概率论复习

Tianxiao Pang

Zhejiang University

December 20, 2021

内容

1. 第一章: 事件及其概率
2. 第二章: 随机变量与分布函数
3. 第三章: 数字特征与特征函数
4. 第四章: 极限定理

事件及其概率

事件: 样本空间Ω的子集.

有了概率的公理化定义后, 我们假定事件都来自事件域, 使得它有概率.

事件的关系和运算, 略.

概率的统计定义: 把频率稳定于的那个常数作为概率大小.

定义（古典概率）

设一试验有n个等可能发生的样本点，而事件A包含其中的m个样本点，则事件A发生的概率$P(A)$定义为

$$P(A)=\frac{m}{n}=\frac{A\text{包含的样本点个数}}{\text{样本空间中样本点总数}}=\frac{\#A}{\#\Omega}.$$

摸球问题(抽签问题)、生日问题.

注: 使用古典概率前，应先验证所考察的事件来自古典概型(满足有限性与等可能性的概率模型).

定义 (几何概率)

假设“等可能性”成立. 记

$$A_g = \{\text{任取一个样本点, 它落在区域} g \subset \Omega\},$$

则定义A_g为

$$\mathsf{P}(A_g) = \frac{g\text{的测度}}{\Omega\text{的测度}} = \frac{m(g)}{m(\Omega)}.$$

会面问题.

注: “等可能性”理解为: 对任意两个区域, 当它们的测度(长度, 面积, 体积, $\cdots$)相等时, 样本点落在这两区域上的概率相等, 而与形状和位置都无关.

概率公理化定义:

定义

概率是定义在$\mathscr{F}$上的实值集函数: $A(\in \mathscr{F}) \mapsto P(A)$, 且满足下列三个条件:

- 非负性: 对任一$A \in \mathscr{F}$, $P(A) \geq 0$;
- 规范性: $P(\Omega) = 1$;
- 可列可加性: 若$A_1, \cdots, A_n, \cdots$是$\mathscr{F}$中两两互不相容事件, 则

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

概率的性质:

1. $\mathsf{P}(\phi) = 0$.

2. 有限可加性: 若$A_iA_j = \phi, i, j = 1, 2, \cdots, n, i \neq j$, 则

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

3. $\mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A)$.

4. 若$B \subset A$, 则$\mathsf{P}(A - B) = \mathsf{P}(A) - \mathsf{P}(B)$.

5. $\mathsf{P}(A - B) = \mathsf{P}(A) - \mathsf{P}(AB)$.

6. 加法公式: $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$.

7. 多还少补(exclusion-inclusion)定理:

$$\mathsf{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{1\le i<j\le n} \mathsf{P}(A_iA_j) + \sum_{1\le i<j<k\le n} \mathsf{P}(A_iA_jA_k) + \cdots + (-1)^{n-1}\mathsf{P}(A_1A_2\cdots A_n).$$

8. 次可加性: $\mathsf{P}(\bigcup_{i=1}^{n} A_i) \le \sum_{i=1}^{n} \mathsf{P}(A_i)$. ($\mathsf{P}(\bigcup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} \mathsf{P}(A_i)$ 也成立)

概率的连续性定理:

定理

如果$A_1, A_2, \cdots$是一列单调增加(或减少)的事件序列, 具有极限$A = \cup_{n=1}^{\infty} A_n$(或$A = \cap_{n=1}^{\infty} A_n$), 则

$$P(A) = \lim_{n\to\infty} P(A_n).$$

定义（条件概率）

对任一事件A和B，若$P(B) \neq 0$，则“在事件B发生的条件下A发生的概率”记作$P(A|B)$，定义为

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

容易验证：这样定义的条件概率满足非负性，规范性和可列可加性．即条件概率具有概率的一切性质．

概率的乘法公式:

- 若$\mathsf{P}(B) \neq 0$, 则$\mathsf{P}(AB) = \mathsf{P}(B)\mathsf{P}(A|B)$;
- 若$\mathsf{P}(A) \neq 0$, 则$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B|A)$;
- 若$\mathsf{P}(A_1A_2\cdots A_{n-1}) \neq 0$, 则

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

若事件组$\{A_1, A_2, \cdots, A_n, \cdots\}$满足下列两个条件:

- $A_i, i = 1, 2, \cdots$,两两互不相容, 且$\mathsf{P}(A_i) > 0$;
- $\sum_{i=1}^{\infty} A_i = \Omega$.

则称$\{A_1, A_2, \cdots, A_n, \cdots\}$是$\Omega$的一个完备事件组.

全概率公式: 若$\{A_1, A_2, \cdots, A_n, \cdots\}$是$\Omega$的一个完备事件组, 则对任意事件$B$, 有

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

贝叶斯公式: 若$\{A_1, A_2, \cdots, A_n, \cdots\}$是$\Omega$的一个完备事件组, 则

$$\mathsf{P}(A_i|B) = \frac{\mathsf{P}(A_iB)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A_i)\mathsf{P}(B|A_i)}{\sum_{k=1}^{\infty} \mathsf{P}(A_k)\mathsf{P}(B|A_k)} \quad i = 1, 2, \cdots.$$

事件的独立性:

两个事件的独立性: 对于事件A和B, 若

$$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B),$$

则称A与B相互独立.

任意有限个事件的独立性: 对于n个事件$A_1, A_2, \cdots, A_n$, 若对其中的任意$k(2 \le k \le n)$个事件:A_{i_1}, $A_{i_2}, \cdots$, A_{i_k}, 有

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

则称n个事件$A_1, A_2, \cdots, A_n$相互独立.

定义（伯努里概型)

若试验E满足下列三个条件:

- 是n次独立重复试验;
- 每次试验只有两个结果: $A, \bar{A}$;
- 每次试验中, A发生的概率是固定不变的: $P(A) = p$.

则称这种试验E为(n重)伯努里概型.

(n重)伯努里概型中A发生的总次数是一个二项分布随机变量.

例1. 加工某一零件需经过三道工序, 各道工序的次品率分别为2%, 3%和5%, 假定各工序互不影响, 求加工后所得零件的次品率.

例1. 加工某一零件需经过三道工序, 各道工序的次品率分别为2%, 3%和5%, 假定各工序互不影响, 求加工后所得零件的次品率.

例2. 甲袋中有a只白球, b只黑球; 乙袋中有c只白球, d只黑球. 某人从甲袋中任取两球投入乙袋, 然后在乙袋中任取两球, 求最后所得两球全为白球的概率.

随机变量与分布函数

定义 (随机变量)

设$\xi(\omega)$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的单值实函数，且对于任一Borel集B，有

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathscr{F}, \tag{2.1}$$

就称$\xi(\omega)$为随机变量(random variable)，而称$P(\xi(\omega) \in B)$，$B \in \mathscr{B}$(一维Borel域)为随机变量$\xi(\omega)$的概率分布(probability distribution).

随机变量分类：离散型、连续型、其它.

定义（离散型随机变量）

若随机变量ξ可能取的值至多可列个(有限个或可列无穷多个), 则称ξ为离散型随机变量(discrete random variables).

对于离散型随机变量ξ, 我们往往关心两方面的内容: (1) 可能取值的集合$\{x_i; i = 1, 2, \cdots\}$; (2) ξ取这些值的概率, 即

$$p_i := p(x_i) := \mathsf{P}(\xi = x_i), i = 1, 2, \cdots .$$

我们称

$$\mathsf{P}(\xi = x_i) = p(x_i), i = 1, 2, \cdots$$

或

ξ	x_1	x_2	$\cdots$	x_n	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

为ξ的**分布列**(distribution sequence), 有时也称为ξ的概率分布.

