姜晓千 2023 年强化班笔记

数学笔记

Weary Bird

2025 年 7 月 4 日

相见欢 · 林花谢了春红

林花谢了春红，太匆匆。无奈朝来寒雨晚来风。胭脂泪，相留醉，几时重。自是人生长恨水长东。

2025 年 7 月 4 日

目录

第一章　事件与概率论

1.1　事件的关系、运算与概率的性质

1. 事件: 样本点的**集合**
2. 事件的关系 (3+1): 包含, 互斥, 对立 + 独立
3. 事件的运算 (3 个): 交, 并, 补

Remark. (事件的运算律)

(1)	交换律	$A \cup B = B \cup A, AB = BA$
(2)	结合律	$A \cup (B \cup C) = (A \cup B) \cup A, A(BC) = (AB)C$
(3)	分配律	$A \cup (BC) = (A \cup B)(A \cup C), A(B \cup C) = (AB) \cup (AC)$
(4)	摩根律	$\overline{A \cup B} = \bar{A}\bar{B}, \bar{(}AB) = \bar{A} \cup \bar{B}$
(5)	吸收律	$A \cup (AB) = A, A(A \cup B) = A$

1. 设 A, B 为随机事件, 且 $P(A) = P(B) = \frac{1}{2}$,$P(A \cup B) = 1$, 则

$$(A)\ A \cup B = \Omega \quad (B)\ AB = \varnothing \quad (C)\ P(\bar{A} \cup \bar{B}) = 1 \quad (D)\ P(A - B) = 0$$

Solution. 由加法公式 $P(A \cup B) = P(A) + P(B) - P(AB) \implies P(AB) = 0$
注意由概率并不能推断事件, 所以 (A)(B) 均不正确
对于 (C) 选项 $P(\bar{A} \cup \bar{B}) = 1 - P(\overline{AB}) = 1$ 正确
对于 (D) 选项, 由减法公式 $P(A - B) = P(A) - P(AB) = \frac{1}{2}$ □

总结

(1) 必然事件发生的概率为 1, 但概率为一的事件不一定是必然事件

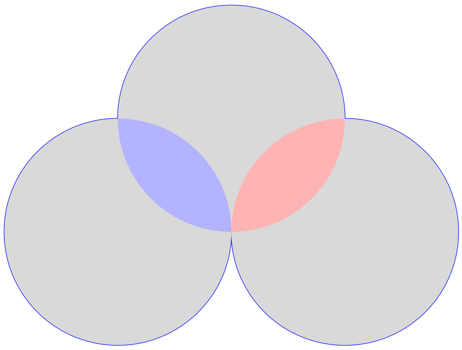
(2) 不可能事件发生的概率为 0, 但概率为零的事件不一定是不可能事件

这两个结论考虑**连续型随机变量**即可

2. (2020, 数一、三) 设 A,B,C 为随机事件, 且 $P(A)=P(B)=P(C)=\frac{1}{4}$,$P(AB)=0$,$P(AC)=P(BC)=\frac{1}{12}$, 则 A,B,C 只有一个事件发生的概率为

$$(A)\ \frac{3}{4}\quad (B)\ \frac{2}{3}\quad (C)\ \frac{1}{2}\quad (D)\ \frac{5}{12}$$

Solution. 这种题一般考虑 Venn 图, 比用公式展开简单很多



则只有一个事件发生的概率为 $(\frac{1}{4}-\frac{1}{12})\times 2+\frac{1}{4}-2\times\frac{1}{12}=\frac{5}{12}$ □

3. 设随机事件 A,B 满足 $AB=\bar{A}\bar{B}$, 且 $0<P(A)<1$,$0<P(B)<1$, 则 $P(A|\bar{B})+P(B|\bar{A})=$ ________

Solution. 根据结论, 有 A,B 互斥, 则 $P(A|\bar{B})=P(B|\bar{A})=1$ □

Corollary 1.1.1. 若 $AB=\bar{A}\bar{B}$, 则 A,B 必然对立

Proof.

$$\begin{aligned}
&AB=\bar{A}\bar{B}\\
&\iff AB\cup\bar{A}B=\bar{A}\bar{B}\cup\bar{A}B\\
&\iff (A\cup\bar{A})B=\bar{A}(\bar{B}\cup B)\\
&\iff B=\bar{A}
\end{aligned}$$

□

4. 设随机事件 A, B, C 两两独立, 满足 $ABC = \varnothing$, 且 $P(A) = P(B) = P(C)$,A, B, C 至少有一个发生的概率为 $\frac{9}{16}$, 则 $P(A) =$

   Solution. 由题意有 $P(A \cup B \cup C) = \frac{9}{16}$, 由加法公式与独立性有

$$\begin{aligned}P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A)P(B) \\ &\quad - P(A)P(C) - P(B)P(C) - P(A)P(B)P(C)\end{aligned}$$

   由 $P(A) = P(B) = P(C)$, 上式化为 $3P(A) - 3P(A)^2 = \frac{9}{16} \implies P(A) = \frac{1}{4}$或$P(A) = \frac{3}{4}$, 显然 $P(A) \neq \frac{3}{4} > P(A \cup B \cup C)$, 故 $P(A) = \frac{1}{4}$ □

5. 设 A, B 为随机事件, 且 $P(A) = \frac{2}{3}$,$P(B) = \frac{1}{2}$, 则 $P(A|B) + P(B|A)$ 的最大值为 ______, 最小值为 ______.

   Solution. 关于概率的不等式基于如下事实, 对于任意一个概率其值均位于 $[0, 1]$ 之间, 事件 AB 的和事件不可能小于单独 A,B 发生概率之和, 事件 AB 的积事件不可能大于任意一个事件单独发生的概率.

$$P(A) + P(B) - 1 <= P(AB) \leq \min\left(P(A), P(B)\right) \leq P(A) + P(B) \leq P(A \cup B)$$

□

1.2 三大概型的计算

> **Remark.** 三大概率模型
>
> 经典概型 – 有限个等可能的样本点, **排列组合问题**
>
> 几何概型 – 使用几何参数度量概率, 比如说长度, 面积, 体积等
>
> 伯努利概型 – 独立重复试验每次成功的概率为 p, 不成功的概率为 $(1 - p)$

6. (2016, 数三) 设袋中有红、白、黑球各 1 个, 从中**有放回地**取球, 每次取 1 个, 直到三种颜色的球都取到为止, 则取球次数恰好为 4 的概率为

   Solution. (古典概型)

$$\frac{\binom{3}{1}\binom{2}{1}\binom{2}{3}}{3^4} = \frac{2}{9}$$

   首先从 3 个颜色中选择一个为第四次抽的颜色, 再从剩下两个颜色中选择一个为出现两次的颜色, 在选择该颜色抽出的次序. □

7. 在区间 $(0,a)$ 中随机地取两个数, 则两数之积小于 $\frac{a^2}{4}$ 的概率为

Solution. (几何概型)

$$\frac{\frac{a}{4}\cdot a+\int_{\frac{a}{4}}^{a}\frac{a^2}{4x}dx}{a^2}=\frac{1}{4}+\frac{1}{2}\ln 2$$

□

8. 设独立重复的试验每次成功的概率为 p, 则第 5 次成功之前至多 2 次失败的概率为

Solution. 失败零次–p^5, 失败一次–$\binom{1}{5}p^4(1-p)p$, 失败两次–$\binom{2}{6}p^4(1-p)^2p$

故第 5 次成功之前至多 2 次失败的概率为

$$p^5+\binom{1}{5}p^4(1-p)p+\binom{2}{6}p^4(1-p)^2p$$

□

1.3 三大概率公式的计算

Remark. 三大概率公式

条件概率公式 $P(A\mid B)=\frac{P(AB)}{P(B)}$

推论 $P(AB)=P(B)P(A\mid B), P(A_1A_2\ldots A_n)=P(A_1)P(A_2\mid P(A_1))P(A_3|P(A_1A_2))\ldots$

全概率公式 $P(A)=\sum_{i=1}^{n}P(AB_i)=\sum_{i=1}^{n}P(B_i)P(A\mid B_i)$

贝叶斯公式 $P(B_j\mid A)=\frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n}P(B_i)P(A|B_i)}$

若称 $P(B_j)$ 为 B_j 的先验概率, 称 $P(B_j\mid A)$ 为 B_j 的后验概率. 则贝叶斯公式专门用于计算后验概率的公式.

9. 设 A,B 为随机事件, 且 $P(A\cup B)=0.6$,$P(B|\bar{A})=0.2$, 则 $P(A)=$ ______

Solution.

$$P(A\cup B)=P(A)+P(B)-P(AB)=0.6, P(B\mid\bar{A})=\frac{P(B)-P(AB)}{1-P(A)}=0.2$$

联立有

$$\frac{0.6-P(A)}{1-P(A)}=0.2$$

, 则 $P(A)=0.5$ □

10. (2018, 数一) 设随机事件 A 与 B 相互独立,A 与 C 相互独立, 满足 $BC=\varnothing$, 且

$$P(A)=P(B)=\frac{1}{2},\quad P(AC|AB\cup C)=\frac{1}{4},$$

则 $P(C)=$______.

Solution.

$$\begin{aligned} P(AC|AB\cup C) &= \frac{P(AC)}{P(AB\cup C)} \\ &= \frac{P(A)P(C)}{P(AB)+P(C)} \\ &= \frac{\frac{1}{2}P(C)}{\frac{1}{4}+P(C)}=\frac{1}{4} \end{aligned}$$

则 $P(C)=\frac{1}{4}$ □

11. (2003, 数一) 设甲、乙两箱装有同种产品, 其中甲箱装有 3 件合格品和 3 件次品, 乙箱装有 3 件合格品。从甲箱中任取 3 件产品放入乙箱,

(1) 求乙箱中次品件数 X 的数学期望;

(2) 求从乙箱中任取一件产品是次品的概率.

Solution. (作为小题来考还可以)

方法一: 用概率

(1) 对于数字特征的题目, 先求概率分布再说, 由于 $P(X=k)=\frac{C_3^kC_3^{3-k}}{C_6^3}$

X	0	1	2	3
$P(x)$	$\frac{1}{20}$	$\frac{9}{20}$	$\frac{9}{20}$	$\frac{1}{20}$

则所求数学期望 $EX=\frac{9}{20}+2\times\frac{9}{20}+\frac{3}{20}=\frac{3}{2}$

(2)

$$\begin{aligned} P(A) &= \sum_{k=0}^{3}P(X=k)P(A\mid x=k) \\ &= \frac{1}{20}\times 0+\frac{9}{20}\times\frac{1}{6}+\frac{9}{20}\times\frac{2}{6}+\frac{1}{20}\times\frac{3}{6} \\ &= \frac{1}{4} \end{aligned}$$

方法二: 超几何分布

(1) $X \sim H(N, M, n), N=6, M=3, n=3$, 则 $EX = \frac{nM}{N} = \frac{3}{2}$

(2)

$$\begin{aligned}
P(A) &= \sum_{k=0}^{3} P(X=k)P(A \mid x=k) \\
&= \sum_{k=0}^{3} P(X=k)\frac{k}{6} \\
&= \frac{1}{6}\sum_{k=0}^{3} P(X=k)k \\
&= \frac{1}{6}EX \\
&= \frac{1}{4}
\end{aligned}$$

□

1.4 事件独立的判定

Remark. (事件独立的充要条件)

$$\begin{aligned}
&P(AB) = P(A)P(B) \\
&\iff P(A \mid B) = P(A) \\
&\iff P(A \mid \bar{B}) = P(A) \iff P(A \mid B) = P(A \mid \bar{B}) \quad (0 < P(B) < 1) \\
&\iff A \text{ 与} \bar{B}, \text{ 或} \bar{A} \text{ 与} B, \text{ 或} \bar{A} \text{ 与} \bar{B} \text{ 相互独立} \\
&\iff P(A \mid B) + P(\bar{A} \mid \bar{B}) = 1, \quad 0 < P(B) < 1
\end{aligned}$$

12. 设 A, B 为随机事件, 且 $0 < P(A) < 1$, 则

(A) 若 $A \supset B$, 则 A, B 一定不相互独立

(B) 若 $B \supset A$, 则 A, B 一定不相互独立

(C) 若 $AB = \varnothing$, 则 A, B 一定不相互独立

(D) 若 $A = \bar{B}$, 则 A, B 一定不相互独立

Solution. (A)(B)(C) 考虑 $\varnothing$ 则都不对

(D) 由于 A 不是必然事件, 则 B 不是不可能事件, 则 $0 < P(A) < 1, 0 < P(B) < 1$, 根据下面的总结 A, B 一定不独立 □

总结

(1) 概率为 0 或 1 的事件与任意事件独立

特别的, 不可能事件与必然事件与任意事件独立

(2) 设 $0 < P(A) < 1, 0 < P(B) < 1$,

A, B 互不相容, 则 A, B 一定不独立

A, B 独立, 则 A, B 一定不互不相容

13. 设 A, B, C 为随机事件,A 与 B 相互独立, 且 $P(C) = 0$, 则 $\bar{A}, \bar{B}, \bar{C}$

(A)相互独立 (B)两两独立, 但不一定相互独立

(C)不一定两两独立 (D)一定不两两独立

Solution. 由 $P(C) = 0$ 知 A, B, C 相互独立, 则 $\bar{A}\bar{B}\bar{C}$ 也相互独立. □

两两独立与相互独立

$$\text{相互独立}\left\{\begin{array}{l} \left.\begin{array}{l} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{array}\right\} \quad \text{两两独立} \\ P(ABC) = P(A)P(B)P(C) \end{array}\right.$$

第二章　一维随机变量

2.1　分布函数的判定与计算

Remark. (分布函数的性质)

(1) $0 \leq F(x) \leq 1, -\infty < x < +\infty, F(-\infty) = 0, F(+\infty) = 1$

(2) (单调不减) 当 $x_1 < x_2$ 时,$F(x_1) < F(x_2)$

(3) (右连续) $F(x+0) = F(x)$

上面三个性质为分布函数的定义, 只要满足上述性质的函数一定是某一个概率分布的分布函数

(4) $P\{a < X \leq b\} = F(b) - F(a)$

(5) $P\{X < x\} = F(x-0), P\{X = x\} = F(x) - F(x-0)$

$P\{a \leq x \leq b\} = P\{x \leq b\} - P\{x < a\} = F(b) - F(a-0)$

$P\{a < x < b\} = P\{x < b\} - P\{x \leq a\} = F(b-0) - F(a)$

1. 设随机变量 X 的分布函数为 $F(x)$,a, b 为任意常数, 则下列一定不是分布函数的是

$$(A)\ F(ax+b) \quad (B)\ F(ax^2+b) \quad (C)\ F(ax^3+b) \quad (D)\ 1-F(-x)$$

总结

对于 $F(ax+b), F(ax^3+b), \ldots$ 只要 $a > 0$ 则这些函数都是分布函数

对于 $F(a^2x+b), F(a^4+b), \ldots$ 都一定不是分布函数

对于 $G(x) = 1 - F(-x)$

若 X 是连续性随机变量则是, 否则不是 ($F(x)$ 不满足左连续, 则 $G(x)$ 不满足右连续)

2. 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}1-|x|, & |x|<1\\ 0, & \text{其他}\end{cases}$$

则 X 的分布函数 $F(x)=$______,$P\{-2<X<\frac{1}{4}\}=$______.

