

概率统计笔记

臧亦驰

2021年10月11日

目录

1	概率	3
1.1	随机事件概率	3
1.1.1	排列组合	3
1.1.2	概率模型	3
1.1.3	条件概率	3
1.2	一维随机变量	4
1.2.1	分布函数	4
1.2.2	概率密度	4
1.2.3	常见离散随机变量	4
1.2.4	常见连续随机变量	4
1.3	多维随机变量	5
1.3.1	分布函数	5
1.3.2	概率密度	5
1.3.3	常见多维随机变量	6
1.3.4	条件分布	6
1.3.5	随机变量的独立性	6
1.3.6	随机变量函数的分布	6
1.4	随机变量数字特征	7
1.4.1	数学期望	7
1.4.2	方差	7
1.4.3	协方差	7
1.4.4	相关系数	7
1.5	极限定理	8
1.5.1	切比雪夫不等式	8
1.5.2	大数定律	8
1.5.3	中心极限定理	8
2	统计	9
2.1	抽样分布理论	9
2.1.1	样本	9
2.1.2	统计量	9
2.1.3	经验分布函数	9
2.1.4	抽样分布	10
2.1.5	常见抽样分布	10
2.1.6	样本均值和样本方差的分布	11
2.2	参数估计	11
2.2.1	参数估计	11
2.2.2	参数的点估计	11

1 概率

1.1 随机事件概率

1.1.1 排列组合

定义式:

$$A_n^m = \frac{n!}{(n-m)!}, \quad C_n^m = \frac{n!}{m!(n-m)!}$$

常用排列组合公式:

$$\begin{cases} C_n^m = C_n^{n-m} \\ C_n^m + C_n^{m-1} = C_{n+1}^m \\ \sum_{k=0}^m C_{n_1}^k C_{n_2}^{m-k} = C_{n_1+n_2}^m \\ \sum_{k=0}^n C_n^k = 2^n \end{cases}$$

1.1.2 概率模型

古典概型: 满足样本空间只有有限个样本点, 且每个样本点出现的可能性相等的随机试验。

超几何概型: N 个物体, 分成 s 类, 每一类分别有 N_i 个物体, 从中不放回地抽取 n 个, 恰好抽到 n_i 个第 i 类物体的概率为

$$P(A) = \frac{\prod_{i=1}^s C_{N_i}^{n_i}}{C_N^n}$$

几何概型: 样本空间是一个可以度量的几何区域, 向上随意抛掷一点, 该点落在区域 A 中的可能性只与 A 的度量 $\mu(A)$ 成比例而与位置和形状无关的随机试验。

1.1.3 条件概率

事件 B 发生的前提下事件 A 发生的概率称作**条件概率**, 其定义式为

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

显然从中可以推出

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

这称为**贝叶斯公式**或**后验概率公式**。

对于一组两两互斥且包含所有可能事件的事件组 A_i , 有

$$P(B) = \sum_i P(A_i)P(B|A_i)$$

这称为**全概率公式**。

1.2 一维随机变量

1.2.1 分布函数

设 X 为随机变量, x 是任意实数, 则 X 的分布函数为

$$F(x) = P\{X \leq x\}$$

1.2.2 概率密度

对于连续型随机变量 X , 它的概率密度函数是这样的 $f(x)$, 使得

$$F(x) = \int_{-\infty}^x f(u)du$$

1.2.3 常见离散随机变量

两点分布 (伯努利分布): 记作 $X \sim B(1, p)$

$$P\{X = 0\} = p, \quad P\{X = 1\} = 1 - p$$

二项分布: 独立重复 n 次伯努利试验, 得到指定结果的次数满足的概率分布, 记作 $X \sim B(n, p)$

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

泊松分布: 记作 $X \sim P(\lambda)$

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

几何分布: 独立重复伯努利试验, 第一次得到指定结果的次数满足的概率分布, 记作 $X \sim Ge(p)$

$$P\{X = k\} = p(1-p)^{k-1}, \quad k = 1, 2, \dots$$

负二项分布 (帕斯卡分布): 独立重复伯努利试验, 得到 n 次指定结果时的试验次数满足的概率分布, 记作 $X \sim Nb(r, p)$

$$P\{X = k\} = C_{k-1}^{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

超几何分布: N 个物品不放回取件 n 次, 得到指定种类物品 (共 M 件) 的次数满足的概率分布, 记作 $X \sim h(n, N, M)$

$$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = 1, 2, \dots, \min\{M, n\}$$

