作业二讲评

任香璇

2024 年 3 月 15 日

分布函数

定义

随机变量 X 和实变量 x，函数 $F(x) = P(X \le x)$ 为 X 的概率分布函数，$x \in R$.

(1) 非负性: $F(x) \in [0, 1]$ 且满足 $F(-\infty) = 0$ 和 $F(\infty) = 1$

(2) 单调性: $F(x)$ 单调不减

(3) 右连续: $\lim_{\epsilon \to 0^+} F(x + \epsilon) = F(x)$

例

1. 设随机变量 X 的绝对值不大于 1，$P(X=-1)=\frac{1}{8}$，$P(X=1)=\frac{1}{4}$，在事件 $\{-1<X<1\}$ 出现的条件下，X 在 $(-1,1)$ 内的任一子区间上取值的条件概率与该区间长度成正比. 试求:

- X 的分布函数 $F(x)$.
- $P(X\leq 0)$.

解.

(1) 根据分布函数的定义 $F(x) = P(X \le x)$，由题设得:
当 $x < -1$ 时，$F(x) = 0$;
当 $x \ge 1$ 时，$F(x) = 1$;
当 $-1 \le x < 1$ 时，$F(x) = P(X \le x) = P(-1 \le X \le x) = P(X = -1) + P(-1 < X \le x) = \frac{1}{8} + P(-1 < X \le x)$，其中

$$\begin{aligned} P(-1 < X \le x) &= P(-1 < X \le x, -1 < X < 1) \\ &= P(-1 < X < 1) \cdot P(-1 < X \le x | -1 < X < 1) \\ &= (1 - \frac{1}{4} - \frac{1}{8}) \cdot \frac{x+1}{2} \\ &= \frac{5(x+1)}{16} \end{aligned}$$

故可得 X 的分布函数为:

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{7}{16} + \frac{5x}{16}, & -1 \le x < 1, \\ 1, & x \ge 1. \end{cases}$$

(2) $P(X \le 0) = F(0) = \frac{7}{16}$

例

2. 从 1，2，3，4，5 五个数中任取三个，按大小排列记为 $x_1 < x_2 < x_3$，令 $X = x_2$，试求：

- X 的分布函数 $F(x)$.
- P(X<2) 及 P(X>4).

解.

(1) （古典概型）因为 X 的分布列：

X	2	3	4
P	3/10	4/10	3/10

所以 X 的分布函数为：

$$F(x) = \begin{cases} 0, & x < 2, \\ 3/10, & 2 \leq x < 3, \\ 7/10, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

(2) $P(X < 2) = F(2-0) = 0$， $P(X > 4) = 1 - F(4) = 1 - 1 = 0.$

例

3. x 轴上有一质点，每经一个单位时间，它分别以概率 p 及 $q = 1 - p$ 向右或向左移动一格，若该质点在时刻 0 从原点出发，而且每次移动是相互独立的，$x = -a$ 和 $x = b$ 处各有一个吸收壁，求质点在 $x = b$ 处被吸收的概率.

解. 默认 a，b 正整数.

设 q_n 为质点在 $x=n$ 时最终被 $x=b$ 吸收的概率，则 $q_b=1$ 且 $q_{-a}=0$.

当 $p=1$ 时，$q_0=1$；

当 $p=0$ 时，$q_0=0$；

当前质点位于 $x=n$，下一时刻质点的位置可能为 $x=n+1$ 或 $x=n-1$，由全概率公式 $q_n=pq_{n+1}+qq_{n-1}$. 进而 $p\left(q_{n+1}-q_n\right)=q\left(q_n-q_{n-1}\right)$. 记 $r=\frac{q}{p}$，上式化为 $\left(q_{n+1}-q_n\right)=r\left(q_n-q_{n-1}\right)$.

