间上的 σ 代数对应到实数轴上左开右闭区间组成的 σ 代数（Borel 集）。

离散型随机变量

- 离散型随机变量取值范围是有限个值或可数个值，设 X 可取的值为 $x_1, x_2, \ldots, x_k, \ldots$。
- 随机变量作为古典概型、独立试验序列概型等的推广，用取值作为基本事件。由前面的定理，分布由基本事件的概率确定。
- 离散型随机变量 X 的每个取值的概率可列表如下

X	x_1	x_2	$\cdots$	x_k	$\cdots$
p	p_1	p_2	$\cdots$	p_k	$\cdots$

称为 X 的**概率分布表**（**或者概率质量函数**）。

常用离散型随机变量

- 两点分布或者 Bernoulli(p);
- 二项分布;
- Poisson 分布;
- 超几何分布;

连续型随机变量与概率密度函数

- **定义：** 对于随机变量 X 称为**连续型随机变量**，如果存在非负可积函数 $p(x)(-\infty < x < \infty)$，使对任意 $a, b(a < b)$ 都有

$$P(a < X \leq b) = \int_a^b p(x)\,dx$$

称 $p(x)$ 为 X 的概率密度函数（probability density funcition, PDF），简称概率密度或密度。

- 由随机变量的定义，全空间上概率为 1，可得：密度函数非负，且

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

连续型随机变量单点概率为零

- 连续型随机变量至少在一个区间内可以取到任意实数值，所以取每个值的概率应该等于零。
- 对正整数 n,

$$
\begin{aligned}
\{X=a\} &\subset \left\{a-\frac{1}{n}<X<a+\frac{1}{n}\right\} \\
P(X=a) &\leq P\left(a-\frac{1}{n}<X<a+\frac{1}{n}\right) \\
&= \int_{a-\frac{1}{n}}^{a+\frac{1}{n}} p(x)\,dx \to 0 \quad (n\to\infty)
\end{aligned}
$$

所以

$$P(X=a)=0$$

常用连续分布

- 均匀分布
- 指数分布 $Exp(\lambda)$

$$p(x;\lambda) = \lambda e^{-\lambda x}, x \geq 0$$

- 正态分布 $N(\mu,\sigma^2)$

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

- 贝塔分布 $\mathrm{Beta}(\alpha,\beta)$

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, 0 \leq x \leq 1.$$

- 伽马分布 $\mathrm{Gamma}(\alpha,\beta)$

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, x \geq 0$$

分布函数

- 由随机变量的定义，对任意 x, $\{\omega | X(\omega) \leq x\} \in \mathscr{F}$ 是事件，故可以取到概率。
- **定义：** 设 X 是一个随机变量，称函数

$$F(x) = P(X \leq x) \quad (-\infty < x < \infty) \tag{4.1}$$

为 X 的分布函数 (distribution function, 或 cumulative distribution function, CDF)。
- 任何一个随机变量都有分布函数。

分布函数的性质

(1). $0 \le F(x) \le 1 \ (-\infty < x < \infty)$;

(2). $F(x)$ 是 x 的单调递增函数;

(3). $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$;

(4). $F(x)$ 是 x 右连续函数;

(5). $P(a < X \le b) = F(b) - F(a)$。

(6). 若 X 为连续型,则 $P(a < X < b) = F(b) - F(a)$。

连续型随机变量的分布函数

- 设 X 是连续型随机变量，有密度 $p(x)$, 分布函数 $F(x)$，则

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(t)\,dt$$

- $F(x)$ 是 $p(x)$ 的可变上限的定积分，也是 $p(x)$ 的一个原函数。
- (1) $F(x)$ 是 $x \in (-\infty, \infty)$ 的连续函数;
- (2) 对于 $p(x)$ 的连续点 x_0 而言有

$$F'(x_0) = p(x_0)$$

(注意密度函数可以修改单个点的函数值而仍为原随机变量密度)

