

随机变量及其分布

Didnelpsun

目录

1	一维随机变量	1
1.1	随机变量概念	1
1.2	分布函数	1
1.2.1	概念	1
1.2.2	性质	1
1.2.3	应用	1
2	一维离散型随机变量	2
2.1	分布律	2
2.2	性质	2
2.3	应用	2
2.4	分布	3
2.4.1	0-1 分布	3
2.4.2	二项分布	3
2.4.3	泊松分布	3
2.4.4	几何分布	3
2.4.5	超几何分布	4
3	一维连续型随机变量	4
3.1	概率密度	4
3.2	性质	4
3.3	应用	4
3.4	分布	5
3.4.1	均匀分布	5

3.4.2	指数分布	5
3.4.3	正态分布	6
4	一维随机变量函数分布	7
4.1	离散型	7
4.2	连续性	8
5	多维随机变量	9
5.1	概念	9
5.2	联合分布函数	9
5.3	边缘分布函数	9
5.4	条件分布函数	10
6	二维离散型随机变量	10
6.1	联合分布律	10
6.2	边缘分布律	10
6.3	条件分布律	11
7	二维连续型随机变量	11
7.1	联合概率密度	11
7.2	边缘概率密度	12
7.3	条件概率密度	12
7.4	二维均匀分布	12
7.5	二维正态分布	12
8	随机变量独立性	13
8.1	概念	13
8.2	充要条件	13
8.3	性质	13
9	二维随机变量函数分布	13
9.1	离散型	14
9.2	连续型	14
9.2.1	和的分布	15
9.2.2	差的分布	15

9.2.3	积的分布	15
9.2.4	商的分布	15
9.2.5	卷积公式	15
9.3	混合型	17

1 一维随机变量

1.1 随机变量概念

定义：随机变量就是其值会随机而定的变量。设随机试验 E 的样本空间 $\Omega = \omega$ ，如果对每一个 ω 都有唯一的实数 $X(\omega)$ 与之对应，并且对任意实数 x ， $\{\omega | X(\omega) \leq x, \omega \in \Omega\}$ 是随机事件，则称定义在 Ω 上的实值单值函数 $X(\omega)$ 为随机变量，记为随机变量 X 。

1.2 分布函数

1.2.1 概念

定义：设 X 为随机变量， x 为任意实数，称函数 $F(x) = P\{X \leq x\}$ ($x \in R$ 且取遍所有实数) 为随机变量 X 的分布函数，或称 X 服从分布 $F(x)$ ，记为 $X \sim F(x)$ 。(随着 x 从 $-\infty$ 到 $+\infty$ ， $X(\omega)$ 到 \emptyset 到 Ω)

1.2.2 性质

同样是分布函数的充要条件：

- $F(x)$ 是 x 的单调不减函数，即对任意实数 $x_1 < x_2$ ，有 $F(x_1) \leq F(x_2)$ 。
- $F(x)$ 是 x 的右连续函数，即对任意 $x_0 \in R$ ，有 $\lim_{x \rightarrow x_0^+} F(x) = F(x_0 + 0) = F(x_0)$ 。(左空心右实心)
- $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ ， $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$ 。

1.2.3 应用

- $P\{X \leq a\} = F(a)$ 。
- $P\{X < a\} = F(a - 0)$ 。是指分布函数下该点左极限的概率值。
- $P(X = a) = F(a) - F(a - 0)$ 。 $\therefore P\{X \leq a\} = P\{X < a \cup X = a\} = P\{X < a\} + P\{X = a\}$ ， $\therefore P\{X = a\} = P\{X \leq a\} - P\{X < a\} = F(a) - F(a - 0)$ 。

2 一维离散型随机变量

定义：若随机变量 X 只可能取有限个或可列各值 x_1, x_2, \dots ，则称 X 为离散型随机变量。

2.1 分布律

定义： $P\{X = x_i\} = p_i, i = 1, 2, \dots$ 为 X 的分布列、分布律或概率分布，记为 $X \sim p_i$ 。

概率分布常用表格或矩阵表示：

X	x_1	x_2	\dots
P	p_1	p_2	\dots

 或 $X \sim \begin{pmatrix} x_1 & x_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$ 。

2.2 性质

数列 $\{p_i\}$ 是离散型随机变量的概率分布的充要条件是 $p_i \geq 0 (i = 1, 2, \dots)$ 且 $\sum_{i=1}^n p_i = 1$ 。

设离散型随机变量 X 的概率分布为 $P\{X = x_i\} = p_i$ ，则 X 的分布函数 $F(x) = P\{X \leq x\} = \sum_{x_i \leq x} P\{X = x_i\}$ ，即离散型随机变量的分布律函数是一个左实右空的阶梯形函数。

$P\{X = x_i\} = P\{X \leq x_i\} - P\{X < x_i\} = F(x_i) - F(x_i - 0)$ ，即某点的概率值为该点分布律值减去该点左极限的分布律值。

对实数轴上的任一集合 B 有 $P\{X \in B\} = \sum_{x_i \in B} P\{X = x_i\}$ ，特别地 $P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\} = F(b) - F(a)$ 。

2.3 应用

例题：已知随机变量 X 的概率分布为：

X	1	2	3
$P\{X = k\}$	θ^2	$2\theta(1 - \theta)$	$(1 - \theta)^2$

且 $P\{X \geq 2\} = \frac{3}{4}$ ，求未知参数 θ 与 X 的分布函数 $F(x)$ 。

解： $\because P\{X \geq 2\} = \frac{3}{4}, \therefore 2\theta(1 - \theta) + (1 - \theta)^2 = \frac{3}{4}$ ，解得 $\theta = \frac{1}{2}, -\frac{1}{2}$ 舍。