Solution.

(方法一变限积分)

$$f(x)=\begin{cases}1+x, & -1<x<0\\ 1-x, & 0\le x<1\\ 0, & \text{其他情况}\end{cases}$$

根据 $F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t$，有

$$F(x)=\begin{cases}0, & x\le -1\\ \int_{-1}^{x}(1+t)\,\mathrm{d}t, & -1<x<0\\ \int_{-1}^{0}(1+t)\,\mathrm{d}t+\int_{0}^{x}(1-t)\,\mathrm{d}t, & 0\le x<1\\ 1, & x\ge 1\end{cases}$$

$$=\begin{cases}0, & x\le -1\\ x+\frac{x^2}{2}+\frac{1}{2}, & -1<x<0\\ x-\frac{x^2}{2}+\frac{1}{2}, & 0\le x<1\\ 1, & x\ge 1\end{cases}$$

$$\begin{aligned}P\{-2<X<\frac{1}{4}\}&=F(\frac{1}{4})-F(-2)\\&=\int_{-2}^{\frac{1}{4}} f(x)\mathrm{d}x\\&=\frac{23}{32}\end{aligned}$$

(方法二定积分)

$$\int f(x)\mathrm{d}x = \begin{cases} C_1, & x < -1 \\ x + \frac{x^2}{2} + C_2, & -1 \le x < 0 \\ x - \frac{x^2}{2} + C_3, & 0 \le x < 1 \\ C_4, & x \ge 1 \end{cases}$$

$$\xlongequal{\text{由分布函数的定义}} \begin{cases} 0, & x \le -1 \\ x + \frac{x^2}{2} + \frac{1}{2}, & -1 < x < 0 \\ x - \frac{x^2}{2} + \frac{1}{2}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

□

2.2 概率密度的判定与计算

Remark. (概率密度的性质)

(1) $f(x) \ge 0, -\infty < x + \infty$

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

上面两条性质为概率密度的定义, 任何满足上面的函数都是某个概率的概率密度函数

(3) $P\{a < X \le b\} = \int_a^b f(x)\mathrm{d}x$

推广 $P\{a < X \le b\} = P\{a \le X \le b\} = P\{a \le X < b\} = P\{a < X < b\} = \int_a^b f(x)\mathrm{d}x$

(4) 在 $f(x)$ **连续点**处有 $F'(x) = f(x)$

3. 设随机变量 X 的概率密度为 $f(x)$, 则下列必为概率密度的是

$$(A)\ f(-x+1) \quad (B)\ f(2x-1) \quad (C)\ f(-2x+1) \quad (D)\ f\left(\frac{1}{2}x-1\right)$$

Solution. 由于 $f(x)$ 已经满足非负性, 故选项的非负性都不需要考虑, 只需要考虑正则性就可以.

(A) $\int_{-\infty}^{+\infty} f(-x+1)dx = \int_{-\infty}^{+\infty} f(u)du = 1$

(B) $\int_{-\infty}^{+\infty} f(2x-1)dx = \frac{1}{2}\int_{-\infty}^{+\infty} f(u)du = \frac{1}{2}$

(C) $\int_{-\infty}^{+\infty} f(-2x+1)dx = \frac{1}{2}\int_{-\infty}^{+\infty} f(u)du = \frac{1}{2}$

(D) $\int_{-\infty}^{+\infty} f(\frac{1}{2}-1)dx = 2\int_{-\infty}^{+\infty} f(u)du = 2$ □

总结

$f(ax+b)$ 为概率密度 $\iff |a|=1$

4. (2011, 数一、三) 设 $F_1(x), F_2(x)$ 为分布函数, 对应的概率密度 $f_1(x), f_2(x)$ 为连续函数, 则下列必为概率密度的是

$(A)\ f_1(x)f_2(x)\quad (B)\ 2f_2(x)F_1(x)\quad (C)\ f_1(x)F_2(x)\quad (D)\ f_1(x)F_2(x)+f_2(x)F_1(x)$

总结

(1) 线性组合

$af_1(x)+bf_2(x), a>0, b>0$ 为概率密度 $\iff a+b=1$

$aF_1(x)+bF_2(x), a>0, b>0$ 为分布函数 $\iff a+b=1$

(2) 乘积

F_1F_2 一定是分布函数

f_1f_2 不一定是概率论密度

(3) 混搭

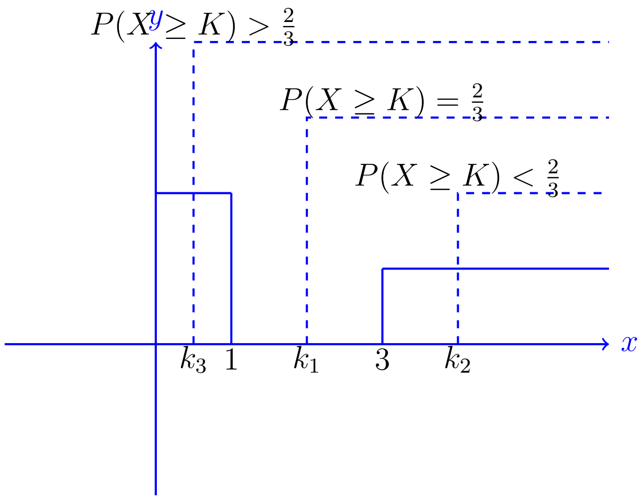
$f_1F_2+f_2F_1, 2f_1F_1, 2f_2F_2$ 是概率密度, 其余都不是.

5. (2000, 三) 设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{3}, & x\in[0,1]\\ \frac{2}{9}, & x\in[3,6]\\ 0, & \text{其他}\end{cases}$$

若 $P\{X\geq k\}=\frac{2}{3}$, 则 k 的取值范围是 ______.

Solution. 如图所示, 当且仅当 $1\leq k\leq 3$ 时候 $P(X\geq k)=\frac{2}{3}$ □



2.3 关于八大分布

Remark. (八大分布的概率分布与数字特征)

(1) 0-1 分布, $X\sim B(1,p)$ $\begin{array}{c|cc} X & 0 & 1 \\ \hline P & 1-p & p \end{array}$, $EX=p, DX=p(1-p)$

(2) 二项分布,$X\sim B(n,p)$

$P\{X=k\}=C_n^k p^k(1-p)^{n-k}, k=0,1,\ldots,n, EX=np, DX=np(1-p)$

(3) 泊松分布,$X\sim P(\lambda)$

$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\ldots,\ EX=\lambda, DX=\lambda$

(4) 几何分布,$X\sim G(p)$

$P=\{X=k\}=p(1-p)^{k-1}, k=1,2,\ldots,\ EX=\frac{1}{p}, DX=\frac{1-p}{p^2}$

(5) 超几何分布,$X \sim H(N, M, n)$

$P=\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k=0,1,\ldots,min(n,M),\ EX=\frac{nM}{N}$

(6) 均匀分布 $X \sim U(a,b)$

$$f(x)=\begin{cases}\frac{1}{b-a}, & a \le x \le b \\ 0, & \text{其他}\end{cases},\ F(x)=\begin{cases}0, & x<a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b\end{cases},\ EX=\frac{a+b}{2}, DX=\frac{(b-a)^2}{12}$$

(7) 指数分布 $X \sim E(\lambda)$

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0 \\ 0, & x \le 0\end{cases},\ \lambda>0\ F(x)=\begin{cases}1-e^{-\lambda x}, & x>0 \\ 0, & x \le 0\end{cases},\ EX=\frac{1}{\lambda}, DX=\frac{1}{\lambda^2}$$

(8) 一般正态分布 $X \sim N(\mu,\sigma^2),\ EX=\mu, DX=\sigma^2$

$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ F(\mu)=\frac{1}{2},\ F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$

标准正态分布 $X \sim N(0,1)\ \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ \Phi(0)=\frac{1}{2},\Phi(-x)=1-\Phi(x)$

正态分布的标准化若 $X \sim N(\mu,\sigma^2)$，则 $\frac{X-\mu}{\sigma} \sim N(0,1)$.

拓展-负二项分布

在一系列独立重复的伯努利试验 (每次试验只有"成功"或"失败"两种结果, 成功概率为 p) 中, 达到 r 次成功所需的试验总次数 X 服从负二项分布。

$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r},\quad k=r,r+1,r+2,\ldots,\quad EX=\frac{r}{p},\quad DX=\frac{r(1-p)}{p^2}$

6. 设随机变量 X 的概率分布为 $P\{X=k\}=C\frac{\lambda^k}{k!},k=1,2,\cdots$, 则 $C=$______.

Solution.

(方法一: 级数) 由概率的规范性可知 $\sum_{k=1}^{\infty} C\frac{\lambda^k}{k!}=1$, 由于 $e^x=\sum_{i=0}^{\infty}\frac{x^n}{n!}$, 故 $C(e^\lambda-1)=1$, 故 $C=\frac{1}{e^\lambda-1}$

(方法二: 泊松分布) 考虑泊松分布 $P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\ldots$ □

7. 设随机变量 X 的概率密度为 $f(x)=Ae^{-\frac{x^2}{2}+Bx}$, 且 $EX=DX$, 则 $A=$___,$B=$___.

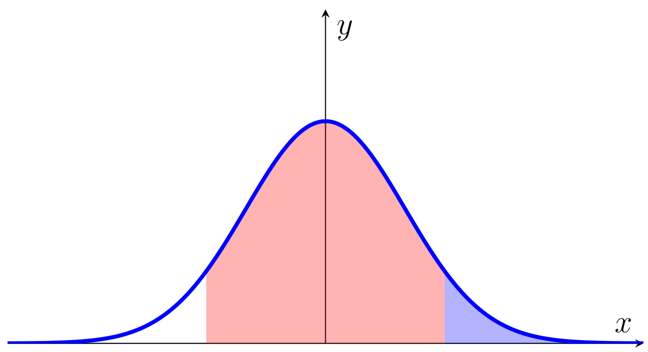
Solution. $f(x)=Ae^{\frac{B^2}{2}}e^{-\frac{(x-B)^2}{2}} \sim N(1,B^2)$, 又 $D(x)=E(x)$ 故 $B^2=1$, 对比正态分布的概率密度函数有 $Ae^{\frac{B^2}{2}}=\frac{1}{\sqrt{2\pi}}$ 故 $A=\frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}$ □

总结

形如 $f(x)=Ae^{ax^2+b+c}, a<0$ 一定可以化成某一个正态分布的概率密度.

8. (2004, 数一、三) 设随机变量 $X\sim N(0,1)$, 对给定的 $\alpha(0<\alpha<1)$, 数 u_α 满足 $P\{X>u_\alpha\}=\alpha$。若 $P\{|X|<x\}=\alpha$, 则 x 等于

$$(A)\ u_{\frac{\alpha}{2}}\quad (B)\ u_{1-\frac{\alpha}{2}}\quad (C)\ u_{\frac{1-\alpha}{2}}\quad (D)\ u_{1-\alpha}$$



Solution. 如图所示,x 右侧的面积为 $\frac{1-\alpha}{2}$ 故 x 是 $\frac{1-\alpha}{2}$ 上侧分位点 □

9. 设随机变量 $X\sim N(2,\sigma^2)$, 且 $P\{2<X<4\}=0.3$, 则 $P\{X<0\}=$______.

Solution. 正态分布的基本套路就是遇事不决标准化 $P\{2<X<4\}=P\{0<\frac{X-2}{\sigma}<\frac{2}{\sigma}\}=0.3$, 故 $P\{X<0\}=P\{\frac{X-2}{\sigma}<\frac{-2}{\sigma}\}=\frac{1}{2}-0.3=0.2$ □

10. 设随机变量 $X\sim N(\mu,\sigma^2)(\mu<0)$,$F(x)$ 为其分布函数,a 为任意常数, 则

$$(A)\ F(a)+F(-a)>1\quad (B)\ F(a)+F(-a)=1$$
$$(C)\ F(a)+F(-a)<1\quad (D)\ F(\mu+a)+F(\mu-a)=\frac{1}{2}$$

Solution. 这道题是比较隐晦的考察了正态分布的对称性, 具体直接看总结. 但要注意先标准化再套结论! □

$$\Phi(a)+\Phi(b)=\begin{cases}1, & a+b=1\\ <1, & a+b<1\\ >1, & a+b>1\end{cases}$$

11. 设随机变量 X 与 Y 相互独立, 均服从参数为 1 的指数分布, 则 $P\{1 < \max\{X,Y\} < 2\} =$______.