分布列性质:

- $p(x_i) \geq 0, i = 1, 2, \cdots;$
- $\sum_{i=1}^{\infty} p(x_i) = 1.$

一些常见的、重要的离散型随机变量:

1. 退化分布(degenerate distribution)

定义

设随机变量ξ只取一个常数值c, 即

$$P(\xi = c) = 1,$$

我们称它为退化分布, 又称为单点分布.

2. 两点分布(two-point distribution)

定义

若一个随机试验只有两个可能值x_1, x_2, 且相应的分布列为

ξ	x_1	x_2
P	p	q

其中$p, q > 0, q = 1 - p$. 则称ξ服从两点分布.

3. 二项分布(binomial distribution)

定义

若随机变量ξ的分布列为

$$P(\xi = k) = C_n^k p^k q^{n-k} =: b(k; n, p), \quad k = 0, 1, 2, \cdots, n,$$

其中$p, q > 0, p + q = 1$. 则称ξ服从参数为n, p的二项分布, 记作$\xi \sim B(n, p)$.

$n = 1$时的二项分布即为$0 - 1(p)$分布.

二项分布具有可加性: 若$\xi_j, j = 1, \cdots, k$ 各自服从二项分布$B(n_j, p)$, 且相互独立, 则

$$\sum_{j=1}^{k} \xi_j \sim B(\sum_{j=1}^{k} n_j, p).$$

4. 泊松分布(Poisson distribution)

定义

若随机变量ξ的分布列为

$$P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda} \ (\lambda > 0), \ \ k = 0, 1, 2, \cdots ,$$

则称ξ服从参数为λ的泊松分布, 记为$\xi \sim P(\lambda)$.

泊松分布具有可加性: $\xi_j, j = 1, \cdots, k$ 各自服从Poisson分布$P(\lambda_j)$, 且相互独立, 则

$$\sum_{j=1}^{k} \xi_j \sim P(\sum_{j=1}^{k} \lambda_j).$$

5. 几何分布(geometric distribution)

定义

若随机变量ξ的分布列为

$$P(\xi = k) = q^{k-1}p, \ \ p, q > 0, \ p + q = 1, \ \ k = 1, 2, \cdots ,$$

则称ξ服从几何分布.

几何分布具有无记忆性: 若Bernoulli试验中前m次失败, 则从第$m+1$次开始直到首次成功的试验次数也服从几何分布(好像把前面的m次失败“忘记”了). 即,

$$\mathrm{P}(\xi = m + k | \xi > m) = \mathrm{P}(\xi = k) = q^{k-1}p.$$

6. 超几何分布(hype-geometric distribution)

定义

若随机变量ξ的分布列为

$$P(\xi=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},\ n\le N, M\le N,\ k=0,1,\cdots,\min(n,M).$$

则称ξ服从超几何分布.

超几何分布的一个典型例子: 在产品质量的不放回抽样中, 若N件产品中有M件次品, 则抽检n件时所得次品数服从超几何分布.

定义（分布函数）

设$\xi(\omega)$为概率空间$(\Omega, \mathscr{F}, P)$上的随机变量．我们称

$$F(x) = P(\xi(\omega) \le x), \quad -\infty < x < \infty$$

为随机变量$\xi(\omega)$的分布函数．

分布函数具有下列性质:

(1) 若$a \leq b$, 则$F(a) \leq F(b)$; (单调不减性)
(2) $F(-\infty) := \lim\limits_{x \to -\infty} F(x) = 0, \ F(\infty) := \lim\limits_{x \to \infty} F(x) = 1;$
(3) $F(x+0) = \lim\limits_{\varepsilon \searrow 0} F(x+\varepsilon) = F(x)$. (右连续性)

离散型随机变量的分布函数

设ξ的分布列为

ξ	x_1	x_2	$\cdots$	x_k	$\cdots$
P	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_k)$	$\cdots$

且$x_1 < x_2 < \cdots < x_k < \cdots$. 则$\xi$的分布函数为

$$F(x) = \begin{cases} 0, & x < x_1, \\ p(x_1), & x_1 \le x < x_2, \\ \vdots & \vdots \\ \sum_{i \le k} p(x_i), & x_k \le x < x_{k+1}, \\ \vdots & \vdots \end{cases}$$

定义 (连续型随机变量)

若随机变量ξ可取某个区间(有限或无限)中的一切值, 并且存在某个非负的可积函数$p(x)$, 使分布函数$F(x)$满足

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

则称ξ为连续型随机变量, 称$p(x)$为ξ的(概率)密度函数.

连续型随机变量的性质:

(1) $F(x)$是连续函数, 在$p(x)$的连续点上$F(x)$可导, 且

$$F^{'}(x) = p(x).$$

(2) ξ落在任何一个一维Borel集B上的概率可通过求定积分来确定:

$$\begin{aligned} \mathsf{P}(a < \xi \leq b) =& F(b) - F(a) \\ =& \int_{-\infty}^{b} p(y)\mathrm{d}y - \int_{-\infty}^{a} p(y)\mathrm{d}y \\ =& \int_{a}^{b} p(y)\mathrm{d}y, \end{aligned}$$

因此,

$$\mathsf{P}(\xi \in B) = \int_{B} p(y)\mathrm{d}y.$$

(3) 对任意常数c,

$$\mathsf{P}(\xi = c) = F(c) - F(c-0) = \lim_{h \searrow 0} \int_{c-h}^{c} p(y)\mathrm{d}y = 0.$$

因此, 连续型随机变量等于任何一个常数的概率都为0. 这与离散型随机变量有本质的区别.

密度函数的性质:

- $p(x) \geq 0$;
- $\int_{-\infty}^{\infty} p(y)\mathrm{d}y = 1$.

常见的连续型随机变量:

1. 均匀分布(uniform distribution)

定义 (均匀分布)

若随机变量ξ具有密度函数

$$p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b,\\ 0, & \text{其它}.\end{cases}$$

则称ξ服从(a,b)上的均匀分布. 记为$\xi\sim U(a,b)$.

分布函数为

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

2. 正态分布(normal distribution)

定义（正态分布）

若随机变量ξ的密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}},\quad -\infty<x<\infty,$$

则称ξ服从参数为(a,σ^2)的正态分布$(-\infty<a<\infty,\sigma>0)$. 记作$\xi\sim N(a,\sigma^2)$.

分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}\mathrm{d}t,\quad -\infty<x<\infty.$$

正态分布的计算:

(1) $\xi \sim N(0,1)$时,

- 当$x \geq 0$时, 查正态分布表(pages 223-224): 每隔一定数值可以查到对应的分布函数$\Phi(x)$的值(特别地, $\Phi(0)=0.5$); 在这些数值之间, 可以用线性插值法求得相应的函数值.
- 当$x < 0$时, 注意到

$$\Phi(x)+\Phi(-x)=1,$$

因此可以先通过查表或线性插值得到$\Phi(-x)$, 再求得

$$\Phi(x)=1-\Phi(-x).$$

(2) $\xi \sim N(a,\sigma^2)$时, $\eta = \frac{\xi-a}{\sigma} \sim N(0,1)$. 因此对于任意的$x<y$, 有

$$\begin{aligned}\mathsf{P}(x<\xi\le y) =& \mathsf{P}(\frac{x-a}{\sigma} \le \eta \le \frac{y-a}{\sigma}) \\ =& \Phi(\frac{y-a}{\sigma}) - \Phi(\frac{x-a}{\sigma}).\end{aligned}$$

3. 指数分布(exponential distribution)

定义（指数分布）

若随机变量ξ具有概率密度

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

其中$\lambda > 0$. 则称ξ服从参数为λ的指数分布. 记为$\xi \sim \text{Exp}(\lambda)$.

