概率论笔记

作者：肖程哲

时间：August 21, 2022

苟日新，日日新，又日新

目录

第 2 章 随机变量

内容提要

- 离散与连续随机变量
- 一元与多元
- cdf, pmf, pdf
- 条件分布
- 独立随机变量
- 随机变量函数的分布
- 次序随机变量

在概率论中，主要关心 X 取值于数值集合 $\mathcal{X}$ 中某个子集 B 的可能性，即希望得到 $\mathbb{P}(\{\omega\in\Omega:X(\omega)\in B\})$。概率论不关心具体的样本点 $\omega\in\Omega$，而关注其集合 $\{\omega\in\Omega:X(\omega)\in B\}$，将其记为 $X^{-1}(B)$。由于 $\mathbb{P}$ 定义在 $\mathscr{F}$ 上，故需 $X^{-1}(B)\in\mathscr{F}$。

定义 2.1 (可测性)

设所有值得关心的 $B\subset\mathcal{X}$ 组成 $\mathscr{F}_{\mathcal{X}}$，且 $\forall B\in\mathscr{F}_{\mathcal{X}}$ 都满足 $X^{-1}(B)\in\mathscr{F}$，则称 X 为 $\mathscr{F}/\mathscr{F}_{\mathcal{X}}$ **可测的**。当 $\mathscr{F}_{\mathcal{X}}$ 不引起混淆时，简记为关于 $\mathscr{F}$ **可测**，写作 $X\in\mathscr{F}$。

由于原像保持交、并、补等集合运算，且 $\mathscr{F}$ 是 σ 代数，可将 $\mathscr{F}_{\mathcal{X}}$ 扩张为合适的最小的 σ 代数，即 $\sigma(\mathscr{F}_{\mathcal{X}})$，因此可测映射的定义不妨只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 代数 的情况.

定义 2.2 (随机变量)

为了表示因随机性而变动的量，称可测映射(measurable mapping)

$$X:(\Omega,\mathscr{F},\mathbb{P})\to(\mathcal{X},\mathscr{F}_{\mathcal{X}}),\quad \omega\in\Omega\mapsto X(\omega)\in\mathcal{X}$$

为**随机元** (random element)，也称**随机变量** (random variable)。其中 $\mathscr{F}_{\mathcal{X}}$

由于只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 代数的情况，可将随机变量看作将原概率空间映射到新概率空间的方式。新样本空间为 $\mathcal{X}$（一般由Borel 点集 构成），新事件域为 $\mathscr{F}_{\mathcal{X}}$，概率测度等于对应原像的。

注 使用随机变量 X 时，有两个可能的含义：

- X 的 (随机) 取值
- X 的分布

定义 2.3 (离散与连续随机变量)

若 $\mathcal{X}$ 是（至多可数的）离散点集，$\mathscr{F}_{\mathcal{X}}$ 由 $\mathcal{X}$ 的所有子集组成，则称 X 为**离散随机变量** (discrete random varible)。若 $\mathcal{X}=\mathbb{R}^n$，$\mathscr{F}_{\mathcal{X}}$ 为 $\{\prod_{i=1}^n(-\infty,x_i]:x_1,\dots,x_n\in\mathbb{R}\}$ 生成的 Borel 代数 (最小的 σ 代数)，则称其为**连续随机变量** (continuous random varible)。

2.1 随机变量的分布

定义 2.4 (分布)

称随机元 X 诱导的概率测度

$$Q(\bullet)=\mathbb{P}\{X\in\bullet\},\ \bullet\in\mathscr{F}_{\mathcal{X}}$$

为 X 的**概率分布** (distribution/law)。

注 对于随机变量，其取值是随机的，但其分布是固定的。

定义 2.5 (分布函数)

