

摘要

我自学使用的教材为李贤平的《概率论基础》，在此笔记中主要记录书中的核心内容，配以心得体会。

目录

1	公理化结构	2
1.1	事件域	2
1.2	概率	2
1.3	概率空间	3
2	条件概率与统计独立性	4
2.1	条件概率	4
2.2	事件独立性	4
2.3	伯努利试验	5
2.4	二项分布与泊松分布	5
3	随机变量与分布函数	6
3.1	随机变量及其分布	6
3.2	正态分布, 指数分布, Γ 分布	7
3.3	随机向量, 联合分布函数	8
3.4	边际分布, 条件分布	9
3.5	随机变量的独立性	11

1 公理化结构

1.1 事件域

定义 1.1.1 (样本空间). 对于(随机)试验,可能出现的结果称为**样本点** ω ,样本点全体构成**样本空间** Ω .

在概率论中一般假定样本空间是给定的,这是必要的抽象,是我们能更好地把我住随机变量的本质,类比于线性空间.

定义 1.1.2. **事件**定义为样本空间 Ω 的一个子集,称事件发生当且仅当它所包含的某一个样本点出现.

在此定义下,集合的包含关系诱导了事件的包含关系,补集对应于逆事件,或称对立事件,两个集合的交意味着两个事件的交意味着两个事件同时发生,并集意味着至少发生一个.

一般不把样本空间 Ω 的一切子集作为事件,这会带来困难,譬如在几何概率中把不可测集也作为事件将会带来不可克服的麻烦. 另一方面,又必须把感兴趣的事件都包括进来,所以要求事件全体 \mathcal{F} 组成一个 σ 代数.

定义 1.1.3 (事件域). 若 \mathcal{F} 是由样本空间 Ω 的一些子集构成的一个 σ 代数,则称它为**事件域**, \mathcal{F} 中的元素称为**事件**, Ω 称为必然事件, \emptyset 称为不可能事件.

需要特别注意的是由给定的 Ω 的一个非空集族 \mathcal{G} ,必定存在由 \mathcal{G} 生成的 σ 代数.这种方法可以定义Borel集.

1.2 概率

定义 1.2.1 (概率). 定义在事件域 \mathcal{F} 上的一个集合函数 P 称为**概率**,如果它满足如下三个要求:

- (i) $P(A) \geq 0, \forall A \in \mathcal{F}$
- (ii) $P(\Omega) = 1$
- (iii) $\forall A_i \in \mathcal{F}, i = 1, 2, \dots$, 若 A_i 两两互不相容,则

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

实际上实变函数中的一般测度在全空间的测度为1时就是概率.

有了可数可加性,我们便有了下连续性.实际上若 $A_i \in \mathcal{F}, i = 1, 2, \dots$ 且 A_i 两两互不相容,则有

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

两边取极限,右边由于是级数和,1显然是其上界,所以级数和存在,于是有:

$$\lim_{n \rightarrow \infty} P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^{\infty} P(A_i) \stackrel{\text{可数可加}}{=} P\left(\sum_{i=1}^{\infty} A_i\right)$$

即对于单调不减的集合列,其概率的极限等于极限集合的概率.考虑补集,可以知道概率也是上连续的.另外,有限可加且下连续与可数可加等价.

1.3 概率空间

定义 1.3.1 (概率空间). Ω 是样本空间, \mathcal{F} 是事件域, P 是概率,则称三元总体 (Ω, \mathcal{F}, P) 为概率空间.

最后这里放个最大似然估计法,因为书上第一章提到了,我觉得比较重要就记下来了,以后有更合适的地方再搬过去吧.

定义 1.3.2. 把概率 $p(n)$ 看作未知参数 n 的函数,称为似然函数,在通过求其最大值而得到 n 的估计,这就是数理统计中的最大似然估计法

2 条件概率与统计独立性

2.1 条件概率

定义 2.1.1 (条件概率). 设 (Ω, \mathcal{F}, P) 为一个概率空间, $B \in \mathcal{F}, P(B) > 0$,则对任意 $A \in \mathcal{F}$,记

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

并称 $P(A|B)$ 为在事件 B 发生的条件下事件 A 发生的条件概率.