分布函数为

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

“无记忆性”：设随机变量ξ ~Exp(λ), 则对任意的$s > 0, t > 0$,

$$\mathsf{P}(\xi \geq s + t | \xi \geq s) = \mathsf{P}(\xi \geq t).$$

随机向量

在很多随机现象中, 对同一个随机试验我们往往需要同时考察几个随机变量.

定义 (n维随机向量)

若随机变量$\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega)$定义在同一概率空间$(\Omega, \mathscr{F}, P)$上, 就称

$$\boldsymbol{\xi}(\omega) = (\xi_1(\omega), \xi_2(\omega), \cdots, \xi_n(\omega))$$

为n维随机向量(n-dimensional random vector)或n维随机变量(n-dimensional random variable).

离散型随机向量:

定义 (联合分布列)

若二维随机向量(ξ, η)所有可能取值为$(x_i, y_j), i, j = 1, 2, \cdots$,且

$$P(\xi = x_i, \eta = y_j) = p_{ij}, \quad i, j = 1, 2, \cdots, \tag{2.2}$$

则称(2.2)为(ξ, η)的(联合)分布列.

(ξ, η)的联合分布列也可以用下列的表格表示:

$\xi \diagdown \eta$	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

联合分布列的性质:

- $p_{ij} \geq 0,\ \ i, j = 1, 2, \cdots;$
- $\displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1.$

由联合分布列可推得边际分布列:

$$\mathsf{P}(\xi = x_i) = \sum_{j=1}^{\infty} p_{ij} =: p_{i\cdot}, \quad i = 1, 2, \cdots .$$

$$\mathsf{P}(\eta = y_j) = \sum_{i=1}^{\infty} p_{ij} =: p_{\cdot j}, \quad j = 1, 2, \cdots .$$

定义 (联合分布函数)

对任意的$(x_1, \cdots, x_n) \in \mathbb{R}^n$, 称$n$元函数

$$F(x_1, \cdots, x_n) = \mathsf{P}(\xi_1(\omega) \le x_1, \cdots, \xi_n(\omega) \le x_n)$$

为随机向量$\boldsymbol{\xi}(\omega) = (\xi_1(\omega), \cdots, \xi_n(\omega))$的(联合)分布函数((joint) distribution function).

二维联合分布函数$F(x,y)$的性质:

- $F(x,y)$对每个变量单调不减;
- $F(x,y)$对每个变量右连续;
- 对任意的(x,y),

$$F(x,-\infty)=F(-\infty,y)=0,\ \ F(\infty,\infty)=1;$$

- 对任意的$a_1<b_1$, $a_2<b_2$,

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

联合分布函数与边际分布函数的关系:

$$
\begin{aligned}
F_\xi(x) &= \mathsf{P}(\xi \le x) = \mathsf{P}(\xi \le x, \eta < \infty) = F(x, \infty), \\
F_\eta(y) &= \mathsf{P}(\eta \le y) = \mathsf{P}(\xi < \infty, \eta \le y) = F(\infty, y).
\end{aligned}
$$

定义（连续型随机向量）

若存在n元可积的非负函数$p(x_1, \cdots, x_n)$, 使得n元分布函数可表示为

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(y_1, y_2, \cdots, y_n) \mathrm{d}y_1 \cdots \mathrm{d}y_n,$$

则称所对应的随机向量为n维连续型随机向量, 称$p(y_1, \cdots, y_n)$为相应的(联合)密度函数.

联合密度函数与边际密度函数的关系, 以二维为例:

设(ξ,η)的密度函数为$p(x,y)$. 则ξ的边际密度函数为

$$p_\xi(x)=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}y.$$

则η的边际密度函数为

$$p_\eta(y)=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}x.$$

两个重要的连续型随机向量:

1. n维均匀分布

定义 (n维均匀分布)

若n维向量$\boldsymbol{\xi}$具有密度函数

$$p(x_1,\cdots,x_n)=\begin{cases}1/S_G, & (x_1,\cdots,x_n)\in G,\\ 0, & \text{其它}.\end{cases}$$

其中, G是$\mathbb{R}^n$中的一个*Borel*集, S_G为G的测度(二维时, S_G表示G的面积; 三维时, S_G表示G的体积).

n维正态分布

一些记号: 设$\boldsymbol{\Sigma}=(\sigma_{ij})$为$n$阶正定对称矩阵, $|\boldsymbol{\Sigma}|$为其行列式, $\boldsymbol{\Sigma}^{-1}$为其逆. 又记

$$\boldsymbol{x}=(x_1,\cdots,x_n)',\ \ \boldsymbol{a}=(a_1,\cdots,a_n)'.$$

定义 (n维正态分布)

若n维向量$\boldsymbol{\xi}$具有概率密度函数

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{a})\Big\},\quad \boldsymbol{x}\in\mathbb{R}^n. \tag{2.3}$$

则称$\boldsymbol{\xi}$服从参数为$\boldsymbol{a}$和$\boldsymbol{\Sigma}$的n维正态分布.

二维正态分布:

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\cdot$$
$$\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-a)^2}{\sigma_1^2}-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right]\right\},$$

记为$(\xi,\eta)\sim N(a,b,\sigma_1^2,\sigma_2^2,r)$.

结论: 二维正态分布的边际分布仍然是正态分布, 且与r无关(若$(\xi,\eta)\sim N(a,b,\sigma_1^2,\sigma_2^2,r)$, 则$\xi\sim N(a,\sigma_1^2)$, $\eta\sim N(b,\sigma_2^2)$). 但反过来不一定正确, 即(ξ,η)的边际分布都是正态分布, 其联合分布未必是二维正态分布.

条件分布, 随机变量的独立性

离散情形: 设(ξ,η)是离散型随机向量. 分布列为

$$\mathsf{P}(\xi=x_i,\eta=y_j)=p_{ij},\quad i,j=1,2,\cdots.$$

若$\mathsf{P}(\xi=x_i)>0$, 则

$$\mathsf{P}(\eta=y_j|\xi=x_i)=\frac{\mathsf{P}(\xi=x_i,\eta=y_j)}{\mathsf{P}(\xi=x_i)}=\frac{p_{ij}}{p_{i\cdot}},\ j=1,2,\cdots.\quad (2.4)$$

我们称(2.4)为在$\xi=x_i$的条件下η的条件分布列.