2.4 分布

2.4.1 0-1 分布

定义：若 X 的概率分布为 $X \sim \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$ ，即 $P\{X=1\}=p$ ， $P\{X=0\}=1-p$ ，则称 X 服从参数为 p 的 0-1 分布，记为 $X \sim B(1, p)$ ($0 < p < 1$)。

0-1 分布基于一次伯努利试验， X 也称为伯努利计数变量。

2.4.2 二项分布

定义：如果 X 的概率分布为 $P\{X=k\}=C_n^k p^k (1-p)^{n-k}$ ($k=0, 1, \dots, n$, $0 < p < 1$)，则称 X 服从参数为 (n, p) 的二项分布，记为 $X \sim B(n, p)$ 。

二项分布基于 n 重伯努利试验。

二项分布的分布律计算，总共进行试验 n 次，已知成功的概率为 p ，若成功了 k 次，则 k 次成功概率为 p^k ，则失败次数为 $n-k$ ，从而 $n-k$ 失败概率为 $(1-p)^{n-k}$ ，因为 n 次试验都是相互独立的，所以将成功的概率与失败的概率乘在一起。又在 n 次中成功 k 次就可以了，进行排列，所以还乘上 C_n^k 。

2.4.3 泊松分布

定义：如果 X 的概率分布为 $P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$ ($k=0, 1, \dots, n$, $\lambda > 0$)，则称 X 服从参数为 λ 的泊松分布，记为 $X \sim P(\lambda)$ 。

泊松分布基于某场合某单位时间内源源不断的质点来流的个数 $X=k$ ， λ 代表质点流动到来的强度。也可以代表稀有事件发生的概率。

2.4.4 几何分布

定义：如果 X 的概率分布为 $P\{X=k\}=(1-p)^{k-1}p$ ($k=0, 1, \dots, n$, $0 < p < 1$)，则称 X 服从参数为 p 的几何分布，记为 $X \sim G(p)$ 。

几何分布与几何无关，代表的是 n 重伯努利试验首次成功就停止试验，试验次数可以为无穷。设 X 表示伯努利试验中事件 A 首次放生所需要的试验次数，则 $X \sim G(p)$ ，其中 $p=P(A)$ 。

从而根据意义，几何分布要求前 $k-1$ 次都失败，从而概率为 $(1-p)^{k-1}$ ，最后一次成功，所以再乘上 p 。

2.4.5 超几何分布

定义：如果 X 的概率分布为 $P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$ ($\max\{0, n - N + M\} \leq k \leq \min\{M, n\}$, M, N, n 为正整数且 $M \leq N$, $n \leq N$, k 为整数), 则称 X 服从参数为 (n, N, M) 的超几何分布, 记为 $X \sim H(n, N, M)$ 。

超几何分布考的可能性很小, 事件数就是古典概型的一个特例。

如有 N 件产品, 其中 M 件正品, 从而 $N - M$ 件次品, 任取 n 个, 则取出 k 件正品的概率就是超几何分布。

3 一维连续型随机变量

定义：若随机变量 X 的分布函数可以表示为 $F(x) = \int_{-\infty}^x f(t) dt$ ($x \in R$ 且取遍所有实数), 其中 $f(x)$ 是非负可积函数, 则 X 为连续型随机变量。

3.1 概率密度

定义： $f(x)$ 称为 X 的概率密度函数, 简称概率密度, 记为 $X \sim f(x)$ 。

3.2 性质

改变 $f(x)$ 有限各点的值 $f(x)$ 仍是概率密度 (因为单个点没有面积), $f(x)$ 为某一随机变量 X 的概率密度的充分必要条件: $f(x) \geq 0$, 且 $\int_{-\infty}^{+\infty} f(x) dx = 1$ 。

若 X 为连续型随机变量, $X \sim f(x)$, 则对任意实数 c 有 $P\{X = c\} = 0$ 。

对实数轴上的任一集合 B 有 $P\{X \in B\} = \int_B f(x) dx$, 特别地 $P\{a < X < b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} = \int_a^b f(x) dx = F(b) - F(a)$ 。

3.3 应用

例题：已知随机变量 X 的概率密度为
$$\begin{cases} Ax, & 1 < x < 2 \\ B, & 2 \leq x < 3 \\ 0, & \text{其他} \end{cases}$$
, 且 $P\{1 < X < 2\} = P\{2 < X < 3\}$, 求常数 AB , 分布函数 $F(x)$ 以及概率 $P\{2 < X < 4\}$ 。

解：由于归一性 $\int_{-\infty}^{+\infty} f(x) dx = 1$, $\therefore \int_1^2 Ax dx + \int_2^3 B dx = 1$ 。

$\therefore \frac{3}{2}A + B = 1$ 。又 $P\{1 < X < 2\} = P\{2 < X < 3\}$ 。

$$\therefore \int_1^2 Ax \, dx = \int_2^3 B \, dx, \text{ 即 } \therefore \int_1^2 Ax \, dx = \int_2^3 B \, dx = \frac{1}{2}, \quad A = \frac{1}{3}, \quad B = \frac{1}{2}.$$

$$f(x) = \begin{cases} \frac{1}{3}x, & 1 < x < 2 \\ \frac{1}{2}, & 2 \leq x < 3 \\ 0, & \text{其他} \end{cases}, \quad \therefore F(x) = \int_{-\infty}^x f(t) \, dt.$$