Solution.

$$\begin{aligned} P\{1<\max\{X,Y\}<2\} &= P\{\max\{X,Y\}<2\} - P\{\max\{X,Y\}\le 1\} \\ &= P\{X<2,Y<2\} - P\{X\le 1, Y\le 1\} \\ &\xlongequal{\text{由独立性}} P\{X<2\}P\{Y<2\} - P\{X\le 1\}P\{Y\le 1\} \\ &= (1-e^{-2\lambda})^2 - (1-e^{-\lambda})^2 \end{aligned}$$

□

12. 设随机变量 X 与 Y 相互独立, 均服从区间 $[0,3]$ 上的均匀分布, 则 $P\{1 < \min\{X,Y\} < 2\} =$______.

Solution.

$$\begin{aligned} P\{1<\min\{X,Y\}<2\} &= P\{\min\{X,Y\}>1\} - P\{\min\{X,Y\}\ge 2\} \\ &= P\{X>1\}P\{Y>1\} - P\{X\ge 2\}P\{Y\ge 2\} \\ &= \frac{1}{3} \end{aligned}$$

□

总结

对于 min 和 max 问题基本按照如下思路:

$$\begin{aligned} &P\{a<\min(X_1,X_2,\dots,X_n)<b\} \\ &= P\{\min(X_1,X_2,\dots,X_n)>a\} - P\{\min(X_1,X_2,\dots,X_n)\ge b\} \end{aligned}$$

$$\begin{aligned} &P\{a<\max(X_1,X_2,\dots,X_n)<b\} \\ &= P\{\max(X_1,X_2,\dots,X_n)<b\} - P\{\min(X_1,X_2,\dots,X_n)\le a\} \end{aligned}$$

13. (2013, 数一) 设随机变量 $Y \sim E(1)$,$a>0$, 则 $P\{Y\le a+1|Y>a\} =$______.

Solution. 由指数分布的无记忆性，有 $P\{Y \leq a+1|Y > a\} = P\{0 < Y < 1\} = \int_0^1 e^{-x}dx = 1-e^{-1}$ □

14. 设随机变量 $X \sim G(p)$,m,n 为正整数，则 $P\{X > m+n|X > m\}$

(A) 与 m 无关，与 n 有关，且随 n 的增大而减少

(B) 与 m 无关，与 n 有关，且随 n 的增大而增大

(C) 与 n 无关，与 m 有关，且随 m 的增大而减少

(D) 与 n 无关，与 m 有关，且随 m 的增大而增大

Solution. 由几何分布的无记忆性，有 $P\{X > m+n|X > m\} = P\{X > n\} = \sum_{i=n+1}^{\infty} p(1-p)^{i-1}$，故随着 n 增大概率反而减少 □

总结

指数分布与几何分布具有无记忆性

$$X \sim E(\lambda)$$

$$P\{x > s+t \mid x > s\} = P\{x > t\}$$

$$P\{x < s+t \mid x > s\} = P\{0 < x < t\}$$

$$X \sim G(p)$$

$$P\{x > n+m \mid x > m\} = P\{x > t\}$$

$$P\{x = n+m \mid x = m\} = P\{x = n\} = p(1-p)^{n-1}$$

2.4 求一维连续型随机变量函数的分布

Remark. 【方法】

设随机变量 X 的概率密度为 $f_X(x)$, 求 $Y = g(X)$ 的分布.

分布函数法

(1) 设 Y 的分布函数为 $F_Y(y)$, 则 $F_Y(y) = P\{Y \le y\} = P\{g(X) \le y\}$.

(2) 求 $Y = g(X)$ 在 X 的正概率密度区间的值域 (α, β) , 讨论 y .

当 $y < \alpha$ 时, $F_Y(y) = 0$;

当 $\alpha \le y < \beta$ 时, $F_Y(y) = \int_{g(x) \le y} f_X(x)\, dx$;

当 $y \ge \beta$ 时, $F_Y(y) = 1$.

(3) 若 Y 为连续型随机变量, 则 Y 的概率密度为 $f_Y(y) = F'_Y(y)$.

公式法

设 $y = g(x)$ 在 X 的正概率密度区间单调, 值域为 (α, β) , 反函数为 $x = h(y)$, 则 Y 的概率密度为

$$f_Y(y) = \begin{cases} f_X(h(y))\,|h'(y)|, \alpha < y < \beta \\ \qquad\qquad 0, \end{cases}$$

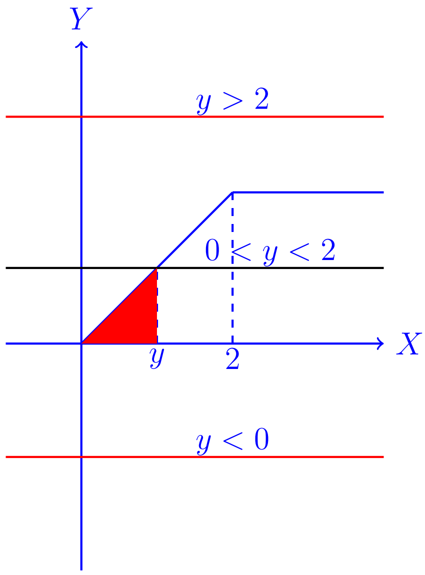
若 $y = g(x)$ 在 X 的正概率密度区间 $[a, b]$ 分段严格单调, 则分段运用公式法, 然后将概率密度相加.

15. 设随机变量 $X \sim E(\lambda)$, 则 $Y = \min\{X, 2\}$ 的分布函数

(A) 为连续函数 (B) 为阶梯函数

(C) 至少有两个间断点 (D) 恰好有一个间断点

Solution. 这是一道比较简单的题目, 主要是用于演示所谓**图像法讨论 y** 的具体操作, 注意画的是 $X - Y$ 图像



故 $F_Y(y)=\min\{X,2\}<y$, 当 $y<0$ 时候 $F_Y(y)=0$,$y\geq 2$,$F_Y(y)=1$, 当 $0\leq y<2$ 时候, 有 $\int_0^y f(x)dx=1-e^{-\lambda y}$, 综上

$$F_Y=\begin{cases}0, & y<0\\ 1-e^{-\lambda y}, & 0\leq y<2\\ 1, & y\geq 2\end{cases}$$

容易发现 $F(2-0)\neq 1$ 故存在一个跳跃间断点 □

16. 设随机变量 X 的概率密度为 $f(x)=\begin{cases}\frac{x^2}{a}, & 0<x<3\\ 0, & \text{其他}\end{cases}$ $Y=\begin{cases}2, & X\leq 1\\ X, & 1<X<2\\ 1, & X\geq 2\end{cases}$

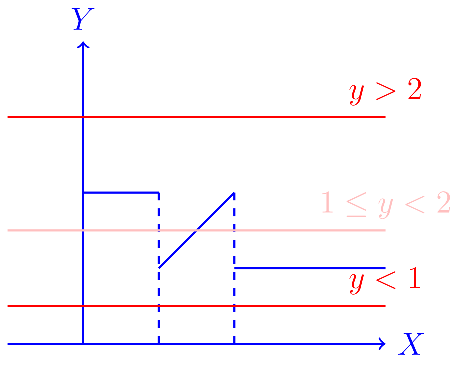
(a) 求 Y 的分布函数;

(b) 求 $P\{X\leq Y\}$.

Solution. 带参数的概率密度第一步就应该根据正则性把这个参数求出来.

$$\int_0^3 f(x)dx=1 \implies a=9$$

然后和上一题一样画 $X-Y$ 图像, 求 $F_Y(y)$, 注意分区域就是.



当 $y<1, F_Y(y)=0; y>2, F_Y(y)=1$

$1 \le y < 2, F_Y(y)=\int_1^y f(x)dx+\int_2^3 f(x)dx=\frac{y^3}{27}+\frac{2}{3}$ □

17. (2021, 数一、三) 在区间 $(0,2)$ 上随机取一点, 将该区间分成两段, 较短一段的长度记为 X, 较长一段的长度记为 Y。

(a) 求 X 的概率密度;

(b) 求 $Z=\frac{Y}{X}$ 的概率密度;

(c) 求 $E\left(\frac{Y}{X}\right)$.

Solution. 有题设容易得到 $X \sim U(0,1), Y=2-X$

(1) 则 $f(x)=\begin{cases} 1, & x\in(0,1) \\ 0, & 其他 \end{cases}$

(2) $Z=\frac{Y}{X}=\frac{2}{X}-1$, 显然 Z 关于 X 是单调的, 可以用公式法直接求出 $f_Z(z)$, 即

$$f_Z(z)=1\cdot\frac{2}{(y+1)^2}=\frac{2}{(y+1)^2}, z\in(1,+\infty)$$

(3)

$$E(Z)=\int_1^{\infty} zf_Z(z)dz=2\ln 2-1$$

或者也可以用

$$E(\frac{2}{x}-1)=\int_0^1(\frac{2}{x}-1)dx=2\ln(2)-1$$

□

第三章　二维随机变量

3.1　联合分布函数的计算

Remark. (联合分布函数的性质)

(1) $0 \le F(x,y) \le 1, -\infty < x < +\infty, F(-\infty,y) = F(x,-\infty) = F(-\infty,-\infty) = 0, F(+\infty,+\infty) = 1$

(2) $F(x,y)$ 关于 x 和 y 均单调不减

(2) $F(x,y)$ 关于 x 和 y 均右连续

(4) $P\{a < X \le b, c < Y \le b\} = F(b,d) - F(b,c) - F(a,d) + F(a,c)$

1. 设随机变量 X 与 Y 相互独立,$X \sim B(1,p)$,$Y \sim E(\lambda)$, 则 (X,Y) 的联合分布函数 $F(x,y) =$_____.

Solution. 由 X 和 Y 相互独立, 则有 $F_{XY}(x,y) = F_X(x)F_Y(y), f(x,y) = f_X(x)F_Y(x)$,$X$ 的概率分布如下:

X	0	1
P	$1-p$	p

则 X 的分布函数为

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

而 $Y \sim E(\lambda)$, 故

$$F_{XY}(x,y) = F_X(x)F_Y(y) = \begin{cases} (1-p)(1-e^{-\lambda y}), & 0 \le x < 1,\ y > 0 \\ 1-e^{-\lambda y}, & x \ge 1,\ y > 0 \\ 0, & \text{其他} \end{cases}$$

□

3.2 二维离散型随机变量分布的计算

2. 设随机变量 X 与 Y 相互独立, 均服从参数为 p 的几何分布。

   (a) 求在 $X+Y=n(n\ge 2)$ 的条件下,X 的条件概率分布;

   (b) 求 $P\{X+Y\ge n\}(n\ge 2)$.

Solution.

(1)

$$\begin{aligned} P\{X+Y=n\} &\xlongequal{\text{几何分布从 1 开始}} \sum_{k=1}^{n-1} P\{X=k, Y=n-k\} \\ &\xlongequal{\text{独立性}} \sum_{k=1}^{n-1} P\{X=k\}P\{Y=n-k\} \\ &= \sum_{k=1}^{n-1} (1-p)^{k-1}p\cdot(1-p)^{n-k-1}p \\ &= \sum_{k=1}^{n-1} (1-p)^{n-2}p^2 \\ &= (n-1)(1-p)^{n-2}p^2 \end{aligned}$$

在 $X+Y=n$ 的条件下,X 的条件概率为

$$\begin{aligned} P\{X=k \mid X+Y=n\} &= \frac{P\{X=k, Y=n-k\}}{P\{X+Y=n\}} \\ &= \frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}} \\ &= \frac{1}{n-2} \end{aligned}$$

$k=1,2\ldots n-1$这个范围千万别忘喽!

(2)

$$
\begin{aligned}
P\{X+Y\geq n\} &= P\{X+Y=n\}+P\{X+Y=n+1\}+\dots \\
&= \sum_{k=n}^{+\infty} P\{X+Y=k\} \\
&= \sum_{k=n}^{+\infty}(k-1)p^2(1-p)^{k-2}
\end{aligned}
$$

不妨先计算级数 $\sum_{k=n}^{\infty}(k-1)x^{k-2}$

$$
\begin{aligned}
\sum_{k=n}^{\infty}(k-1)x^{k-2} &= \sum_{k=n}^{\infty}(x^{k-1})' \\
&= \left(\frac{\sum_{n=k}^{\infty}}{x}\right)' \\
&= \frac{(n-1)x^{n-2}(1-x)+x^{n-1}}{(1-x)^2}
\end{aligned}
$$

故当 $x=1-p$ 的时有

$$
\begin{aligned}
P\{X+Y\geq n\} &= p^2\frac{(n-1)(1-p)^{n-2}p+(1-p)^{n-1}}{p^2} \\
&= (1-p)^{n-2}(np-2p+1)
\end{aligned}
$$

□

3.3 二维连续型随机变量分布的计算

Remark. 主要内容

联合概率密度的性质

(1) $f(x,y)\geq 0, -\infty<x<+\infty, -\infty<y<+\infty$；

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

(3) $P\{(X,Y)\in D\}=\iint_D f(x,y)\,dxdy$；

(4) 在 $f(x,y)$ 的连续点处有 $\frac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y)$．

边缘概率密度

(1) (X,Y) 关于 X 的边缘概率密度 $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy$

(2) (X,Y) 关于 Y 的边缘概率密度 $f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,dx$

条件概率密度

(1) 在 $Y=y$ 的条件下, X 的条件概率密度 $f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}$

(2) 在 $X=x$ 的条件下, Y 的条件概率密度 $f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}$

3. (2010, 数一、三) 设二维随机变量 (X,Y) 的概率密度为

$$f(x,y)=Ae^{-2x^2+2xy-y^2},\quad -\infty<x<+\infty,-\infty<y<+\infty$$

求常数 A 及条件概率密度 $f_{Y|X}(y|x)$.

Solution.

