基础部分
第四课 概率统计
第2章 随机变量及其分布

本章内容

§2.1 随机变量与分布

§2.2 重要概率分布

本章提要(略，见大纲)

§2.1 随机变量与分布函数

正确理解对概率论研究和发展起重大推动作用的两个最基本概念：“随机变量”和“分布函数”.

2.1.1 随机变量和分布函数的定义和分类

1. rv和 df的定义

定义 2.1.1 设$(\Omega, \mathscr{F}, P)$为概率空间，X为Ω上的实值函数，满足对任意的 $x\in R$， $(X\le x):=\{\omega : X(\omega)\le x\}\in\mathscr{F}$

则称 X为随机变量,简记 rv. 而称实变量的实值函数

$$F_X(x) := P(X\le x),\quad x\in R$$

为 X的分布函数，简记 df.

2. rv与 df的关系

rv给定则 df是存在且唯一决定的.

3. rv和 df的分类

定义 2.1.2 至多取可列多个值的rv [或相应的$F(x)$]，称为离散型的. 设$\{x_i\}$是rv X可能取的值的全体，

$$p_i := P(X=x_i),\quad i=1,2,\cdots(,n)$$

称实数列$\{p_i\}$为离散型X的分布. 称两行矩阵

$$\begin{pmatrix} x_1 & x_2 & \cdots & (x_n) \\ p_1 & p_2 & \cdots & (p_n) \end{pmatrix}$$

为 X的分布列. 其中最后一列表示列数为有限的 n或为可列无穷多的情形. □

定义 2.1.3 在一个有限或无限区间取值的 rv X，如存在非负可积函数 $f(x)$ 使 X在$(-\infty, x]$ 的概率可写成

$$F_X(x)=P(X\le x)=P(-\infty<X\le x)=\int_{-\infty}^{x} f(y)dy,\qquad \forall x\in R$$

则称X [或$F(x)$]为连续型的，称$f(x)$为X [或$F(x)$]的概率密度函数，简记为 pdf. 也常记为 $f_X(x)$. □

2.1.2 分布函数，分布和密度函数

1. 离散型和连续型 df

例 2.1.1 本节引例中，如该厂生产的电子元件的等级数 Y 有分布列

$$Y\sim\begin{pmatrix} 1 & 2 & 3 \\ 0.3 & 0.6 & 0.1 \end{pmatrix}.$$

求 Y的 df

【 $F_Y(y)=\begin{cases} 0 & y<1 \\ 0.3 & 1\le y<2 \\ 0.9 & 2\le y<3 \\ 1 & y\ge 3. \end{cases}$ 】

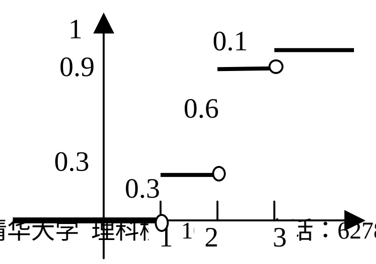

图 2.1.2 离散型分布函数图象

例 2.1.2 设 X 的 pdf 为，

$$f_X(x)=\begin{cases}1/(b-a) & x\in(a,b]\\ 0 & \text{otherwise}\end{cases}$$ ，求 X 的 df.　　【$F_X(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x<b\\ 1 & x\ge b\end{cases}$.】

2．df 的基本性质

性质 1　rv X 的 df $F(x)$ 有下述基本性质：

$F1$）非降性，即 $F(x)\le F(y)$，$\forall$ $x<y$；

$F2$）边界极端性，即

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

$F3$）右连续性，即 $F(x+0)$:= $\lim_{y\downarrow x}F(y)=F(x)$.

性质 2（存在定理）满足性质 $F1$)至 $F3$)的任意一个实变量的实值函数，都可作为一个 df.