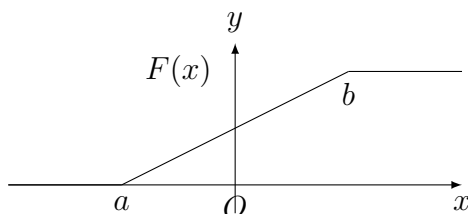
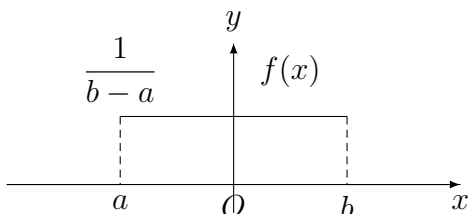
$$\therefore F(x) = \begin{cases} 0, & x < 1 \\ \int_1^x \frac{1}{3}t \, dt = \frac{x^2}{6} - \frac{1}{6}, & 1 \leq x < 2 \\ \int_1^2 \frac{1}{3}x \, dx + \int_2^x \frac{1}{2} \, dx = \frac{1}{2}x - \frac{1}{2}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

3.4 分布

3.4.1 均匀分布

定义：如果 X 的概率密度或分布函数分别为 $f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$,

$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$, 则称 X 在区间 (a, b) 上服从均匀分布, 记为 $X \sim U(a, b)$ 。



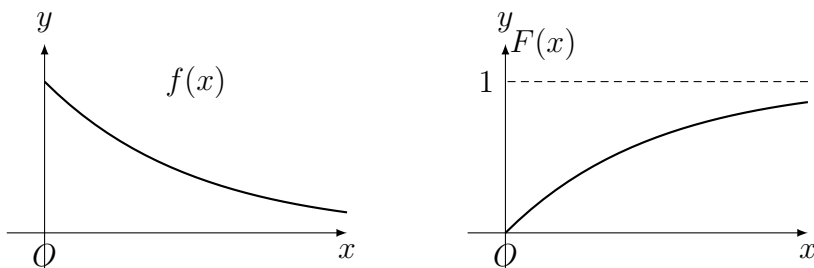
几何概型在一维情况下就是几何分布。

若 X 在区间 I 上的任一子区间取值的概率与该子区间的长度成正比, 则 $X \sim U(a, b)$ 。

3.4.2 指数分布

定义：如果 X 的概率密度或分布函数分别为 $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{其他} \end{cases}$,

$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$, 则称 X 在区间 (a, b) 上服从参数为 λ 的指数分布, 记为 $X \sim E(\lambda)$ 。



指数分布中 λ 代表失效率，往往用来代表一个事物毁坏的过程，如灯泡毁坏。

定理：无记忆性：若 X 服从指数分布，则 $P\{X > s+t | X > s\} = P\{X > t\}$ 。

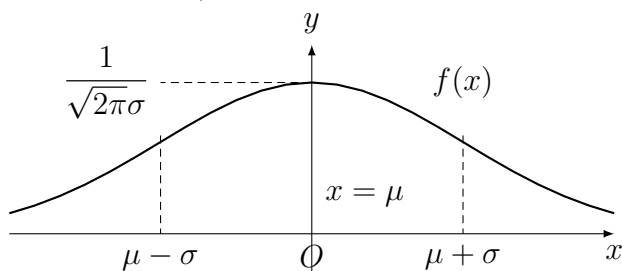
即在指数分布下事情发生的概率与前面所经过的时间无关，如果 T 是某一元件的寿命，已知元件使用了 t 小时，它总共使用至少 $s+t$ 小时的条件概率，与从开始使用时算起它使用至少 s 小时的概率相等。

$$\begin{aligned} \text{证明：} P\{X > s+t | X > s\} &= \frac{P\{X > s+t\}}{P\{X > s\}} = \frac{1 - P\{X \leq s+t\}}{1 - P\{X \leq s\}} \\ &= \frac{1 - F(s+t)}{1 - F(s)} = \frac{1 - (1 - e^{-\lambda(s+t)})}{1 - (1 - e^{-\lambda s})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - (1 - e^{-\lambda t}) \\ &= 1 - F(t) = 1 - P\{X \leq t\} = P\{X > t\}. \end{aligned}$$

3.4.3 正态分布

定义：如果 X 的概率密度为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ ($-\infty < x < +\infty$, $-\infty < \mu < +\infty$, $\sigma > 0$)，则称 X 服从参数为 (μ, σ^2) 的正态分布，称 X 为正态变量，记为 $X \sim N(\mu, \sigma^2)$ 。

$f(x)$ 的图形关于 $x = \mu$ 对称，即 $f(\mu - x) = f(\mu + x)$ ，并在 $x = \mu$ 处有唯一最大值 $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$ 。 $\mu - \sigma$ 和 $\mu + \sigma$ 为拐点。



当 $\mu = 0$, $\sigma = 1$ 时的正态分布 $N(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 为标准正态分布，记为 $\phi(x)$ ，分布函数为 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ 。 $\phi(x)$ 为偶函数， $\Phi(0) = \frac{1}{2}$ ， $\Phi(-x) = 1 - \Phi(x)$ 。

若 $X \sim N(0, 1)$ ， $P\{X > \mu_\alpha\} = \alpha$ ，则称 μ_α 为标准正态分布的上侧 α 分位数/上 α 分位点。