(方法一正常求) 首先通过规范性求出参数 A

$$\begin{aligned}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} Ae^{-2x^2+2xy-y^2}\mathrm{d}x\mathrm{d}y\\ &= A\int_{-\infty}^{+\infty} e^{-x^2}\mathrm{d}x\int_{-\infty}^{+\infty} e^{-(y-x)^2}\mathrm{d}y\\ &\xlongequal{\text{Possion 积分}} A\pi=1 \implies A=\frac{1}{\pi}\end{aligned}$$

X 的边缘分布函数为

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x,y)dy &= \int_{-\infty}^{+\infty}\frac{1}{\pi}e^{-2x^2+2xy-y^2}\mathrm{d}y\\ &= \frac{1}{\pi}e^{-x^2}\int_{-\infty}^{+\infty} e^{-(y-x)^2}\\ &= \frac{1}{\sqrt{\pi}}e^{-x^2}, x\in\mathbf{R}\end{aligned}$$

则在 $X=x$ 的条件下,Y 的条件概率为

$$\begin{aligned}f_{Y|X}(y\mid x) &= \frac{f(x,y)}{f_X(x)}\\ &= \frac{1}{\sqrt{\pi}}e^{-(y-x)^2}\end{aligned}$$

(方法二, 通过二维正态分布) 形如 $f(x,y)=Ae^{ax^2+bxy+cy^2}$ 的函数如果是概率密度, 则其一定是某个二维正态的概率密度函数, 故

$$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$$

通过下一节讲的确定系数的办法, 可以很快的确定

$$(X,Y)\sim N(0,0;\frac{1}{2},1;\frac{\sqrt{2}}{2})$$

故 $A=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}=\frac{1}{\pi}, f_X(x)=\frac{1}{\sqrt{\pi}}e^{-x^2}$ □

4. 设随机变量 $X\sim U(0,1)$, 在 $X=x(0<x<1)$ 的条件下, 随机变量 $Y\sim U(x,1)$。

   (a) 求 (X,Y) 的联合概率密度;

   (b) 求 (X,Y) 关于 Y 的边缘概率密度 $f_Y(y)$;

   (c) 求 $P\{X+Y>1\}$.

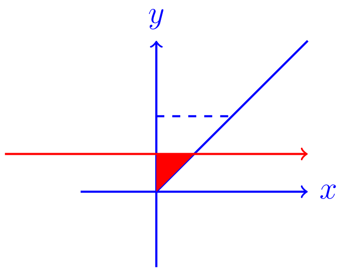
Solution.

(1) 在 $X=x$ 的条件下,Y 的条件概率密度为

$$f_Y(y)=\begin{cases}\frac{1}{1-x}, & x\le y\le 1\\ 0, & \text{其他}\end{cases}$$

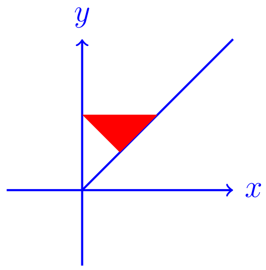
故 $f(x,y)=f_{Y|X}(y\mid x)f_X(x)=\begin{cases}\frac{1}{1-x}, & 0<x<1, x<y<1\\ 0, & \text{其他}\end{cases}$

(2) 通过概率密度求边缘密度的时候, 需要画出 x-y 图, 并且确定要求的那个参数的范围, 比如说这里是 $y\in(0,1)$, 让后再从 [0,1] 上面去做偏积分, 具体如图所示



$f_Y(y)=\int_{+\infty}^{-\infty}f(x,y)\mathrm{d}x=\begin{cases}-\ln(1-y), & 0<y<1\\ 0, & \text{其他}\end{cases}$

(3) 根据性质 (3) 有 $P\{X+Y>1\}=\iint\limits_{x+y<1}f(x,y)\mathrm{d}x\mathrm{d}y$ 此时 x-y 的可行范围为



$$
\begin{aligned}
\text{原式} &= \int_{1/2}^{1} \mathrm{d}y \int_{1-y}^{y} \frac{1}{1-x} \mathrm{d}x \\
&= \int_{1/2}^{1} [\ln y - \ln(1-y)]\, \mathrm{d}y \\
&= [y \ln y - (1-y)\ln(1-y)]\Big|_{1/2}^{1} \\
&= \ln 2
\end{aligned}
$$

□

3.4 关于二维正态分布

Remark. 二维正态分布的性质　设 $(X,Y) \sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 则

(1) $X \sim N\left(\mu_1,\sigma_1^2\right), Y \sim N\left(\mu_2,\sigma_2^2\right)$, 反之不成立 (独立的时候反之成立);

(2) X 与 Y 相互独立 $\Leftrightarrow$ X 与 Y 不相关 $(\rho=0)$;

(3) $aX+bY \sim N\left(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2+2ab\rho\sigma_1\sigma_2\right)$; 特别地, 若 X 与 Y 相互独立, $X \sim N\left(\mu_1,\sigma_1^2\right), Y \sim N\left(\mu_2,\sigma_2^2\right)$, 则 $aX+bY \sim N\left(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2\right)$;

(4) 若 $U=aX+bY, V=cX+dY$, 即 $\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$, 则 (U,V) 服从二维正态分布 $\Leftrightarrow \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$.

5. 设二维随机变量 $(X,Y) \sim N(1,2;1,4;-\frac{1}{2})$, 且 $P\{aX+bY \leq 1\} = \frac{1}{2}$, 则 (a,b) 可以为

$$
(A)\ \left(\frac{1}{2},-\frac{1}{4}\right) \quad (B)\ \left(\frac{1}{4},-\frac{1}{2}\right) \quad (C)\ \left(-\frac{1}{4},\frac{1}{2}\right) \quad (D)\ \left(\frac{1}{2},\frac{1}{4}\right)
$$

Solution. 由性质 (3) 可知 $aX+bY \sim N$, 而由正态分布的对称性可知,$\mu=1 \Longrightarrow a+2b=1$ 故选择 (D) □

6. (2020, 数三) 设二维随机变量 $(X,Y)\sim N(0,0;1,4;-\frac{1}{2})$, 则下列随机变量服从标准正态分布且与 X 相互独立的是

$$(A)\ \frac{\sqrt{5}}{5}(X+Y)\quad (B)\ \frac{\sqrt{5}}{5}(X-Y)\ (C)\ \frac{\sqrt{3}}{3}(X+Y)\quad (D)\ \frac{\sqrt{3}}{3}(X-Y)$$

Solution. 这道题选择出来并不困难, 但要证明其与 X 相互独立还是有点说法的.

第一步, 先求 $X+Y$ 和 $X-Y$ 的标准化

由性质三可知 $X+Y\sim N(0,3), X-Y\sim N(0,7)$, 故 $\frac{\sqrt{3}}{3}(X+Y)\sin N(0,1)$; $\frac{\sqrt{7}}{7}\sim N(0,1)$;

这里其时就已经可以选出答案喽

第二步证明独立性

考虑 $(X+Y,X)=\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}X\\ Y\end{pmatrix}$, 且 $\begin{vmatrix}1 & 1\\ 1 & 0\end{vmatrix}=-1\neq 0$

由性质 (4) 可知,$(X+Y,X)$ 服从二维正态分布, 由性质 (2) 可知, 只需要证明二者的相关系数为 0 即可, 证明二者独立.

$\square$

7. (2022, 数一) 设随机变量 $X\sim N(0,1)$, 在 $X=x$ 的条件下, 随机变量 $Y\sim N(x,1)$, 则 X 与 Y 的相关系数为

$$(A)\ \frac{1}{4}\quad (B)\ \frac{1}{2}\quad (C)\ \frac{\sqrt{3}}{3}\quad (D)\ \frac{\sqrt{2}}{2}$$

Solution.

(方法一传统方法计算)

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}=\frac{EXY-EXEY}{\sqrt{DX}\sqrt{DY}}$$

问题转换为求 EXY,DY, 由题设可知, 在 $X=x$ 的条件下,Y 的概率密度函数为

$$f_{Y|X}(y\mid x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-x)^2}{2}}$$

故 (X,Y) 的概率密度函数为

$$f(x,y)=\frac{1}{2\pi}e^{-x^2+xy-\frac{y^2}{2}}$$

故 y 的边缘分布函数为

$$\int_{+\infty}^{-\infty}f(x,y)dx=\frac{1}{2\sqrt{\pi}}e^{-\frac{y^2}{4}}$$

即 $Y \sim N(0,2)$, 故 $EY=0, DY=2$ 而 EXY 根据方差的定义可以计算

TODO: 计算 EXY

$$EXY = \int_{+\infty}^{-\infty}\int_{+\infty}^{-\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y = 1$$

故 $\rho = \frac{\sqrt{2}}{2}$

(2) 通过二维正态参数的结论直接求出 ρ, 由上述可知 $f(x,y) = \frac{1}{2\pi}e^{-x^2+xy-\frac{y^2}{2}}$, 对比二维正态概率密度的公式

$$f(x,y) = \frac{1}{2\sigma_1\sigma_2\sqrt{1-\rho^2}}exp\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)}{\sigma_2^2}\right]\}$$

容易得出 $(X,Y) \sim N(0,0;1,2;\frac{\sqrt{2}}{2})$, 具体如总结所示. □

总结

对于形如 $Ae^{-ax^2+bxy+cy^2}$ 的式子, 若其是概率密度, 则必然是某个二维正态的概率密度 (由规范性) 且满足

(1) $b^2 = 4\rho^2a^2c^2 \implies \rho^2 = \frac{b^2}{4a^2c^2}$

(2) rho 的符号与 xy 系数的符号一致

3.5 求二维离散型随机变量函数的分布

8. 设随机变量 X 与 Y 相互独立,$X \sim P(\lambda_1)$,$Y \sim P(\lambda_2)$, 求 $Z = X + Y$ 的概率分布.

Solution. 这道题是参数可加性的直接考察, 可以先证明一下

$$
\begin{aligned}
P\{Z=n\} &= P\{X+Y=n\} \\
&= \sum_{k=0}^{n} P\{X=k, Y=n-k\} \\
&\xlongequal{\text{独立性}} \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\} \\
&= \sum_{k=0}^{n} \frac{\lambda_1^k}{k!}e^{-\lambda_1}\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2} \\
&= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} \\
&\xlongequal{\text{上下同乘}k!} e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{n(n-1)\dots(n-k+1)}{k!}\lambda_1^k\lambda_2^{n-k} \\
&= e^{-(\lambda_1+\lambda_2)}\frac{1}{n!}\sum_{k=0}^{n}C_n^k\lambda_1^k\lambda_2^{n-k} \\
&\xlongequal{\text{二项式定理}} \frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}
\end{aligned}
$$

□

参数可加性

当 X, Y 独立的时候

(1) $X \sim B(m,p), Y \sim B(n,p) \implies X+Y \sim B(n+m,p)$

(2) $X \sim P(\lambda_1), Y \sim P(\lambda_2) \implies X+Y \sim P(\lambda_1+\lambda_2)$

(3) $X \sim N(\mu_1,\sigma_1^2), Y \sim N(\mu_2,\sigma_2^2) \implies X+Y \sim (\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

(4) $X \sim \chi^2(m), Y \sim \chi^2(n), \implies X+Y \sim \chi^2(n+m)$

(5) $X \sim E(\lambda_1), Y \sim E(\lambda_2) \implies \min(X,Y) \sim E(\lambda_1+\lambda_2)$

3.6 求二维连续型随机变量函数的分布

Remark. 问题描述

设二维随机变量 (X,Y) 的联合概率密度为 $f(x,y)$, 求 $Z=g(X,Y)$ 的概率密度 $f_Z(z)$.

分布函数法

(1) 设 Z 的分布函数为 $F_Z(z)$, 则 $F_Z(z)=P\{Z\le z\}=P\{g(X,Y)\le z\}$.

(2) 求 $Z=g(X,Y)$ 在 (X,Y) 的正概率密度区域的值域 (α,β) , 讨论 z .

$z<\alpha$ 时, $F_Z(z)=0$;

当 $\alpha\le z<\beta$ 时, $F_Z(z)=\iint_{g(x,y)\le z} f(x,y)\,dxdy;$

当 $z\ge\beta$ 时, $F_Z(z)=1$.

(3) Z 的概率密度为 $f_Z(z)=F_Z'(z)$.

卷积公式

(1) 设 $Z=aX+bY$, 则 $f_Z(z)=\int_{-\infty}^{+\infty}\frac{1}{|b|}f\left(x,\frac{z-ax}{b}\right)dx=\int_{-\infty}^{+\infty}\frac{1}{|a|}f\left(\frac{z-by}{a},y\right)dy;$

(2) 设 $Z=XY$, 则 $f_Z(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f\left(x,\frac{z}{x}\right)dx=\int_{-\infty}^{+\infty}\frac{1}{|y|}f\left(\frac{z}{y},y\right)dy$;

(3) 设 $Z=\frac{Y}{X}$, 则 $f_Z(z)=\int_{-\infty}^{+\infty}|x|\,f(x,xz)\,dx;$

(4) 设 $Z=\frac{X}{Y}$, 则 $f_Z(z)=\int_{-\infty}^{+\infty}|y|\,f(yz,y)\,dy$

9. 设二维随机变量 (X,Y) 的联合概率密度为 $f(x,y)=\begin{cases}1, & 0<x<1, 0<y<2x\\ 0, & 其他\end{cases}$ 求:

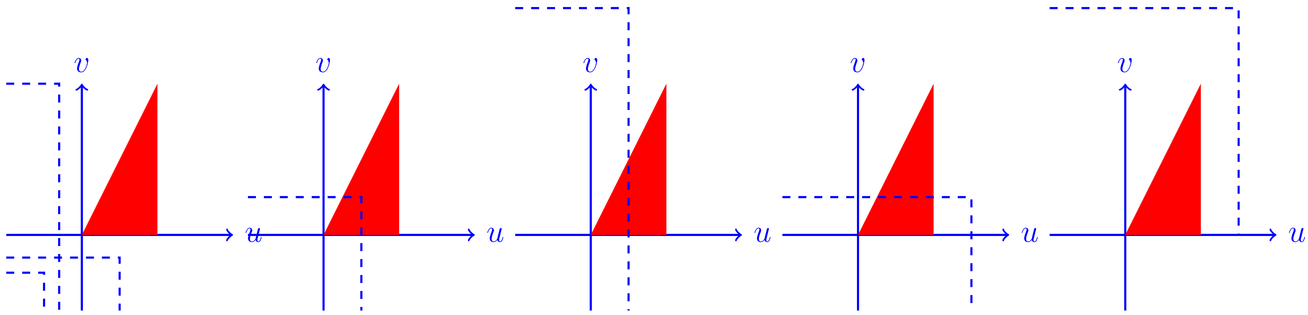
(a) (X,Y) 的联合分布函数 $F(x,y)$;

(b) (X,Y) 的边缘概率密度 $f_X(x), f_Y(y)$;

(c) 条件概率密度 $f_{X|Y}(x|y), f_{Y|X}(y|x)$;

(d) $P\left\{Y\le\frac{1}{2}|X\le\frac{1}{2}\right\}$,$P\left\{Y\le\frac{1}{2}|X=\frac{1}{2}\right\}$;

(e) $Z=2X-Y$ 的概率密度 $f_Z(z)$.