当 $p=q$ 时，有 $q_{n+1}-q_n=q_{-a+1}-q_{-a}$，

得 $q_n=q_n-q_{n-1}+\ldots+q_{-a+1}-q_{-a}=(n+a)(q_{-a+1}-q_{-a})$

由边界条件 $q_b=1$ 和 $q_{-a}=0$，有 $q_n=\frac{n+a}{a+b}$. 令 $n=0$，得到 $q_0=\frac{a}{a+b}$；

当 $p\neq q$ 时，有 $q_{n+1}-q_n=r^{n+a}\left(q_{-a+1}-q_{-a}\right)$，

得 $q_n=q_n-q_{n-1}+\ldots+q_{-a+1}-q_{-a}=\sum_{i=0}^{n+a-1}r^i\left(q_{-a+1}-q_{-a}\right)=\frac{1-r^{n+a}}{1-r}\left(q_{-a+1}-q_{-a}\right)$，由边界条件 $q_b=1$ 和 $q_{-a}=0$，有

$q_n=\frac{1-r^{n+a}}{1-r^{a+b}}=\frac{1-\left(\frac{q}{p}\right)^{n+a}}{1-\left(\frac{q}{p}\right)^{a+b}}$. 令 $n=0$，得到 $q_0=\frac{1-\left(\frac{q}{p}\right)^{a}}{1-\left(\frac{q}{p}\right)^{a+b}}$.

例

4.(泊松分布的分解) 若每条蚕的产卵数服从泊松分布，参数为 λ，而每个卵变为成虫的概率为 p，且各卵是否变为成虫彼此独立，求每条蚕养活 k 只小蚕的概率.

解. 以 A_n 记蚕产出 n 个卵，以 B_k 记每条蚕养活 k 只小蚕，依独立性假定，若蚕产出 n 个卵，则这 n 个卵中成活的成虫数服从二项分布. 记 $q = 1 - p$，则所求概率为

$$\begin{aligned} P(B_k) &= \sum_{n=k}^{\infty} P(A_n)P(B_k|A_n) \\ &= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{k} p^k q^{n-k} \\ &= \frac{p^k e^{-\lambda}}{k!} \sum_{j=0}^{\infty} \frac{\lambda^{k+j}}{j!} q^j \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda p}. \end{aligned}$$

二项分布的极限分布

定理

对于二项分布 $P_{B,n}(X=k) \sim B(n, p_n)$, 考虑 $n \to \infty$ 且 $\lim_{n\to\infty} np_n = \lambda$ 则对于任意 $k = 0, 1, \ldots,$ 有 $\lim_{n\to\infty} P_{B,n}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

例

5. 一个工厂出产的产品中废品率为 0.005，任意取来 1000 件，解答以下问题：

- 求其中至少有两件废品的概率.
- 求其中不超过 5 件废品的概率.
- 能以 90% 的概率希望废品件数不超过多少？

解. 本题用泊松逼近，有 $\lambda = 1000 \times 0.005 = 5$.

(1) 记 “至少有两件废品” 为事件 A,

$$P(A) = 1 - P(\bar{A}) = 1 - P(\xi = 0) - P(\xi = 1) \approx 0.96.$$

(2) 记 “不超过 5 件废品” 为事件 B,

$$P(B) = \sum_{k=0}^{5} P(\xi = k) \approx 0.62.$$

(3) 对照泊松分布数值表，可知废品件数以 90% 概率不超过 8.
$\left(F(7) = 0.867,\ F(8) = 0.932\right)$

定量理解二项式分布的泊松和高斯近似

例

6. 设随机变量 X 与 Y 同分布, X 的密度函数

$$f(x)=\begin{cases}\frac{3}{8}x^2, & 0<x<2,\\ 0, & \text{其他}.\end{cases}$$

已知事件 $A=\{X>a\}$ 与 $B=\{Y>a\}$ 相互独立, 且 $P\{A\cup B\}=\frac{3}{4}$, 求常数 a.

解.
$P(A)=P(X>a)=1-P(X\le a)=1-F_X(a)$,
$P(B)=P(Y>a)=1-P(Y\le a)=1-F_Y(a)$,
X 与 Y 同分布, $P(A)=P(B)$.
由加法公式和独立性,
$\frac{3}{4}=P(A\cup B)=P(A)+P(B)-P(A)P(B)=2P(A)-[P(A)]^2$.
解得 $\frac{1}{2}=P(A)=P(X>a)=\int_a^2\frac{3}{8}x^2\,\mathrm{d}x=1-\frac{1}{8}a^3$, 所以 $a=\sqrt[3]{4}$.

随机变量的函数分布

定理

设 X 的概率密度为 $f_X(x)$，函数 $g(x)$ 严格单调且其反函数 $h(y)$ 有连续导数，则 $Y = g(X)$ 也是一个随机变量，且概率密度为

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)| & y \in (\alpha, \beta) \\ 0 & else \end{cases}$$

其中 $\alpha = \min\{g(-\infty), g(+\infty)\}$，$\beta = \max\{g(-\infty), g(+\infty)\}$

例

7. 设随机变量 X 的密度函数为

$$f(x) = \begin{cases} Ae^{-x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

求出 A 并求以下 Y 的密度函数:

- $Y = 2X + 1$.
- $Y = e^{x}$.
- $Y = X^2$.