若函数 $F:\mathbb{R}\to[0,1]$ 满足以下性质，则称为单变量分布函数：

单调性 $F(x_1)\le F(x_2),\quad \forall x_1<x_2$

有界性 $\lim_{n\to-\infty}F(x)=0,\quad \lim_{n\to\infty}F(x)=1$

右连续性 $\lim_{x\to x_0^+}F(x)=F(x_0)$

命题 2.1

对每个分布 Q 都存在唯一一个分布函数 F_Q 使得 $F_Q(x)=Q[(-\infty,x]],\quad \forall x\in\mathbb{R}$ 成立。

证明

$$\begin{aligned}
&\because \forall x_1>x_2\in\mathbb{R},\ (-\infty,x_1]=(-\infty,x_2]\cup(x_2,x_1]\\
&\therefore \forall x_1>x_2\in\mathbb{R},\ Q[(-\infty,x_1]]=Q[(-\infty,x_2]]+Q[(x_2,x_1]]\\
&\Rightarrow F_Q(x_1)=F_Q(x_2)+\mathbb{P}\{X\in(x_2,x_1]\}\ge F_Q(x_2)
\end{aligned}$$

即 F_Q 满足单调性。

由于 $F_Q(x)=\mathbb{P}\{X\in(-\infty,x]\}$，所以 $0\le F(x)\le 1$。由 $F(x)$ 的单调性知，对任意整数 m 和 n，有

$$\lim_{x\to-\infty}F(x)=\lim_{m\to-\infty}F(m),\ \lim_{x\to+\infty}F(x)=\lim_{n\to+\infty}F(n)$$

又由概率的可列可加性得

$$\begin{aligned}
1=P(-\infty<X<+\infty)&=P\left(\bigcup_{i=-\infty}^{+\infty}\{i-1<X\le i\}\right)\\
&=\sum_{i=-\infty}^{+\infty}P(i-1<X\le i)=\lim_{\substack{n\to+\infty\\ m\to-\infty}}\sum_{i=m}^{n}P(i-1<X\le i)\\
&=\lim_{n\to+\infty}F(n)-\lim_{m\to-\infty}F(m)
\end{aligned}$$

所以：

$$1\le\lim_{n\to+\infty}F(n)=1+\lim_{m\to-\infty}F(m)\le 1 \tag{2.1}$$

$$\lim_{n\to+\infty}F(n)=1,\quad \lim_{m\to-\infty}F(m)=0 \tag{2.2}$$

即 F_Q 满足有界性。

因为 $F(x)$ 是单调有界非降函数，所以其任一点 x_0 的右极限 $F(x_0+0)$ 必存在。右连续性的充要条件为：对任意单调下降的数列 $x_1>x_2>\cdots>x_n>\cdots>x_0$，当 $x_n\to x_0(n\to+\infty)$ 时，有 $\lim_{n\to+\infty}F(x_n)=F(x_0)$。因为

$$\begin{aligned}
F(x_1)-F(x_0)=P(x_0<X\le x_1)&=P\left(\bigcup_{r=1}^{+\infty}\{x_{i+1}<X\le x_i\}\right)\\
&=\sum_{i=1}^{+\infty}P(x_{i+1}<X\le x_i)=\sum_{i=1}^{+\infty}F(x_i)-F(x_{i+1})\\
&=\lim_{n\to\infty}[F(x_1)-F(x_n)]=F(x_1)-\lim_{n\to\infty}F(x_n)
\end{aligned}$$

即 F_Q 满足右连续性。

笔记 由于 $\{x_1\le x_0\}\ne\bigcup_{r=1}^{+\infty}\{x_i<X\le x_{i+1}\}$，例 $(-1,0]\ne\bigcup_{r=1}^{+\infty}(\frac{-1}{i},\frac{-1}{i+1}]=(-1,0)$。若将分布函数定义中右连续性改为左连续性，只需将命题改为 $F_Q(x)=Q[(-\infty,x)],\quad \forall x\in\mathbb{R}$ 即可。

命题 2.2

对每个分布函数 F 都存在唯一一个分布 Q_F 使得 $Q_F[(-\infty,x]]=F(x),\quad \forall x\in\mathbb{R}$ 成立。