同理, 若$\mathsf{P}(\eta=y_j)>0$, 那么在$\eta=y_j$的条件下ξ的条件分布列为

$$\mathsf{P}(\xi=x_i|\eta=y_j)=\frac{\mathsf{P}(\xi=x_i,\eta=y_j)}{\mathsf{P}(\eta=y_j)}=\frac{p_{ij}}{p_{\cdot j}},\ i=1,2,\cdots.$$

连续情形：若$p_\xi(x)>0$，我们称$\dfrac{p(x,y)}{p_\xi(x)}$为$\xi=x$时η的条件概率密度函数，记为$p_{\eta|\xi}(y|x)$．即

$$p_{\eta|\xi}(y|x)=\frac{p(x,y)}{p_\xi(x)}.$$

同理，若$p_\eta(y)>0$，定义$\eta=y$时ξ的条件概率密度函数为

$$p_{\xi|\eta}(x|y)=\frac{p(x,y)}{p_\eta(y)}.$$

定义 (随机变量的独立性)

设ξ与η为定义在同一概率空间上的随机变量. 若对一切$x, y \in \mathbb{R}$, 都有

$$P(\xi \leq x, \eta \leq y) = P(\xi \leq x)P(\eta \leq y),$$

即

$$F(x, y) = F_\xi(x)F_\eta(y), \ \forall x, y \in \mathbb{R}. \tag{2.5}$$

则称随机变量ξ与η相互独立(independent). 否则, 称随机变量ξ与η相依(dependent).

若(ξ,η)为二维离散型随机向量, 则(2.5)等价于

$$\mathsf{P}(\xi = x_i, \eta = y_j) = \mathsf{P}(\xi = x_i)\mathsf{P}(\eta = y_j), \ \ \forall i, j = 1, 2, \cdots .$$

即等价于

$$p_{ij} = p_{i\cdot}p_{\cdot j}, \ \ \forall i, j = 1, 2, \cdots . \tag{2.6}$$

若(ξ,η)为二维连续型随机向量, 则(2.5)等价于

$$p(x, y) = p_\xi(x)p_\eta(y) \ \ a.e. \tag{2.7}$$

结论: 设$(\xi,\eta) \sim N(a,b,\sigma_1^2,\sigma_2^2,r)$, 则$\xi,\eta$独立的充要条件为$r=0$.

随机变量的函数及其分布:

假设ξ是一个随机变量, $y = g(x)$是Borel函数, 那么$\eta = g(\xi)$是随机变量. η的概率分布是什么? η的概率分布与ξ的概率分布有何关系?

- 若ξ是离散型随机变量: 先求出$\eta = g(\xi)$的所有可能取值, 再求其对应的概率大小.
- 若ξ是连续型随机变量: 先求出$\eta = g(\xi)$的分布函数, 再求其密度函数.

定理

若$f(x)$严格单调, 其反函数$f^{-1}(y)$有连续导函数. 则$\eta = f(\xi)$也是连续型随机变量, 其密度函数为

$$g(y) = \begin{cases} p(f^{-1}(y)) \cdot \left|\left(f^{-1}(y)\right)'\right|, & y \in f(x)\text{的值域}, \\ 0, & \text{其它}. \end{cases}$$

推论

若$y=f(x)$在不相重叠的区间$I_1, I_2, \cdots$上逐段严格单调, 在各段的反函数$h_1(y), h_2(y), \cdots$ 有连续导函数, 则$\eta=f(\xi)$是连续型随机变量, 其密度函数为

$$g(y)=\begin{cases} \sum_i p(h_i(y))\cdot|h_i^{'}(y)|, & y\in \text{各}h_i(y)\text{的定义域}, \\ 0, & \text{其它}. \end{cases}$$

(连续型)随机向量的函数:

1. $\eta = \xi_1 + \xi_2$.

η的概率密度函数为

$$p_\eta(z) = \int_{-\infty}^{\infty} p(x, z-x)\mathrm{d}x$$
$$= \int_{-\infty}^{\infty} p(z-y, y)\mathrm{d}y.$$

若ξ_1与ξ_2相互独立, 则有卷积公式(convolution formula):

$$p_\eta(z) = \int_{-\infty}^{\infty} p_{\xi_1}(x)p_{\xi_2}(z-x)\mathrm{d}x$$
$$= \int_{-\infty}^{\infty} p_{\xi_1}(z-y)p_{\xi_2}(y)\mathrm{d}y.$$

2. $\eta = \xi_1/\xi_2$.

$$p_\eta(z) = \int_{-\infty}^{\infty} p(zy, y)|y|\mathrm{d}y.$$

3. 次序统计量(order statistic)的分布.

设$\xi_1, \xi_2, \cdots, \xi_n$独立同分布, 分布函数为$F(x)$.

$$\xi_1^* \leq \xi_2^* \leq \cdots \leq \xi_n^*.$$

$$\xi_1^* = \min\{\xi_1, \xi_2, \cdots, \xi_n\}, \quad \xi_n^* = \max\{\xi_1, \xi_2, \cdots, \xi_n\}.$$

(1) ξ_n^*的分布函数

$$\begin{aligned} \mathsf{P}(\xi_n^* \le x) =& \mathsf{P}(\xi_1 \le x, \xi_2 \le x, \cdots, \xi_n \le x) \\ =& \mathsf{P}(\xi_1 \le x)\mathsf{P}(\xi_2 \le x)\cdots\mathsf{P}(\xi_n \le x) \\ =& [F(x)]^n. \end{aligned}$$

(2) ξ_1^*的分布函数

$$\begin{aligned} \mathsf{P}(\xi_1^* > x) =& \mathsf{P}(\xi_1 > x, \xi_2 > x, \cdots, \xi_n > x) \\ =& \mathsf{P}(\xi_1 > x)\mathsf{P}(\xi_2 > x)\cdots\mathsf{P}(\xi_n > x) \\ =& [1 - F(x)]^n, \end{aligned}$$

所以

$$\mathsf{P}(\xi_1^* \le x) = 1 - [1 - F(x)]^n.$$

(3) (ξ_1^*, ξ_n^*)的分布函数

$$
\begin{aligned}
F(x,y) =& \mathsf{P}(\xi_1^* \le x, \xi_n^* \le y) \\
=& \mathsf{P}(\xi_n^* \le y) - \mathsf{P}(\xi_1^* > x, \xi_n^* \le y) \\
=& [F(y)]^n - \mathsf{P}(\bigcap_{i=1}^{n} \{x < \xi_i \le y\}) \\
=& \begin{cases} [F(y)]^n - [F(y) - F(x)]^n, & x < y, \\ [F(y)]^n, & x \ge y. \end{cases}
\end{aligned}
$$

四. (连续型)随机向量的变换

设$(\xi_1,\cdots,\xi_n)$的密度函数为$p(x_1,\cdots,x_n)$. 现有m个Borel函数:

$$\eta_1=f_1(\xi_1,\cdots,\xi_n),\cdots,\eta_m=f_m(\xi_1,\cdots,\xi_n),$$

则$(\eta_1,\cdots,\eta_m)$是随机向量, 其联合分布函数为

$$\begin{aligned}G(y_1,\cdots,y_m)=&\mathsf{P}(\eta_1\le y_1,\cdots,\eta_m\le y_m)\\=&\int\cdots\int_D p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n,\end{aligned}$$

这里, D是n维区域:

$$\{(x_1,\cdots,x_n):\ f_1(x_1,\cdots,x_n)\le y_1,\cdots,f_m(x_1,\cdots,x_n)\le y_m\}.$$

关于$(\eta_1, \cdots, \eta_n)$的联合密度函数, 我们有下列的结论:

定理

如果$m = n$, $\{f_j, j = 1, \cdots, n\}$有唯一的反函数组:

$$x_i = x_i(y_1, \cdots, y_n), i = 1, \cdots, n,$$

且

$$J = \frac{\partial(x_1, \cdots, x_n)}{\partial(y_1, \cdots, y_n)} \neq 0,$$

则$(\eta_1, \cdots, \eta_n)$是连续型随机向量. 当$(y_1, \cdots, y_n) \in (f_1, \cdots, f_n)$的值域时, 其密度函数为

$$q(y_1, \cdots, y_n) = p(x_1(y_1, \cdots, y_n), \cdots, x_n(y_1, \cdots, y_n)) \cdot |J|,$$

其它情况$q(y_1, \cdots, y_n) = 0$.

