期中复习(1-3章)

2018年4月29日　　15:17

第一章:

- 对立事件与互斥事件的区别:

  对立事件:A,B中必有一个发生，且不能同时发生(样本点必定属于A,B之一);

  互斥事件:A,B不能同时发生(可以有样本点既不属于a又不属于b)
- 推论:对于任意事件A,B有:

  P(B-A)=P(B)-P(A.B)
- 不放回抽样问题:

  要不都考虑顺序，要不都不考虑顺序

- 区分排列问题与组合问题:P11，P25.11

  （1）将n只球随机放入N(N≥ n)个盒子中去，试求每个盒子至多有一只球的概率

错解: $P=1-\frac{C_N^n}{N^n}$

正解: $P=1-\frac{A_N^n}{N^n}$ ⟶ 这样更好理解:

$$P=1-\frac{N\cdot(N-1)(N-2)\cdots(N-n+1)}{N^n}$$

- 独立事件

  定义:若P(AB)=P(A)*P(B),则称A,B独立。

  推论:若A,B独立，则A与$\bar{B}$、$\bar{A}$与B也相互独立

（2）袋中有a只白球，b只红球，k个人**依次取球**（k≤ a + b），**不放回抽样**，求第i人取到白球的概率(P12例5)

↳ 排列问题

第i个人从a只白球中取一只，有a种可能;剩下k-1个人再从剩下的a+b-1只球中各取一只:

$$p=\frac{a\cdot A_{a+b-1}^{k-1}}{A_{a+b}^{k}}=\frac{a}{a+b}$$

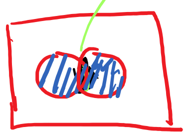
排列问题

若进行**放回抽样**，第i个人取到白球的概率**依然是**:

$$p=\frac{a}{a+b}$$

- 求和事件发生的概率时,不要忘记**减去交集**(P13例6;P22例2)

减去交集



$$A=A_1A_2\cup A_3A_4$$

$$\text{则 } P(A)=P(A_1A_2)+P(A_3A_4)-P(A_1A_2A_3A_4)$$

$$P(\bar{A}\bar{B})=P(\overline{A\cup B})=1-P(A\cup B)$$
$$=1-[P(A)+P(B)-P(AB)]$$

- 平均分配问题(P13例7):

  15名学生随机平均分配到3个班，分法总数为:(不是排列问题！因为分班时班级只关心人数)

$$C_{15}^5\cdot C_{10}^5\cdot C_5^5$$

- 条件概率:
  （1）乘法公式的应用:P16例3
  （2）贝叶斯公式的应用:P19例5，**P22例3**

- 第一章章节习题
  事件的划分、乘法公式：P26.18;
  事件的划分+全概率公式+乘法公式：P26.20
  事件的多次划分:P26.24
  事件的集合:P28.37-3
  贝叶斯公式的应用P29.40

- 第二章

- n重伯努利试验的条件:各次实验之间相互独立
  a只白球，b只黑球，从中依次抽取n只球（**不放回**），抽到白球的数量，这就**不是**n重伯努利实验;
  若是**放回抽样**，则**是**n重伯努利实验

- 二项分布：P36例4

- 泊松分布:

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

泊松定理:

$$\lim_{n\to\infty} C_n^k \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!} \quad (其中 \lambda = n \cdot p_n)$$

当实验次数足够多时，且各次实验结果发生概率足够小时，二项分布可用泊松分布来近似:

$n\to\infty,\ p\to 0 \qquad n\cdot p=\lambda$（常数）

- 随机变量的分布函数
  (离散随机变量的分布函数和连续随机变量的分布函数都满足以下性质)

$$F(x)=P\{X\leqslant x\}=\sum_{x_k\leqslant x} P\{X=x_k\}=\sum_{x_k\leqslant x} p_k$$

  (1)不减函数
  （2） $0 \le F(x) \le 1$，且$F(-\infty)=0;F(+\infty)=1$
  (3) F(X)右连续
  反之，满足以上三条性质的函数必是某个随机变量的分布函数。

- 连续随机变量
  例题:P41例2

- 概率密度函数
  性质:

(1) $f(x)\geqslant 0$

(2) $\int_{-\infty}^{+\infty} f(x)\,dx=1$

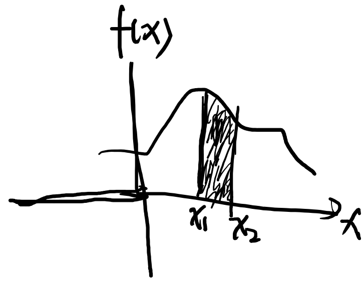
$f(x)$

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

(3) 对 $x_1 \leq x_2$

有 $P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\, dx$



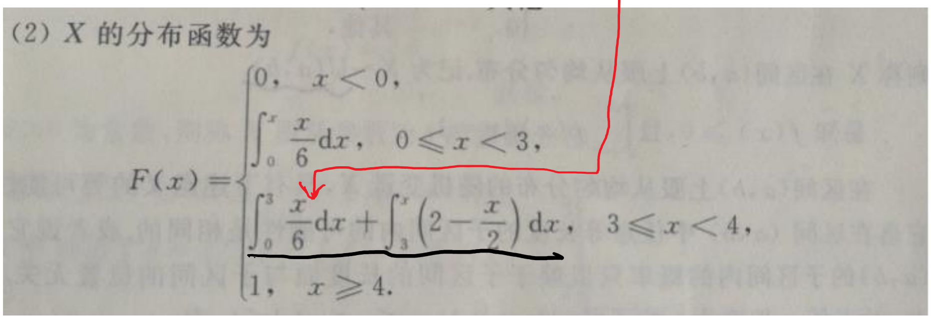
(4) 若f(x)在x处连续，则有 $F'(x) = f(x)$

$$f(x) = \lim_{\Delta x \to 0^+} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0^+} \frac{P\{x \leq X \leq x+\Delta x\}}{\Delta x}, \quad P\{x < X \leq x+\Delta x\} \approx f(x)\Delta x$$

注意:

(1)改变概率密度在个别点的函数值不影响分布函数的值，因此，**概率密度不唯一**

(2)在用积分求连续随机变量的分布函数时，要注意，不要忘记加上“历史积分值”



(2) X 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ \int_0^x \frac{x}{6} dx, & 0 \leqslant x < 3, \\ \int_0^3 \frac{x}{6} dx + \int_3^x \left(2 - \frac{x}{2}\right) dx, & 3 \leqslant x < 4, \\ 1, & x \geqslant 4. \end{cases}$$

- 常见分布的数学符号描述

- 离散型随机变量

(1)二项分布:

$X \sim b(n, p)$

(2) 泊松分布

$X \sim \pi(\lambda)$

- 连续型随机变量

(1) 指数分布 $f(x) = \begin{cases} \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}}, & x > 0 \\ 0 & 其他 \end{cases}$ 无记忆性

(2)均匀分布

$X \sim U(a, b)$, $f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & 其他 \end{cases}$

(3) 正态分布

$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $X \sim N(\mu, \sigma^2)$

$f(x) = \frac{\quad}{\sqrt{2\pi}\cdot\sigma} e$, $X \sim N(\mu, \sigma)$

$f(x)_{max} = f(x)|_{x=\mu} = \frac{1}{\sqrt{2\pi}\sigma}$.

- 标准正态分布

$(\mu=0, \sigma=1)$

概率密度: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

分布函数性质: $\Phi(-x) = 1 - \Phi(x)$

一般正态分布与标准正态分布之间的关系:

$X \sim N(\mu, \sigma^2)$

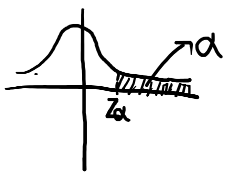
$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

标准正态分布的分位点:

$X \sim N(0,1)$

若 $P\{X > z_\alpha\} = \alpha$, $0<\alpha<1$

点 z_α 称为 α 分位点.



- 随机变量的函数的分布

例 2 设随机变量 X 具有概率密度

$$f_X(x) = \begin{cases} \frac{x}{8}, & 0<x<4, \\ 0, & \text{其他}. \end{cases}$$

求随机变量 $Y=2X+8$ 的概率密度.

解 分别记 X,Y 的分布函数为 $F_X(x), F_Y(y)$. 下面先来求 $F_Y(y)$.

$$F_Y(y) = P\{Y \leqslant y\} = P\{2X+8 \leqslant y\}$$
$$= P\left\{X \leqslant \frac{y-8}{2}\right\} = F_X\left(\frac{y-8}{2}\right).$$

①

将 $F_Y(y)$ 关于 y 求导数,得 $Y=2X+8$ 的概率密度为

$$f_Y(y) = f_X\left(\frac{y-8}{2}\right)\left(\frac{y-8}{2}\right)'$$
$$= \begin{cases} \frac{1}{8}\left(\frac{y-8}{2}\right)\cdot\frac{1}{2}, & 0<\boxed{\frac{y-8}{2}}<4, \\ 0, & \text{其他} \end{cases}$$
$$= \begin{cases} \frac{y-8}{32}, & 8<y<16, \\ 0, & \text{其他}. \end{cases}$$

②

定理:

定理 设随机变量 X 具有概率密度 $f_X(x)$，$-\infty<x<\infty$，又设函数 $g(x)$ 处处可导且恒有 $g'(x)>0$（或恒有 $g'(x)<0$），则 $Y=g(X)$ 是连续型随机变量，其概率密度为

$$f_Y(y)=\begin{cases} f_X[h(y)]\,|h'(y)|, & \alpha<y<\beta, \\ 0, & 其他, \end{cases} \tag{5.2}$$

其中 $\alpha=\min\{g(-\infty),g(\infty)\}$，$\beta=\max\{g(-\infty),g(\infty)\}$，$h(y)$ 是 $g(x)$ 的反函数.