性质 3　df 的凸组合，还是 df，即如 $F_i(x)$ 是 df，$i=1,2,\cdots,n$，则对任意实数 $a_i\ge 0,\ i=1,2,\ldots,n,\ \sum_{i=1}^n a_i=1$，$F(x):=\sum_{i=1}^n a_i\ F_i(x)$ 仍是 df.

2.2.3．分布与密度函数的性质

性质 1（基本性质）分布 $\{p_i\}$ 满足 $p_i\ge 0,\ \forall i$，且 $\sum_i p_i=1$

而 pdf 满足 $f(x)\ge 0$，$\forall$ $x\in R$，且 $\int_{-\infty}^{+\infty}f(y)dy=1$.

性质 2　1）对离散型 rv，如其分布为 $\{p_i\}$ 则

$$F_X(x)=\sum_{i:\ x_i\le x}p_i,\quad \forall x\in R$$

2）对有 pdf $f(x)$ 的连续型 rvX，

$$F_X(x)=\int_{-\infty}^{x}f(y)dy,\quad \forall x\in R$$

性质 3　1）凡离散型 rv 有最可能值，即存在 x_m，rv X 取该值的概率不小于取其它值的概率：$P(X=x_m)=p_m\ge p_i$，$\forall$ i.

2）连续型分布取任意一固定值的概率为零，即对每个固定的实数 x，$P(X=x)=0$.

$f(x)dx$ 为 X 在 x 点微分邻域的概率． 由此

$$P(X\in(a,b])=\int_a^b f_X(x)dx=\int_{(a,b]}f_X(x)dx .$$

对更一般的实数集合 D 有 $\boxed{P(X\in D)=\int_D f_X(x)dx}$

[例题精选]

- 分布与 df 的概念

例 2.1.3 将 3 个球逐个随机放入 4 个分别编号为 1、2、3 和 4 的盒子.令 X 是“有球盒子的最小号码”，求 X 的分布列.　　【$\begin{pmatrix}1 & 2 & 3 & 4\\ 37/64 & 19/64 & 7/64 & 1/64\end{pmatrix}$】

例 2.1.4 设 rvX 的 pdf 为 $f(x)=\begin{cases}1/3 & x\in[0,1]\\ 2/9 & x\in[3,6]\\ 0 & otherwise\end{cases}$，若 k 使得 $P(X\ge k)=2/3$，则 k 的取值范围是_________.　　【[1，3]】

- 分布与 df 的性质

例 2.1.5 试确定 a 值，使下一函数为 pdf，$f(x)=a\ e^{-3(x-1)}I_{(1,\ \infty)}(x)$.

例 2.1.6 设 $F_i(x)$ 是 X_i 的 df，i=1,2，为使 $F(x)= aF_1(x)-bF_2(x)$ 是 df，下列给定各组数值中应取

A) a = 3/5，b = -2/5.　　B) a = 2/3，b = 2/3.

C) a = -1/2，b = 3/2.　　D) a = 1/2，b = -3/2.

- **综合题**

例 2.1.7 设某电子元件寿命的 pdf 为 $f(x)=\dfrac{a}{x^2}\quad I(x>100)$

1）试确定 a 值;

2）某台设备装有三个这种电子元件．问在开始使用的 150 小时中它们中恰有一个要替换和至少有一个要替换的概率各是多少？

【 1）$1=\int_{-\infty}^{\infty} f(x)dx=\int_{100}^{\infty} a/x^2\,dx=\dfrac{a}{100}$, 所以 $a=100$。

2）每个元件的寿命有两个可能结果：大于或不大于 150 小时，即可看为 Ber-E，从而三个元件中寿命小于 150 小时（因此要替换）的个数，服从二项分布 B(3，p)，　其中

$$p=\int_{-\infty}^{150} f(x)dx=\int_{100}^{150} 100/x^2\,dx=100\cdot[\frac{1}{x}\Big|_{150}^{100}]=\frac{1}{3}\,.$$