五. **数理统计中的几个重要分布**(了解, 不作考试内容)

χ^2分布, t分布, F分布.

例1. 设(ξ, η)服从矩形区域$D: \{0 < x < 1, 0 < y < 2\}$上的均匀分布.

(1)写出联合密度;

(2)求边际密度;

(3)求联合分布函数;

(4)求P$(\xi + \eta < 1)$.

例1. 设(ξ,η)服从矩形区域$D:\{0<x<1,0<y<2\}$上的均匀分布.
(1)写出联合密度;
(2)求边际密度;
(3)求联合分布函数;
(4)求P$(\xi+\eta<1)$.

例2. 对圆的直径D作近似测量, 设其值在$[a,b]$内均匀分布, 求圆面积S的密度函数.

数字特征与特征函数

一. 数学期望的定义

定义（一般随机变量的数学期望）

若随机变量ξ的分布函数为$F(x)$, 则$\int_{-\infty}^{\infty}|x|\mathrm{d}F(x)<\infty$, 则称

$$E\xi=\int_{-\infty}^{\infty}x\mathrm{d}F(x)$$

为ξ的数学期望. 若$\int_{-\infty}^{\infty}|x|\mathrm{d}F(x)=\infty$, 则称$\xi$的数学期望不存在.

离散情形: $\mathrm{E}\xi=\sum_k x_k p_k$.

连续情形: $\mathrm{E}\xi=\int_{-\infty}^{\infty}xp(x)\mathrm{d}x$.

常见的离散型随机变量的数学期望:

(1) $\xi \equiv a$, 则E$\xi = a$.

(2) $\xi \sim B(n,p)$, 则E$\xi = np$. (熟练掌握把一个“复杂”随机变量分解成“简单”随机变量之和的技巧)

(3) $\xi \sim P(\lambda)$, 则E$\xi = \lambda$.

(4) ξ服从几何分布, 分布列为

$$\mathrm{P}(\xi = k) = pq^{k-1}, \;\; k = 1, 2, \cdots, \; 0 < p < 1, \; q = 1 - p.$$

则E$\xi = 1/p$.

常见的连续型随机变量的数学期望:

(1) $\xi \sim U[a,b]$, 则$\mathsf{E}\xi = (a+b)/2$.

(2) $\xi \sim \mathrm{Exp}(\lambda)$, 则$\mathsf{E}\xi = \frac{1}{\lambda}$.

(3) $\xi \sim N(a,\sigma^2)$, 则$\mathsf{E}\xi = a$.

二. 随机变量函数的数学期望

定理

设ξ是随机变量, $f(x)$是一元*Borel*函数, 记$\eta = f(\xi)$, ξ和η的分布函数分别为$F_\xi(x)$和$F_\eta(x)$. 则

$$Ef(\xi) = \int_{-\infty}^{\infty} f(x)\mathrm{d}F_\xi(x). \tag{3.1}$$

推广:

定理

设$(\xi_1,\cdots,\xi_n)$的分布函数为$F(x_1,\cdots,x_n)$, 而$g(x_1,\cdots,x_n)$是n元Borel函数. 则

$$Eg(\xi_1,\cdots,\xi_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n).$$

特别地, 有

$$E\xi_i=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_i\mathrm{d}F(x_1,\cdots,x_n)=\int_{-\infty}^{\infty}x\mathrm{d}F_i(x),$$

其中$F_i(x)$是ξ_i的边际分布函数. 对于二元分布函数$F(x,y)$, 有

$$E(\xi\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy\mathrm{d}F(x,y),\ E\xi^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^2\mathrm{d}F(x,y).$$

三. 数学期望的基本性质

性质1. 若$a \le \xi \le b$, 则$\mathsf{E}\xi$存在且$a \le \mathsf{E}\xi \le b$. 特别地, 若$\xi = c$, 则$\mathsf{E}\xi = \mathsf{E}c = c$.

性质2. 若$\mathsf{E}\xi_1, \cdots, \mathsf{E}\xi_n$都存在, 则对任意常数$c_1, \cdots, c_n$及$b$, $\mathsf{E}\big(\sum_{i=1}^n c_i\xi_i + b\big)$存在, 且

$$\mathsf{E}\Big(\sum_{i=1}^n c_i\xi_i + b\Big) = \sum_{i=1}^n c_i\mathsf{E}\xi_i + b.$$

特别地, 我们有

$$\mathsf{E}\Big(\sum_{i=1}^n \xi_i\Big) = \sum_{i=1}^n \mathsf{E}\xi_i, \quad \mathsf{E}(c\xi) = c\mathsf{E}\xi.$$

性质3. 若$\xi_1, \cdots, \xi_n$相互独立, 各$\mathsf{E}\xi_i$ $(i = 1, \cdots, n)$存在, 则

$$\mathsf{E}(\xi_1 \cdots \xi_n) = \mathsf{E}\xi_1 \cdots \mathsf{E}\xi_n.$$

定理（Markov不等式)

设ξ是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量，$f(x)$是$[0, \infty)$上非负单调不减函数. 则对任意的$x > 0$,

$$P(|\xi| \geq x) \leq \frac{Ef(|\xi|)}{f(x)}.$$

定理（Cauchy-Schwarz不等式）

对任意的随机变量ξ与η都有

$$|E(\xi\eta)| \le \sqrt{E\xi^2 \cdot E\eta^2}. \tag{3.2}$$

等号成立当且仅当

$$P(\eta = t_0\xi) = 1, \tag{3.3}$$

这里t_0是某一个常数.

四. 条件数学期望(conditional expectation)

- 设在$\xi = x$的条件下, η有条件概率密度函数$p_{\eta|\xi}(y|x)$. 如果$\int_{-\infty}^{\infty} |y| p_{\eta|\xi}(y|x)\mathrm{d}y < \infty$. 则它的条件数学期望为

$$\mathsf{E}(\eta|\xi = x) = \int_{-\infty}^{\infty} y p_{\eta|\xi}(y|x)\mathrm{d}y. \tag{3.4}$$

- 设在$\xi = x$的条件下, η有条件分布列$\mathsf{P}(\eta = y_j|\xi = x_i)$, $j = 1, 2, \cdots$. 如果$\sum_{j=1}^{\infty} |y_j| \mathsf{P}(\eta = y_j|\xi = x_i) < \infty$. 则它的条件数学期望为

$$\mathsf{E}(\eta|\xi = x_i) = \sum_{j=1}^{\infty} y_j \mathsf{P}(\eta = y_j|\xi = x_i). \tag{3.5}$$

一个重要的结论:

$$\mathsf{E}[\mathsf{E}(\eta|\xi)] = \mathsf{E}\eta. \tag{3.6}$$

当ξ是离散型随机变量时，记$p_i = \mathsf{P}(\xi = x_i)$，则(3.6)成为

$$\mathsf{E}\eta = \sum_{i=1}^{\infty} p_i \mathsf{E}(\eta|\xi = x_i). \tag{3.7}$$

我们称上式为全数学期望公式(total expectation formula).