1.2.4 常见连续随机变量

均匀分布: 记作 $X \sim U[a, b]$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{others} \end{cases}$$

指数分布: 记作 $X \sim E(\lambda)$

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

正态分布: 由大量微小的独立随机因素叠加而成的现象通常近似符合正态分布, 记作 $X \sim N(\mu, \sigma^2)$

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

伽马分布: 记作 $X \sim Ga(\alpha, \lambda)$

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

其中 $\Gamma(\alpha)$ 为伽马函数, 定义为

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

贝塔分布: 记作 $X \sim Be(a, b)$

$$f(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{others} \end{cases}$$

其中 $B(a, b)$ 为贝塔函数, 有如下定义和性质

$$\begin{aligned} B(a, b) &= B(b, a) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \\ &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \end{aligned}$$

1.3 多维随机变量

1.3.1 分布函数

设 (X_1, X_2, \dots, X_n) 是多维随机变量, x_1, x_2, \dots, x_n 是任意实数, 则随机变量的联合分布函数 (简称分布函数) 为

$$F(x_1, x_2, \dots, x_n) = P\{\forall k, X_k \leq x_k\}$$

关于 X_k 的边缘分布函数为

$$F_{X_k}(x) = P\{X_k \leq x\}$$

1.3.2 概率密度

对于多维连续型随机变量 (X_1, X_2, \dots, X_n) , 它的联合概率密度是这样的函数 $f(x_1, x_2, \dots, x_n)$, 使得

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty < y_k < x_k} f(y_1, y_2, \dots, y_n) d^n V$$

X_k 的边缘密度函数 $f_{X_k}(x)$, 使得

$$F_{X_k}(x) = \int_{-\infty}^x f_{X_k}(u) du$$

1.3.3 常见多维随机变量

多项分布：二项分布的推广， r 维随机变量，记作 $M(n, p_1, p_2, \dots, p_r)$

$$P\{\forall k, X_k = x_k\} = \frac{n!}{\prod_{i=1}^r (x_i!)} \prod_{j=1}^r p_j^{x_j}$$

多维均匀分布：记作 $U(D)$

$$f(x_1, x_2, \dots, x_n) = \begin{cases} \frac{1}{\mu(D)}, & (x_1, x_2, \dots, x_n) \in D \\ 0, & \text{others} \end{cases}$$

多维正态分布：记作 $N(\mu, \Sigma)$ ，其中 $\Sigma = (\sigma_{ij})$ ， $\mu = (\mu_1, \mu_2, \dots, \mu_n)^T$ ，并且记 $x = (x_1, x_2, \dots, x_n)^T$

$$f(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right\}$$

1.3.4 条件分布

定义条件概率为

$$P\{X_i = x_i | X_j = x_j\} = \frac{P\{X_i = x_i, X_j = x_j\}}{P\{X_j = x_j\}}$$

或

$$P\{X_i \leq x_i | x_j - \epsilon < X_j \leq x_j + \epsilon\} = \frac{P\{X_i \leq x_i, x_j - \epsilon < X_j \leq x_j + \epsilon\}}{P\{x_j - \epsilon < X_j \leq x_j + \epsilon\}}$$

如果 $\epsilon \rightarrow 0$ 时上式极限存在，则将其记为 $F_{X_i|X_j}(x_i|x_j)$ 。条件密度函数是这样的函数 $f_{X_i|X_j}(x_i|x_j)$ ，使得

$$F_{X_i|X_j}(x_i|x_j) = \int_{-\infty}^{x_i} f_{X_i|X_j}(u|x_j) du$$

1.3.5 随机变量的独立性

若 $\forall x_i$ ，都有

$$F(x_1, x_2, \dots) = \prod_i F_{X_i}(x_i)$$

则称这些随机变量互相独立。亦可以等价地写为

$$f(x_1, x_2, \dots) = \prod_i f_{X_i}(x_i)$$

1.3.6 随机变量函数的分布

对于二维情况，假设随机变量 X, Y 独立，随机变量函数 $Z = X + Y$ 在离散情况下的分布律为

$$P\{Z = z\} = \sum_{k=0}^z p(k)q(z-k)$$

这是一维离散卷积公式。在连续情况下概率密度为

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-y)dx = \int_{-\infty}^{\infty} f_X(z-x)f_Y(y)dy$$