若 $X \sim N(\mu, \sigma^2)$, 则

- $F(x) = P\{X \leq x\} = P\left\{\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right\} = \Phi\left(\frac{x - \mu}{\sigma}\right)$ 。(标准化)
- $F(\mu - x) + F(\mu + x) = 1$ 。
- $P\{a < X < b\} = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$ 。(标准化得到)
- $aX + b \sim N(a\mu + b, a^2\sigma^2)$ ($a \neq 0$)。

4 一维随机变量函数分布

设 X 为随机变量, 函数 $y = g(x)$, 则以随机变量 X 作为自变量的函数 $Y = g(X)$ 也是随机变量, 称为随机变量 X 的函数。

如 $Y = aX^2 + bX + c$ 等等。

4.1 离散型

设 X 为离散型随机变量, 其概率分布为 $p_i = P\{X = x_i\}$ ($i = 1, 2, \dots$), 则 X 的函数 $Y = g(X)$ 也是离散型随机变量, 其概率分布为 $P\{Y = g(x_i)\} = p_i$ ($i = 1, 2, \dots$)。

$$\text{即 } Y \sim \begin{pmatrix} g(x_1) & g(x_2) & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix}.$$

若有若干个 $g(x_i)$ 相同, 则合并为一项, 并将对应概率相加作为 Y 取 $g(x_i)$ 的概率。

离散型一维随机变量函数分布单独考的可能性很低。

例题: 设 X 是仅可能取 6 个值的离散型随机变量, 分布为:

X	-2	-1	0	1	2	3
P	0.05	0.15	0.20	0.25	0.20	0.15

求 $Y = 2X + 1$, $Z = X^2$ 的概率分布。

因为 $Y = 2X + 1$ 是线性的, 所以改变 X 变为 Y , 所对应的 P 不变:

Y	-3	-1	1	3	5	7
P	0.05	0.15	0.20	0.25	0.20	0.15

对于 $Z = X^2$ 是一个平方, 导致 Z 的值有些是一样的, 所以概率合在一起:

Z	0	1	4	9
P	0.20	0.40	0.25	0.15

4.2 连续性

设 X 为离散型随机变量, 其分布函数、概率密度为 $F_X(x)$ 与 $f_X(x)$, 随机变量 $Y = g(X)$ 也是 X 的函数, 则 Y 的分布函数或概率密度可用分布函数法得到 $F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = \int_{g(X) \leq y} f_X(x) dx$ 。

若 $F_Y(y)$ 连续, 且除有限个点外, $F'_Y(y)$ 存在且连续, Y 的概率密度为 $f_Y(y) = F'_Y(y)$ 。

首先已知 X 的概率密度函数为 $f_X(x)$, 分布函数为 $F_X(x)$, 已知 $Y = g(X)$, 即 Y 对 X 的映射关系。现在要求 Y 的概率规律, 即要求 Y 的概率密度 $f_Y(y)$ 与分布函数 $F_Y(y)$ 。

先求分布函数 $F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\}$, 然后用 y 来表示 X , 这是连续性随机变量函数分布的重点。

即 X 在以 y 表示的一个区间上, $X \in I_y$, 所以解得 Y 分布函数 $\int_{I_y} f_X(x) dx$ 。

例题: 设随机变量 X 的概率密度为 $f_X(x) = \begin{cases} 1+x, & -1 \leq x < 0 \\ 1-x, & 0 \leq x \leq 1 \\ 0, & \text{其他} \end{cases}$, 求随

机变量 $Y = X^2 + 1$ 的分布函数。

解: 求随机变量 $Y = X^2 + 1$ 的分布函数即求 $F_Y(y) = P\{Y \leq y\} = P\{X^2 + 1 \leq y\}$ 。即将 $X^2 + 1$ 与 y 的概率关系解出, 即求曲线 $X^2 + 1$ 与直线 y 的关系。

根据 X 的概率密度函数, 所以只有 $x \in [-1, 1]$ 才有正概率, 其他区间概率为 0, 即不能取, 将 Y 的取值范围划在 $[-1, 1]$ 中。

又由 $f_X(x)$ 的关系知道 Y 的函数, 是在 $[-1, 1]$ 的属于 $[1, 2]$ 的抛物线。

当 $y < 1$ 时, $Y = X^2 + 1 > 1$ 恒成立, 所以 $X^2 + 1 \leq y$ 不可能发生, 概率为 0, 所以 $F_Y(y) = 0$ 。

当 $y > 2$ 时, $Y = X^2 + 1$ 在 $X \in [-1, 1]$ 时 $Y \in [1, 2]$, 所以 $X^2 + 1 \leq y$ 必然成立, 所以 $F_Y(y) = 1$ 。

当 $1 < y < 2$ 时, 解出 $X^2 + 1 \leq y$ 为 $X = \pm\sqrt{y-1}$, 所以 $F_Y(y) = P\{-\sqrt{y-1} \leq X \leq \sqrt{y-1}\} = \int_{-\sqrt{y-1}}^{\sqrt{y-1}} f(x) dx = \int_{-\sqrt{y-1}}^0 1+x dx + \int_0^{\sqrt{y-1}} 1-x dx = 2\sqrt{y-1} - y + 1$ 。