- 若 $p(x)$ 只有至多有限个间断点，则对非间断点的 x

$$p(x) = F'(x)$$

离散型随机变量的函数

- 若 X 是离散型随机变量，取值 $x_1, x_2, \ldots, x_k, \ldots$，
$P(x_k) = p_k(k = 1, 2, \ldots)$，
- 则 $Y = f(X)$ 取值 $f(x_1), f(x_2), \ldots, f(x_k), \ldots$，若各 $f(x_k), k = 1, 2, \ldots$
互不相同则

$$P(Y = f(x_k)) = p_k \qquad (*)$$

(*) 即为 Y 的概率分布。
- 若 $f(x_k), k = 1, 2, \ldots$ 有重复值，设所有互不相等的值为 $y_1, y_2, \ldots$，
则

$$P(Y = y_k) = \sum_{f(x_j) = y_k} p_j, \quad k = 1, 2, \ldots$$

求连续型随机变量函数的密度的一般方法

- 若 $Y=f(X)$, 如果 $y=f(x)$ 是一一映射函数 (不妨设为严格单调上升函数)，若反函数为 $x=g(y)$, 那么

$$\begin{aligned} F_Y(y) &= P(Y=f(X)\leq y) \\ &= P(X=g(Y)\leq g(y)) = F_X(g(y)) \end{aligned}$$

求导得到

$$q_Y(y)=p_X(g(y))g'(y)$$

- 上述过程用到了如下定理：
- **定理（习题七 16 题）** 如果 X 的分布函数 $F(x)$ 具有连续的导函数 $F'(x)$，则 X 为连续随机变量，且 $F'(x)$ 是 X 的密度函数。
- **定理（习题七 17 题）** 如果 X 的分布函数 $F(x)$ 连续, 在除去有限个点之外的区间上 $F'(x)$ 存在且连续，那么

$$f(x)=\begin{cases} F^x & F'(x)\text{存在时} \\ 0 & F'(x)\text{不存在时} \end{cases}$$

X 为连续随机变量，且 $f(x)$ 是 X 的密度函数。

一般情况

- **定理：** 设 X 有密度函数 $f(x)$, $D \subset \mathcal{R}$, $Y = g(X)$, $P(Y \in D) = 1$. 如果存在函数 $h_i(y)$ 使得
  1. 对 $y \in D$, $\{Y = y\} = \bigcup_{i=1}^{n}\{X = h_i(y)\}$,
  2. 每个 $h_i(y)$ 是 D 到其值域 D_i 的可逆映射, 有连续的导数,
  3. 值域 $D_1, D_2, \cdots, D_n$ 互不相交,

  则 Y 有密度函数

$$f_Y(y) = \begin{cases} \sum_{i=1}^{n} f(h_i(y))|h_i'(y)|, & y \in D, \\ 0, & y \in \overline{D}. \end{cases}$$

本节目录

离散随机变量的期望

- **定义：** 设离散型随机变量 X 的概率分布为

$$P(X=x_k)=p_k, k=1,2,\dots$$

如果 $\sum_k |x_k|p_k$ 收敛，则称

$$\sum_k x_k p_k$$

为随机变量 X 的**期望**（或 **数学期望**），记作 EX 或 $E(X)$。

- EX 是随机变量取值的加权平均，按照概率的频率定义，这更符合多次重复观测后随机变量的平均表现。也叫做 X 的**均值**, 或 X 的分布的均值。

常用离散分布的期望

- 两点分布 $X \sim Ber(p)$, $EX = p$.
- 二项分布 $X \sim B(n, p)$, $EX = np$.
- 泊松分布 $X \sim Poisson(\lambda)$, $EX = \lambda$.
- 超几何分布 $X \sim HyperGeo(N, M, n)$, $EX = \frac{nM}{N}$

连续型随机变量的期望

- **定义：** 设连续型随机变量的密度为 $p(x)$，如果积分 $\int_{-\infty}^{+\infty} |x| p(x)\, dx$ 收敛，则称

$$\int_{-\infty}^{\infty} xp(x)\, dx \tag{2.1}$$

为 X 的期望（或均值），记作 EX(或 $E(X)$)。

- 解释：连续型随机变量的 $p(x)$ 是一个比例系数（权重），不是概率。

常用连续型随机变量的期望

- 均匀分布 $X \sim U(a,b)$, $EX = \frac{a+b}{2}$.
- 指数分布 $X \sim Exp(\lambda)$, $EX = 1/\lambda$.
- 正态分布 $X \sim N(\mu, \sigma^2)$, $EX = \mu$.
- 伽马分布 $X \sim Gamma(\alpha, \beta)$, $EX = \frac{\alpha}{\beta}$.