解.

(1) 因为 $Y=2X+1$ 的可能取值范围是 $(1,+\infty)$, 且 $y=g(x)=2x+1$ 是严格单调增函数, 其反函数为 $x=h(y)=(y-1)/2$, 及 $h'(y)=1/2$, 所以 Y 的密度函数为

$$f_Y(y)=\begin{cases} f_X((y-1)/2)|1/2|, & y>1 \\ 0, & \text{其他} \end{cases} = \begin{cases} \frac{1}{2}\mathrm{e}^{-\frac{y-1}{2}}, & y>1, \\ 0, & \text{其他}. \end{cases}$$

(2) 因为 $Y=\mathrm{e}^X$ 的可能取值范围是 $(1,+\infty)$, 且 $y=g(x)=\mathrm{e}^x$ 是严格单调增函数, 其反函数为 $x=h(y)=\ln y$, 及 $h'(y)=1/y$, 所以 Y 的密度函数为

$$f_Y(y)=\begin{cases} f_X(\ln y)|1/y|, & y>1 \\ 0, & \text{其他} \end{cases} = \begin{cases} \frac{1}{y^2}, & y>1, \\ 0, & \text{其他}. \end{cases}$$

(3) 因为 $Y=X^2$ 的可能取值范围是 $(0,+\infty)$, 且 $y=g(x)=x^2$ 在 $(0,+\infty)$ 上是严格单调增函数, 其反函数为 $x=h(y)=\sqrt{y}$, 及 $h'(y)=1/(2\sqrt{y})$, 所以 Y 的密度函数为

$$f_Y(y)=\begin{cases} f_X(\sqrt{y})|1/(2\sqrt{y})|, & y>0 \\ 0, & \text{其他} \end{cases} = \begin{cases} \frac{1}{2\sqrt{y}}\mathrm{e}^{-\sqrt{y}}, & y>0, \\ 0, & \text{其他}. \end{cases}$$

例

8. 设随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{9}x^2, & 0 < x < 3, \\ 0, & \text{其他}. \end{cases}$$

令随机变量 $Y = \begin{cases} 2, & X \leqslant 1, \\ X, & 1 < X < 2, \\ 1, & X \geqslant 2. \end{cases}$

- 求 Y 的分布函数.
- 求概率 $P\{X \leqslant Y\}$.

解.

(1) 由题设知，$P(1 \leqslant Y \leqslant 2) = 1$. 记 Y 的分布函数为 $F_Y(y)$，则:
当 $y < 1$ 时，$F_Y(y) = 0$;
当 $y \geqslant 2$ 时，$F_Y(y) = 1$;
当 $1 \leqslant y < 2$ 时，

$$\begin{aligned} F_Y(y) &= P(Y \leqslant y) = P(Y = 1) + P(1 < Y \leqslant y) \\ &= P(X \geqslant 2) + P(1 < X \leqslant y) = \int_2^3 \frac{x^2}{9}\,\mathrm{d}x + \int_1^y \frac{x^2}{9}\,\mathrm{d}x = \frac{y^3+18}{27}. \end{aligned}$$

所以 Y 的分布函数为

$$F_Y(y) = \begin{cases} 0, & y < 1, \\ \frac{y^3+18}{27}, & 1 \leqslant y < 2, \\ 1, & y \geqslant 2. \end{cases}$$

(2) $P(X \leqslant Y) = P(X < 2) = \int_0^2 \frac{x^2}{9}\,\mathrm{d}x = \frac{8}{27}$.

例

9. 设 N 是正整数，X 服从 $[0, N^2]$ 的均匀分布.

- 求 $\sqrt{X}$ 的密度函数.
- 求 $[\sqrt{X}]$ 的分布列，这里 $[x]$ 表示不超过 x 的最大整数.
- 求 $\sqrt{X} - [\sqrt{X}]$ 的分布函数

解.

(1) 由连续型随机变量函数分布定理，所求密度函数为

$$f_{\sqrt{X}}(y)=\begin{cases}2y/N^2, & y\in[0,N],\\ 0, & \text{其他}.\end{cases}$$

(2)

$$\begin{aligned}P([\sqrt{X}]=k)&=P(k^2\le X<(k+1)^2)\\ &=\begin{cases}(2k+1)/N^2 & k\in[0,N-1]\cap Z\\ 0, & \text{其他}.\end{cases}\end{aligned}$$

解.