Solution.

(1) 由定义可知 $F(x,y)=\int_{-\infty} x\int_{-\infty} y f(u,v)\mathrm{d}u\mathrm{d}v$, 其中 x,y 的可行域如下图所示, 分为五个部分故

$$F(x,y)=\begin{cases}\int_0^y \mathrm{d}v\int_{\frac{v}{2}}^x \mathrm{d}u, & 0<x<1,0<y<2x\\ \int_0^x \mathrm{d}u\int_0^{2u}\mathrm{d}v, & 0<x<1,y\geq 2x\\ \int_0^y \mathrm{d}v\int_{\frac{v}{2}}^1 \mathrm{d}u, & x>1,0<y<2\\ 1, & x\geq 1,y\geq 2x\\ 0, & \text{其他}\end{cases}=\begin{cases}\frac{y^2}{4}-xy, & 0<x<1,0<y<2x\\ x^2, & 0<x<1,y\geq 2x\\ y-\frac{y^2}{4}, & x>1,0<y<2\\ 1, & x\geq 1,y\geq 2x\\ 0, & \text{其他}\end{cases}$$

(2) 由定义可知

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\begin{cases}2x, & 0<x<1;\\ 0, & \text{其他}\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\begin{cases}1-\frac{y}{2}, & 0<y<2\\ 0, & \text{其他}\end{cases}$$

(3) 当0<x<1在 $X=x$ 的条件下,Y 的条件概率密度为

$$f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}=\begin{cases}\frac{1}{2x}, & 0<y<2x\\ 0, & \text{其他}\end{cases}$$

当0<y<2在 $Y=y$ 的条件下,X 的条件概率密度为

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}\frac{2}{2-y}, & \frac{y}{2}<x<1\\ 0, & \text{其他}\end{cases}$$

(4) 对于 $P\left\{Y \le \frac{1}{2}|X \le \frac{1}{2}\right\}$ 可以采用条件概率公式,

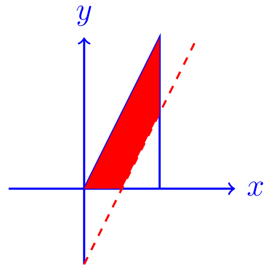
$$P\left\{Y \le \frac{1}{2}|X \le \frac{1}{2}\right\} = \frac{\iint\limits_{y\le\frac{1}{2},x\le\frac{1}{2}} f(x,y)\mathrm{d}x\mathrm{d}y}{\int_0^{\frac{1}{2}} f_X(x)\mathrm{d}x} = \frac{3}{4}$$

而对于 $P\left\{Y \le \frac{1}{2}|X = \frac{1}{2}\right\}$ 则不能采用条件概率公式, 因为 $P\{X = \frac{1}{2}\} = 0$ 不能做分母, 此时就体现出来条件概率的用处

$$P\left\{Y \le \frac{1}{2}|X = \frac{1}{2}\right\} = \int_0^{\frac{1}{2}} f_{Y|X}(y \mid x)\mathrm{d}y$$

将 $X = \frac{1}{2}$ 带入, 求出该条件概率为 $\frac{1}{2}$

(5) 方法一: 分布函数法



$F_Z(z) = P\{2X - Y \ge Z\} = \iint\limits_{2x-y\le z} f(x,y)\mathrm{d}x\mathrm{d}y$, 绘制 $y \ge 2x - z$, 讨论截距, 如图所示, 其结果如下

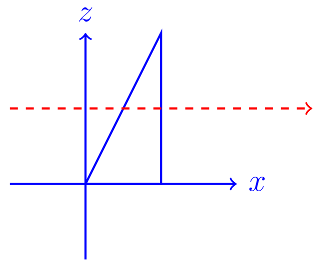
$$F_Z(z) = \begin{cases} 0, & z < 0 \\ z - \frac{z^2}{4}, & 0 \le z < 2 \\ 1, & z \ge 2 \end{cases}$$

方法二: 卷积公式

由卷积公式有 $f_Z(z) = -\int_{-\infty}^{+\infty} f(x, 2x - z)dx$, 此时把 $f(x,y)$ 中的 y 全部转换为 z 并确定 z 的取值范围即

$$f(x, 2x - z) = \begin{cases} 1, & 0 < x < 1, 0 < 2x - z < 2x \implies, 0 < x < 1, 0 < z < 2x \\ 0, & \text{其他} \end{cases}$$

此时再对 x 进行偏积分即可, 绘制 $x - z$ 图像, 首先确认 z 的范围, 再从 z 上对 x 进行积分



如图, 最终

$$f_Z(z) = \begin{cases} 1 - \frac{z}{2}, & 0 \le z < 2; \\ 0, & \end{cases}$$

□

3.7　求一离散一连续随机变量函数的分布

10. (2020, 数一) 设随机变量 X_1, X_2, X_3 相互独立,X_1 与 X_2 均服从标准正态分布,X_3 的概率分布为 $P\{X_3 = 0\} = P\{X_3 = 1\} = \frac{1}{2}$,$Y = X_3X_1 + (1 - X_3)X_2$。

(1) 求 (X_1, Y) 的联合分布函数 (结果用标准正态分布函数 $\Phi(x)$ 表示);

(2) 证明 Y 服从标准正态分布.

Solution. 一离散加一连续的基本方法就是" 全概率公式 + 独立性"

(1)

$$\begin{aligned} F(X_1, Y) &= P\{X \le x, Y \le y\} \\ &= P\{X_1 \le x, X_3X_1 + (1 - X_3)X_2 \le y\} \\ &\xlongequal{\text{全概率公式}} P\{X_1 \le x, X_2 \le y, X_3 = 0\} + P\{X_1 \le x, X_1 \le y, X_3 = 1\} \\ &\xlongequal{\text{独立性}} P\{X_1 \le x\}P\{X_2 \le y\}\frac{1}{2} + \frac{1}{2}P\{X_1 \le \min(x, y)\} \\ &= \frac{1}{2}\Phi(x)\Phi(y) + \frac{1}{2}\Phi(\min(x, y)) \end{aligned}$$

(2) 方法一, 通过 Y 的分布函数确定

$$\begin{aligned} F_Y(y) = P\{Y \le y\} &= P\{X_3X_1 + (1 - X_3)X_2 \le y\} \\ &= (\text{和 (1) 完全一致省去})\ldots \\ &= \Phi(y) \end{aligned}$$

方法二, 直接求边缘分布函数

$$
\begin{aligned}
&F_X(x) = P\{X \leq x\} = F(X, +\infty) \\
&F_Y(y) = P\{Y \leq y\} = F(+\infty, Y) \\
&F_Y(y) = F(\infty, y) = \frac{1}{2}\Phi(y) + \frac{1}{2}\Phi(y) = \Phi(y)
\end{aligned}
$$

故 $Y \sim N(0,1)$ □

第四章 数字特征

4.1 期望与方差的计算

Remark. 期望与方差

期望的定义

(1) 设随机变量 X 的概率分布为 $P\{X=x_i\}=p_i, i=1,2,\ldots,$ 则 $EX=\sum_i x_i p_i$

推广: 若 $Y=g(X)$ 则 $EY=\sum_i g(x_i)p_i$

(2) 设随机变量 X 的概率密度为 $f(x)$ 则 $EX=\int_{-\infty}^{+\infty} f(x)\mathrm{d}x$

推广: 若 $Y=g(X)$ 则 $EY=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$

(3) 设二维随机变量 (X,Y) 的联合概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij}, i,j=1,2,\ldots$ 则

$EZ=\sum_i\sum_j g(x_i,y_j)p_{ij}$

(4) 设二维随机变量 (X,Y) 的联合概率密度为 $f(x,y), Z=g(X,Y)$ 则

$EZ=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$

特别的 $EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)\mathrm{d}x\mathrm{d}y$, $EY=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)\mathrm{d}x\mathrm{d}y$

期望的性质

(1) $E(aX+bY+c)=aE(X)+bE(Y)+c$

(2) $EXY=EX\cdot EY \iff$ X 与 Y 不相关

特别的若 X 与 Y 相互独立, 由 $EXY=EXEY$

方差的定义

(1) $DX=E(X-EX)^2=EX^2-(EX)^2$

方差的性质

(1) $D(aX+c)=a^2DX$

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

推论 $D(X\pm Y)=D(X)+D(Y) \iff$ X 与 Y 不相关

特别的, 若 X 与 Y 独立, 则有 $D(X\pm Y)=D(X)+D(Y)$

(3) 若 X 与 Y 独立, 则 $DXY=DXDY+(EX)^2DY+(EY)^2DX$

1. 设随机变量 X 的概率密度为 $f(x)=\frac{1}{\pi(1+x^2)}$,$-\infty<x<\infty$, 则 $E[\min\{|X|,1\}]=$______.

Solution.

$$
\begin{aligned}
E\left[\min\left(|X|,1\right)\right] &= \int_{-\infty}^{+\infty}\min\left(|x|,1\right)f(x)\mathrm{d}x \\
&= 2\int_{0}^{+\infty}\min\left(|x|,1\right)f(x)\mathrm{d}x \\
&= 2(\int_{0}^{1}xf(x)\mathrm{d}x+\int_{1}^{+\infty}f(x)\mathrm{d}x) \\
&= \frac{1}{\pi}\ln\left(1+x^2\right)|_0^1+\frac{2}{\pi}\arctan x\ |_1^{+\infty} \\
&= \frac{1}{\pi}\ln 2+\frac{1}{2}
\end{aligned}
$$

□

2. (2016, 数三) 设随机变量 X 与 Y 相互独立,$X\sim N(1,2)$,$Y\sim N(1,4)$, 则 $D(XY)=$

(A) 6 (B) 8 (C) 14 (D) 15

Solution.

(方法一) 通过计算方法做

$$
\begin{aligned}
DXY &= E(XY)^2-(EXY)^2 \\
&= EX^2\cdot EY^2-(EXEY)^2 \\
&= [DX+(EX)^2][DY+(EY)^2]-(EXEY)^2 \\
&= 3\times 5-1=14
\end{aligned}
$$

(方法二) 用结论

$$
\begin{aligned}
DXY &= DXDY+(EX)^2DY+(EY)^2DX \\
&= 8+4+2=14
\end{aligned}
$$

□

3. 设随机变量 X 与 Y 同分布, 则 $E\left(\frac{X^2}{X^2+Y^2}\right)=$______

Solution. 由轮换对称性有

$$E\left(\frac{X^2}{X^2+Y^2}\right)=E\left(\frac{Y^2}{X^2+Y^2}\right)=\frac{1}{2}E\left(\frac{X^2+Y^2}{X^2+Y^2}\right)=\frac{1}{2}$$

□

总结

若 X,Y 同分布, 则 X,Y 具有相同的 F,f,E,D, 上题的推广结论

$$\text{若}X_1,X_2\ldots,X_n\text{同分布, 则}E\left(\frac{X_1^2}{X_1^2+\ldots+X_n^2}\right)=\frac{1}{n}$$

4. 设随机变量 X 与 Y 相互独立,$X\sim P(\lambda_1)$,$Y\sim P(\lambda_2)$, 且 $P\{X+Y>0\}=1-e^{-1}$, 则 $E(X+Y)^2=$______.

Solution. 利用参数可加性可知,$X+Y\sim P(\lambda_1+\lambda_2)$, 由 $P\{X+Y>0\}=1-e^{-1}=1-P\{X=0\}\implies\lambda_1+\lambda_2=1$, 则 $E(X+Y)^2=D(X+Y)+(E(X+Y))^2=1+1=2$ □

5. 设随机变量 X 与 Y 相互独立,$X\sim E(\frac{1}{3})$, $Y\sim E\left(\frac{1}{6}\right)$, 若 $U=\max\{X,Y\}$, $V=\min\{X,Y\}$, 则 $EU=$______,$EV=$______.