随机变量函数的期望公式

- 对 $X \sim p(x)$，$Y = f(X)$，

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)\,dx$$

(要求右边绝对收敛)

- 若 $P(X = x_i) = p_i, i = 1, 2, \ldots,$

$$E[f(X)] = \sum_i f(x_i)p_i$$

(要求右边的级数绝对收敛)

- 这样的公式免去了求 $Y = f(X)$ 的分布的过程。

期望的简单性质

- 对常数 c, k, b 和随机变量 X, 有

  (1). $E(c) = c$;
  (2). $E(kX) = kE(X)$;
  (3). $E(X + b) = E(X) + b$;
  (4). $E(kX + b) = kE(X) + b$.

离散随机变量的方差

- **定义：** 设离散型随机变量的概率分布为

$$P(X = x_k) = p_k,\ k = 1, 2, \ldots$$

则称

$$\sum_k [x_k - E(X)]^2 p_k$$

为 X 的**方差**（要求级数求和收敛），记作 $D(X)$ 或 $\mathrm{Var}(X)$。

- 为了方差存在，期望 $E(X)$ 必须先存在；
- 方差总是非负的：

连续随机变量的方差

- **定义：** 设连续型随机变量 X 的密度是 $p(x)$，则称

$$\int_{-\infty}^{\infty} [x - E(X)]^2 p(x)\, dx$$

为 X 的方差（要求积分收敛），记作 $D(X)$ 或 $\mathrm{Var}(X)$。

- 为了方差存在，期望 $E(X)$ 必须先存在；
- 方差总是非负的：

$$D(X) \geq 0$$

常用离散分布的方差

- 两点分布 $X \sim Ber(p)$, $Var(X) = p(1-p)$;
- 二项分布 $X \sim B(n,p)$, $Var(X) = np(1-p)$;
- 泊松分布 $X \sim Poisson(\lambda)$, $Var(X) = \lambda$.

常用离散分布的方差

- 均匀分布 $X \sim U(a,b)$, $Var(X) = (b-a)^2/12$;
- 指数分布 $X \sim Exp(\lambda)$, $Var(X) = 1/\lambda^2$;
- 正态分布 $X \sim N(\mu,\sigma^2)$, $Var(X) = \sigma^2$;
- 伽马分布 $X \sim Gamma(\alpha,\beta)$, $Var(X) = \alpha/\beta^2$

方差的简单性质

- 当 k, b, c 为常数时

  (1). $D(c) = 0$;
  (2). $D(kX) = k^2 D(X)$;
  (3). $D(X + b) = D(X)$;
  (4). $D(kX + b) = k^2 D(X)$.

切比雪夫不等式

- **定理：** 设随机变量 X 存在均值 $E(X)$ 与方差 $D(X)$，则有

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2} \quad (\varepsilon > 0) \tag{5.1}$$

- 在切比雪夫不等式中取 $\varepsilon = k\sqrt{D(X)}$, 则

$$P(|X - E(X)| \geq k\sqrt{D(X)}) \leq \frac{1}{k^2}$$

  其中 $\sqrt{D(X)}$ 叫做 X 的标准差。
- 特别地，$k = 3$ 时

$$P(|X - E(X)| \geq 3\sqrt{D(X)}) \leq \frac{1}{9}$$

- 对比正态分布，这个比例是小于千分之三，也说明切比雪夫不等式给的比较松散。

本节目录

二维离散型随机向量

- **定义：** 如果二维随机向量 $\boldsymbol{\xi} = (X, Y)$ 可能取的值只有有限个或可数个，则称 $\boldsymbol{\xi}$ 为**离散型随机向量**。
- 若 $\boldsymbol{\xi} = (X, Y)$ 是离散型，则两个分量 X, Y 都是离散型；反之亦然。
- 设 X 的取值范围是 $\{x_i, i = 1, 2, \dots\}$，Y 的取值范围是 $\{y_j, j = 1, 2, \dots\}$，则 (X, Y) 的取值范围是 $\{(x_i, y_j) : i = 1, 2, \dots, j = 1, 2, \dots\}$ (其中有些组合可能是不可能事件)。
- 二维随机变量 $\boldsymbol{\xi} = (X, Y)$ 的概率分布：

$$P((X, Y) = (x_i, y_j)) = p_{ij}, \quad i = 1, 2, \dots, j = 1, 2, \dots \tag{1.1}$$