这是一维连续卷积公式。

随机变量函数 $Z = \max\{X, Y\}$ 和 $Z = \min\{X, Y\}$ 的分布函数分别为

$$F_{max}(z) = F_X(z)F_Y(z)$$

$$F_{min}(z) = 1 - [1 - F_X(z)][1 - F_Y(z)]$$

1.4 随机变量数字特征

1.4.1 数学期望

若对于离散型随机变量 X ，级数 $\sum_i x_i P(X = x_i)$ 绝对收敛，则称其为 X 的数学期望，即

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

若对于连续型随机变量 X ，积分 $\int_{-\infty}^{+\infty} xf(x)dx$ 绝对收敛，则称其为 X 的数学期望，即

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

数学期望对不同随机变量是线性的。

1.4.2 方差

对于一个随机变量 X ，定义它的方差如下

$$D(X) = E[(X - E(X))^2]$$

可以将其写成

$$D(X) = E(X^2) - E^2(X)$$

1.4.3 协方差

两个随机变量的和/差的方差为

$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

其中 $Cov(X, Y)$ 称为协方差，定义为

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

可以将其写成

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

随机变量独立是协方差为零的充分不必要条件。

1.4.4 相关系数

对于两个方差大于零的随机变量，定义

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为 X 和 Y 的相关系数。相关系数的绝对值反映了两个随机变量之间存在线性关系的概率。

1.5 极限定理

1.5.1 切比雪夫不等式

有随机变量 X ，则对于任意正数 ϵ ，有

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

1.5.2 大数定律

若对于随机变量序列 $\{X_n\}$ ，存在一个常数 μ ，使得对于任意正数 ϵ ，都有

$$\lim_{n \rightarrow \infty} P\{|X_n - \mu| < \epsilon\} = 1$$

则称该随机变量序列依概率收敛于 μ ，记为 $X_n \xrightarrow{P} \mu$ 。

设随机变量序列 $\{X_n\}$ 互相独立，且 $\{D(X_i)\}$ 存在上确界，则对于任意正数 ϵ ，有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \epsilon\right\} = 1$$

这被称为切比雪夫大数定律，可以通过对随机变量序列的平均数应用切比雪夫不等式得到。

设 μ_n 是 n 次伯努利试验中事件 A 发生的次数， p 是事件 A 发生的概率，则对于任意正数 ϵ ，有

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

这被称为伯努利大数定律，它为频率的稳定值作为概率提供了理论依据。

对于服从相同分布且相互独立的随机变量序列 $\{X_i\}$ ，其数学期望为 μ ，则对于任意正数 ϵ ，有

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$$

这被称为辛钦大数定律。

1.5.3 中心极限定理

对于服从相同分布且相互独立的随机变量序列 $\{X_i\}$ ，其数学期望为 μ ，方差为 σ^2 ，则对于任意实数 x ，有

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \Phi(x)$$

这称为林德伯格-列维中心极限定理，其中 $\Phi(x)$ 是标准正态分布的分布函数。这表明，若 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ，则当 n 充分大时

$$\bar{X} \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

对于相互独立的随机变量序列 $\{X_i\}$ ，有 $E(X_i) = \mu_i$ ， $D(X_i) = \sigma_i^2$ ，记 $B_n = \sqrt{\sum_{i=1}^n \sigma_i^2}$ ，如果存在正数 δ ，使得 $n \rightarrow \infty$ 时

$$\frac{1}{B_n^{2+\delta}} \sum_{i=1}^n E(|X_i - \mu_i|^{2+\delta}) \rightarrow 0$$

则对于任意 x , 有

$$\lim_{n \rightarrow \infty} P\left\{\frac{1}{B_n} \sum_{i=1}^n (X_i - \mu_i) \leq x\right\} = \Phi(x)$$

这被称为李雅普诺夫中心极限定理。这个定理说明无论各个随机变量服从什么分布, 只要符合定理条件, 那么 n 足够大时, 他们的和近似地服从正态分布, 即

$$\sum_{i=1}^n X_i \overset{\text{approx}}{\sim} N\left(\sum_{i=1}^n \mu_i, B_n^2\right)$$

2 统计

2.1 抽样分布理论

2.1.1 样本

从总体中抽取 n 个个体 X_1, X_2, \dots, X_n 称为总体的一个**样本**, 样本是随机变量, 但一旦具体抽取后, 就能得到一组确定的数值, 用 x_i 表示, 成为**样本观测值**。一般抽样时采取简单随机抽样方法, 特点是样本的每一个分量 X_i 与总体 X 具有相同的概率分布, 且样本之间互相独立。采用简单随机抽样方法得到的样本称为**简单随机样本**。