我们只证 $g'(x)>0$ 的情况.此时 $g(x)$ 在 $(-\infty,\infty)$ 严格单调增加，它的反函数 $h(y)$ 存在，且在 (α,β) 严格单调增加，可导.分别记 X，Y 的分布函数为 $F_X(x)$，$F_Y(y)$.现在先来求 Y 的分布函数 $F_Y(y)$.

- 第二章章节习题

P58.26:灵活转换概率表达式:

$$P\{|X|>2\}=1-P\{|X|\leqslant 2\}$$

P59.31：条件概率、混合型随机变量（灵活题）

- 第三章

二维随机变量的**分布函数**的性质:

1. $F(x,y)$ 是 x 和 y 的不减函数

2. $0\leqslant F(x,y)\leqslant 1$；

   for fixed y, $F(-\infty,y)=0$；

   for fixed x, $F(x,-\infty)=0$；

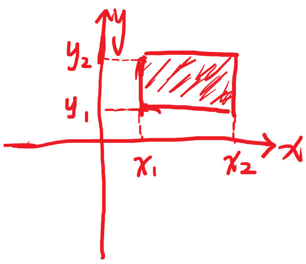
   $F(-\infty,-\infty)=0$；

   $F(\infty,\infty)=1$；

3. $F(x,y)$ 关于 x 右连续，关于 y 也右连续

4. $F(x_2,y_2)+F(x_1,y_1)-F(x_2,y_1)-F(x_1,y_2)\geqslant 0$

   $= P\{x_1\leqslant X\leqslant x_2,\ y_1\leqslant Y\leqslant y_2\}$



对于离散型随机变量:

$$F(x,y)=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}p_{ij}$$

对于连续型随机变量

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

二维随机变量的**概率密度函数**的性质:

1. $f(x,y)\geqslant 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=F(\infty,\infty)=1$
3. G是xOy上的区域，点(X,Y)落在G内的概率为
$$P\{(X,Y)\in G\}=\iint_G f(x,y)\,dx\,dy$$
4. 在f(x,y)的连续点处有
$$\frac{\partial^2 F(x,y)}{\partial x\,\partial y}=f(x,y)$$，当 $\Delta x,\Delta y$ 很小时
$$P\{x\leqslant X\leqslant x+\Delta x,\ y<Y\leqslant y+\Delta y\}\approx f(x,y)\cdot\Delta x\cdot\Delta y$$

边缘分布:

$$F_X(x)=P\{X\leqslant x\}=P\{X\leqslant x,\ Y<\infty\}=F(x,\infty)$$

- 注意:求边缘概率分布时也要先画出积分区间，再进行求解，否则很容易错！！！

对于离散型随机变量：$F_X(x)=\sum\limits_{x_i\leqslant x}\sum\limits_{j=1}^{\infty}p_{ij}$

对于连续型随机变量：

$$F_X(x)=F(x,\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty} f(x,y)\,dy\right]dx$$

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

- 二维正态分布

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

- 二维随机变量的条件概率分布

离散：在 $X=x_i$ 条件下Y的分布律为：

$$P\{Y=y_j\mid X=x_i\}=\frac{P\{X=x_i,Y=y_j\}}{P\{X=x_i\}}=\frac{p_{ij}}{p_{i\cdot}},\ j=1,2\cdots$$

连续：$f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$

$$F_{X|Y}(x|y)=P\{X\leq x|Y=y\}=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}\cdot dx$$

易错题:

1.P69例2 （注意是否需要乘以组合系数）

2.P71例3（求边缘分布时注意积分区间和变量的范围）

独立随机变量的性质:

① $F(x,y)=F_X(x)\cdot F_Y(y)$

② $f(x,y)=f_X(x)\cdot f_Y(y)$（连续）

③ $P\{X=x_i,Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}$（离散）

巧妙计算联合概率分布（不需计算积分）:P74例题

- 定理:

若 $(X_1,X_2\cdots,X_m)$ 和 $(Y_1,Y_2\cdots Y_n)$ 相互独立，

h,g 连续，

那么 $X_i\ (i=1,2\cdots m)$ 和 $Y_j\ (j=1,2\cdots n)$ 也相互独立

$h(X_1,X_2\cdots X_m)$ 和 $g(Y_1,Y_2\cdots Y_n)$ 相互独立

- 两个随机变量的函数的分布

一. $Z=X+Y$ 的分布

$$f_{X+Y}(z)=\int_{-\infty}^{\infty}f(z-y,y)\,dy$$
$$=\int_{-\infty}^{\infty}f(x,z-x)\cdot dx$$

