一、事件和概率

二、随机变量和分布函数

三、数字特征

四、极限定理

一、事件和概率

1. 概率空间
2. 条件概率
3. 独立性

1. 概率空间

概率空间是随机现象的数学模型，它由三个基本要素组成，通常写作$(\Omega, \mathcal{F}, P)$。具体含义如下，

Ω是样本空间，包含随机现象的所有可能基本结果；

$\mathcal{F}$是σ-事件域，具有下列基本性质：

(1) $\emptyset, \Omega \in \mathcal{F}$

(2) 如果$A \in \mathcal{F}$，那么$\bar{A} \in \mathcal{F}$

(3) 如果$A_n, n \geq 1$是一列事件，$A_n \in \mathcal{F}$，那么

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$

P 是概率，即定义在$\mathcal{F}$上、取值于$[0,1]$的函数，满足下列性质：

(1) 规范性

$$P(\emptyset)=0, \qquad P(\Omega)=1$$

(2) 可数可加性

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

例1. 令Ω为样本空间，$A \subset \Omega$。定义

$$\mathcal{F} = \{\emptyset, \Omega, A, \bar{A}\}$$

$$P(\emptyset) = 0, \quad P(\Omega) = 1$$

$$P(A) = p, \quad P(\bar{A}) = 1 - p, \quad 0 < p < 1$$

那么，$(\Omega, \mathcal{F}, P)$是概率空间，最简单的非平凡概率空间。

例2. 古典概率模型

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}, \quad N < \infty$$

$$\mathcal{F} = 2^{\Omega}$$

所有子集构成的σ-域

$$P(A) = \frac{|A|}{N}$$

那么，$(\Omega, \mathcal{F}, P)$是一个古典概率空间。
等价地，每个基本结果等可能地发生。

例3. 几何概率模型

假设Ω是$\mathbb{R}^d$, $d \geq 1$ 的一个可测区域, $|\Omega| < \infty$。

$$\mathcal{F} = \mathcal{B}$$

$$P(A) = \frac{|A|}{|\Omega|}, \quad A \in \mathcal{F}$$

那么，$(\Omega, \mathcal{F}, P)$是一个几何概率空间。

等价地，每个点等可能地发生。

2. 条件概率

初等概率论的主要内容就是研究如何计算事件的概率，特别给定一个概率空间，如何从简单事件的概率，计算出复杂事件的概率。一个重要的技巧是利用条件概率。

假设$(\Omega, \mathcal{F}, P)$是一个概率空间，$A, B \in \mathcal{F}$。假设$P(A) > 0$，那么给定A发生的条件下，B发生的条件概率为

$$P(B|A) = \frac{P(AB)}{P(B)}$$

固定A，$P(\cdot|A)\mathcal{F} \mapsto [0,1]$ 是一个概率。

乘法公式：

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

链式法则：

$$P(A_1A_2\cdots A_m) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1)\cdots P(A_m|A_{m-1}\cdots$$

全概率公式：

假设$B_1, B_2, \cdots, B_n$是一列互不相交的事件，$\sum_{i=1}^{n} B_i = \Omega$，那么

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

Bayes公式：

假设$B_1, B_2, \cdots, B_n$是一列互不相交的事件，$\sum_{i=1}^n B_i = \Omega$，那么

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^n P(B_i)P(A|B_i)}$$

3. 独立性

假设$(\Omega, \mathcal{F}, P)$是一个概率空间。

- 两个事件独立：

假设$A, B \in \mathcal{F}$，如果

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

那么称A和B相互独立。

如果$P(B) > 0$，那么A和B相互独立当且仅当

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

如果A和B相互独立，那么$\bar{A}$和B相互独立；A和$\bar{B}$相互独立；$\bar{A}$和$\bar{B}$相互独立。

- m个事件相互独立：

假设$A_1, A_2, \cdots, A_m \in \mathcal{F}$，如果对任意$1 \leq k \leq m$，任意指标$1 \leq i_1, i_2, \cdots, i_k \leq m$

