第一章　一维随机变量

1.1　分布函数的判定与计算

Remark. (分布函数的性质)

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

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

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

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

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

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

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

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)

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

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

□

1.2 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

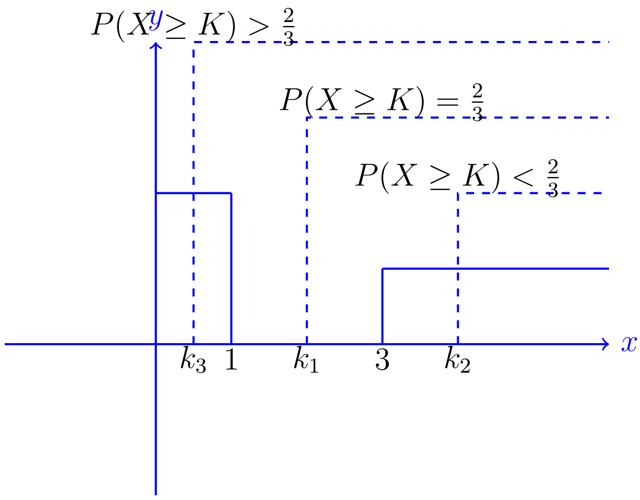
$f_1F_2+f_2F_1, 2f_1F_1, 2f_2F_2$ 是概率密度, 其余都不是.

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 $P(X\geq k)=\frac{2}{3}$ □



1.3 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,$X \sim H(N,M,n)$

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

(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。

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

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

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 $A=$___,$B=$___.

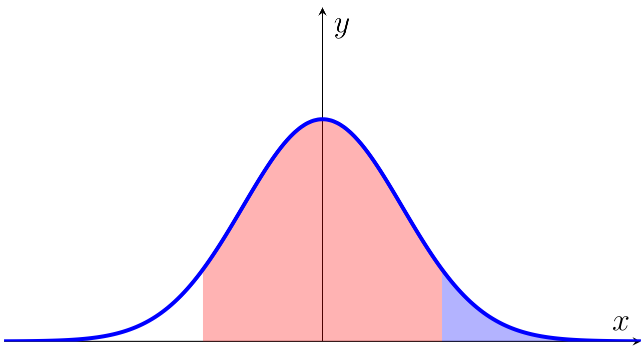
Solution. $f(x)=Ae^{\frac{B^2}{2}}e^{-\frac{(x-B)^2}{2}}\sim N(1,B^2)$, 又 $D(x)=E(x)$ 故 $B^2=1$, 对比正态分布的概率密度函数有 $Ae^{\frac{B^2}{2}}=\frac{1}{\sqrt{2\pi}}$ 故 $A=\frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}$ □

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.

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

□

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.

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

□

总结

对于 min 和 max 问题基本按照如下思路:

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

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

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

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

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

$$
\begin{aligned}
&X \sim E(\lambda) \\
&P\{x > s+t \mid x > s\} = P\{x > t\} \\
&P\{x < s+t \mid x > s\} = P\{0 < x < t\} \\
&X \sim G(p) \\
&P\{x > n+m \mid x > m\} = P\{x > t\} \\
&P\{x = n+m \mid x = m\} = P\{x = n\} = p(1-p)^{n-1}
\end{aligned}
$$

1.4 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 $F_Y(y)=P\{Y\leq y\}=P\{g(X)\leq y\}$.

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

当 $y\geq\beta$ 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

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

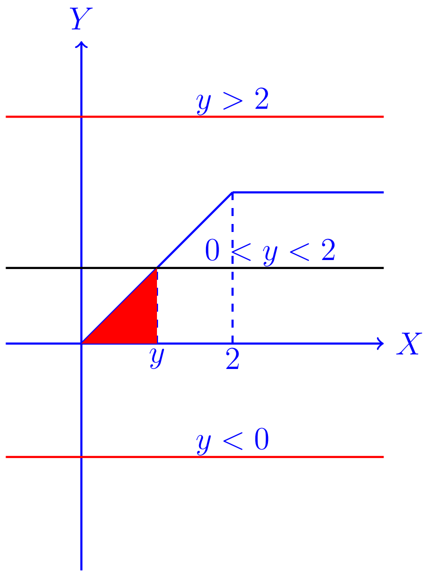
若 $y=g(x)$ 在 X 的正概率密度区间 $[a,b]$ 分段严格单调, 则分段运用公式法, 然后将概率密度相加.

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 $f(x) = \begin{cases} \frac{x^2}{a}, & 0 < x < 3 \\ 0, & 其他 \end{cases}$ $Y = \begin{cases} 2, & X \leq 1 \\ X, & 1 < X < 2 \\ 1, & X \geq 2 \end{cases}$

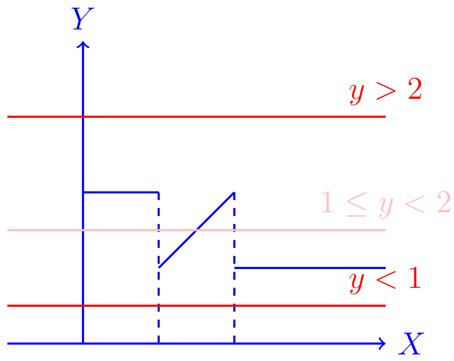
(a) 求 Y 的分布函数;

(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 $y<1, F_Y(y)=0; y>2, F_Y(y)=1$

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

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

(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)

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

或者也可以用

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

□