(3) 当 $y<0$ 时, $F_{\sqrt{X}-[\sqrt{X}]}(y)=0$;
当 $y\geqslant 1$ 时, $F_{\sqrt{X}-[\sqrt{X}]}(y)=1$;
当 $0\le y<1$ 时,

$$\begin{aligned}
F_{\sqrt{X}-[\sqrt{X}]}(y) &= P(\sqrt{X}-[\sqrt{X}]\le y)\\
&=\sum_{i=0}^{N-1} P(\sqrt{X}-[\sqrt{X}]\le y,[\sqrt{X}]=i)\\
&=\sum_{i=0}^{N-1} P(\sqrt{X}\le y+i, i\le\sqrt{X}<i+1)\\
&=\sum_{i=0}^{N-1} P(i\le\sqrt{X}\le y+i)\\
&=\frac{1}{N^2}\sum_{i=0}^{N-1}(y^2+2yi)\\
&=\frac{y^2+(N-1)y}{N}
\end{aligned}$$

例

10.

- 利用课上讲的证明“二项分布的极限是泊松分布”的办法，论证几何分布的极限和指数分布的关系.
- 利用课上讲的证明“二项分布的极限是泊松分布”的办法，论证负二项分布的极限和伽马分布的关系（这里负二项分布的参数 r 限制为整数）.

证明.

(1) 将单位时间 n 等分，第一次成功需要试验次数服从几何分布，设每一次试验成功的概率为 λ/n，每一次试验的耗时为 $1/n$，考虑试验第一次成功时间的分布函数.

$$
\begin{aligned}
P(X \le t) &= \lim_{n\to\infty} \sum_{i=1}^{[nt]} (\lambda/n)(1-\lambda/n)^{i-1} \\
&= \lim_{n\to\infty} 1-(1-\lambda/n)^{[nt]} \\
&= 1-\lim_{n\to\infty} (1-\lambda/n)^{-n/\lambda\cdot-(\lambda t)} \\
&= 1-e^{-\lambda t}.
\end{aligned}
$$

此为指数分布.

□

证明.

(2)

$$P(X \le t) = \lim_{n\to\infty} \sum_{i=0}^{[nt]-r} \binom{i+r-1}{r-1} \left(\frac{\lambda}{n}\right)^r \left(1-\frac{\lambda}{n}\right)^i$$

成功 r 次时，试验次数不超过 $[nt]$，等价于在 $[nt]$ 次试验中，至少成功 r 次的概率. 因此有

$$\begin{aligned} P(X \le t) &= \lim_{n\to\infty} \sum_{i=r}^{[nt]} \binom{[nt]}{i} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{[nt]-i} \\ &= \lim_{n\to\infty} \left(1-\sum_{i=0}^{r-1} \binom{[nt]}{i} \left(\frac{\lambda t}{nt}\right)^i \left(1-\frac{\lambda t}{nt}\right)^{[nt]-i}\right) \\ &= 1-\sum_{i=0}^{r-1} \lim_{n\to\infty} \binom{[nt]}{i} \left(\frac{\lambda t}{nt}\right)^i \left(1-\frac{\lambda t}{nt}\right)^{[nt]-i} \end{aligned}$$

由二项分布的极限分布为泊松分布得

$$F(t) = P(X \le t) = 1-\sum_{i=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

证明.

伽马分布函数（$\alpha = r$ 取整时为爱尔朗分布）:

$$
\begin{aligned}
F(t) &= 1 - \int_t^{\infty} \frac{x^{r-1}\lambda^r e^{-rx}}{\Gamma(r)}\mathrm{d}x \\
&= 1 - \int_t^{\infty} \frac{e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}\mathrm{d}(\lambda x) \\
&= 1 - \Big(\int_t^{\infty} \frac{e^{-\lambda x}(\lambda x)^{r-2}}{\Gamma(r-1)}\mathrm{d}(\lambda x) + \frac{e^{-\lambda t}(\lambda t)^{r-1}}{\Gamma(r)}\Big) \\
&= 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^k}{\Gamma(k+1)} \\
&= 1 - \sum_{k=0}^{r-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}
\end{aligned}
$$

因此负二项分布的极限为伽马分布. $\square$

谢谢大家！