证明 参考知乎

定理 2.1

分布函数可以唯一决定概率分布，即：

$$Q_{F_Q}=Q,\quad F_{Q_F}=F$$

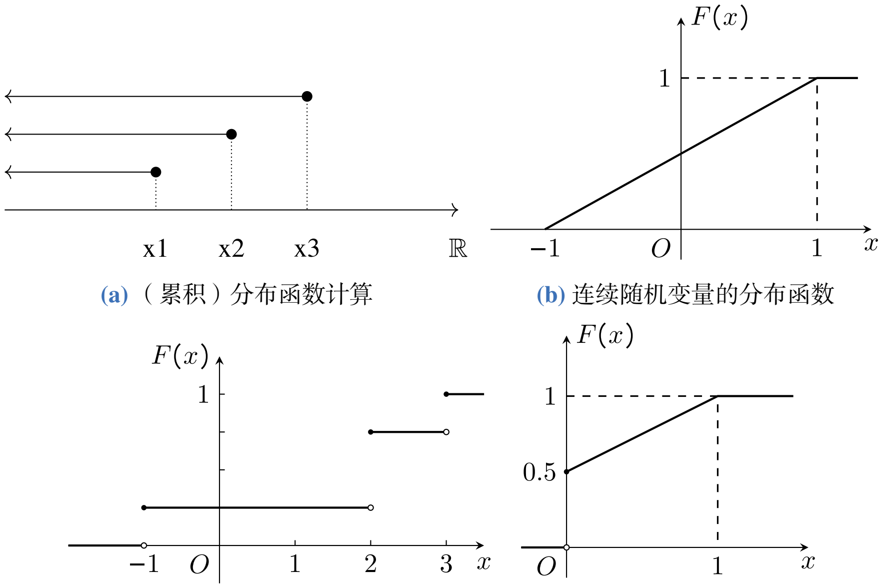
由此，概率分布与分布函数具有等同意义。所以可有以下定义：

定义 2.6 (随机变量的分布函数)

设 X 是一个随机变量，其分布由分布函数：

$$F(x)=P(X\leq x),\ \forall x\in\mathbb{R}$$

唯一刻画，称为随机变量 X 的**（累积）分布函数** (cumulative distribution function, c.d.f.)。把随机变量 X 服从分布函数 $F(x)$，简记作 $X\sim F(x)$。



(a)（累积）分布函数计算 (b) 连续随机变量的分布函数 (c) 离散随机变量的分布函数 (d) 既非离散、又非连续的分布函数

图 2.1: 累积分布函数

利用分布函数表示概率

$$\begin{aligned}
&P(a<X\leq b)=F(b)-F(a),\\
&P(X=a)=F(a)-F(a-0),\\
&P(X\geq b)=1-F(b-0),\\
&P(X>b)=1-F(b),\\
&P(a<x<b)=F(b-0)-F(a),\\
&P(a\leq X\leq b)=F(b)-F(a-0),\\
&P(a\leq X<b)=F(b-0)-F(a-0).
\end{aligned}$$

定义 2.7 (概率质量函数)

当且仅当函数 $p(x)$ 满足下述条件时, 被称为**概率质量函数** (probability mass function, p.m.f.):

- $p(x) \geq 0$
- $\sum_{x \in \mathcal{X}} p(x) = 1$

当 X 是离散型随机变量, 设 $\mathscr{F}_X$ 由 $\mathcal{X}$ 的所有子集组成, 此时 X 的分布由

$$p_X(x) = \mathbb{P}\{X = x\} = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}), \quad x \in \mathcal{X}$$

唯一刻画. 其与分布函数间的关系为:

- $F(x) = \sum_{t \leq x} P(X = t) = \sum_{t \leq x} p(t)$
- $p(x) = P(X = x) = F(x) - F(x-)$

定义 2.8 (概率密度函数)

当且仅当函数 $f(x)$ 满足下述条件时, 被称为**概率密度函数** (probability density function, p.d.f.):