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

那么称$A_1, A_2, \cdots, A_m \in \mathcal{F}$相互独立。

如果$A_1, A_2, \cdots, A_m \in \mathcal{F}$相互独立，那么其中任意$r < m$个事件也相互独立。

- m个事件两两独立：

假设$A_1, A_2, \cdots, A_m \in \mathcal{F}$，如果对任意$1 \le i, j \le m$，

$$P(A_iA_j) = P(A_i)P(A_j)$$

那么称$A_1, A_2, \cdots, A_m \in \mathcal{F}$两两独立。

m个事件相互独立可推出m个事件两两独立；反之，不然。

如果$A_1, A_2, \cdots, A_m \in \mathcal{F}$相互独立，$I, J \subset \{1, 2, \cdots, m\}$并且$I \cap J = \emptyset$，那么由$A_i, i \in I$的事件通过并、交、差等运算得到的事件和由$A_j, j \in J$的事件通过并、交、差等运算得到的事件相互独立。

- σ-域独立：

假设$(\Omega, \mathcal{F}, P)$是一个概率空间，$\mathcal{A}_1, \mathcal{A}_2$是$\mathcal{F}$的两个子$\sigma$-域；如果任意$A_1 \in \mathcal{A}_1$和$A_2 \in \mathcal{A}_2$相互独立，那么称$\mathcal{A}_1$和$\mathcal{A}_2$是两个相互独立的$\sigma$-域。

多个情形类似定义。

二、随机变量和分布函数

1. 随机变量
2. 分布函数
3. 随机变量函数的分布
4. 随机向量和联合分布函数
5. 条件分布函数
6. 独立性

1. 随机变量

假设$(\Omega, \mathcal{F}, P)$是一个概率空间，$X : \Omega \mapsto \mathbb{R}$是一个函数，满足下列可测性质：

$$\{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}, \quad \forall A \in \mathcal{B}$$

其中$\mathcal{B}$ 为$\mathbb{R}$ 上的Borel σ-域，那么称X为实值随机变量。可测性要求也可写成

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}, \quad \forall x \in \mathbb{R}$$

2. 分布函数

假设$(\Omega, \mathcal{F}, P)$是一个概率空间，$X : \Omega \mapsto \mathbb{R}$是一个随机变量，定义

$$F_X(x) = P(\omega \in \Omega : X(\omega) \leq x), \quad x \in \mathbb{R}$$

称$F_X(x)$为X的分布函数。

分布函数具有下列性质：

(1)

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$$

(2) $F_X(x)$关于x单调不减函数

(3) $F_X(x)$关于x右连续函数

离散型随机变量只取有限个或者可列个值，其分布函数为阶梯型函数；
连续型随机变量在一个区间上取值，具有概率密度函数，其分布函数为光滑连续曲线
存在既不是离散型，也不是连续型的随机变量。
任给一个满足上述三条的函数$F(x)$，一定存在一个概率空间$(\Omega, \mathcal{F}, P)$和随机变量$X: \Omega \mapsto \mathbb{R}$使得

$$P(\omega \in \Omega : X(\omega) \leq x) = F(x), \quad \forall x \in \mathbb{R}$$

3. 随机变量函数的分布

常见的基本随机变量包括两点分布、二项分布、Poisson 分布、几何分布、超几何分布、均匀分布、指数分布、正态分布、Chi-方分布、t-分布、F-分布等等。但在实际问题中，所遇到的随机变量往往是这些基本随机变量的函数。初等概率论的一个基本内容就是研究如何计算随机变量函数的分布函数。更具体地说，

假设$(\Omega, \mathcal{F}, P)$是概率空间，$X : \Omega \mapsto \mathbb{R}$是随机变量，具有分布函数$F_X(x)$。假设$f : \mathbb{R} \mapsto \mathbb{R}$是一个Borel可测函数，那么$Y =: f(X)$是一个随机变量，问题在于如何计算$Y$的分布函数。

4. 随机向量和联合分布函数

假设$(\Omega, \mathcal{F}, P)$是概率空间，$(X, Y) : \Omega \mapsto \mathbb{R}^2$是一个二元函数。如果对任意$x, y \in \mathbb{R}$，

$$\{\omega \in \Omega : X(\omega) \le x, Y(\omega) \le y\} \in \mathcal{F}$$

那么称(X, Y)为二元随机向量。

联合分布函数：

定义

$$F(x, y) = P(\omega \in \Omega : X(\omega) \le x, Y(\omega) \le y)$$

称$F(x, y)$为(X, Y)的联合分布函数。

边际分布函数：

称X和Y的分布函数$F_X(x)$和$F_Y(y)$为边际分布函数。

联合分布函数唯一确定边际分布函数；但反之，不然。

条件分布函数：

假设$(\Omega, \mathcal{F}, P)$是概率空间，(X, Y)是二元随机向量，具有联合分布函数$F(x, y)$。

给定$X = x$，Y的条件分布函数为

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

给定$Y = y$，X的条件分布函数为

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

5. 独立性

假设$(\Omega, \mathcal{F}, P)$是概率空间，(X, Y)是二元随机向量，具有联合分布函数$F(x, y)$，边际分布函数$F_X(x)$和$F_Y(y)$。