概率论的重要课题之一就是通过简单事件的概率推算处复杂事件的概率,这里全概率公式起着重要作用.

定义 2.1.2 (全概率公式). 设事件 $A_1, A_2, \dots, A_n, \dots$ 是样本空间 Ω 的一个分割,亦称完备事件组,即 A_i 两两互不相容,而且 $\sum_{i=1}^{\infty} A_i = \Omega$,这样便有 $B = \sum_{i=1}^{\infty} A_i B$,由概率的可加性与条件概率定义可得

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

此公式称为**全概率公式**

定义 2.1.3 (Bayes公式). 若 B 总是与两两互不相容的事件 A_1, A_2, \dots 之一同时发生,即 $B = \sum_{i=1}^{\infty} B A_i$,结合条件概率的定义与全概率公式得

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

此公式称为**Bayes公式**

Bayes公式得应用场景十分广泛,比如医生为了诊断病人是患了 A_1, \dots, A_n 这几个疾病中的哪一种,可以先对病人进行检查确定指标 B ,此时利用指标 B ,可以计算相关概率.

Bayes公式中的 $P(A_i)$ 称为**先验概率**,反映了各种原因发生的可能性大小,实际应用中一般是以往经验的总结,试验前便已知道.条件概率 $P(A_i|B)$ 称为**后验概率**,它反映了试验发生后各种原因发生的可能性大小的条件概率.

2.2 事件独立性

定义 2.2.1 (事件独立性). 对 n 个事件 A_1, A_2, \dots, A_n ,若对于所有可能的组合 $1 \leq i < j < k < \dots \leq n$ 成立着

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

$$P(A_i A_j A_k) = P(A_i)P(A_j)P(A_k)$$

...

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

则称 A_1, A_2, \dots, A_n 相互独立

事件独立性的每个条件都是必要的,与集合不同,集合两两相交为空集便有任意的交为空集,但是这与事件独立性不同,下面是三个事件两两独立,但三个事件一起并不独立的例子.

考虑一个均匀的正四面体,第一面染成1色,第二面染成2色,第三面染成3色,第四面同时染上1,2,3三种颜色,现在记事件 A, B, C 分别为投一次四面体出现1,2,3色朝下的事件,因此 $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(AB) = P(BC) = P(AC) = \frac{1}{4}$,即此时事件 A, B, C 两两独立,但是 $P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$,即三个事件一起并不一致独立.

有了事件的独立性之后可以定义试验的独立性与重复独立试验,不在这里写了.

2.3 伯努利试验

定义 2.3.1 (伯努利试验). 把事件域 \mathcal{A} 取为 $\{\emptyset, A, \bar{A}, \Omega\}$, 并称出现 A 为成功, 出现 \bar{A} 为失败, 这种只有两个可能结果的试验称为伯努利试验. 考虑重复进行 n 次独立的伯努利试验, 这种试验称作 n 重伯努利试验

定义 2.3.2 (二项分布). 记伯努利试验中 $P(A) = p, P(\bar{A}) = q = 1 - p$, 记 n 重伯努利试验中事件 A 出现 k 次的概率为 $b(k; n, p)$, 那么有 $b(k; n, p) = \binom{n}{k} p^k q^{n-k}, k = 1, 2, \dots, n$, $b(k; n, p)$ 称为二项分布

二项分布的结果也可以容易地推广到 n 次重复独立试验且每次试验可能有若干有限个结果的情形, 此时称为多项分布.

2.4 二项分布与泊松分布

二项分布定义如上, 我们下面考虑二项分布的性质, 由于

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

于是 $k = [(n+1)p]$ 为最可能成功次数, 这也是使得 $b(k; n, p)$ 最大的项, 此时也称为中心项.