2.1.2 统计量

若关于样本的一个函数其中不含任何未知参数, 则称这个函数是一个**统计量**。以下是一些常见统计量

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

称为**样本均值**。

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

称为**样本方差**。

$$R = \max_{1 \leq i \leq n} \{X_i\} - \min_{1 \leq i \leq n} \{X_i\}$$

称为**样本极差**。

$$C_v = \frac{S}{|\bar{X}|}$$

称为**样本变异系数**。

2.1.3 经验分布函数

通过样本观测值可以估计总体的分布函数, 将样本观测值按从小到大的顺序排列, 则函数

$$F_n(x) = \begin{cases} 0, & x < x_1 \\ \frac{k}{n}, & x_k \leq x < x_{k+1} \\ 1, & x \geq x_n \end{cases}$$

称为样本的经验分布函数。有

$$P\{\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0\} = 1$$

2.1.4 抽样分布

统计量的分布称为抽样分布。

对于总体和给定正数 α ，若存在实数 x_α ，满足

$$P\{X \geq x_\alpha\} = \alpha$$

则称 x_α 为 X 的上侧 α 分位数。

2.1.5 常见抽样分布

以下抽样分布是以标准正态分布为基石而构造的。

χ^2 分布：来自标准正态分布的 n 个样本，它们平方和 Y 所服从的分布为自由度为 n 的 χ^2 分布，记为 $Y \sim \chi^2(n)$ ，密度函数为

$$f(y) = \begin{cases} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

可以发现这是一种特殊的伽马分布， $\chi^2(n) = Ga(\frac{n}{2}, \frac{1}{2})$ 。 χ^2 分布的期望值和方差分别为 $E(Y) = n, D(Y) = 2n$ 。

t 分布：设随机变量 $X \sim N(0, 1), Y \sim \chi^2(n)$ ，则随机变量

$$t = \frac{X}{\sqrt{Y/n}}$$

服从的分布为自由度为 n 的 t 分布，记为 $t \sim t(n)$ ，密度函数为

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + \frac{t^2}{n})^{-\frac{n+1}{2}}, \quad -\infty < t < \infty$$

n 大于一时期望值为 $E(t) = 0$ ， $n = 1$ 时期望值不存在， n 大于二时方差为 $D(t) = \frac{n}{n-2}$ ， $n = 1, 2$ 时方差不存在。

F 分布：设随机变量 $Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$ ，则随机变量

$$F = \frac{Y_1/n_1}{Y_2/n_2}$$

服从的分布为第一自由度为 n_1 第二自由度为 n_2 的 F 分布，记作 $F \sim F(n_1, n_2)$ ，密度函数为

$$f(x) = \begin{cases} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} (\frac{n_1}{n_2})^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1} (1 + \frac{n_1}{n_2}x)^{-\frac{n_1+n_2}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

2.1.6 样本均值和样本方差的分布

对于来自均值为 μ ，方差为 σ^2 的某总体的一组样本，当 n 充分大时，近似有

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

对于来自正态分布的样本，则有

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) \\ \frac{(n-1)S^2}{\sigma^2} &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1) \\ \frac{\bar{X} - \mu}{S/\sqrt{n}} &\sim t(n-1)\end{aligned}$$

且样本均值与样本方差独立。

对于分别来自正态分布 $N(\mu_1, \sigma^2)$ 和 $N(\mu_2, \sigma^2)$ 且相互独立的样本 X_i, Y_j ，有

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

其中

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

又有

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

2.2 参数估计

2.2.1 参数估计

假设总体的分布函数中有一个未知参数 θ ，通过样本构造统计量 $\hat{\theta}(X_i)$ 来估计参数 θ ，称为参数的估计量，将样本观测值带入得到一个具体的值称为估计值。

构造一个统计量作为未知参数的估计量，就称其为参数的点估计，构造两个统计量而用他们的区间作为未知参数可能取值范围的一种估计，称为这个参数的区间估计，也称为置信区间。

2.2.2 参数的点估计

样本均值和样本方差分别可以用来估计总体均值和总体方差，即

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

矩估计法：对于随机变量 X ，若 $E(|X|^k)$ 存在，则将 $a_k = E(X^k)$ 称为总体的 k 阶原点矩，若 $E(|X - E(X)|^k)$ 存在，则将 $b_k = E((X - E(X))^k)$ 称为总体的 k 阶中心矩。对于样本而言，同样有样本原点矩和样本中心矩。

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

对于原点矩和中心矩，有 $\hat{a}_k = A_k, \hat{b}_k = B_k$ 。若用未知参数表示总体的矩，再带入矩的估计值，则可以求出对应参数的估计值，这称为矩估计。