如果对任意x, y，

$$F(x, y) = F_X(x)F_Y(y)$$

那么称随机变量X, Y 相互独立。

X, Y相互独立当且仅当对任意Borel集$A, B \in \mathcal{B}$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

特别，如果X, Y相互独立，$f, g : \mathbb{R} \mapsto \mathbb{R}$是两个Borel可测函数，那么$f(X), g(Y)$相互独立。

例1. 假设(X, Y)是矩形区域$(a, b) \times (c, d)$的均匀分布，那么X是区间(a, b)上的均匀分布，Y是区间(c, d)上的均匀分布；并且X, Y相互独立。

例2. 假设(X, Y)是单位圆$\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$的均匀分布，那么$X$和$Y$服从$(0, 1)$上的半圆律，即密度函数为

$$p(x) = \frac{1}{\pi}\sqrt{1 - x^2}, \quad |x| < 1$$

但X和Y并不相互独立。

例3. 假设(X, Y)是二元联合正态随机变量$N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$。那么

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

X, Y相互独立的充要条件为$\rho = 0$；

给定$X = x$，Y的条件分布为$N()$

给定$Y = y$，X的条件分布为$N()$

三、数字特征

1. 数学期望
2. 方差和Chebyschev 不等式
3. 相关系数
4. 特征函数

1. 数学期望

假设X是一个离散型随机变量，其分布列为

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n \cdots \\ p_1 & p_2 & \cdots & p_n \cdots \end{pmatrix}$$

如果

$$\sum_{n=1}^{\infty} |x_n| p_n < \infty$$

那么称X的数学期望存在，记为

$$EX = \sum_{n=1}^{\infty} x_n p_n$$

假设X是一个连续型随机变量，其密度函数为$p(x)$。如果

$$\int_{-\infty}^{\infty} |x|p(x)dx < \infty$$

那么称X的数学期望存在，记为

$$EX = \int_{-\infty}^{\infty} xp(x)dx$$

假设X是一个随机变量，其分布函数为$F(x)$。如果

$$\int_{-\infty}^{\infty} |x|dF(x) < \infty$$

其中积分为Riemann-Stieltjes积分，那么称X的数学期望存在，记为

$$EX = \int_{-\infty}^{\infty} xdF(x)$$

数学期望具有以下基本性质：

(1) 如果$P(X=c)=0$，那么$EX=c$

(2) 如果$a\le X\le b$，那么$a\le EX\le b$

(3) (线性性质) $E(aX+b)=aEX+b$

(4) (加法定理) $E(X+Y)=EX+EY$

(5) 如果$X\sim F(x)$，$f:\mathbb{R}\mapsto\mathbb{R}$是Borel 可测函数，那么

$$Ef(X)=\int_{-\infty}^{\infty}f(x)dF(x)$$

(假定上述等式两边存在)

(6) (全期望公式)

$$EX=E(E(X|Y))$$

2. 方差

假设X是随机变量，具有二阶矩，即$EX^2 < \infty$，那么定义其方差为

$$Var(X) = E(X - EX)^2 = EX^2 - (EX)^2$$

方差的基本性质：

(1) 如果$P(X=c)=1$，那么$Var(X)=0$

(2) 如果$Var(X)=0$，那么$P(X=EX)=1$

(3)

$$Var(ax+b)=a^2Var(X)$$

(4) (加法公式) 如果X,Y相互独立，那么

$$Var(X+Y)=Var(X)+Var(Y)$$

(5) 如果$a\neq EX$，那么

$$Var(X)<E(X-a)^2$$

3. Chebyschev 不等式

对任意$x > 0$

$$P(|X - EX| \geq x) \leq \frac{Var(X)}{x^2}$$

Markov 不等式

假设f是一个单调非负函数，那么对任意$x > 0$

$$P(X \geq x) \leq \frac{Ef(X)}{f(x)}$$

4. 相关系数

假设(X, Y) 是随机向量，具有二阶矩$EX^2 < \infty$，$EY^2 < \infty$。那么称(EX, EY) 为均值向量，

$Cov(X, Y) = E(X - EX)(Y - EY)$ 为协方差，

协方差矩阵为

$$\begin{pmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{pmatrix}$$

相关系数为

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

Cauchy-Schwarz 不等式

$$|Cov(X,Y)| \leq \sqrt{Var(X) \cdot Var(Y)}$$

$$-1 \leq \rho_{X,Y} \leq 1$$

如果$\rho_{X,Y} = 0$，那么称X,Y不相关。更确切地说，X,Y不线性相关。

如果X,Y相互独立，那么X,Y不相关；但反之不然。

如果$(X,Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$，那么$X,Y$相互独立当且仅当$\rho = 0$。