Solution. EV 是比较好求的, 由参数可加性有 $V\sim E(\frac{1}{2})$

方法一利用二维概率密度计算:

由 X,Y 独立, 知 $f(x,y)=f_X(x)f_Y(y)$, 则

$$EU=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\max(x,y)f(x,y)\mathrm{d}x\mathrm{d}y=\ldots=7$$

方法二求 U 的概率密度:

由 $U=\max(X,Y)$ 知 $F_U(u)=F_1F_2\implies f_u=f_1F_2+F_1f_2$

$$EU=\int_{-\infty}^{+\infty}uf_u\mathrm{d}u=\ldots=7$$

方法三利用性质

$$E(U+V)=E(X+Y)=EX+EY=3+6=9$$

$EV=2\implies EU=7$ □

总结

若 $U=\max\{X,Y\}$,$V=\min\{X,Y\}$, 则 $E(U+V)=E(X+Y),E(UV)=E(XY)$

独立同分布随机变量的最大值与最小值的分布函数, 由如下结果

令 $Z=\max\left(X_1,X_2,\ldots,X_n\right)$, 则

$$F_Zz=F_{X_1}F_{X_2}\ldots F_{X_n}$$

令 $Z=\min\left(X_1,X_2,\ldots,X_n\right)$, 则

$$F_Zz=1-[(1-F_{(X_2)})][(1-F_{(X_2)})]\ldots[(1-F_{(X_n)})]$$

6. (2017, 数一) 设随机变量 X 的分布函数为 $F(x)=0.5\Phi(x)+0.5\Phi\left(\frac{x-4}{2}\right)$, 其中 $\Phi(x)$ 为标准正态分布函数, 则 $EX=$______

Solution.

(方法一) $f(x)=\frac{1}{2}\phi(x)+\frac{1}{2}\phi(\frac{x-4}{2})$, 则 $EX=\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=2$

(方法二) 考虑 $F(X_1)=0.5\Phi(x),F(X_2)=0.5\Phi(\frac{x-4}{2})$, 则由第二章的结论 $aF_1+bF_2,(a,b>0,a+b=1)$ 的时候也是分布函数, 故 $EX=\frac{1}{2}EX_1+\frac{1}{2}EX_2=0+\frac{4}{2}=2$ □

7. 设随机变量 $X\sim N(0,1)$, 则 $E|X|=$______,$D|X|=$______.

Solution.

$$\begin{aligned}E|X|&=\int_{-\infty}^{+\infty}|x|\phi(x)\mathrm{d}x\\&=2\int_0^{+\infty}x\phi(x)\mathrm{d}x\\&=\frac{-2}{\sqrt{2\pi}}\int_0^{+\infty}e^{-\frac{x^2}{2}}\mathrm{d}(-\frac{x^2}{2})\\&=\sqrt{\frac{2}{\pi}}\end{aligned}$$

$$\begin{aligned}D|X|&=E(|X|)^2-(E|X|)^2\\&=EX^2-(E|X|)^2\\&=DX+(EX)^2-(E|X|)^2\\&=1-\frac{2}{\pi}\end{aligned}$$

□

总结

(1) 若 $X \sim N(0,1)$, 则 $E|X| = \sqrt{\frac{2}{\pi}}, D|X| = 1 - \frac{2}{\pi}$

(2) 若 $X \sim N(0,\sigma^2)$, 则 $E|X| = \sqrt{\frac{2}{\pi}} \cdot \sigma, D|X| = (1 - \frac{2}{\pi}) \cdot \sigma^2$

(3) 若 $X \sim N(\mu,\sigma^2)$, 则 $E|X-\mu| = \sqrt{\frac{2}{\pi}} \cdot \sigma, D|X| = (1 - \frac{2}{\pi}) \cdot \sigma^2$

8. 设随机变量 X 与 Y 相互独立, 均服从 $N(\mu,\sigma^2)$, 求 $E[\max\{X,Y\}]$,$E[\min\{X,Y\}]$.

Solution. 由 X,Y 独立, 有 $X-Y \sim N(0,2\sigma^2)$,$E\left|X-Y\right| = \frac{2\sigma}{\sqrt{\pi}}$

由下述总结, 可知所求期望为

$$E\left[max\{X,Y\}\right] = \frac{1}{2}\left[E(X)+E(Y)+E\left|X-Y\right|\right] = \mu + \frac{\sigma}{\sqrt{\pi}}$$
$$E\left[min\{X,Y\}\right] = \frac{1}{2}\left[E(X)+E(Y)-E\left|X-Y\right|\right] = \mu - \frac{\sigma}{\sqrt{\pi}}$$

□

总结

关于最大值最小值函数的拆法

$max\{X,Y\} = \frac{X+Y+|X-Y|}{2}$

$min\{X,Y\} = \frac{X+Y-|X-Y|}{2}$

9. 设独立重复的射击每次命中的概率为 p,X 表示第 n 次命中时的射击次数, 求 EX,DX.

Solution. Pascal 分布 (负二项分布), 关键在于分解随机变量, 设 X_i 表示第 $i-1$ 次命中到 i 命中所需要的射击次数, 则有 $X_1,X_2,\ldots$ 之间相互独立, 且 $X_i \sim G(p)$, 对于 $X = X_1 + X_2 \ldots X_n$, 故

$$EX = EX_1 + EX_2 + \ldots + EX_n = \frac{n}{p}$$
$$DX = DX_1 + DX_2 + \ldots + DX_n = \frac{n(1-p)}{p^2}$$

□

10. (2015, 数一、三) 设随机变量 X 的概率密度为 $f(x) = \begin{cases} 2^{-x}\ln 2, & x > 0 \\ 0, & x \le 0 \end{cases}$, 对 X 进行独立的观测, 直到第 2 个大于 3 的观测值出现时停止, 记 Y 为观测次数。

(a) 求 Y 的概率分布;

(b) 求 EY.

Solution. 不妨令 $p = P\{X > 3\} = \int_3^{+\infty} 2^{-x} \ln 2 \mathrm{d}x = \frac{1}{8}$

(1)

$$
\begin{aligned}
P\{Y = k\} &= C_{k-1}^1 p^2 (1-p)^{k-2} \\
&= (k-1)(\frac{1}{8})^2(\frac{7}{8})^{k-2}, k = 2, 3, \ldots
\end{aligned}
$$

(2)

$$
\begin{aligned}
EY &= \sum_{k=2}^{\infty} kP\{Y = k\} \\
&= p^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\
&\xlongequal{\text{幂级数求和}} \ldots \\
&= 16
\end{aligned}
$$

也可以用 Pascal 分布的结论直接得出 $EX = \frac{2}{\frac{1}{8}} = 16$ □

4.2 协方差的计算

Remark. 协方差

协方差的定义 $Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EX \cdot EY$

协方差的性质

(1) $Cov(X, Y) = Cov(Y, X), Cov(X, X) = DX$

(2) $Cov(aX + bY + c, Z) = aCov(X, Z) + bCov(Y, Z)$

11. 设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的简单随机样本。若 $DX = 4$, 正整数 $s \leq n, t \leq n$, 则

$$
\mathrm{Cov}\left(\frac{1}{s}\sum_{i=1}^{s} X_i, \frac{1}{t}\sum_{j=1}^{t} X_j\right) =
$$

(A) $4\max\{s, t\}$ (B) $4\min\{s, t\}$ (C) $\frac{4}{\max\{s, t\}}$ (D) $\frac{4}{\min\{s, t\}}$

Solution.

$$
\begin{aligned}
\operatorname{Cov}\left(\frac{1}{s}\sum_{i=1}^{s}X_i,\frac{1}{t}\sum_{j=1}^{t}X_j\right) &= \frac{1}{st}[Cov(X_1,X_1)+Cov(X_1,X_2)+\dots \\
&+Cov(X_2,X_1)+\dots+Cov(X_s,X_t)] \\
&\xlongequal{Cov(X_i,X_i)=DX_i,Cov(X_i,X_j)=0} \frac{\min(s,t)}{st}\cdot DX \\
&= \frac{4}{\max(s,t)}
\end{aligned}
$$

来自总体 X 的简单随机样本必然是独立同分布的. □

12. (2005, 数三) 设 $X_1,X_2,\cdots,X_n(n>2)$ 为来自总体 $N(0,\sigma^2)$ 的简单随机样本, 样本均值为 $\bar{X}$。记 $Y_i=X_i-\bar{X}$,$i=1,2,\cdots,n$。
   (1) 求 Y_i 的方差 DY_i,$i=1,2,\cdots,n$;
   (2) 求 Y_1 与 Y_n 的协方差 $Cov(Y_1,Y_n)$;
   (3) 若 $c(Y_1+Y_n)^2$ 为 σ^2 的无偏估计量, 求常数 c.

Solution.

(1) 方法一:

$$
\begin{aligned}
DY_i &= D(X_i-\bar{X}) \\
&= DX_i+D\bar{X}-2Cov(X_i,\bar{X}) \\
&\xlongequal{E\bar{X}=\mu,D\bar{X}=\sigma^2/n} \sigma^2+\frac{\sigma^2}{n}-2Cov(X_i,\frac{1}{n}\sum_{i=1}^{n}X_i) \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

方法二:

$$
\begin{aligned}
DY_i &= D(\frac{n-1}{n}X_i-\frac{1}{n}\sum_{i=j}^{n}X_j(j\neq i)) \\
&= (\frac{n-1}{n})^2\sigma^2-\frac{n-1}{n^2}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

(2)

$$
\begin{aligned}
Cov(Y_1, Y_n) &= Cov(X_1, \bar{X}, X_n - \bar{X}) \\
&= Cov(X_1, X_n) - Cov(X_1, \bar{X}) - Cov(X_n - \bar{X}) + D\bar{X} \\
&= \frac{-\sigma^2}{n}
\end{aligned}
$$

(3) 由无偏性有 $cE(Y_1+Y_n)^2 = \sigma^2 \implies c = \frac{\sigma^2}{E(Y_1+Y_n)^2}$

$$
\begin{aligned}
E(Y_1+Y_n)^2 &= D(Y_1+Y_n) + (EY_1EY_n)^2 \\
&= DY_1 + DY_n + 2Cov(Y_1, Y_n) + 0 \\
&= \frac{2(n-2)}{n}\sigma^2
\end{aligned}
$$

故 $c = \frac{n}{2(n-2)}$ □

4.3 相关系数的计算

Remark. 相关系数

相关系数的定义 $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}} = \frac{EXY - EXEY}{\sqrt{DX}\sqrt{DY}}$

相关系数的性质

(1) $|\rho_{XY}| \leq 1$

(2) $\rho_{XY} = 0 \iff Cov(X,Y) = 0 \iff EXY = EXEY \iff D(X+Y) = DX + DY$

(3) $\rho_{XY} = 1 \iff P\{Y = aX + b\} = 1(a > 0); \rho_{XY} = -1 \iff P\{Y = aX + b\} = 1(a < 0)$

13. (2016, 数一) 设试验有三个两两互不相容的结果 A_1, A_2, A_3, 且三个结果发生的概率均为 $\frac{1}{3}$。将试验独立重复地做两次,X 表示两次试验中 A_1 发生的次数,Y 表示两次试验中 A_2 发生的次数, 则 X 与 Y 的相关系数为

$(A)\ -\frac{1}{2}$ $(B)\ -\frac{1}{3}$ $(C)\ \frac{1}{3}$ $(D)\ \frac{1}{2}$

Solution.

(方法一) 由题意有 X, Y 均服从 $B(2, \frac{1}{3})$, 而 $P\{XY = 1\} = PX = 1, Y = 1 = C_2^1(\frac{1}{3})^2$, 且 $P\{XY = 0\} = \frac{7}{9}$, 故 XY 的概率分布如下所示

XY	0	1
P	$\frac{7}{9}$	$\frac{2}{9}$

故 $EXY=\frac{2}{9}$, 进而可以求出 $\rho_{XY}=\frac{-\frac{2}{9}}{\frac{4}{9}}=-\frac{1}{2}$

(方法二) 设 Z 为"A_3 在两次试验中发生的次数"

由题意有 $Z\sim B(2,\frac{1}{3}), X+Y+Z=2$ 而 $D(X+Y)=DX+DY+2Cov(X,Y)=\frac{8}{9}+2Cov(X,Y)$, 其中 $D(X+Y)=D(2-Z)=DZ=\frac{4}{9}$, 故 $Cov(X,Y)=\frac{-2}{9}$

(方法三)

$$\begin{aligned}Cov(X,X+Y+Z)&=DX+Cov(X,Y)+Cov(X,Z)\\&\xlongequal{\text{轮换对称性}}\frac{4}{9}+2Cov(X,Y)\\&=Cov(X,2)=0\implies Cov(X,Y)=-\frac{2}{9}\end{aligned}$$

□

14. 设随机变量 $X\sim B\left(1,\frac{3}{4}\right)$,$Y\sim B\left(1,\frac{1}{2}\right)$, 且 $\rho_{XY}=\frac{\sqrt{3}}{3}$。

(a) 求 (X,Y) 的联合概率分布;

(b) 求 $P\{Y=1|X=1\}$.

Solution. 这道题比较简单, 直接给答案

X/Y	0	1	P_i
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
P_j	$\frac{1}{2}$	$\frac{1}{2}$	1

$P\{Y=1|X=1\}=\frac{2}{3}$　□

4.4 相关与独立的判定

Remark. 相关与独立性

(1) 一般来说独立是强于不相关的条件, 即 独立 $\implies$ 不相关

(2) 对于二维正态分布有 独立 $\iff$ 不相关

(3) 对于 0-1 分布有 独立 $\iff$ 不相关

Remark. 判断是否独立的基本方法

(1) $P(AB)=P(A)P(B)$, 对于离散型选点, 对于连续型选区间

(2) 三个充要条件 $\forall(x,y)$或$(i,j)F(x,y)=F_XF_Y, f(x,y)=f_Xf_Y, P(ij)=P_{\cdot i}P_{j\cdot}$

(3) $\rho_{XY}\neq 0 \implies X,Y$ 不独立

15. 设二维随机变量 (X,Y) 服从区域 $D=\{(x,y)|x^2+y^2\leq a^2\}$ 上的均匀分布, 则

(A) X 与 Y 不相关, 也不相互独立 (B) X 与 Y 相互独立

(C) X 与 Y 相关 (D) X 与 Y 均服从 $U(-a,a)$

Solution. 这道题可以记结论, 对于均匀分布若其区域不为 $(a,b)\times(c,d)$ 的矩形, 则必然不独立, 其中 $X\in(a,b), Y\in(c,d)$

正常来做的话, 步骤如下

$$f(x,y)=\begin{cases}\frac{1}{\pi a^2}, & (x,y)\in D\\ 0, & (x,y)\notin D\end{cases}$$

$$EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\mathrm{d}x\mathrm{d}y \xlongequal{\text{对称性}} 0$$

同理根据对称性可知 $EXY=EX=EY=0$, 故 X,Y 一定不相关, 现在求 X,Y 的边缘分布概率密度, 有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\frac{2}{\pi a^2}\sqrt{a^2-x^2}, & x\in(-a,a)\\ 0, & x\notin(-a,a)\end{cases}$$

同理可以求出

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x=\begin{cases}\frac{2}{\pi a^2}\sqrt{a^2-y^2}, & y\in(-a,a)\\ 0, & y\notin(-a,a)\end{cases}$$

显然 $f_Yf_X\neq f(x,y)$ 故 X,Y 不独立. □

16. 设随机变量 X 的概率密度为 $f(x)=\frac{1}{2}e^{-|x|}, -\infty<x<+\infty$。

(a) 求 X 的期望与方差;

(b) 求 X 与 $|X|$ 的协方差, 问 X 与 $|X|$ 是否不相关?