若 X,Y 独立，

则 $f(x,z-x)=f_X(x)\cdot f_Y(z-x)$

$$f_{X+Y}(z)=\int_{-\infty}^{\infty}f_X(x)\cdot f_Y(z-x)\cdot dx$$
$$=f_X(x)*f_Y(y)$$

卷积

相互独立的随机变量的和的概率密度等于它们各自概率密度的卷积

- 定理:

有限个相互独立的正态随机变量的线性组合仍然服从正态分布

例题:明确**积分区间**，**函数自变量**取值范围，搞清楚积分区域的横坐标，纵坐标分别是哪个，**P77例2**

二. $Z=\frac{Y}{X}$，$Z=X\cdot Y$ 的分布

$$f_{\frac{Y}{X}}(z)=\int_{-\infty}^{\infty}|x|\cdot f(x,xz)\cdot dx$$

$$f_{X\cdot Y}(z)=\int_{-\infty}^{\infty}\frac{1}{|x|}\cdot f\left(x,\frac{z}{x}\right)\cdot dx$$

三、$M=\max\{X,Y\}$ 及 $N=\min\{X,Y\}$ 的分布

$$F_{\max}(z)=P\{M\leq z\}=P\{X\leq z,Y\leq z\}$$ （最大值小于z，它们都小于z）

若 X,Y 相互独立

则 $F_{\max}(z)=F_X(z)\cdot F_Y(z)$

$$F_{\min}(z)=P\{N\leq z\}=1-P\{N>z\}$$ （最小值大于，它们都大于）

$$=1-P\{X\geq z,Y\geq z\}$$

若 X,Y 独立，

则 $F_{\min}(z)=1-P\{X\geq z\}\cdot P\{Y\geq z\}$

$$=1-[1-F_X(z)]\cdot[1-F_Y(z)]$$

- 第三章习题

P84.1：区分条件概率和**积事件**的概率

“第一次取出的是正品、第二次取出的是次品”是一个**积事件**，

可以用乘法公式来计算概率(不放回抽样);也可以用**排列**来算

P85.11:二项分布的变形:

$$(1)\ P\{X=n\}=\sum_{m=0}^{n}P\{X=n,Y=m\}$$

$$=\sum_{m=0}^{n}\frac{e^{-14}(7.14)^m(6.86)^{n-m}}{m!(n-m)!}$$

$$=\frac{e^{-14}}{n!}\sum_{m=0}^{n}\boxed{\frac{n!}{m!(n-m)!}}(7.14)^m(6.86)^{n-m}\rightarrow C_n^m$$

$$=\frac{e^{-14}}{n!}(7.14+6.86)^n=\frac{14^n e^{-14}}{n!},n=0,1,2,\cdots.$$

P86.13(3)条件概率的概率密度

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}$$

$$P\left\{Y\geq\frac{1}{4}\middle|X=\frac{1}{2}\right\}=\int_{\frac{1}{4}}^{\infty}f_{Y|X}\left(y\middle|x=\frac{1}{2}\right)dy$$

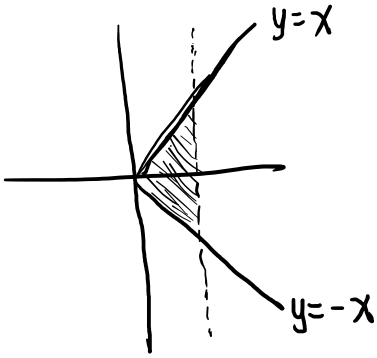
条件概率密度的积分

条件概率分布 = 条件概率密度的积分

P86.14:积分区间不要弄错

$|y| < x$, $0<x<1$

$$|y| = x \begin{cases} y = x & y>0 \\ y = -x & y<0 \end{cases}$$



易懵逼：P87.21

随机变量的函数的概率分布

步骤:以Z=X+Y为例

1.把其中一个随机变量用另外一个表示:

Y=Z-X

2.替换随机变量的约束条件

0<y<1→ $0 < z - x < 1$

3.根据新的约束条件画出积分区间(这时候变量就是Z和X了)

0<X<1;

0<Z-X<1;

4.写出分布函数的表达式，进行积分（**X的积分区间依赖于Z,先确定z的范围，再确定X的积分区间**）

$$\underbrace{f_{X+Y}(z)}_{\text{概率密度}} = \int_{-\infty}^{\infty} \underbrace{f(x, z-x)}_{\text{概率密度}} \cdot dx$$