也称为 (X, Y) 的**联合分布**。

二维离散随机变量

- 二维概率分布表也可以排列成

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

- 多项分布 $(X_1, X_2, \cdots, X_s) \sim Multinomial(n; \alpha_1, \alpha_2, \cdots, \alpha_s)$

$$p(n_1, n_2, \cdots, n_s) = \frac{n!}{n_1! n_2! \cdots n_s!} \alpha_1^{n_1} \alpha_2^{n_2} \cdots \alpha_s^{n_s}$$

其中 $n_1 + n_2 + \cdots + n_s = n.$

联合分布与边缘分布

- 二维随机变量 (X, Y) 的分布也称为**联合分布**，分量 X 的概率分布称为 (X, Y) 的关于 X 的**边缘分布**；分量 Y 的概率分布称为 (X, Y) 的关于 Y 的**边缘分布**。
- 设

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

- 则

$$\begin{aligned}
P(X = x_i) &= \sum_j P(X = x_i, Y = y_j) \quad (\text{全概公式}) \\
&= \sum_j p((X, Y) = (x_i, y_j)) \\
&= \sum_j p_{ij} \quad (\text{这是 } X \text{ 的边缘分布}) \\
P(Y = y_j) &= \sum_i p_{ij} \quad (\text{这是 } Y \text{ 的边缘分布})
\end{aligned}$$

- 多项分布的边缘分布是二项分布。

二维连续型随机变量的密度

- **定义：** 对于二维随机向量 $\boldsymbol{\xi}=(X,Y)$, 如果存在非负函数 $p(x,y)(-\infty<x<\infty,-\infty<y<\infty)$, 使对任意 $a<b,c<d$ 及 $D=\{(x,y):a<x<b,c<y<d\}$ 有

$$P((X,Y)\in D)=\iint\limits_{D} p(x,y)\,dxdy$$

则称随机向量 $\xi=(X,Y)$ 为连续型的，并称 $p(x,y)$ 为 $\boldsymbol{\xi}$ 的分布密度，也称 $p(x,y)$ 为 (X,Y) 的**联合分布密度** (简称联合密度)。
- 用微元法，二维随机变量在 $[x,x+dx]\times[y,y+dy]$ 上的概率为 $p(x,y)\,dxdy$, 其中 dx,dy 为无穷小微元。
- 连续型随机向量属于更一般的平面子集 D 的概率为

$$P((X,Y)\in D)=\iint\limits_{D} p(x,y)\,dxdy$$

但集合 D 的要求涉及到勒贝格积分，这里不做讨论。一般的开集、并集及其有限运算都符合条件。

连续型二维分布的边缘分布

- **定义：** 对于随机向量 (X, Y), 作为其分量的随机变量 X（或 Y）的密度函数 $p_X(x)$ (或 $p_Y(y)$)，称为 (X, Y) 的关于 X（或 Y）的**边缘分布密度**。
- 连续型二维随机向量的分量一定是连续型随机变量。
- **定理 1.1** 若 (X, Y) 的联合密度是 $p(x, y)$, 则

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

分别是 X, Y 的分布密度。

随机变量的独立性

- 随机变量的独立性，就是关于随机变量的事件的独立性。
- **定义：** 设 X, Y 是两个随机变量，如果对任意 $a < b,\ c < d$，事件 $\{a < X < b\}$ 与事件 $\{c < Y < d\}$ 相互独立，则称 X 与 Y 是**相互独立**的，简称独立。
- **定理：** 设连续型随机变量 X, Y 分别有分布密度 $p_X(x), p_Y(y)$, 则 X 与 Y 相互独立的充分必要条件是：二元函数

$$p_X(x)p_Y(y)$$

是随机向量 (X, Y) 的联合密度。
- **定理：** 设 X 可能取的值是 $x_1, x_2, \ldots$(有限个或可列个), Y 可能取的值是 $y_1, y_2, \ldots$(有限个或可列个), 则 X 与 Y 相互独立的充分必要条件是：对一切 i, j 成立

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

二维正态分布

- 二维正态分布是最常见最重要的多维分布。
- **定义：** 称 $\boldsymbol{\xi} = (X, Y)$ 服从二维正态分布，如果其密度联合密度为

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

- 其中 $\mu_1, \mu_2, \sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1$ 是 5 个参数。
- $p(x, y)$ 称为二维正态密度。

二维正态分布的性质

- 二维正态分布的边缘分布是正态分布。若 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，那么 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$；
- **性质** 若 (X, Y) 服从二维正态分布（参数为 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$），则 X 与 Y 相互独立 $\Longleftrightarrow \rho = 0$。