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) = 1$

当 X 是连续型随机变量, 且 $F_X : \mathbb{R}^n \to [0,1]$ 可微 (或者更一般地, 绝对连续), 此时 X 的分布由

$$f_X := \frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}$$

唯一刻画. 其与分布函数间的关系为:

- $F(x) = \int_{-\infty}^{x} f(t)dt$
- $f(x) = \frac{d}{dx} F(x)$

注 即使对于 $f(x) > 0$ 的 x, $P(X = x) = x \int_x^x f(t)dt = 0$, 即连续型随机变量在实轴上任意一点的概率测度为零. 概率密度函数 $f(x)$ 代表的是在此位置上单位长度的概率, 可能是一个很大的值.

2.2 多元随机变量

定义 2.9 (随机向量)

若随机变量 $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ 定义在同一概率空间$(\Omega, \mathscr{F}, \mathbb{P})$ 上, 则称

$$X(\omega) = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$$

构成一个 n 维**随机向量**, 亦称 n 维随机变量.

命题 2.3

若 B_n 为 $\mathbb{R}^n$ 上任一博雷尔点集, 有

$$\{X(\omega) \in B_n\} \in \mathscr{F}$$

定义 2.10

称 n 元函数

$$F(x_1, x_2, \cdots, X_n) = \mathbb{P}\{X_1(\omega) < x_1, X_2(\omega) < x_2, \cdots, X_n(\omega) < x_n\}$$

为随机向量 $X(\omega)$ 的**联合分布函数** (joint cdf).

当 $n = 2$ 时, 有

$$\mathbb{P}((a_1, b_1) \leq X < (a_2, b_2)) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \tag{2.3}$$

性质 多元分布函数的一些性质:

1. 单调性：关于每个变元是单调不减函数；
2.

$$F(x_1, x_2, \cdots, -\infty, \cdots, X_n) = 0$$

$$F(+\infty, +\infty, \cdots, , +\infty) = 1$$

3. 关于每个变元右连续.
4. 在二元场合，还应该有：对任意 $a_1 < b, , a_2 < b_2$ ，都有

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0$$

为保证1.3式中的概率的非负性，性质 4 是必须的，而且由性质 4 可以推出单调性，但存在着反例说明，由单调性并不能保证性质 4 的成立 (见习题 12)． 这是多元场合与一元场合的不同之处.

2.2.1 边际分布

定义 2.11

对于多维随机变量 X, 只考虑其中一个分量的分布时, 称其为 X 的**边际分布或边缘分布**. 对于分量 X_i, 其**边缘分布函数** (marginal cdf) 为:

$$F_{X_i}(x_i) = \mathbb{P}\{X_i \leq x_i\} = F(\infty, \cdots, x_i, \cdots, \infty)$$

♣

2.2.2 条件分布

定义 2.12

对一切使 $P\left(Y = y_j\right) = p_{\cdot j} = \sum_{i=1}^{+\infty} p_{ij} > 0$ 的 y_j, 称

$$p_{i|j} = P\left(X = x_i | Y = y_j\right) = \frac{P\left(X = x_i, Y = y_j\right)}{P\left(Y = y_j\right)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \ldots$$

为给定 $Y = y_j$ 条件下 X 的条件分布列. 若 $p_X(x) = 0$, 则定义其为 0.

♣

设二维连续随机变量 (X, Y) 的联合密度函数为 $p(x, y)$, 边际密度函数为 $p_X(x), p_Y(y)$.