(c) 问 X 与 $|X|$ 是否相互独立? 并说明理由.

Solution.

(1)

$$
\begin{aligned}
EX &= \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = 0 \\
EX^2 &= \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_{0}^{+\infty} x^2 e^{-x}\mathrm{d}x = 2 \\
DX &= EX^2 - (EX)^2 = 2
\end{aligned}
$$

(2)

$$
E(X|X|) = \int_{-\infty}^{+\infty} |X|Xf(x)\mathrm{d}x = 0 = EXE|X| \implies \rho_{X|X|} = 0, Cov(X, |X|) = 0
$$

(3) 设 $A = \{0 < X < 1\}, B = \{|X| < 1\}$, 故

$$
P(AB) = P\{0 < X < 1, |X| < 1\} = P\{0 < X < 1\} = P(A)
$$

而 $P(B) < 1$ 是显然的, 故 $P(AB) \neq P(A)P(B)$, 即 $X|X|$ 不独立 $\square$

第五章　大数定律与中心极限定理

Remark. 三个考点

(1) 切比雪夫不等式

$$P\{|X-EX|\geq\epsilon\}\leq\frac{DX}{\epsilon^2},\text{或者}P\{|X-EX|<\epsilon\}>1-\frac{DX}{\epsilon^2}$$

(2) 大数定理

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

(3) 中心极限定理

$$\sum_{i=1}^{n}X_i\sim N(n\mu,n\sigma^2)$$

1. (2022, 数一) 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立同分布,$\mu_k=E(X_i^k)(k=1,2,3,4)$。由切比雪夫不等式, 对任意 $\varepsilon>0$, 有 $P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i^2-\mu_2\right|\geq\varepsilon\right\}\leq$

$$(A)\ \frac{\mu_4-\mu_2^2}{n\varepsilon^2}\quad(B)\ \frac{\mu_4-\mu_2^2}{\sqrt{n}\varepsilon^2}\quad(C)\ \frac{\mu_2-\mu_1^2}{n\varepsilon^2}\quad(D)\ \frac{\mu_2-\mu_1^2}{\sqrt{n}\varepsilon^2}$$

Solution. 首先需要确定 $E(\frac{1}{n}\sum_{i=1}^{n}X_i^2)$ 是否等于 μ_2 显然, 所以这个式子满足切比雪夫不等式, 故根据切比雪夫不等式有

$$\text{原式}\geq\frac{D(\frac{1}{n}\sum_{i=1}^{n}X_i^2)}{\epsilon^2}=\frac{\mu_4-\mu_2^2}{n\epsilon^2}$$

□

2. (2022, 数一) 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立同分布,X_i 的概率密度为

$$f(x)=\begin{cases}1-|x|, & |x|<1\\ 0, & \text{其他}\end{cases}$$

则当 $n\to\infty$ 时,$\frac{1}{n}\sum_{i=1}^{n}X_i^2$ 依概率收敛于?.

Solution. 由大数定理有 $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} EX_i^2$, 又期望的定义有

$$EX_i^2 = 2\int_0^1 x^2(1-x)\mathrm{d}x = \frac{1}{6}$$

□

3. (2020, 数一) 设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的简单随机样本,$P\{X=0\} = P\{X=1\} = \frac{1}{2}$,$\Phi(x)$ 表示标准正态分布函数。利用中心极限定理得 $P\left\{\sum_{i=1}^{100} X_i \le 55\right\}$ 的近似值为

$$(A)\ 1-\Phi(1) \quad (B)\ \Phi(1) \quad (C)\ 1-\Phi(0.2) \quad (D)\ \Phi(0.2)$$

Solution. 由中心极限定理有 $\sum_{i=1}^{100} X_i \sim N(50, 25)$ 标准化后所求概率为

$$P\{\frac{X-50}{5} \le 1\} \implies \Phi(1)$$

□

第六章　统计初步

6.1　求统计量的抽样分布

Remark. 样本均值与方差

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

$E\bar{X}=\mu, D\bar{X}=\frac{\sigma^2}{n}, ES^2=\sigma^2$来自同一总体的样本均值与方差是独立的

有偏估计量 $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 其 $ES_n^2=\frac{n-1}{n}\sigma^2$

统计的三大分布

χ^2 分布的定义

设随机变量 $X_1,X_2,\ldots,X_n$ 相互独立, 均服从 $N(0,1)$ 称 $\chi^2=X_1^2+X_2^2+\ldots+X_n^2$ 服从自由度为 n 的 χ^2 分布, 记 $\chi^2\sim\chi^2(n)$, 特别的若 $X\sim N(0,1)$, 则 $\chi^2\sim\chi^2(1)$

χ^2 分布的性质

(1) 参数可加性　　设 χ_1^2 与 χ_2^2 相互独立, 且 $\chi_1^2\sim\chi^2(n),\chi_2^2\sim\chi^2(m)$ 则 $\chi_1^2+\chi_2^2\sim\chi^2(n+m)$

(2) 设 $\chi^2\sim\chi^2(n)$ 则 $E\chi^2=n, D\chi^2=2n$

F 分布的定义

设随机变量 X 和 Y 相互独立, 且 $X\sim\chi^2(n_1), Y\sim\chi^2(n_2)$, 称 $F=\frac{X/n_1}{Y/n_2}$ 服从自由度为 n_1,n_2 的 F 分布, 记作 $F\sim F(n_1,n_2)$

F 分布的性质

(1) 设 $F\sim F(n_1,n_2)$, 则 $\frac{1}{F}\sim F(n_2,n_1)$

(2) $F_{1-\alpha}(n_2,n_1)=\frac{1}{F_\alpha(n_1,n_2)}$

t 分布的定义　　设随机变量 X 和 Y 相互独立,$X\sim N(0,1), Y\sim\chi^2(n)$, 则称 $T=\frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 记作 $T\sim t(n)$

t 分布的性质

(1) 设 $T \sim t(n)$, 则 $T^2 \sim F(1,n)$, $\frac{1}{T^2} \sim F(n,1)$

(2) $t_{1-\alpha}(n) = -t_\alpha(n)$

Remark. 单正态总体与双正态总体

单正态总体

设 $X_1, X_2, \ldots, X_n$ 为来自总体 $X \sim N(\mu, \sigma^2)$ 的简单随机样本,$\bar{X}$ 与 S^2 分别为样本均值与样本方差, 则

(1) $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, 即 $\bar{X} \sim N(\mu, \sigma^2/n)$

(2) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ 即 $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$, 且 $\bar{X}$ 与 S^2 相互独立

(3) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

双正态总体

设总体 $X \sim N\left(\mu_1, \sigma_1^2\right)$ ，总体 $Y \sim N\left(\mu_2, \sigma_2^2\right), X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别为来自总体 X 与 Y 的简单随机样本且相互独立, 样本均值分别为 $\bar{X}, \bar{Y}$ ，样本方差分别为 S_1^2, S_2^2 ，则

(4) $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$;

(5) $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F\left(n_1-1, n_2-1\right)$;

(6) 当 $\sigma_1^2 = \sigma_2^2$ 时, $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t\left(n_1+n_2-2\right)$ ，其中 $S_\omega = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$.

1. (2013, 数一) 设随机变量 $X \sim t(n), Y \sim F(1,n)$。给定 $\alpha(0<\alpha<0.5)$, 常数 c 满足 $P\{X>c\}=\alpha$, 则 $P\{Y>c^2\}=$

$$(A)\ \alpha \quad (B)\ 1-\alpha \quad (C)\ 2\alpha \quad (D)\ 1-2\alpha$$

Solution. 这道题考察的是 t 分布的对称性, 由题有

$$Y = \frac{\chi^2(1)}{\chi^2(n)} \qquad X = \frac{N(0,1)}{\sqrt{\chi^2(n)/n}}$$

则有 $X^2 = Y$, 所求概率就变成 $P\{X^2 > c^2\}$ 由 t 分布的对称性有 $P\{X^2 > c^2\} = 2\alpha$ □

总结

正态分布与 t 分布具有相似的概率密度图像,F 分布与 χ^2 分布也有类似的图像.

2. 设 $X_1, X_2, \cdots, X_9$ 为来自总体 $N(\mu, \sigma^2)$ 的简单随机样本,$Y_1 = \frac{1}{6}(X_1+X_2+\cdots+X_6)$,$Y_2 = \frac{1}{3}(X_7+X_8+X_9)$,$S^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - Y_2)^2$, 求 $\frac{\sqrt{2}(Y_1-Y_2)}{S}$ 的分布.

Solution. 这种题就是一步一步反推, 注意凑题目要求的结果即可

$Y_1 = \frac{1}{6}\sum_{i=1}^{6} X_i \sim N(\mu, \frac{\sigma^2}{6})$ 同理 $Y_2 \sim N(\mu, \frac{\sigma^2}{3})$

由 Y_1, Y_2 独立, 知道 $Y_1 - Y_2 \sim N(0, \frac{\sigma^2}{2}) \implies \frac{Y_1-Y_2}{\sigma/\sqrt{2}} \sim N(0,1)$

又有 $\frac{2s^2}{\sigma^2} \sim \chi^2(2)$, 故

$$\frac{Y_1 - Y_2}{\sqrt{\sigma^2/2}\sqrt{\frac{2s^2}{\sigma^2}/2}} = \frac{\sqrt{2}(Y_1 - Y_2)}{s} \sim t(2)$$

□

6.2 求统计量的数字特征

3. 设 $X_1, X_2, \cdots, X_n$ 为来自总体 $N(\mu, \sigma^2)$ 的简单随机样本, 则

$$E\left[\sum_{i=1}^{n} X_i \cdot \sum_{j=1}^{n}\left(nX_j - \sum_{k=1}^{n} X_k\right)^2\right] =$$

Solution. 这道题就是个凑系数化简, 过程省去 原式 $= n^3(n-1)\mu\sigma^2$ □

4. 设 $X_1, X_2, \cdots, X_9$ 为来自总体 $N(0, \sigma^2)$ 的简单随机样本, 样本均值为 $\bar{X}$, 样本方差为 S^2。

(1) 求 $\frac{9\bar{X}^2}{S^2}$ 的分布

(2) 求 $E[(\bar{X}^2 S^2)^2]$;

Solution.

(1) 和例题 3 一致, 过程省去 $\frac{9\bar{X}^2}{S^2} \sim F(1,8)$

(2) 对于这种高幂次的一般都需要考虑用 χ^2 的结论

$$\begin{aligned} E\left[(\bar{X}^2 S^2)^2\right] &= E\bar{X}^4 \cdot ES^4 \\ &= \left[D\bar{X}^2 + (E\bar{X}^2)^2\right]\left[DS^2 + (ES^2)^2\right] \\ &= \frac{5}{107}\sigma^8 \end{aligned}$$

又 $\frac{9\bar{X}^2}{\sigma^2} \sim \chi^2(1) \implies D\bar{X}^2 = \frac{2\sigma^4}{81}$ 同理有 $DS^2 = \frac{\sigma^4}{4}$ □

第七章　参数估计

7.1　求矩估计与最大似然估计

Remark. 矩估计与最大似然估计

矩估计

令 $EX^k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 或者 $E(X-EX)^k = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k, k=1,2,\dots$ 得到 $\theta_1,\theta_2\dots$ 的矩估计量

$$\begin{cases} EX = \bar{X}, & \text{一个参数} \\ EX^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 & \text{两个参数} \end{cases}$$

最大似然估计

(1) 对样本点 $x_1,x_2\dots,x_n$, 似然函数为 $L(\theta)\begin{cases}\prod_{i=1}^{n} p(x_i;\theta) \\ \prod_{i=1}^{n} f(x_i;\theta)\end{cases}$

(2) 似然函数两端取对数求导

(3) 令 $\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta} = 0$ 就可以得到 θ 的最大似然估计值

一个关于规范的小提示, 如果问估计值用小写字母 (样本值), 问估计量用大写字母 (随机变量)

1. (2002, 数一) 设总体 X 的概率分布为

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中 $0<\theta<\frac{1}{2}$ 为未知参数, 利用总体 X 的如下样本值 $3,1,3,0,3,1,2,3$, 求 θ 的矩估计值与最大似然估计值。

Solution.