二维随机变量的分布函数

- **定义：** 设 (X, Y) 是二维随机变量，称函数

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

为它的**分布函数**或**联合分布函数**。

- 若 $\boldsymbol{\xi} = (X, Y)$ 的分布函数有二阶连续偏导数，则 $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ 就是 $\boldsymbol{\xi}$ 的分布密度。

单个随机变量函数

- 设 (X, Y) 联合密度为 $p(x, y)$, $Z = f(X, Y)$,
- (1) 求 Z 的分布函数

$$P(f(X, Y) \leq z)$$

- (2) 对 $p(x, y)$ 积分

$$P(f(X, Y) \leq z) = \iint\limits_{f(x,y) \leq z} p(x, y)\, dxdy$$

并进行积分变换，最终化为

$$F_Z(z) = \int_{-\infty}^{z} p_Z(u)\, du$$

的形式，或 $F_Z(z)$ 可导的形式。
- 随机变量和的分布。若 (X, Y) 的密度函数为 $p(x, y)$, 那么和 $Z = X + Y$ 的概率密度为卷积

$$\int_{-\infty}^{+\infty} p(x, z - x)\, dx$$

随机变量变换公式

- **定理：** 设 (X, Y) 有联合密度 $p(x, y)$，且区域 A(可以是全平面) 满足 $P((X, Y) \in A) = 1$。又函数 $f(x, y), g(x, y)$ 满足：
- (1) 对任意实数 u, v，方程组

$$\begin{cases} f(x, y) = u \\ g(x, y) = v \end{cases} \tag{2.4}$$

  在 A 中至多有一个解 $x = x(u, v), y = y(u, v)$;
- (2) f, g 在 A 中有连续偏导数；
- (3) 雅可比行列式 $\frac{\partial(f,g)}{\partial(x,y)}$ 在 A 中处处不等于 0。
- 设 $U = f(X, Y), V = g(X, Y)$，

$$G = \{(u, v) : \text{方程组 (2.4) 在 } A \text{ 中有解}\}$$

则

$$q(u, v) = \begin{cases} p(x(u, v), y(u, v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| & \text{当}(u, v) \in G \\ 0 & \text{当}(u, v) \notin G \end{cases}$$

是 (U, V) 的联合密度。

随机变量函数的均值公式

- 设随机向量 (X, Y) 有密度 $p(x, y)$, 对随机变量函数 $Z = f(X, Y)$，有

$$E(Z) = E[f(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)p(x, y)\,dxdy$$

- 要求绝对可积，即 $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x, y)|p(x, y)\,dxdy$ 存在
- 容易验证，这和先求出 $Z = f(X, Y)$ 的密度，然后按照密度的定义给出期望定义是一致的。即若 $Z = f(X, Y)$ 的密度函数为 $q_Z(z)$,

$$\begin{aligned} E(Z) &= \int_{-\infty}^{+\infty} zq_Z(z)\,dz \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)p(x, y)\,dxdy \end{aligned}$$

随机变量的均值与方差

- 设 (X, Y) 的联合密度为 $p(x, y)$，分量的边缘密度为 $p_X(x)$ 和 $p_Y(y)$，则

$$E(X) = \int_{-\infty}^{+\infty} x p_X(x)\, dx$$

$$E(Y) = \int_{-\infty}^{+\infty} y p_Y(y)\, dy$$

$$D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 p_X(x)\, dx$$

$$D(Y) = \int_{-\infty}^{+\infty} [y - E(Y)]^2 p_Y(y)\, dy$$

- 由随机向量函数的期望公式 (3.1) 又有

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xp(x,y)\,dxdy$$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)\,dxdy$$

$$D(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [x-E(X)]^2 p(x,y)\,dxdy$$

$$D(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [y-E(Y)]^2 p(x,y)\,dxdy$$

协方差

- **定义：** 称向量 $(E(X), E(Y))$ 为随机向量 (X, Y) 的均值，称数值 $E\{[X-E(X)][Y-E(Y)]\}$ 为 X, Y 的**协方差**，记为 $\mathrm{Cov}(X, Y)$ 或 σ_{XY}。
- 容易验证

$$\begin{aligned} &\mathrm{Cov}(X, Y) = \sigma_{XY} \\ = &\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(X)][y-E(Y)]p(x, y)\,dxdy \end{aligned}$$