在离散随机变量场合, 其条件分布函数为 $P(X \leq x|Y = y)$. 但是, 因为连续随机变量取某个值的概率为零, 即 $P(Y = y) = 0$, 所以无法用条件概率直接计算 $P(X \leq x|Y = y)$, 一个很自然的想法是: 将 $P(X \leq x|Y = y)$ 看成是 $h \to 0$ 时 $P(X \leq x|y \leq Y \leq y + h)$ 的极限, 即

$$\begin{aligned} P(X \leq x|Y = y) &= \lim_{h\to 0} P(X \leq x|y \leq Y \leq y + h) \\ &= \lim_{h\to 0} \frac{P(X \leq x, y \leq Y \leq y + h)}{P(y \leq Y \leq y + h)} \\ &= \lim_{h\to 0} \frac{\int_{-\infty}^{x} \int_{y}^{y+h} p(u, v) dv du}{\int_{y}^{y+h} p_Y(v) dv} \\ &= \lim_{h\to 0} \frac{\int_{-\infty}^{x} \left\{\frac{1}{h} \int_{y}^{y+h} p(u, v) dv\right\} du}{\frac{1}{h} \int_{y}^{y+h} p_Y(v) dv}. \end{aligned}$$

当 $p_Y(y), p(x,y)$ 在 y 处连续时, 由积分中值定理可得

$$\lim_{h\to 0}\frac{1}{h}\int_y^{y+h} p_Y(v)dv = p_Y(y),$$
$$\lim_{h\to 0}\frac{1}{h}\int_y^{y+h} p(u,v)dv = p(u,y).$$

所以

$$P(X\le x|Y=y) = \int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}du$$

至此, 我们可以定义连续随机变量的条件分布如下.

定义 2.13

对一切使 $p_Y(y)>0$ 的 y, 给定 $Y=y$ 条件下 X 的**条件分布函数**和 **条件密度函数**分别为

$$F(x|y) = \int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}du,$$
$$p(x|y) = \frac{p(x,y)}{p_Y(y)}.$$

同理对一切使 $p_Y(y)>0$ 的 x, 给定 $X=x$ 条件下 Y 的条件分布函数和条件密度函数分别为

$$F(y|x) = \int_{-\infty}^{y}\frac{p(x,v)}{p_X(x)}dv$$
$$p(y|x) = \frac{p(x,y)}{p_X(x)}.$$

注 对于每一个固定的 x, $p_{Y|X}(y|x)$ 是一个关于 y 的概率质量函数; $f_{Y|X}(y|x)$ 是一个关于 y 的概率密度函数与概率三定理的对应:

乘法法则	$p_{XY}(x,y)=p_{Y\|X}(y\|x)p_X(x),$	$f_{XY}(x,y)=f_{Y\|X}(y\|x)f_X(x)$
全概率公式	$p_Y(y)=\sum_x p_{Y\|X}(y\|x)p_X(x),$	$f_Y(y)=\int_{-\infty}^{+\infty} f_{Y\|X}(y\|x)f_X(x)dx$
Bayes 原理	$p_{X\|Y}(x\|y)=\frac{p_{Y\|X}(y\|x)p_X(x)}{\sum_x p_{Y\|X}(y\|x)p_X(x)},$	$f_{X\|Y}(x\|y)=\frac{f_{Y\|X}(y\|x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y\|X}(y\|x)f_X(x)dx}$

2.2.3 独立

定义 2.14 (独立随机变量)

若随机变量 $X(\omega)=(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))$ 联合分布函数可分解成各分量边缘分布函数的乘积, 即:

$$F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n),\quad \forall x_1,x_2,\cdots,x_n\in\mathbb{R}$$

则称随机变量 X 各分量相互**独立**

注 对于一般的多元随机变量, 其各分量边缘分布不足以描述联合分布的情况. 但若其各分量独立则可以.

定理 2.2

对于连续情况:

$$F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$
$$\Leftrightarrow f(x_1,x_2,\cdots,x_n)=f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

对于离散情况:

$$F(x_1,x_2,\cdots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$
$$\Leftrightarrow p(x_1,x_2,\cdots,x_n)=p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n)$$

2.3 随机变量的函数

在统计学中，常需要转化原始数据以获取其中信息, 由此引出了研究随机变量的函数的需要.