5 多维随机变量

5.1 概念

多维随机变量**定义**：如果 X_1, X_2, \dots, X_n 是定义在同一个样本空间上的 n 个随机变量，则称 (X_1, X_2, \dots, X_n) 为 n 维随机变量或 n 维随机向量， X_i 为第 i 个分量。

当 $n = 2$ 时， (X, Y) 为二维随机变量/二维随机向量。

5.2 联合分布函数

定义：对任意 n 个实数 x_1, x_2, \dots, x_n 称为 n 元函数 $F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$ 为 n 为随机变量 (X_1, X_2, \dots, X_n) 的联合分布函数。

当 $n = 2$ 时对任意的实数 xy 称二元函数 $F(x, y) = P\{X \leq x, Y \leq y\}$ 为二维随机变量 (X, Y) 的联合分布函数，简称分布函数，记为 $(X, Y) \sim F(x, y)$ 。

性质：

- 单调性： $F(x, y)$ 是 xy 的单调不减函数。
- 右连续性： $F(x, y)$ 在右边连续。
- 有界性：当 x 或 y 趋向负无穷时值为 0，当 x 和 y 趋向正无穷时值为 1。
- 非负性：对任意 $x_1 < x_2, y_1 < y_2$ 有 $P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0$ 。（由定义画图可知）

5.3 边缘分布函数

定义：设二维随机变量 (X, Y) 的联合分布函数为 $F(x, y)$ ，随机变量 X, Y 的分布函数 $F_X(x)$ 与 $F_Y(y)$ 分别称 (X, Y) 关于 X 与关于 Y 的边缘分布函数。

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y < +\infty\} = \lim_{y \rightarrow +\infty} P\{X \leq x, Y \leq y\} = \lim_{y \rightarrow +\infty} F(x, y) = F(x, +\infty)。$$

同理 $F_Y(y) = F(+\infty, y)。$

求边缘分布函数的口诀：求谁不积谁，不积先定限，限内画条线，先交写下限，后交写上限。

即若求 x 则积 y ，然后根据 $f(x, y)$ 联合概率密度的图形求 x 的取值区间。然后在 $f(x, y)$ 联合概率密度共同的取值区间内画一条水平的线 $y = y_0$ （若求 y ）

或垂直的线 $x = x_0$ (若求 x)，则从左到右从下到上先与所画线相交的区间边缘的线就是边缘概率密度的下限，后面相交的就是上限。最后将被积函数的联合概率密度加到中间就是最后的边缘分布函数。

5.4 条件分布函数

定义：对于二维随机变量 (X, Y) ，可以考虑在其中一个随机变量取得可能的固定值的条件下，另一随机变量的概率分布，这样得到的 X 或 Y 的概率分布叫做条件概率分布，简称条件分布。

定理：条件概率密度 $= \frac{\text{联合概率密度}}{\text{边缘概率密度}}$ 。

6 二维离散型随机变量

定义：若二维随机变量 (X, Y) 只能取有限或可列对值 $(x_1, y_1), (x_2, y_2), \dots$ 则称 (X, Y) 为二维离散型随机变量。

6.1 联合分布律

定义： $p_{ij} = P\{X = x_i, Y = y_j\}$, $i, j = 1, 2, \dots$ 为 (X, Y) 的分布律或随机变量 X 和 Y 的联合分布律，记为 $(X, Y) \sim p_{ij}$ 。

数列 $\{p_{ij}\}$, $i, j = 1, 2, \dots$ 是某一二维离散型随机变量的概率分布的充要条件是 $p_{ij} \geq 0$, $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$ 。

定义：若 p_{ij} , $i, j = 1, 2, \dots$ 为 (X, Y) 的概率分布，则 (X, Y) 的联合分布函数为 $F(x, y) = P\{X \leq x, Y \leq y\} = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij}$ 。

联合分布函数是以 (x, y) 为定点的左下角平面上 (X, Y) 所有可能取值的概率的和。

设 G 是平面上某个区域，则 $P\{(X, Y) \in G\} = \sum_{(x_i, y_j) \in G} p_{ij}$ 。

6.2 边缘分布律

定义：对于同一个 x 值的所有 y 取值的概率的和，就是该 x 值的边缘分布律。同理对于同一个 y 值的所有 x 取值的概率的和，就是该 y 值的边缘分布律。

即 $p_{i\cdot} = P\{X = x_i\} = \sum_{j=1}^{\infty} P\{X = x_i, Y = y_j\} = \sum_{j=1}^{\infty} p_{ij}$ ($i = 1, 2, \dots$)。

$$p_{\cdot j} = P\{Y = y_j\} = \sum_{i=1}^{\infty} P\{X = x_i, Y = y_j\} = \sum_{i=1}^{\infty} p_{ij} \quad (j = 1, 2, \dots).$$

6.3 条件分布律

条件分布律类比随机事件概率中的条件概率。

定义：如果 $(X, Y) \sim p_{ij}$ ($i, j = 1, 2, \dots$)，对固定的 j ，如果 $p_{\cdot j} = P\{Y = y_j\} > 0$ ，则称 $P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{\cdot j}}$ ($i = 1, 2, \dots$) 为 X 在 $Y = y_j$ 条件下的条件分布。

同理**定义：** Y 在 $X = x_i$ 条件下的条件分布为 $P\{Y = y_j | X = x_i\} = \frac{p_{ij}}{p_{i\cdot}}$ ($j = 1, 2, \dots$)。