(矩估计) 这道题只有一个参数, 只需要用一阶矩估计 $EX = 2\theta(1-\theta)+2\theta^2+3-6\theta = \bar{X}$, 其中 $\bar{X} = \frac{16}{8} = 2$, 故 θ 的矩估计值 $\hat{\theta} = \frac{1}{4}$

(最大似然估计) 对于样本 $3,1,3,0,3,1,2,3$, 似然估计函数为

$$L(\theta) = 4\theta^6(1-\theta)^2(1-2\theta)^4$$

令 $\frac{\mathrm{d}\ln\theta}{\mathrm{d}\theta} = 0$ 有 $\theta = \frac{7+\sqrt{13}}{12}$ 又 $0<\theta<\frac{1}{2}$, 故最终 $\theta = \frac{7-\sqrt{13}}{12}$ □

2. (2011, 数一) 设 $X_1, X_2, \cdots, X_n$ 为来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本, 其中 μ 已知,$\sigma^2>0$ 未知, 样本均值为 $\bar{X}$, 样本方差为 S^2。

(1) 求 σ^2 的最大似然估计量 $\hat{\sigma}^2$;

(2) 求 $E(\hat{\sigma}^2)$ 与 $D(\hat{\sigma}^2)$。

Solution.

(1) 对于样本 $X_1,\ldots,X_n$ 其最大似然函数为

$$L(\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

注意参数为 σ^2, 令 $\frac{\mathrm{d}\ln\sigma^2}{\mathrm{d}\sigma^2} = 0$, 有 $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1} n(X_i-\mu)^2$

(2) 这种题优先考虑 χ^2 分布的期望与方差结论, 有题 (1) 有

$$\frac{X_i-\mu}{\sigma} \sim N(0,1) \implies \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$

故 $E(\hat{\sigma}^2) = \sigma^2, D(\hat{\sigma}^2) = \frac{2\sigma^4}{n}$ □

3. (2022, 数一、三) 设 $X_1,X_2,\cdots,X_n$ 为来自期望为 θ 的指数分布总体的简单随机样本,$Y_1,Y_2,\cdots,Y_m$ 为来自期望为 2θ 的指数分布总体的简单随机样本, 两个样本相互独立。利用 $X_1,X_2,\cdots,X_n$ 与 $Y_1,Y_2,\cdots,Y_m$,

(1) 求 θ 的最大似然估计量 $\hat{\theta}$;

(2) 求 $D(\hat{\theta})$。

Solution. 这是双总体, 但基本上和单总体一致, 不要被唬住了哦!

(1) 由题有 $X \sim E(\frac{1}{\theta}), Y \sim E(\frac{1}{2\theta})$, 故其概率密度分别为

$$f_X(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \le 0 \end{cases} \qquad f_Y(y) = \begin{cases} \frac{1}{2\theta}e^{-\frac{y}{2\theta}}, & y > 0 \\ 0, & x \le 0 \end{cases}$$

则对于样本 $X_1, X_2, \ldots, X_n$ 与 $Y_1, Y_2, \ldots, Y_n$, 最大似然估计函数为

$$L(\theta) = (\frac{1}{2})^m \theta^{-(m+n)} e^{-\frac{1}{\theta}(\sum_{i=1}^n X_i + \frac{1}{2}\sum_{j=1}^m Y_j)}$$

则令 $\frac{\mathrm{d}\ln\theta}{\mathrm{d}\theta} = 0$, 有 $\hat{\theta} = \frac{1}{n+m}(\sum_{i=1}^n X_i + \frac{1}{2}\sum_{j=1}^m Y_j)$

(2)

$$\begin{aligned} D(\hat{\theta}) &= (\frac{1}{m+n})^2 D(\sum_{i=1}^n X_i + \frac{1}{2}\sum_{j=1}^m Y_j) \\ &= \frac{\theta^2}{m+n} \end{aligned}$$

□

7.2 估计量的评价标准

Remark. 估计量的评价标准

(1) (无偏性) 设 $\hat{\theta}$ 为 θ 的估计量, 若 $E\hat{\theta} = \theta$ 则称其为 θ 无偏估计量

(2) (有效性) 设 $\hat{\theta}_1, \hat{\theta}_2$ 为 θ 的无偏估计, 若 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$ 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效

(3) 设 $\hat{\theta}$ 为 θ 的估计量, 若 $\hat{\theta}$ 依概率收敛于 θ, 则称 $\hat{\theta}$ 为 θ 一致 (相合) 估计量

一致性的考点在于—$\frac{1}{n}\sum \square \xrightarrow{P} E\square$

4. 设总体 X 的概率密度为

$$f(x) = \begin{cases} 2e^{-2(x-\theta)}, & x > \theta \\ 0, & x \le \theta \end{cases}$$

其中 $\theta > 0$ 为未知参数,$X_1, X_2, \cdots, X_n$ 为来自总体 X 的简单随机样本。

(1) 求 θ 的最大似然估计量 $\hat{\theta}$;

(2) 问 $\hat{\theta}$ 是否为 θ 的无偏估计量? 并说明理由。

Solution.

(1) 对于样本 $X_1, X_2, \ldots, X_n$ 的最大似然估计函数为

$$L(\theta) = \prod_{i=1}^{n} 2e^{-2(x_i-\theta)} = 2^n e^{-\sum_{i=1}^{n}(x_i-\theta)}$$

显然 $L(\theta)$ 关于 θ 是单调递增的, 则根据最大似然的定义, 应该取使得 $L(\theta)$ 最大的值, 而由题目有 $X_1 > \theta, X_2 > \theta, \ldots$, 故 $\hat{\theta} = \min\{X_1, X_2 \ldots, X_n\}$

(2) 由概率密度函数有 $F_X(x) = \int_{-\infty}^{x} f(t)dt$, 故

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 1 - e^{-2(x-\theta)}, & x > \theta \\ 0, & x \le \theta \end{cases}$$

故 $F_{min} = 1 - [1 - F_X(x)]^n$ 即

$$F_{min} = \begin{cases} 1 - e^{-2n(x-\theta)}, & x > \theta \\ 0, x \le \theta \end{cases}$$

故

$$f_{min} = \begin{cases} 2ne^{-2n(x-\theta)}, & x > \theta \\ 0, & x \le \theta \end{cases}$$

由期望的定义有

$$E\hat{\theta} = \int_{\theta}^{+\infty} 2nxe^{-2n(x-\theta)} = \theta + \frac{1}{2n}$$

□

5. (2010, 数一) 设总体 X 的概率分布为

X	1	2	3
P	$1-\theta$	$\theta-\theta^2$	θ^2

其中参数 $\theta \in (0,1)$ 未知,N_i 表示来自总体 X 的简单随机样本 (样本容量为 n) 中等于 i 的个数 $(i = 1, 2, 3)$ 求常数 a_1, a_2, a_3 使得 $T = \sum_{i=1}^{3} a_i N_i$ 为 θ 的无偏估计量, 并求 T 的方差.

Solution. 由题可知 $N_i \sim B(n,p)$, 具体来说有

$$\begin{cases} N_1 \sim B(n, 1-\theta) \\ N_2 \sim B(n, \theta-\theta^2) \\ N_3 \sim B(n, \theta^2) \end{cases}$$

且有 $N_1 + N_2 + N_3 = n$

故 $ET = \sum_{i=1}^{3} a_i EN_i = n\left[a_1 + (a_2 - a_1)\theta + (a_3 - a_2)\theta^2\right] = \theta$, 只需要令

$$\begin{cases} a_1 = 0 \\ a_2 = \frac{1}{n} \\ a_3 = \frac{1}{n} \end{cases}$$

求 $DT = \frac{1}{n^2} D(n - N_1) = \frac{1}{n^2} DN_1 = \frac{\theta(1-\theta)}{n}$ □

7.3 区间估计与假设检验

第八章　补充知识-概率论

补充知识来自于

(1) 概率论与数理统计　　茆诗松

(2) 做题总结

8.1　配对问题

问题描述: 在一个有 n 个人参加的晚会, 每个人带来一件礼物, 且规定每个人带的礼物都不相同. 晚会期间各人从放在一起的 n 件礼物中随机抽取一件, 问至少有一个人自己抽到自己的礼物的概率是多少?

Solution. (配对问题)

设 A_i 为事件: 第 i 个人自己抽到自己的礼物,$i = 1, 2, \dots, n$ 所求概率为

$$
\begin{aligned}
&P(A_1) = P(A_2) = \dots = P(A_n) = \frac{1}{n} \\
&P(A_1A_2) = P(A_1A_3) = \dots = P(A_{n-1}A_n) = \frac{1}{n(n-1)} \\
&P(A_1A_2A_3) = P(A_1A_2A_4) = \dots = P(A_{n-2}A_{n-1}A_n) = \frac{1}{n(n-1)(n-2)} \\
&\dots \\
&P(A_1A_2A_3\dots A_n) = \frac{1}{n!}
\end{aligned}
$$

再由概率的加法公式 (容斥原理) 得

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \ldots \cup A_n) &= \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} P(A_i A_{i+1}) + \sum_{i=1}^{n-2} P(A_i A_{i+1} A_{i+2}) \\
&+ \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n) \\
&= C_n^1 \frac{1}{n} - C_n^2 \frac{1}{n(n-1)} + \ldots + (-1)^{n-1} C_n^n \frac{1}{n!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} + \ldots + (-1)^{n-1} \frac{1}{n!}
\end{aligned}
$$

当 $n \to \infty$, 上述概率由 $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, 则

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = 1 - e^{-1} \approx 0.6321$$

□

8.2 几个概率的不等式

1. $P(AB) \geq P(A) + P(B) - 1$
2. $P(A_1 A_2 \ldots A_n) \geq P(A_1) + P(A_2) + \ldots + P(A_n) - (n-1)$ (Boole 不等式)
3. $|P(AB) - P(A)P(B)| \leq \frac{1}{4}$

Proof. 相关证明如下:

(1) 由于 $P(A \cup B) = P(A) + P(B) - P(AB) \leq 1 \implies P(AB) \geq P(A) + P(B) - 1$

(2) 采用数学归纳法证明, 对于 n = 2, 即不等式 (1) 已经证明, 不妨假设对于 $n = k$ 个事件, 不等式成立, 即

$$P(A_1 A_2 \ldots A_k) \geq P(A_1) + P(A_2) + \ldots + P(A_k) - (k-1)$$

考虑 $n = k+1$ 个事件 $A_1 A_2 \ldots A_{k+1}$, 不妨令 $B = A_1 A_2 \ldots A_k$, 则

$$P(A_1 A_2 \ldots A_k A_{k+1}) = P(BA_{k+1}) \geq P(B) + P(A_{k+1}) - 1 \geq P(A_1) + P(A_2) + \ldots + P(A_{k+1}) - (k)$$

由数学归纳法可知, 原不等式成立

(3) 由 $P(A) \geq P(AB), P(B) \geq P(AB)$, 则 $P(A)P(B) \geq P(AB)^2$, 则

$$P(AB) - P(A)P(B) \leq P(AB) - P(AB)^2 = P(AB)(1 - P(AB))$$

令 $x = P(AB)$, 则 $f(x) = x(1-x)$, 当 $x = \frac{1}{2}$ 时, 取得 $f(x)_{max} = \frac{1}{4}$ 即

$$P(AB) - P(A)P(B) \leq \frac{1}{4}$$

由于 $P(AB) + P(A\bar{B}) = P(A)$, 即 $P(AB) = P(A) - P(A\bar{B})$ 则

$$P(A)P(B) - P(AB) = P(A)P(B) - P(A) + P(A\bar{B}) = P(A\bar{B}) - P(A)P(\bar{B}) \leq \frac{1}{4}$$

即

$$P(AB) - P(A)P(B) \geq \frac{1}{4}$$

综上原不等式成立 □

8.3 轮流射击模型

问题描述: 有两名选手比赛设计, 轮流对同一个目标进行射击, 甲命中目标的概率为 α, 乙命中的概率为 β. **甲先射**, 谁先设中谁获胜. 问甲乙两人获胜的概率各是多少?

Solution.

(方法一) 记事件 A_i 为第 i 次射中目标,$i = 1, 2, \ldots$, 因为甲先射, 所以甲获胜可以表示为

$$A_1 \cup \bar{A}_1\bar{A}_2A_3 \cup \ldots.$$

由于事件独立, 则甲获胜的概率为

$$\begin{aligned} P(\text{甲获胜}) &= \alpha + (1-\alpha)(1-\beta)\alpha + (1-\alpha)^2(1-\beta)^2\alpha^2 \ldots \\ &= \alpha \sum_{i=0}^{\infty} (1-\alpha)^i(1-\beta)^i \\ &= \frac{\alpha}{1-(1-\alpha)(1-\beta)} \end{aligned}$$

同理, 乙获胜的概率为

$$\begin{aligned} P(\text{乙获胜}) &= (1-\alpha)\beta + (1-\alpha)(1-\beta)(1-\alpha)\beta + \ldots \\ &= \beta(1-\alpha) \sum_{i=0}^{\infty} (1-\alpha)^i(1-\beta)^i \\ &= \frac{\beta(\alpha-1)}{1-(1-\alpha)(1-\beta)} \end{aligned}$$

(方法二) 由于射击是独立, 所有有如下条件

$$P(\text{甲获胜}) = \alpha + (1-\alpha)(1-\beta)P(\text{甲获胜})$$

前面失败的情况并不影响后续获胜 (无记忆性), 则可以直接解出甲获胜的概念

$$P(\text{甲获胜}) = \frac{\alpha}{1-(1-\alpha)(1-\beta)}$$

$$P(\text{乙获胜}) = 1 - P(\text{甲获胜}) = \frac{\beta(\alpha-1)}{1-(1-\alpha)(1-\beta)}$$

□

8.4 补充: 随机变量的矩

设 (X,Y) 是二维随机变量, 如果 $E(X^kY^l)$ 存在, 则称 $E(X^k),(k=1,2\ldots)$ 为 X 的 k 阶原点矩; 称 $E(X-EX)^k, k=(2,3,\ldots)$ 为 X 的 k 阶中心矩; 称 $E(X^kY^l),(k,l=1,2,\ldots)$ 为 X 与 Y 的 $k+l$ 阶混合原点矩; 称 $E[(X-EX)^k(Y_EY)^l,(k,l=1,2,\ldots)]$ 为 X,Y 的 $k+l$ 阶混合中心矩