定理 2.4.1 (二项分布的泊松逼近). 在独立试验中, 以 p_n 代表事件 A 在试验中出现的概率, 它与试验总数 n 有关, 如果 $n \rightarrow \infty$ 时 $np_n \rightarrow \lambda$, 则有

$$b(k; n, p_n) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

其中 $p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$ 称为泊松分布

3 随机变量与分布函数

3.1 随机变量及其分布

定义 3.1.1 (随机变量). 设 $\xi(\omega)$ 是定义于概率空间 (Ω, \mathcal{F}, P) 上的单值实函数, 其中 ω 是样本点, 如果对于直线上任一Borel点集 B , 有

$$\{\omega : \xi(\omega) \in B\} \in \mathcal{F}$$

则称 $\xi(\omega)$ 为**随机变量**, 而 $P\{\omega : \xi(\omega) \in B\}$ 称为随机变量 $\xi(\omega)$ 的**概率分布**.

随机变量其实就是测度空间上的可测函数.

定义 3.1.2 (分布函数). 称

$$F(x) = P\{\omega : \xi(\omega) < x\}, \quad -\infty < x < \infty$$

为随机变量 $\xi(\omega)$ 的**分布函数**.

为了书写方便, 通常把“随机变量 $\xi(\omega)$ 服从分布函数 $F(x)$ ”简记作 $\xi(\omega) \sim F(x)$. 由分布函数的定义立刻得到

$$P\{a \leq \xi(\omega) < b\} = F(b) - F(a)$$

由此定义的分布函数具有**左连续性**, 即 $F(x-0) = F(x)$, 这里左连续而不一定右连续是因为分布函数的定义中不取等号导致的, 具体来说更本质的原因是 $[x_n, x)$ 在 $x_n \rightarrow x$ 时会趋近空集. 分布函数的定义还能推导出以下等式:

$$P\{\xi(\omega) = a\} = F(a+0) - F(a)$$

$$P\{\xi(\omega) \leq a\} = F(a+0)$$

这些公式基本都来源于**概率的下连续性**, 即**概率的极限等于极限集合的概率**

定义 3.1.3 (离散型随机变量). 设 $\{x_i\}$ 为离散型随机变量的所有可能值, 而 $p(x_i)$ 是 ξ 取 x_i 的概率, 即

$$P\{\xi = x_i\} = p(x_i), \quad i = 1, 2, \dots$$

那么称 $\{p(x_i), i = 1, 2, \dots\}$ 为随机变量 ξ 的**概率分布**.

由此我们可以求出分布函数

$$F(x) = P\{\xi(\omega) < x\} = \sum_{x_k < x} p(x_k)$$

此时分布函数为一个跳跃的阶梯函数. 另外, 常用**分布列表**出离散型随机变量的概率分布.

定义 3.1.4 (连续型随机变量). 连续型随机变量 ξ 可取某个区间 $[c, d]$ 或者 $(-\infty, \infty)$ 中的一切值, 而且其分布函数 $F(x)$ 是绝对连续函数, 即存在可积函数 $p(x)$, 使得

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

此时称 $p(x)$ 为 ξ 的**密度函数**.

这里的绝对连续是实变函数中的概念, $p(x)$ 在零测集上作改动也不影响分布函数, 所以对于密度函数 $p(x)$ 的论断通常都是在“几乎处处”的意义上成立. 这些都是实变函数中老生常谈的内容.

3.2 正态分布, 指数分布, Γ 分布

定义 3.2.1 (正态分布). 密度函数为

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

其中 $\sigma > 0$, μ 与 σ 均为常数, 相应的分布函数为

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

此分布称为**正态分布**, 简记为 $N(\mu, \sigma^2)$.

为了验证如上定义的 $p(x)$ 确实是密度函数, 非负是显然的, 于是只需验证其在 \mathbb{R} 上的积分为1, 这是经典数分题目, 考虑原积分的平方, 将其化成一个累次积分的形式, 进而化为二元重积分, 再利用极坐标换元便可直接积出来.