定义 2.15 (可测函数)

设 $y = g(x)$ 是 $\mathbb{R}$ 到 $\mathbb{R}$ 上的一个映照，若对于一切 $\mathbb{R}$ 中的 Borel 点集 B_1 均有

$$\{x : g(x) \in B_1\} \in \mathcal{B}_1$$

其中 $\mathcal{B}_1$ 为 $\mathbb{R}$ 上 Borel σ 域，则称 $g(x)$ 是一元博雷尔函数, 也称为一元**可测函数**

定理 2.3 (事件法)

设 $Y = g(X)$ 是随机向量 $X = (X_1, X_2, \cdots, X_n)$ 的函数, 则 Y 的分布由 X 的分布通过下式决定:

$$\mathbb{P}\{Y \in B\} = \mathbb{P}\{X \in A\}, \quad A = \{\omega | g(X(\omega)) \in B\}$$

此法是其他方法的基础, 但使用不便, 常用于离散随机变量.

例题 2.1 已知随机变量 X, Y 的联合概率质量函数为 $p(x, y)$, 求 $Z = X + Y$ 的分布

解

$$p_Z = P(Z = z) = P(X + Y = z) = \sum_{x=-\infty}^{\infty} p(x, z - x)$$

若 X, Y 独立, 则 $p_Z(z) = \sum_{-\infty}^{\infty} p_X(x) p_Y(z - x) dx$, 为 p_X 与 p_Y 的卷积

定理 2.4 ()

若随机变量 X, Y 独立, 则其变换 $Z = g(X), W = h(Y)$ 也独立.

泛化情况: 若随机向量 $\{X\}_n$ 各分类独立, 则其变换 $\{Y\}_n = g(\{X\}_n)$ 各分类也独立.

证明

2.3.1 分布函数法

通过下式获取随机变量的函数的分布:

$$F_Y(y) = \begin{cases} \int_{A_y} f_X(x) dx \\ \sum_{x \in A_y} p_X(x) \end{cases}, \quad A_y = \{x | g(x) \leq y\}$$

对每一个变换分别运用上式则可得到向量函数的分布.

例题 2.2 已知随机变量 X 的概率密度函数 $f_X(x)$ 与分布函数 $F_X(x)$, 求 $Y = X^2$ 的分布

解

$$F_Y(y) = P(Y \leq y) = P(-\sqrt{y} < X \leq \sqrt{y}) + P(X = \sqrt{y} = F_X(\sqrt{y}) - F_X(-\sqrt{y})) + 0 \quad y \geq 0$$

$$\begin{aligned} f_Y(y) = \frac{d}{dy} F_Y(y) &= \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(\sqrt{y}) \\ &= f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \frac{-1}{2\sqrt{y}} \\ &= \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \end{aligned}$$

例题 2.3 已知随机变量 X, Y 的联合概率密度函数 $f(x, y)$, 求 $Z = X + Y$ 的分布

解

$$F_Y(y) = P(Y \leq y) = P(-\sqrt{y} < X \leq \sqrt{y}) + P(X = \sqrt{y} = F_X(\sqrt{y}) - F_X(-\sqrt{y})) + 0 \quad y \geq 0$$

$$\begin{aligned}
F_Z(z) &= P(Z \le z) = P(X+Y \le z)\\
&= \iint_{x+y\le z} f(x,y)dxdy\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f(x,y)dxdy\\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x,v-x)dxdv, \quad y = v-x
\end{aligned}$$

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \int_{-\infty}^{\infty} f(x,z-x)dx$$

若 X,Y 独立, 则 $f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$, 为 f_X 与 f_Y 的卷积, 与1.1类似