方差、协方差与相关系数

一. 方差

定义（方差）

若$E(\xi-E\xi)^2<\infty$, 就称它是随机变量ξ的方差(*variance*), 记作$Var\xi$(或$D\xi$), 即

$$Var\xi=E(\xi-E\xi)^2.$$

方差的计算公式:

$$
\begin{aligned}
\mathsf{Var}\xi &= \int_{-\infty}^{\infty}(x-\mathsf{E}\xi)^2 \mathrm{d}F_\xi(x) \\
&= \begin{cases} \sum_i (x_i-\mathsf{E}\xi)^2 \mathsf{P}(\xi=x_i) & (\text{离散型}), \\ \int_{-\infty}^{\infty}(x-\mathsf{E}\xi)^2 p_\xi(x)\mathrm{d}x & (\text{连续型}). \end{cases}
\end{aligned}
$$

另外,

$$\mathsf{Var}\xi = \mathsf{E}(\xi-\mathsf{E}\xi)^2 = \mathsf{E}\left[\xi^2 - 2\xi\mathsf{E}\xi + (\mathsf{E}\xi)^2\right] = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2.$$

一些重要的随机变量的方差:

(1) 设$\xi \sim 0-1(p)$分布, 则$\mathsf{Var}\xi = p(1-p)$.

(1) 设$\xi \sim B(n,p)$分布, 则$\mathsf{Var}\xi = np(1-p)$.

(3) 设$\xi \sim P(\lambda)$分布, 则$\mathsf{Var}\xi = \lambda$.

(4) 设ξ服从几何分布, 分布列为

$$\mathsf{P}(\xi = k) = pq^{k-1}, \quad k = 1, 2, \cdots, \ 0 < p < 1, \ q = 1 - p.$$

则$\mathsf{Var}\xi = \frac{1-p}{p^2}$.

(5) 设$\xi \sim U[a,b]$分布, 则$\mathsf{Var}\xi = (b-a)^2/12$.

(6) 设$\xi \sim \mathsf{Exp}(\lambda)$分布, 则$\mathsf{Var}\xi = 1/\lambda^2$.

(6) 设$\xi \sim N(a,\sigma^2)$分布, 则$\mathsf{Var}\xi = \sigma^2$.

定理（Chebyshev不等式）

设ξ为随机变量，则对任意给定的$\varepsilon > 0$，恒有

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{Var\xi}{\varepsilon^2}.$$

方差的性质:

性质1: $\mathsf{Var}\xi = 0$的充要条件是$\mathsf{P}(\xi = c) = 1$, 其中c是某常数.

性质2: 设b, c都是常数, 则$\mathsf{Var}(c\xi + b) = c^2\mathsf{Var}\xi$.

性质3: 若$c \neq \mathsf{E}\xi$, 则$\mathsf{Var}(\xi) < \mathsf{E}(\xi - c)^2$.

性质4:

$$\mathsf{Var}(\sum_{i=1}^{n} \xi_i) = \sum_{i=1}^{n} \mathsf{Var}(\xi_i) + 2 \sum_{1 \leq i < j \leq n} \mathsf{Cov}(\xi_i, \xi_j).$$

特别地, 若$\xi_1, \cdots, \xi_n$两两独立, 则

$$\mathsf{Var}(\sum_{i=1}^{n} \xi_i) = \sum_{i=1}^{n} \mathsf{Var}(\xi_i).$$

二. 协方差

定义（协方差）

记ξ_i和ξ_j的联合分布函数为$F_{ij}(x,y)$. 若

$$E|(\xi_i - E\xi_i)(\xi_j - E\xi_j)| < \infty,$$

就称

$$E(\xi_i - E\xi_i)(\xi_j - E\xi_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E\xi_i)(y - E\xi_j)\mathrm{d}F_{ij}(x,y)$$

为ξ_i和ξ_j的协方差(*covariance*), 记作$Cov(\xi_i, \xi_j)$.

协方差的性质:

性质1: $\mathsf{Cov}(\xi,\eta)=\mathsf{Cov}(\eta,\xi)=\mathsf{E}(\xi\eta)-\mathsf{E}\xi\cdot\mathsf{E}\eta$.

性质2: 设a,b是常数, 则$\mathsf{Cov}(a\xi,b\eta)=ab\mathsf{Cov}(\xi,\eta)$.

性质3: $\mathsf{Cov}(\sum_{i=1}^{n}\xi_i,\eta)=\sum_{i=1}^{n}\mathsf{Cov}(\xi_i,\eta)$.

性质4: 设

$$\boldsymbol{\xi}=(\xi_1,\cdots,\xi_n)',\quad \boldsymbol{C}=\begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mn} \end{pmatrix},$$

则$\boldsymbol{C\xi}$的协方差阵为$\boldsymbol{CBC}'$, 其中$\boldsymbol{B}$是$\boldsymbol{\xi}$的协方差阵.

三. 相关系数

定义（相关系数)

称

$$r_{\xi\eta} = Cov(\xi^*, \eta^*) = E(\xi^*\eta^*) = \frac{E(\xi - E\xi)(\eta - E\eta)}{\sqrt{Var\xi \cdot Var\eta}}$$

为ξ, η的相关系数(correlation coefficient).

相关系数的性质:

性质1: 对相关系数$r_{\xi\eta}$, 恒有$|r_{\xi\eta}| \leq 1$. $r_{\xi\eta} = 1$当且仅当

$$\mathsf{P}\left(\frac{\xi - \mathsf{E}\xi}{\sqrt{\mathsf{Var}\xi}} = \frac{\eta - \mathsf{E}\eta}{\sqrt{\mathsf{Var}\eta}}\right) = 1;$$

$r_{\xi\eta} = -1$当且仅当

$$\mathsf{P}\left(\frac{\xi - \mathsf{E}\xi}{\sqrt{\mathsf{Var}\xi}} = -\frac{\eta - \mathsf{E}\eta}{\sqrt{\mathsf{Var}\eta}}\right) = 1.$$

性质2: 对随机变量ξ与η, 下列事实等价:

(1) $\mathsf{Cov}(\xi,\eta)=0$; (2) ξ与η不相关;
(3) $\mathsf{E}(\xi\eta)=\mathsf{E}\xi\cdot\mathsf{E}\eta$; (4) $\mathsf{Var}(\xi+\eta)=\mathsf{Var}\xi+\mathsf{Var}\eta$.

性质3: 若ξ与η独立, 则ξ与η不相关.

性质4: 对于二元正态分布, 两个分量不相关与相互独立是等价的.

四. 矩(moment)

定义（原点矩)

对正整数k, 称

$$m_k = E\xi^k$$

为ξ的k阶(原点)矩(origin moment).

定义（中心矩)

对正整数k, 称

$$c_k = E(\xi - E\xi)^k$$

为ξ的k阶中心矩(center moment).