- $D(X)$, $D(Y)$ 也可以记成 σ_{XX}, σ_{YY}。
- 当 X, Y 相互独立时 $\mathrm{Cov}(X, Y) = 0$。
- 若 $\mathrm{Cov}(X, Y) = 0$，则称 X, Y **不相关** 。
- 独立必不相关；不相关不一定独立。

相关系数

- **定义：** 称

$$\frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$$

为 X, Y 的**相关系数**（要求分母不等于 0），记作 ρ_{XY} 或 ρ。即

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$$

- 对二元正态分布，参数 ρ 是两个分量的相关系数。

联合密度与边缘密度

- **定义：** 对于 n 维随机向量 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$, 如果存在非负函数 $p(x_1, x_2, \ldots, x_n)$，使对于任意 n 维长方体

$$
\begin{aligned}
D = &\big\{(x_1, x_2, \ldots, x_n) : a_1 < x_1 < b_1, a_2 < x_2 < b_2, \ldots, \\
&\quad a_n < x_n < b_n\big\}
\end{aligned}
$$

均成立

$$
P(\boldsymbol{\xi} \in D) = \underset{D}{\iint \cdots \int} p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n \tag{4.1}
$$

则称 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$ 是连续型的，并称 $p(x_1, x_2, \ldots, x_n)$ 为 $\boldsymbol{\xi}$ 的分布密度，或称联合分布密度（简称联合密度）。

n 维正态分布

- **定义：** 称随机向量 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布 (也称 $\boldsymbol{\xi}$ 是 n 维正态随机向量), 如果它有分布密度

$$
\begin{aligned}
& p(x_1, x_2, \ldots, x_n) \\
= & \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\Big\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \Big\}
\end{aligned}
$$

  其中 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ 是非随机的常数向量， $\Sigma = (\sigma_{ij})_{n\times n}$ 是 n 阶正定常数矩阵。
- 记 $\boldsymbol{\xi} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$。
- 高维正态分布生成：给定独立同分布 $X_1, \cdots, X_n \sim N(0, 1)$, 设有 n 维可逆矩阵 B, 定义 $Y = BX + \mu$, 其中 $\Sigma = BB^T$.

独立性

- **定义：** 设 $X_1, X_2, \ldots, X_n$ 是 n 个随机变量，如果对任意 $a_i < b_i(i = 1, 2, \ldots, n)$，事件 $\{a_1 < X_1 < b_1\}, \{a_2 < X_2 < b_2\}, \ldots, \{a_n < X_n < b_n\}$ 相互独立，则称 $X_1, X_2, \ldots, X_n$ 是**相互独立的**。
- **定理：** 设 $X_1, X_2, \ldots, X_n$ 的分布密度分别是 $p_1(x_1), p_2(x_2), \ldots, p_n(x_n)$，则 $X_1, X_2, \ldots, X_n$ 相互独立的充分必要条件是

$$p_1(x_1)p_2(x_2)\ldots p_n(x_n)$$

是 $(X_1, X_2, \ldots, X_n)$ 的联合分布密度。

条件分布

- **定义** 设对任意 $\varepsilon > 0$, $P(y-\varepsilon < Y \le y+\varepsilon) > 0$. 若极限

$$\lim_{\varepsilon\to 0} P(X \le x | y-\varepsilon < Y \le y+\varepsilon)$$

存在，则称此极限为 $Y = y$ 的条件下 X 的**条件分布函数**，记作 $P(X \le x | Y = y)$ 或 $F_{X|Y}(x|y)$，即

$$F_{X|Y}(x|y) = \lim_{\varepsilon\to 0} P(X \le x | y-\varepsilon < Y \le y+\varepsilon)$$

- 条件分布函数也是分布函数：单调上升右连续，$-\infty$ 极限为 0，$+\infty$ 极限为 1。
- 条件分布与联合分布有关，被联合分布决定。

离散型的条件分布

- 设 (X, Y) 是二维离散型随机向量，

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

  其中 $P(Y = y_j) > 0, j = 1, 2, \ldots$。
- 则在 $Y = y_j$ 的条件下 X 的条件分布为

$$\begin{aligned} P(X = x_i | Y = y_j) &= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\ &= \frac{p_{ij}}{\sum_k p_{kj}} \quad (i = 1, 2, \ldots) \end{aligned}$$