特别地, 当 $\mu = 0, \sigma = 1$, 这时分布称为**标准正态分布**, 记为 $N(0, 1)$, 相应的密度函数与分布函数分别记为 $\varphi(x)$ 与 $\Phi(x)$.

可以验证, 若随机变量 ξ 服从正态分布 $N(\mu, \sigma^2)$, 简记作 $\xi \sim N(\mu, \sigma^2)$, 则随机变量 $\zeta = \frac{\xi - \mu}{\sigma}$ 服从 $N(0, 1)$.

有了以上的关系式, 一般的正态分布可以化为标准正态分布来处理:
若 $\xi \sim N(\mu, \sigma^2)$, 则

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

$$P\{a \leq \xi < b\} = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P\{|\xi - \mu| < k\sigma\} = P\left\{-k < \frac{\xi - \mu}{\sigma} < k\right\} = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

定义 3.2.2 (指数分布). 分布密度函数为

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

分布函数为

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

这里 $\lambda > 0$ 是参数, 这分布称为**指数分布**, 简记为 $\text{Exp}(\lambda)$.

定义 3.2.3 (Γ 分布). 称密度函数为

$$f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

的分布为 **Γ 分布**, 其中 $\lambda > 0, r > 0$ 为参数. 简记作 $\Gamma(r, \lambda)$.

3.3 随机向量, 联合分布函数

定义 3.3.1 (随机向量). 若随机变量 $\xi_1(\omega), \dots, \xi_n(\omega)$ 定义在同一概率空间 (Ω, \mathcal{F}, P) 上, 则称

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

构成一个 **n 维随机向量**, 亦称 **n 维随机变量**.

对于任意 n 个实数 x_1, \dots, x_n ,

$$\{\omega : \xi_1(\omega) < x_1, \dots, \xi_n(\omega) < x_n\} = \bigcap_{i=1}^n \{\xi_i(\omega) < x_i\} \in \mathcal{F}$$

即对于 \mathbb{R}^n 中的 n 维矩形 $C_n = \prod_{i=1}^n (-\infty, x_i)$, 有 $\{\boldsymbol{\xi}(\omega) \in C_n\} \in \mathcal{F}$ 于是随机向量也可良好定义(联合)分布函数:

定义 3.3.2 (联合分布函数). 称 **n 元函数**

$$F(x_1, \dots, x_n) = P\{\omega : \xi_1(\omega) < x_1, \dots, \xi_n(\omega) < x_n\}$$

为随机向量 $\boldsymbol{\xi}(\omega) = (\xi_1(\omega), \dots, \xi_n(\omega))$ 的**(联合)分布函数**.

多元分布函数的一些性质:

(i) 单调性: 关于每个变元是单调不减函数.

$$(ii) \quad F(x_1, \dots, -\infty, \dots, x_n) = 0 \\ F(+\infty, \dots, +\infty) = 1$$

(iii) 关于每个变元左连续.

特别地,在二元场合,还有: (类似的结论可以推广到 n 元)

(iv) 对任意 $a_1 < b_1, a_2 < b_2$,都有

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) = P\{a_1 \leq \xi_1 < b_1, a_2 \leq \xi_2 < b_2\} \geq 0$$

随机向量也有不同类型,最常见的也是离散型与连续型两类.

离散型场合中,概率分布集中在有限或可列个点上,其中比较重要的有多项分布与多元超几何分布

在连续型场合,存在非负函数 $p(x_1, \dots, x_n)$,使得

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(t_1, \dots, t_n) dt_1 \cdots dt_n$$

这里的 $p(x_1, \dots, x_n)$ 称为(多元分布)密度函数

定义 3.3.3 (多元正态分布). 若 $\Sigma = (\sigma_{ij})$ 是 n 阶正定对称矩阵,以 $\Sigma^{-1} = (\gamma_{ij})$ 表示 Σ 的逆矩阵, $\det \Sigma$ 表示 Σ 的行列式的值. $\mu = (\mu_1, \dots, \mu_n)$ 是任意实值行向量,则由密度函数

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^{\top}}$$

定义的分布称为 n 元正态分布,简记为 $N(\mu, \Sigma)$.