设$\xi \sim N(0,\sigma^2)$. 此时, $\mathsf{E}\xi = 0$且

$$m_n = c_n = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x^n \exp\Big\{-\frac{x^2}{2\sigma^2}\Big\}\mathrm{d}x.$$

显然, n为奇数时, $m_n = 0$; n为偶数时,

$$\begin{aligned}
m_n =& \sqrt{\frac{2}{\pi}}\int_0^{\infty}\frac{x^n}{\sigma}\exp\Big\{-\frac{x^2}{2\sigma^2}\Big\}\mathrm{d}x\\
=& \sqrt{\frac{2}{\pi}}\sigma^n 2^{\frac{n-1}{2}}\int_0^{\infty} z^{\frac{n-1}{2}}e^{-z}\mathrm{d}z\\
=& \sqrt{\frac{2}{\pi}}\sigma^n 2^{\frac{n-1}{2}}\Gamma\Big(\frac{n+1}{2}\Big)\\
=& 1\times 3\times\cdots\times(n-1)\sigma^n.
\end{aligned}$$

特征函数

一、定义

定义（特征函数）

设ξ为实随机变量，称

$$f(t) = Ee^{\mathrm{i}t\xi} = \int_{-\infty}^{\infty} e^{\mathrm{i}tx}\mathrm{d}F(x)$$

为ξ的特征函数(characteristic function)，这里t是任意实数.

离散情形：$f(t) = \mathrm{E}e^{\mathrm{i}t\xi} = \sum_{n=1}^{\infty} p_n e^{\mathrm{i}tx_n}$.

连续情形：$f(t) = \mathrm{E}e^{\mathrm{i}t\xi} = \int_{-\infty}^{\infty} e^{\mathrm{i}tx} p(x)\mathrm{d}x$.

一些重要分布的特征函数:

(1) 退化分布$\mathrm{P}(\xi=c)=1$的特征函数为$f(t)=e^{\mathrm{i}tc}$.

(2) 二项分布$B(n,p)$的特征函数为$f(t)=(pe^{\mathrm{i}t}+q)^n$.

(3) $0-1(p)$分布的特征函数为$f(t)=pe^{\mathrm{i}t}+q$.

(4) 泊松分布$P(\lambda)$的特征函数为$f(t)=e^{\lambda(e^{\mathrm{i}t}-1)}$.

(5) 指数分布$\mathrm{Exp}(\lambda)$的特征函数为$f(t)=(1-\frac{\mathrm{i}t}{\lambda})^{-1}$.

(6) 均匀分布$U[a,b]$的特征函数为$f(t)=\frac{e^{\mathrm{i}tb}-e^{\mathrm{i}ta}}{(b-a)\mathrm{i}t}$. 特别地, $U[-b,b]$的特征函数为$f(t)=\frac{\sin(bt)}{bt}$.

(7) 正态分布$N(a,\sigma^2)$的特征函数为$f(t)=e^{\mathrm{i}at-\sigma^2t^2/2}$.

二. 逆转公式与唯一性定理

定理（逆转公式）

设分布函数$F(x)$的特征函数为$f(t)$，又x_1, x_2是$F(x)$的两个连续点，则

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\mathrm{i}tx_1} - e^{-\mathrm{i}tx_2}}{\mathrm{i}t} f(t)\mathrm{d}t. \quad (3.8)$$

定理（唯一性定理）

分布函数可由特征函数唯一确定.

例1. 流水作业线上生产的每个产品为不合格品的概率为p, 当生产出k个不合格品时即检修一次. 求两次检修期间产品总数的数学期望.

例1. 流水作业线上生产的每个产品为不合格品的概率为p, 当生产出k个不合格品时即检修一次. 求两次检修期间产品总数的数学期望.

例2. 设随机变量$\xi_1, \xi_2, \cdots, \xi_{m+n}$ $(n > m)$相互独立, 有相同分布, 且方差存在, 求$S = \xi_1 + \cdots + \xi_n$与$T = \xi_{m+1} + \cdots + \xi_{m+n}$之间的相关系数.

极限定理

一. 分布函数弱收敛

定义（分布函数弱收敛）

设F是一分布函数，$\{F_n, n \geq 1\}$是一列分布函数，如果对F的每个连续点$x \in \mathbb{R}$，都有$F_n(x) \to F(x)$ $(n \to \infty)$，则称F_n弱收敛(weak convergence)于F，记作$F_n \xrightarrow{w} F$.

定义（依分布收敛）

设ξ为一随机变量，$\{\xi_n, n \geq 1\}$是一列随机变量，如果ξ_n的分布函数弱收敛于ξ的分布函数，则称ξ_n依分布收敛(convergence in distribution)于ξ，记作$\xi_n \xrightarrow{d} \xi$.

定理（Lévy连续性定理/正极限定理)

设F是一分布函数, $\{F_n, n \geq 1\}$是一列分布函数. 如果$F_n \xrightarrow{w} F$, 则相应的特征函数$\{f_n(t), n \geq 1\}$收敛于F的特征函数$f(t)$, 且在t的任一有限区间内收敛是一致的.

定理（逆极限定理)

设$\{f_n(t), n \geq 1\}$是分布函数$\{F_n(x), n \geq 1\}$的特征函数，如果对每一t，$f_n(t) \to f(t)$，且$f(t)$在$t=0$处连续，则$f(t)$一定是某个分布函数F的特征函数，且$F_n \xrightarrow{w} F$.

此定理常被用来证明随机变量序列依分布收敛于某个随机变量，或者依概率收敛于某个常数.

二. 性质

除连续性定理外, 分布函数弱收敛还有下列性质:

性质1: 设$\{F_n, n \geq 1\}$是一列分布函数, 如果$F_n \xrightarrow{w} F$, F是一连续的分布函数, 则F_n在$\mathbb{R}$上一致收敛于$F(x)$.

性质2: 设ξ是一随机变量, $\{\xi_n, n \geq 1\}$是一列随机变量, $g(x)$是$\mathbb{R}$上的连续函数. 如果$\xi_n \xrightarrow{d} \xi$, 则$g(\xi_n) \xrightarrow{d} g(\xi)$.

性质3: 设$\{a_n, n \geq 1\}$和$\{b_n, n \geq 1\}$是两列常数, F是一分布函数. $\{F_n, n \geq 1\}$是一列分布函数. 如果$a_n \to a$, $b_n \to b$, $F_n \xrightarrow{w} F$, 则

$$F_n(a_n x + b_n) \to F(ax + b),$$

其中x使得$ax + b$是F的连续点.

三. 中心极限定理(Central Limit Theorem/CLT)

定义（中心极限定理)

设$\{\xi_n, n\geq 1\}$是一列随机变量. 如果存在常数列$B_n>0$以及A_n, 使得

$$\frac{\sum_{k=1}^n \xi_k - A_n}{B_n} \xrightarrow{d} N(0,1),$$

则称$\{\xi_n, n\geq 1\}$满足中心极限定理(Central Limit Theorem).

定理（Lindeberg-Lévy CLT）

设$\{\xi_n, n \geq 1\}$是一列独立同分布的随机变量. 记

$$S_n = \sum_{k=1}^{n} \xi_k, \ E\xi_1 = a, \ \ Var\xi_1 = \sigma^2,$$

则CLT成立，即

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

依概率收敛与弱大数定律

一. 依概率收敛

定义（依概率收敛）

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量. 如果对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0,$$

或

$$\lim_{n\to\infty} P(|\xi_n - \xi| < \varepsilon) = 1,$$

则称$\{\xi_n, n \geq 1\}$依概率收敛(convergence in probability)于ξ, 记作$\xi_n \xrightarrow{P} \xi$.

关于依分布收敛和依概率收敛的关系, 我们有下列的定理.

定理

设ξ和$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量.