连续型条件分布

- 设随机向量 (X, Y) 有联合分布函数 $F(x,y)$，联合密度 $p(x,y)$。
- 对 $p(x,y)$ 加一些条件（实际中通常可以满足）后给出连续型条件分布的表达式。

$$F_{X|Y}(x|y) = \frac{\lim\limits_{\varepsilon\to 0} \frac{F(x,y+\varepsilon)-F(x,y-\varepsilon)}{2\varepsilon}}{\lim\limits_{\varepsilon\to 0} \frac{F_Y(y+\varepsilon)-F_Y(y-\varepsilon)}{2\varepsilon}}$$

$$= \frac{\int_{-\infty}^{x} p(u,y)\,du}{p_Y(y)} = \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)}\,du \qquad (\text{当 } p_Y(y) > 0 \text{ 时})$$

- 于是 $Y = y$ 的条件下 X 的**条件分布密度**，记作 $p_{X|Y}(x|y)$ 为

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$$

- 二元正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 的条件密度为

$$Y|X = x \sim \mathrm{N}\left(\mu_2 + \rho\sigma_2 \frac{x-\mu_1}{\sigma_1}, \sigma_2^2(1-\rho^2)\right)$$

条件期望

- **定义** 设 X, Y 是两个随机变量，有联合密度 $p(x, y)$, 设 $Y = y$ 的条件下 X 有条件分布密度 $p_{X|Y}(x|y)$，则
$$\int_{-\infty}^{\infty} x p_{X|Y}(x|y)\, dx$$
叫做 $Y = y$ 的条件下 X 的**条件期望**, 记作 $E(X|Y = y)$。
- 要求上面的积分绝对收敛。
- 对离散型条件分布类似定义条件期望。
- 性质：
$$\begin{aligned}
E[E(X|Y)] &= E(g(Y)) \\
&= \int_{\{y: p_Y(y) > 0\}} g(y) p_Y(y)\, dy \\
&= \int_{\{y: p_Y(y) > 0\}} E(X|Y = y) p_Y(y)\, dy \\
&= E(X)
\end{aligned}$$

大数定律

- **定理 (弱大数定律)** 设 $X_1, X_2, \ldots, X_n, \ldots$ 是独立同分布的随机变量列，且 $E(X_1)$, $D(X_1)$ 存在，则对任何 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) = 0$$

  其中 $S_n = X_1 + X_2 + \cdots + X_n$; 称上述同分布的随机变量列 $X_1, X_2, \ldots, X_n, \ldots$ 叫做服从**弱大数定律**

- 更多数学讨论可以证明：只要 $E(X_1)$ 存在，不管 $D(X_1)$ 是否存在，就有

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = E(X_1)\right) = 1$$

  称同分布的随机变量列 $X_1, X_2, \ldots, X_n, \ldots$ 叫做服从**强大数定律**

中心极限定理

- **定理 (中心极限定理)** 设 $X_1, X_2, \ldots, X_n, \ldots$ 是独立同分布随机变量列，而且 $E(X_1)$, $D(X_1)$ 存在，$D(X_1) > 0$，则对一切实数 $a < b$，有

$$\lim_{n\to\infty} P\left(a < \frac{S_n - nE(X_1)}{\sqrt{nD(X_1)}} < b\right)$$
$$= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du = \int_a^b \varphi(u)\, du$$

- 其中 $\varphi(\cdot)$ 是标准正态分布密度。

三种收敛方式

设 $\xi_n, n=1,2,\ldots$ 为随机变量序列，ξ 为随机变量（可以为常数）

- 若

$$\lim_{n\to\infty} P\left(|\xi_n-\xi|>\varepsilon\right)=0, \forall\varepsilon>0.$$

 则称 ξ_n **依概率收敛**到 ξ，记作 $\xi_n \xrightarrow{\mathrm{p}} \xi, \quad (n\to\infty).$

- 若

$$P\left(\lim_{n\to\infty}\xi_n=\xi\right)=1$$

 则称 ξ_n **以概率 1 收敛**到 ξ，记作 $\lim\limits_{n\to\infty}\xi_n=\xi$, a.s.

- 设 $f_n(x)$ 是 ξ_n 的分布函数，$f(x)$ 是 ξ 的分布函数，若

$$\lim_{n\to\infty} f_n(x)=f(x), \forall x\in R.$$

 则称 ξ_n **依分布收敛**到 ξ，记作 $\xi_n \xrightarrow{d} \xi$.