5. 特征函数

特征函数对于研究随机变量的分布函数起着重要作用。

假设X是一个随机变量，具有分布函数$F(x)$。定义

$$\phi(t) = Ee^{itX} = \int_{-\infty}^{\infty} e^{itx} dF(x), \quad t \in \mathbb{R}$$

称$\phi(t)$为X的特征函数。

任何随机变量的特征函数总是存在的。

特征函数具有下列基本性质：

(1) $\phi(0)=1$；

(2) $|\phi(t)|\le 1$；

(3) $\phi(t)$ 在$\mathbb{R}$上一致连续；

(4) $\phi(t)$是非负定的，即对任意实数$t_1,t_2,\cdots,t_m$和复数$z_1,z_2,\cdots,z_m$，

$$\sum_{i,j=1}^{n} z_i\bar{z}_j\phi(t_i-t_j)\ge 0$$

(5) 如果$E|X|^k < \infty$，那么$\phi(t)$ k次可微，且在0处可进行Taylor展开：

$$\phi(t) = 1 + \sum_{m=1}^{k} \frac{i^m}{m!}\alpha_m t^m + \theta\frac{\beta_k}{(k+1)!}t^{k+1}$$

其中α_m为m-th 半不变累积量，β_k为k-th阶绝对矩，$|\theta| \le 1$。

(6) 如果X, Y相互独立，那么

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

唯一性定理：分布函数和特征函数相互唯一确定。

(1) 逆转公式

(2) 如果随机变量X的特征函数$\phi(t)$绝对可积，即

$$\int_{-\infty}^{\infty} |\phi(t)| dt < \infty$$

那么X具有密度函数，

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt$$

(3) 如果随机变量X的特征函数$\phi(t)$具有下列形式：

$$\phi(t) = \sum_{n=1}^{\infty} a_n e^{int}$$

并且

$$a_n \geq 0, \quad \sum_{n=1}^{\infty} a_n = 1$$

那么X是离散型随机变量，具有分布

$$P(X = n) = a_n$$

四、极限定理

1. 依概率收敛
2. 大数定律
3. 依分布收敛
4. 中心极限定律
5. 几乎处处收敛
6. 强大数律
7. 均方收敛

1. 依概率收敛

假设$(\Omega,\mathcal{F},P)$是概率空间，$X, X_n, n\geq 1$是一列随机变量。

如果对任意$\varepsilon>0$

$$P(\omega : |X_n(\omega)-X(\omega)|>\varepsilon)\to 0$$

那么称X_n依概率收敛到X，记作$X_n \xrightarrow{P} X$。

依概率收敛的基本性质：

(1) 如果$X_n \xrightarrow{P} X$，$X_n \xrightarrow{P} Y$，那么

$$P(X=Y)=1$$

(2) 如果$X_n \xrightarrow{P} X$，$X_n - Y_n \xrightarrow{P} 0$，那么

$$Y_n \xrightarrow{P} X$$

(3) 如果$X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y$，那么

$$X_n \pm Y_n \xrightarrow{P} X \pm Y$$

并且

$$X_n \cdot Y_n \xrightarrow{P} X \cdot Y$$

(4) 如果$X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y$，那么

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$$

(5) 如果$X_n \xrightarrow{P} X$，f是连续函数，那么

$$f(X_n) \xrightarrow{P} f(X)$$

- 大数律:

1. Bernoulli大数律

假设$S_n \sim B(n,p), n \geq 1, 0 < p < 1$。那么

$$\frac{S_n}{n} \xrightarrow{P} p$$

2. Chebyschev 大数律

假设$\xi_n, n \geq 1$是一列随机变量，$E\xi_n^2 < \infty$。令$S_n = \sum_{k=1}^n \xi_k$

如果

$$\frac{Var(S_n)}{n^2} \to 0$$

那么

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0$$

3. Khinchine大数律

假设$\xi_n, n \geq 1$是一列独立同分布随机变量，$E\xi_1 = \mu$。

令$S_n = \sum_{k=1}^n \xi_k$，那么

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

2. 依分布收敛

假设$F(x), F_n(x), n \geq 1$是一列分布函数，如果对每一个F的连续点x，都有

$$F_n(x) \to F(x)$$

那么称F_n弱收敛到F，记作$F_n \xrightarrow{w} F$。

假设随机变量X, X_n分别定义在概率空间$(\Omega, \mathcal{F}, P)$，$(\Omega_n, \mathcal{F}_n, P_n)$上，$n \geq 1$。如果相应的分布函数弱收敛，那么称$X_n$依分布收敛到$X$，记作$X_n \xrightarrow{d} X$。

依分布收敛的基本性质

(1) 如果$X_n \xrightarrow{P} X$，那么$X_n \xrightarrow{d} X$

(2) 如果$X_n \xrightarrow{d} c$，那么$X_n \xrightarrow{P} c$

(3) 如果$X_n - Y_n \xrightarrow{P} 0$，并且$X_n \xrightarrow{d} X$，那么$Y_n \xrightarrow{d} Y$

(4) 如果$X_n \xrightarrow{d} X$，f是连续函数，那么$f(X_n) \xrightarrow{d} f(X)$

Levy连续性定理

假设$X, X_n, n \geq 1$是一列随机变量，特征函数分别为$\phi(t)$，$\phi_n(t)$。那么$X_n \xrightarrow{d} X$ 当且仅当对每个t，$\phi_n(t) \longrightarrow \phi(t)$。

假设$X_n, n \geq 1$是一列随机变量，特征函数为$\phi_n(t)$。如果对每个t，$\phi_n(t)$ 收敛到某函数$\phi(t)$，并且ϕ在$t = 0$处连续，那么$\phi(t)$一定是某随机变量X的特征函数，进而$X_n \xrightarrow{d} X$ 。

1. De Moivre-Laplace中心极限定理

假设$S_n \sim B(n,p), n \geq 1, 0 < p < 1$。那么

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

2. Levy-Feller 中心极限定理

假设$\xi_n, n \geq 1$是一列独立同分布随机变量，$E\xi_1 = \mu$，$Var(\xi_1) = \sigma^2 < \infty$。那么

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)$$

3.几乎处处收敛

假设$X, X_n, n \geq 1$是定义在概率空间$(\Omega, \mathcal{F}, P)$上得一列随机变量，如果存在$\Omega_0 \in \mathcal{F}$，使得

$$P(\Omega_0) = 0$$

并且对任意$\omega \in \Omega \setminus \Omega_0$，

$$X_n(\omega) \longrightarrow X(\omega)$$

那么称X_n几乎处处收敛到X，记作$X_n \longrightarrow X$ a.s.

几乎处处收敛的基本性质：

1. 如果$X_n \longrightarrow X$ a.s.，那么$X_n \xrightarrow{P} X$。

2. 如果对任意$\varepsilon > 0$，

$$\sum_{n=1}^{\infty} P(\omega : |X_n(\omega) - X(\omega)| > \varepsilon) < \infty$$

那么$X_n \longrightarrow X$ a.s.

1. Borel强大数律：

令$\xi_n, n \geq 1$是一列独立同分布的随机变量，分布为

$$P(\xi_1 = 1) = p, \qquad P(\xi_1 = 0) = 1 - p$$

定义

$$S_n = \sum_{k=1}^{\infty} \xi_k$$

那么

$$\frac{S_n}{n} \longrightarrow p, \quad a.s.$$

2. Kolmogorov强大数律:

令$\xi_n, n \geq 1$是一列独立同分布的随机变量，$E\xi_1 = \mu$。定义

$$S_n = \sum_{k=1}^{\infty} \xi_k$$

那么

$$\frac{S_n}{n} \longrightarrow \mu, \quad a.s.$$

4. 均方收敛

假设$X, X_n, n \geq 1$是一列定义在概率空间$(\Omega, \mathcal{F}, P)$上的随机变量，具有有限r阶矩$(r > 0)$。如果

$$E|X_n - X|^r \longrightarrow 0$$

那么称X_n r阶均方收敛到X，记作$X_n \xrightarrow{L^r} X$。

如果$X_n \xrightarrow{L^r} X$，那么$X_n \xrightarrow{P} X$。