因此，使用到 150 小时它们中恰有一个要替换的概率

$$C_3^1 p(1-p)^2=3\times\frac{1}{3}\times\left(\frac{2}{3}\right)^2=\frac{4}{9}\approx 0.44\,.$$

“至少有一个要替换”概率是　$1-\left(\dfrac{2}{3}\right)^3=\dfrac{19}{27}\approx 0.701\,.$　】

§2.2　重要概率分布

本节从两类随机试验，Poisson 流和误差问题，介绍几类最重要的 rv 及其分布．掌握这些重要分布的定义、性质、产生的背景以及它们间关系．

2.2.1 重要分布的产生与定义

1．Bernoulli 试验及有关分布

1)　Bernoulli 分布

2)　n 重 Ber-试验及其产生的 B(n，p)

3)　可列重 Ber-试验及其产生的 Ge(p)

2．Poisson 流及有关分布

1)　Poisson 流与 Poisson 定理

定理 2.3.1 (Poisson)　设 $\xi_{(0,t]}$，$t \ge 0$ 是 Poisson 流，则存在某正数λ，使

$$p_k(t)=P(\xi_{(0,t]}=k)\;=\;\frac{(\lambda t)^k}{k!}e^{-\lambda t},\quad k=0,1,...$$

Poisson 定理中的λ称为强度．

2)　Poission 流产生离散型的 P(λ)分布

3)　Poisson 流产生的连续型分布：Ex(λ)

误差问题产生的分布：U(a,b)与N(μ，σ^2)

2.2.2 重要分布间的关系和性质

1．重要分布间的关系

2．重要分布的性质

性质 1 重要离散型分布的最可能值

设 X ~ B(n，p)，则 X 的最可能值是 [(n+1)p]．如 (n+1)p 是整数，则[(n+1)p]　1 ＝np-q 也是最可能值．这里 [·]为取整函数．

设 X ~ Ge(p)，则 X 的最可能值是 1.

设 X ~ P(λ)，则 X 的最可能值在[λ]；如 λ＝[λ]，即 λ 是正整数时，则 λ　1 也是最可能值.

性质 2 B(n, p)的 Poisson 逼近.

定理 2.3.1 (Poisson 逼近)　设 X_n ~B(n, p_n)，即对固定的 n 次试验中，每次试验成功的概率是 p_n. 又设存在极限 $\lim_{n\to\infty} np_n$ =λ > 0，则对任意非负整数 k，有

$$P(X_n=k)=C_n^k p_n^{\,k}(1-p_n)^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda},\ n\to\infty\ .$$

性质 3 几何分布和指数分布的无记忆性：

几何分布和指数分布的都有无记忆性:

当 X ~ Ge(p) 时 P(X >n+k | X >n) = P(X >k).

反之，有无记忆性的离散型分布，必为几何分布.

当 X~ Ex(λ)时 P(X >s+t |X >s) = P(X >t)，0 ≤ s，0 < t.

反之，有无记忆性的连续型分布，必为指数分布.

均匀分布和正态分布的性质

性质 4 1) 遵从[a, b]上均匀分布的 rv 的均匀性，使其值落在[a, b]内任一子区间的概率与此子区间长度成正比. 精确地说　$P(X\in D)=L(D)/(b-a)$，其中 L(D)表 D 的长度，而 D 是[a, b]的任意一个(开、闭或半开半闭)子区间，也可以是一些子区间的并集.

2) 正态分布的对称性，使 pdf 是关于直线 x = μ 对称的，

$$\phi(\mu-x;\mu,\sigma)=\ \phi(\mu+x;\mu,\sigma)\ .$$

由此，$\Phi(\mu-x;\mu,\sigma)$= 1　　$\Phi(\mu+x;\mu,\sigma)$.

性质 5 正态分布的其它性质

1) $\phi(x;\mu,\sigma)$>0，任意阶导函数 $\phi^{(n)}(x;\mu,\sigma)$，∀ n，存在且连续.