3.4 边际分布, 条件分布

下面的讨论将对二维场合进行, n 维时这些结论仍然成立,设随机向量为 (ξ, η) .

定义 3.4.1 (离散型随机变量的边际分布). 对于离散型分布,设 ξ 取值 x_1, x_2, \dots , η 取值 y_1, y_2, \dots ,显然有

$$p_1(x_i) := P\{\xi = x_i\} = \sum_j p(x_i, y_j) \\ p_2(y_j) := P\{\eta = y_j\} = \sum_i p(x_i, y_j)$$

这里 $p_1(x_i)$ 与 $p_2(y_j)$ 称为 $p(x_i, y_j)$ 的**边际分布**或**边缘分布**

定义 3.4.2 (连续型随机变量的边际分布). 若 (ξ, η) 是二维随机变量, 其分布函数为 $F(x, y)$, 我们能由 $F(x, y)$ 得出 ξ 和 η 的分布函数. 事实上,

$$F_1(x) = P\{\xi < x\} = P\{\xi < x, \eta < +\infty\} = F(x, +\infty)$$

同理

$$F_2(x) = P\{\eta < y\} = F(+\infty, y)$$

$F_1(x)$ 及 $F_2(x)$ 称为 $F(x, y)$ 的**边际分布函数**.

若 $F(x, y)$ 是连续型分布函数, 有密度函数 $p(x, y)$ 那么

$$F_1(x) = \int_{-\infty}^x \int_{-\infty}^{+\infty} p(u, y) dy du$$

因此 $F_1(x)$ 是连续型分布函数, 其密度函数为

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

$p_1(x)$ 称为**边际(分布)密度函数**. $F_2(x), p_2(x)$ 同理.

定义 3.4.3 (离散型随机变量的条件分布). 对于离散型随机变量, 若已知 $\xi = x_i (p_1(x_i) > 0)$, 则事件 $\{\eta = y_j\}$ 的条件概率为

$$P\{\eta = y_j | \xi = x_i\} = \frac{P\{\xi = x_i, \eta = y_j\}}{P\{\xi = x_i\}} = \frac{p(x_i, y_j)}{p_1(x_i)}$$

此式定义了随机变量 η 关于随机变量 ξ 的**条件分布**.

定义 3.4.4 (连续型随机变量的条件分布). 对于连续型随机变量, $p(x, y), p_1(x)$ 为其密度函数与边际密度函数, 那么其**条件分布函数**定义为

$$P\{\eta < y | \xi = x\} = \frac{\int_{-\infty}^y p(x, v) dv}{p_1(x)} = \int_{-\infty}^y \frac{p(x, v)}{p_1(x)} dv$$

其分布密度函数为

$$p(y|x) = \frac{p(x, y)}{p_1(x)}$$

定义 3.4.5 (二元正态分布). 函数

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

称为**二元正态(分布)密度函数**, 其中 $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$, 简记为 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

定理 3.4.1 (二元正态分布的典型分解). 其实也不是定理, 就是一个重要结论. 二元正态密度函数具有如下分解式:

$$p(x, y) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \times \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))]^2}{2\sigma_2^2(1-\rho^2)}}$$

对于此分解, 第一部分为 $N(\mu_1, \sigma_1^2)$ 的密度函数, 第二部分为 $N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$ 的密度函数

并且二元正态分布的边际分布为正态分布:

$$p_1(x) = \int_{-\infty}^{\infty} p(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

二元正态分布的条件分布仍然是正态分布:

$$p(y|x) = \frac{p(x, y)}{p_1(x)} = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))]^2}{2\sigma_2^2(1-\rho^2)}}$$

因此, 此典型分解式的涵义就完全清楚了, 第一部分是边际密度 $p_1(x)$, 第二部分是条件密度 $p(y|x)$, 整个式子形如 $p(x, y) = p_1(x)p(y|x)$

3.5 随机变量的独立性