- 如果$\xi_n \xrightarrow{P} \xi$, 则$\xi_n \xrightarrow{d} \xi$.
- 如果$\xi_n \xrightarrow{d} \xi$, 则不一定有$\xi_n \xrightarrow{P} \xi$.
- 如果$\xi_n \xrightarrow{d} c$, c是常数, 则$\xi_n \xrightarrow{P} c$.

引理（Slutsky引理）

如果$\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} c$, 则$\xi_n + \eta_n \xrightarrow{d} \xi + c$.

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在同一概率空间上的随机变量序列. 则有

- 若$\xi_n \xrightarrow{P} \xi$, $\xi_n \xrightarrow{P} \eta$, 则$P(\xi = \eta) = 1$.
- 若$\xi_n \xrightarrow{P} \xi$, f是$(-\infty, \infty)$上的连续函数, 则$f(\xi_n) \xrightarrow{P} f(\xi)$.

定理

- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n \pm \eta_n \xrightarrow{P} \xi \pm \eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} \eta$, 则$\xi_n\eta_n \xrightarrow{P} \xi\eta$.
- 若$\xi_n \xrightarrow{P} \xi$, $\eta_n \xrightarrow{P} c$, c为一常数, 假设η_n与c都不为零, 则$\xi_n/\eta_n \xrightarrow{P} \xi/c$.

二. 弱大数定律

定义（弱大数定律/WLLN)

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列，如果存在常数列$\{a_n, n \geq 1\}$和$\{b_n, n \geq 1\}$使得

$$\frac{1}{a_n}\sum_{k=1}^{n}\xi_k - b_n \xrightarrow{P} 0.$$

则称$\{\xi_n, n \geq 1\}$服从弱大数定律(Weak Law of Large Numbers)，简称$\{\xi_n, n \geq 1\}$服从大数定律.

定理（Khinchine大数定律）

设$\{\xi_n, n \geq 1\}$是一列独立同分布随机变量序列, $E|\xi_1| < \infty$. 记$E\xi_1 = \mu$, $S_n = \sum_{k=1}^{n} \xi_k$, 则$\{\xi_n, n \geq 1\}$服从大数定律, 即

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

概率1收敛与强大数定律

一. 以概率1收敛

定义（以概率1收敛）

设ξ和$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列. 如果存在$\Omega_0 \in \mathscr{F}$, $P(\Omega_0) = 0$, 且对任意的$\omega \in \Omega \setminus \Omega_0$, 都有$\xi_n(\omega) \to \xi(\omega)$, 则称$\xi_n$以概率1收敛(converges with probability one)或几乎必然收敛(converges almost surely)于ξ, 记作$\xi_n \to \xi \ \ a.s.$

概率1收敛与强大数定律

一. 以概率1收敛

定义（以概率1收敛）

设ξ和$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列. 如果存在$\Omega_0 \in \mathscr{F}$, $P(\Omega_0) = 0$, 且对任意的$\omega \in \Omega \backslash \Omega_0$, 都有$\xi_n(\omega) \to \xi(\omega)$, 则称$\xi_n$以概率1收敛(converges with probability one)或几乎必然收敛(converges almost surely)于ξ, 记作$\xi_n \to \xi \ \ a.s.$

注: 以概率1收敛/a.s.收敛意味着最多除去一个零概率点集外, ξ_n逐点收敛于ξ.

以概率1收敛的判别准则:

以概率1收敛的判别准则:

定理

设ξ和$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列. 则$\xi_n \to \xi \ \ a.s.$当且仅当对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P(\cup_{k\geq n}\{|\xi_k - \xi| \geq \varepsilon\}) = 0,$$

或者等价地

$$\lim_{n\to\infty} P(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon) = 0.$$

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与*a.s.*收敛互不蕴含;

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与a.s.收敛互不蕴含;
- r-阶平均收敛与a.s.收敛都蕴含依概率收敛, 但依概率收敛不蕴含r-阶平均收敛与a.s.收敛;

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与*a.s.*收敛互不蕴含;
- r-阶平均收敛与*a.s.*收敛都蕴含依概率收敛, 但依概率收敛不蕴含r-阶平均收敛与*a.s.*收敛;
- 依概率收敛蕴含依分布收敛, 但依分布收敛不蕴含依概率收敛;

关于随机变量序列的依分布收敛, 依概率收敛, r-阶平均收敛, a.s.收敛, 我们有如下的结论:

定理

设ξ与$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$的随机变量序列.

- r-阶平均收敛与a.s.收敛互不蕴含;
- r-阶平均收敛与a.s.收敛都蕴含依概率收敛, 但依概率收敛不蕴含r-阶平均收敛与a.s.收敛;
- 依概率收敛蕴含依分布收敛, 但依分布收敛不蕴含依概率收敛;
- 对于退化的随机变量c, 有$\xi_n \xrightarrow{d} c \Leftrightarrow \xi_n \xrightarrow{P} c$.

二. 强大数定律

定义 (强大数定律/SLLN)

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的随机变量序列, 如果存在常数列$\{a_n, n \geq 1\}$和$\{b_n, n \geq 1\}$使得

$$\frac{1}{a_n}\sum_{k=1}^{n}\xi_k - b_n \to 0 \ \ a.s.,$$

则称$\{\xi_n, n \geq 1\}$服从强大数定律(Strong Law of Large Numbers).

第一个强大数定律是Borel在1909年针对Bernoulli试验场合给出的.

定理（Borel强大数定律）

设$\{\xi_n\}$是一列独立同分布的随机变量序列，$P(\xi_i = 1) = p$，$P(\xi_i = 0) = 1 - p$, $0 < p < 1$，记$S_n = \sum_{k=1}^{n} \xi_k$. 则

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

Kolmogorov在1930年将Borel强大数定律进行了推广，使之适用于一般随机变量.

定理（Kolmogorov强大数定律）

设$\{\xi_n, n \geq 1\}$是定义在概率空间$(\Omega, \mathscr{F}, P)$上的独立同分布随机变量序列. 记$S_n = \sum_{k=1}^{n} \xi_k$, 则

$$\frac{S_n}{n} \to \mu \ \ a.s.$$

成立的充要条件是$E|\xi_1| < \infty$且$\mu = E\xi_1$.

例1. 某计算机系统有120个终端.
(1)每个终端有5%时间在使用, 若各终端使用与否是相互独立的, 求有10个或更多终端在使用的概率;
(2)若每个终端有20%时间在使用, 求解上述问题.

例1. 某计算机系统有120个终端.
(1)每个终端有5%时间在使用, 若各终端使用与否是相互独立的, 求有10个或更多终端在使用的概率;
(2)若每个终端有20%时间在使用, 求解上述问题.

例2. 求证: 当$n \to \infty$时, $e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} \to \frac{1}{2}$.

例1. 某计算机系统有120个终端.
(1)每个终端有5%时间在使用, 若各终端使用与否是相互独立的, 求有10个或更多终端在使用的概率;
(2)若每个终端有20%时间在使用, 求解上述问题.

例2. 求证: 当$n \to \infty$时, $e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} \to \frac{1}{2}$.

例3. 设$\{\xi_n\}$独立同分布, 都服从$U[0,1]$分布, 令

$$\eta_n = \Big(\prod_{k=1}^{n}\xi_k\Big)^{1/n}.$$

求证: $\eta_n \xrightarrow{P} c$(常数), 并求出c.