7 二维连续型随机变量

定义：如果二维随机变量 (X, Y) 的联合分布函数 $F(x, y)$ 可表示为 $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$, ($(x, y) \in R^2$)，其中 $f(x, y)$ 为非负可积函数，则称 (X, Y) 为二维连续型随机变量， $f(x, y)$ 为 (X, Y) 的概率密度，记为 $(X, Y) \sim f(x, y)$ 。

二元函数 $f(x, y)$ 是概率密度的充要条件 $f(x, y) \geq 0$, $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$ 。

改变 $f(x, y)$ 有限个点值 (仍取非负值)， $f(x, y)$ 仍是概率密度。

7.1 联合概率密度

定义：设 (X, Y) 的联合分布函数为 $F(x, y)$ ，概率密度为 $f(x, y)$ ，则

- $F(x, y)$ 为 (x, y) 的二元连续函数，且 $F(x, y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv$ 。
- 设 G 为平面上某个区域，则 $P\{(X, Y) \in G\} = \iint_G f(x, y) dx dy$ 。
- 若 $f(x, y)$ 在点 (x, y) 处连续，则 $\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$ 。
- 若 $F(x, y)$ 连续可导，则 (X, Y) 是连续型随机变量，则 $\frac{\partial^2 F(x, y)}{\partial x \partial y}$ 是其概率密度。

7.2 边缘概率密度

定义： 设 $(X, Y) \sim f(x, y)$ ，则 X 的边缘分布函数为 $F_X(x) = F(x, +\infty) = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f(u, v) dv \right] du$ ，所以 X 为连续型随机变量，其概率密度 $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$ ，称 $f_X(x)$ 为 (X, Y) 关于 X 的**边缘概率密度**。同理 Y 也为连续型随机变量，其概率密度为 $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$ 。

7.3 条件概率密度

定义： 设 $(X, Y) \sim f(x, y)$ ，边缘概率密度 $f_X(x) > 0$ ，则称 $f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$ 为 Y 在 $X = x$ 条件下的**条件概率密度**。同理 X 在 $Y = y$ 条件下的条件概率密度为 $f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$ 。

若 $f_X(x) > 0, f_Y(y) > 0$ ，则有概率密度乘法公式 $f(x, y) = f_X(x)f_{Y|X}(y|x) = f_Y(y)f_{X|Y}(x|y)$ 。

定义： Y 在 $X = x$ 条件下的**条件分布函数**为 $F_{Y|X}(y|x) = \int_{-\infty}^y f_{Y|X}(y|x) dy = \int_{-\infty}^y \frac{f(x, y)}{f_X(x)} dy$ ，同理 X 在 $Y = y$ 条件下的条件分布函数为 $F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(x|y) dx = \int_{-\infty}^x \frac{f(x, y)}{f_Y(y)} dx$ 。

7.4 二维均匀分布

定义： 若 (X, Y) 的概率密度为 $f(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D \\ 0, & \text{其他} \end{cases}$ ， S_D 为区域

D 的面积，则称 (X, Y) 在平面有界区域 D 上服从**均匀分布**。

二维均匀分布就是几何概型的二维情况。

7.5 二维正态分布

定义： 若 (X, Y) 的概率密度为：

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

其中 $\mu_1, \mu_2 \in R, \sigma_1, \sigma_2 > 0, -1 < \rho < 1$ ，则称 (X, Y) 服从参数为 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 的**二维正态分布**，记为 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ 。此时：

- $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ ， ρ 为 X 与 Y 的相关系数，即 $\rho = \frac{Cov(X, Y)}{\sqrt{DX}\sqrt{DY}} = \frac{Cov(X, Y)}{\sigma_1\sigma_2}$ 。
- X, Y 的条件分布都是正态分布。

- $aX + bY$ ($a \neq 0$ 或 $b \neq 0$) 服从正态分布。
- XY 相互独立的充要条件是 XY 不相关, 即 $\rho = 0$ 。

8 随机变量独立性

8.1 概念

定义: 设随机变量 X, Y 的联合分布函数为 $F(x, y)$, 边缘分布函数为 $F_X(x)$, $F_Y(y)$, 若对任意实数 x, y , 有 $P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}$, 即 $F(x, y) = F_X(x)F_Y(y)$, 则称随机变量 X 和 Y 相互独立。

8.2 充要条件

若 X, Y 独立, 则

- 对于二维离散型随机变量, $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ 。
- 对于二维连续型随机变量, $f(x, y) = f_X(x) \cdot f_Y(y)$ 。
- 对于二维随机变量, $F(x, y) = F_X(x) \cdot F_Y(y)$ 。(联合分布函数等于各自边缘函数乘积)

8.3 性质

若 X, Y 独立, 则 $f(X)$ 与 $g(Y)$ 也相互独立。

若对所有 x_1, x_2, \dots, x_n 有 $F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n)$, 则 X_1, X_2, \dots, X_n 相互独立。

若 X_1, X_2, \dots, X_n 相互独立, 其中任意 k 个随机变量也相互独立。

若 X_1, X_2, \dots, X_n 相互独立, 则其各自的函数 $g_1(X_1)g_2(X_2) \cdots g_n(X_n)$ 也相互独立。

9 二维随机变量函数分布

定义: 设 X, Y 为随机变量, $g(x, y)$ 为二元函数, 则以随机变量 X, Y 作为变量的函数 $U = g(X, Y)$ 也是随机变量, 称为随机变量 X, Y 的函数。