2) $\phi(x;\mu,\sigma)$ 在 (∞, μ)中单调升，在 x = μ 处达极大值 1/ ($\sqrt{2\pi}\sigma$)，而在 (μ, ∞) 时下降. 参数 μ 决定它的对称位置；σ 越大 pdf 越平缓(参看图 2.2.7)，概率分布越分散.

3) 如 X ~ N(μ, σ^2)则其标准化 $X^*\equiv(X-\mu)/\sigma$ ~ N(0, 1).

4) 3σ法则. 正态变量离中心位置 μ 的距离超过 3σ 的概率不到千分之三，依此在正态性统计判别和产品质量管理中形成很有用的 3σ 法则.

性质 6 独立和的分布与分布的可加性

可加性的证明方法:

(1). 由分布产生的背景，立即可得上述结论: 例如 B(n,p)、F(r,p)和Γ(r,p)的可加性(当 r 为正整数时)，以及关于 Ge(p)、Ex(λ)的结论.

(2). 利用全概率公式，例如 B(n,p)、F(r,p)、P(λ)和Γ(r,p)的可加性;

(3). 利用求独立和的 df 或者密度的卷积公式

[典型例题]

例 2.2.1 设某车间需要安排维修工人负责对一批相同型号设备进行保全维修，有两种建议方案. 方案 A：1 人维修固定的 20 台.　方案 B：3 人维修固定的 80 台. 设每台设备的故障率为 0.01，哪种方案较好，即出现设备需要维修而得不到维修（维修人员正忙于其它设备的维修）的概率较小？

解 Y_n: n台中的故障数，　则 Y_n~B(n, p)，

$$p_a=P(Y_{20}>1)=1-P(Y_{20}=0)-P(Y_{20}=1)$$
$$=1-q^{20}-C_{20}^1pq^{19}\approx 0.0169$$

用 Poisson 近似，λ = 0.2，则　$p_a=1-e^{-0.2}-0.2\times e^{-0.2}\approx 0.0175$

$$p_b=P(Y_{80}>3)\approx 1-\sum_{i=0}^3\frac{(80\times 0.01)^i}{i!}\mathrm{e}^{-(80\times 0.01)}\approx 0.0091\ .$$

p_b > p_a，方案B较好.

例 2.2.2 一大批产品，其次品率为 p，采取下列方法抽样检查：抽样直至抽到一个次品时为止，或一直抽到 10 个产品时就停止检查．设 X 为停止检查时抽样的个数．求 X 分布列．

【$P(X=k)=q^{k-1}p,\quad k=1,\ 2,\ \ldots,\ 9$，$P(X=10)=q^9$】

例 2.2.3（非中心的指数分布） 设某流水线上一类电子元件寿命(小时) X 的 pdf 为 $f_X(x)=\lambda e^{-\lambda(x-10)}\quad I(x>a)$，其中 $\lambda>0$ 是常数．试求常数 a；如令 $y=x-a$，将 $f_X(x)$ 作平移，得到新的函数是否仍然为 pdf？能判断它是什么类型分布吗?

例 2.2.4 已知 $X\sim N(\mu,\sigma^2)$．

1）求 P($a\le X\le b$)；

2）设 μ=20，σ^2=40^2，求P($|X|\le 20$)的值，并找点 x_0，使P($X>x_0$) = 0.05.

【$\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$；$\Phi(0)-\Phi(-1)=0.5-0.1587=0.3413$，$x_0=85.6$】

例 2.2.5 对某射手打靶考核，有两次命中 6 环以下(不含 6 环)时，立即淘汰出局．如果此射手每次命中 6 环及其以上的概率是 0.8，则他在第 4 次射击后即被淘汰的概率是_______.

【$p_2:=P(X=2)=C_{4-1}^{2-1}p^2q^{4-2}$，$p=0.2$】