9.1 离散型

对于 (离散型, 离散型) 随机变量函数分布也是离散型。

例题：将两封信投入 3 个信箱，设 X_1, X_2 分别表示第一个和第二个信箱投进的信的数量。

(1) (X_1, X_2) 的联合分布，边缘分布并判断其独立性。

(2) 在条件 $X_2 = 1$ 下， X_1 的条件分布。

(3) 随机变量 $Y_1 = X_1 + X_2$ 和 $Y_2 = X_1 - X_2$ 的分布。

(1) 解：因为该题目是二维离散型随机变量，所以分析：

$x_1 \backslash x_2$	0	1	2	X_1 边缘
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
X_2 边缘	4/9	4/9	1/9	1

从而联合分布就是表格最中心的部分，边缘分布就是两边。

如 $p_{11} = \frac{1}{9} \neq p_{1\cdot} p_{\cdot 1} = \frac{4}{9} \frac{4}{9} = \frac{16}{81}$ 。所以 $X_1 X_2$ 不独立。

(2) 解： $P\{X_1 = 0 | X_2 = 1\} = \frac{P\{X_1 = 0, X_2 = 1\}}{P\{X_2 = 1\}} = \frac{p_{01}}{p_{\cdot 1}} = \frac{2/9}{4/9} = \frac{1}{2}$ 。

$P\{X_1 = 1 | X_2 = 1\} = \frac{P\{X_1 = 1, X_2 = 1\}}{P\{X_2 = 1\}} = \frac{p_{11}}{p_{\cdot 1}} = \frac{2/9}{4/9} = \frac{1}{2}$ 。

$P\{X_1 = 2 | X_2 = 1\} = \frac{P\{X_1 = 2, X_2 = 1\}}{P\{X_2 = 1\}} = \frac{p_{21}}{p_{\cdot 1}} = \frac{0}{4/9} = 0$ 。

(3) 解：根据表格， $Y_1 = X_1 + X_2$ 的可能取值为 0, 1, 2。

即 Y_1 的概率分布为 $Y_1 \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{pmatrix}$ 。

$Y_2 = X_1 - X_2$ 的可能取值为 -2, -1, 0, 1, 2。

Y_2 的概率分布为 $Y_2 \sim \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} & \frac{1}{9} \end{pmatrix}$ 。

9.2 连续型

对于 (连续型, 连续型) 随机变量函数分布也是连续型。可以采用分布函数法和卷积公式。

9.2.1 和的分布

定理: 设 $(X, Y) \sim f(x, y)$, 则 $Z = X + Y$ 的概率密度为 $f_Z(z) = \int_{-\infty}^{+\infty} f(x, z - x) dx = \int_{-\infty}^{+\infty} f(z - y, y) dy$ 。

证明: 利用定义法 $F_Z(z) = P\{Z \leq z\} = P\{X + Y \leq z\} = \iint_D f(x, y) dx dy$,
 $D = x + y \leq z$ 。

$$\begin{aligned} &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x, y) dy. \text{ 又 } f_Z(z) = F'_Z(z), = \left(\int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x, y) dy \right)'_z \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f(x, y) dy \right)'_z dx = \int_{-\infty}^{+\infty} f(x, z - x) dx \end{aligned}$$

9.2.2 差的分布

定理: 设 $(X, Y) \sim f(x, y)$, 则 $Z = X - Y$ 的概率密度为 $f_Z(z) = \int_{-\infty}^{+\infty} f(x, x - z) dx = \int_{-\infty}^{+\infty} f(z + y, y) dy$ 。

9.2.3 积的分布

定理: 设 $(X, Y) \sim f(x, y)$, 则 $Z = XY$ 的概率密度为 $f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) dx = \int_{-\infty}^{+\infty} \frac{1}{|y|} f\left(\frac{z}{y}, y\right) dy$ 。

9.2.4 商的分布

定理: 设 $(X, Y) \sim f(x, y)$, 则 $Z = \frac{X}{Y}$ 的概率密度为 $f_Z(z) = \int_{-\infty}^{+\infty} |y| f(yz, y) dy$ 。

这里不积 y 只积 x , 因为 $x = zy$, 所以积分简单从而只积 x 。

9.2.5 卷积公式

积谁不换谁, 换完求偏导。

当 XY 相互独立时, 对于联合概率密度就能变成边缘概率密度的乘积, 得到卷积公式**定义:**

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y) dy. \\ f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(x - z) dy = \int_{-\infty}^{+\infty} f_X(y + z) f_Y(y) dx. \\ f_Z(z) &= \int_{-\infty}^{+\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right) dx = \int_{-\infty}^{+\infty} \frac{1}{|y|} f_X\left(\frac{z}{y}\right) f_Y(y) dy. \\ f_Z(z) &= \int_{-\infty}^{+\infty} f_X(zy) f_Y(y) dy. \end{aligned}$$

例题: 设随机变量 X, Y 相互独立, 其概率密度分别为

$$f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{其他} \end{cases}, \quad f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{其他} \end{cases},$$

(1) 求 (X, Y) 概率密度。

(2) 求 $Z = X + Y$ 的概率密度。

(1) 解: 因为 X, Y 相互独立, 所以 $f(X, Y) = f_X(x) \cdot f_Y(y)$ 。

$$\therefore f(X, Y) = \begin{cases} 2e^{-x}y, & x > 0 \cap 0 < y < 1 \\ 0, & \text{其他} \end{cases}。$$

(2) 解:

第一种方法是分布函数法, 即使用定义来解决随机变量函数分布。

$F_Z(z) = P\{Z \leq z\} = P\{X + Y \leq z\}$ 。 XY 的联合分布图形是一个高为 1 长无穷的长方形。利用直线 $x = x_0$ 与定义区间与 $y \leq -x + z$ 对 z 进行范围讨论。

若 $z < 0$, $X + Y \leq z$ 不可能, 即 D 与 $y \leq -x + z$ 交集为空, 所以 $F_Z(z) = 0$ 。

若 $0 \leq z < 1$, 则 $y \leq -x + z$ 与区间相交于 $(0, z)$ 和 $(z, 0)$, 利用直线 $x = x_0$ 对区间求积分 xy 上下限, $F_Z(z) = \int_0^z dx \int_0^{z-x} 2e^{-x}y dy = z^2 - 2z - 2e^{-z} + 2$ 。

若 $z > 1$, 则 $y \leq -x + z$ 与区间交集是一个梯形, 若先积 y 后积 x , 积分区间要分为两个部分来计算, 所以改变积分顺序, 先积 y 再积 x , 用 $y = y_0$ 来计算积分上下限得到 $F_Z(z) = \int_0^1 dy \int_0^{z-y} 2e^{-x}y dx = 1 - 2e^{-z}$ 。

$$\therefore f_Z(z) = F'_Z(z) = \begin{cases} 2z + 2e^{-z} - 2, & 0 < z < 1 \\ 2e^{-z}, & z > 1 \\ 0, & \text{其他} \end{cases}。 \quad (\text{等号在哪里无所谓})$$

第二种方法是卷积公式法。 $f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx$, 现在已知 $f_X(x)$, 所以只用求 $f_Y(z-x)$ 。

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

$$f_X(x)f_Y(z-x) = \begin{cases} 2e^{-x}(z-x), & 0 < x < z < x+1 \\ 0, & \text{其他} \end{cases}, \quad \text{所得的区域是梯形。}$$

当 $z < 0$, 则 $z = z_0$ 与图形区间无交集, 所以 $f_Z(z) = 0$ 。

当 $0 < z < 1$ 时, $f_Z(z) = \int_0^z 2e^{-x}(z-x) dx = 2z + 2e^{-z} - 2$ 。

当 $z > 1$ 时, $f_Z(z) = \int_{z-1}^z 2e^{-x}(z-x) dx = 2e^{-z}$ 。

例题: 设二维随机变量 (X, Y) 在矩形区域 $D = \{(x, y) | 0 \leq x \leq 2, 0 \leq y \leq 1\}$ 上服从均匀分布, 求边长为 X 和 Y 的矩形面积 Z 的概率密度。

$$\text{解: } (X, Y) \text{ 在矩形区域 } D \text{ 服从均匀分布, } \therefore f(x, y) = \begin{cases} \frac{1}{2}, & (x, y) \in D \\ 0, & \text{其他} \end{cases}。$$

根据区间: $0 \leq x \leq 2$, $0 \leq \frac{z}{x} = y \leq 1$, $0 \leq z \leq x \leq 2$, 所以区域就是底 2 高 2 的下三角。

$$\text{又 } Z = XY \sim f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) dx, \text{ 且 } f\left(x, \frac{z}{x}\right) = \frac{1}{2}.$$

$$\text{当 } 0 < z < 2 \text{ 时, } f_Z(z) = \int_z^2 \frac{1}{x} \frac{1}{2} dx = \frac{1}{2}(\ln 2 - \ln z).$$

$$\text{即 } f_Z(z) = \begin{cases} \frac{1}{2}(\ln 2 - \ln z), & 0 < z < 2 \\ 0, & \text{其他} \end{cases}.$$

9.3 混合型

对于 (离散型, 连续型) 随机变量函数分布是连续型。可以采用全概率公式。

即使用全集分解思想解决, 若 A 是离散型随机变量, B 是连续型随机变量, 则 $P(B) = P(B\Omega) = P(BA_1 \cup BA_2 \cup \cdots) = P(BA_1) + P(BA_2) \cdots = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots$ 。这就是全概率公式。

例题: 设随机变量 X 与 Y 相互独立, 其中 X 概率分布为 $\begin{pmatrix} 1 & 2 \\ 0.3 & 0.7 \end{pmatrix}$, 而 Y 的概率密度为 $f(y)$, 求随机变量 $U = X + Y$ 的概率密度 $g(u)$ 。

解: $G(u) = P\{U \leq u\} = P\{X + Y \leq u\}$ 。

$= P\{X + Y \leq u, \Omega\} = P\{X + Y \leq u, \{X = 1\} \cup \{X = 2\}\} = P\{(X + Y \leq u, X = 1) \cup (X + Y \leq u, X = 2)\} = P\{X + Y \leq u, X = 1\} + P\{X + Y \leq u, X = 2\} = P(X = 1)P\{X + Y \leq u|X = 1\} + P(X = 2)P\{X + Y \leq u|X = 2\} = 0.3P\{Y \leq u - 1|X = 1\} + 0.7P\{Y \leq u - 2|X = 2\}$, 又 XY 相互独立, 所以 X 的条件不影响 Y 的概率, $= P\{Y \leq u - 1\} + P\{Y \leq u - 2\} = 0.3F_Y(u - 1) + 0.7F_Y(u - 2)$ 。

$$g(u) = G'(u) = 0.3f_Y(u - 1) \cdot 1 + 0.7f_Y(u - 2) \cdot 1.$$