例题 2.4 已知随机变量 X,Y 的联合概率密度函数 $f(x,y)$, 求 $Z=\frac{Y}{X}$ 的分布

解

$$Q_z = \{(x,y): y/x \le z\} = \{(x,y): x<0, y\ge zx\} \cup \{(x,y): x>0, y\le zx\}$$

$$\begin{aligned}
F_Z(z) = \iint_{Q_z} f(x,y)dxdy &= \int_{-\infty}^{0}\int_{xz}^{\infty} + \int_{0}^{\infty}\int_{-\infty}^{xz} f(x,y)dydx\\
&= \int_{-\infty}^{0}\int_{z}^{-\infty} + \int_{0}^{\infty}\int_{-\infty}^{z} xf(x,xv)dvdx \quad (\text{set} y = xv)\\
&= \int_{-\infty}^{0}\int_{-\infty}^{z} (-x)f(x,xv)dvdx + \int_{0}^{\infty}\int_{-\infty}^{z} xf(x,xv)dvdx\\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} |x|f(x,xv)dxdv
\end{aligned}$$

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \int_{-\infty}^{\infty} |x|f(x,xz)dx$$

2.3.2 Copula

定义 2.16

设连续型 实值随机变量 X 有分布函数 F, 易见 F 在 $\overline{\mathbb{R}} = [-\infty,+\infty]$ 上从 0 递增到 1. 定义相应的**分位数函数** (quantile function) 为

$$F^{-1}(p) := \inf\{x\in\mathbb{R} : F(x)\ge p\}, \quad p\in[0,1].$$

注 当 F 严格递增时, 这与一般的反函数定义相同.

定理 2.5

设连续型 实值随机变量 X 有分布函数 F, 则 $F(X) \sim \mathrm{Uniform}([0,1])$

证明

$$\mathbb{P}\{F(X)\le p\} = \mathbb{P}\{X \le F^{-1}(p)\} = F(F^{-1}(p)) = p, \quad \forall p\in[0,1].$$

定理 2.6

设连续型 实值随机变量 X 有分布函数 F, 且设 $U\sim \mathrm{Uniform}([0,1])$, 则 $F^{-1}(U) \stackrel{d}{=} X$, 其中 $\stackrel{d}{=}$ 表示分布相同 (equal in distribution).

定理 2.7 (Sklar 定理)

考虑多个连续型 实值随机变量 $X_1, \dots, X_k$, 记 X_i 的分布函数为 F_i. 我们称 $(F_1(X_1), \dots, F_k(X_k))$ 的分布函数 $C: [0,1]^k \to [0,1]$ 为相应的 **Copula**, 适合

$$C(F_1(x_1), \dots, F_k(x_k)) = \mathbb{P}\{X_1 \le x_1, \dots, X_k \le x_k\}, \quad \forall x_1, \dots, x_k$$

这个结果称为 **Sklar 定理**

这个结果, 在金融统计中有颇多应用. 稍作诠释的话, Copula 提取了变量间的相关性, 通过粘合边际能够恰好地表示总体. 人们可以构造各种各样的 Copula, 对真实世界进行建模.

2.3.3 概率密度函数法

定理 2.8

设连续随机变量 X 的概率密度函数为 $f_X(x)$. 令 $Y = g(X)$, 其中 g 为可微函数, 且严格单调, 则当 $y = g(x)$ 有定义时:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

否则为 0

若 g 为分段单调函数, 则分段 计算上是结果, 再进行相加

定理 2.9

设连续随机向量 $\mathbf{X}$ 的概率密度函数为 $f_{\mathbf{X}}(\mathbf{x})$. 令 $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, 其中 $\mathbf{g}$ 为双射, 定义其逆函数为:

$$\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y}) = \mathbf{w}(\mathbf{y})$$

若 $\mathbf{w}$ 存在连续偏导数, 则当 $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ 有定义时:

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y})) \left| \frac{\partial \mathbf{w}}{\partial \mathbf{y}} \right|$$

否则为 0

注 若 $\mathbf{Y}$ 的维数 k 小于 $\mathbf{X}$ 的维数 n, 可增补 $n-k$ 维的函数 $\mathbf{Z} = \mathbf{h}(\mathbf{X})$, 使得 $(\mathbf{Y}, \mathbf{Z})$ 满足条件, 再通过积分获取 $\mathbf{Y}$ 的概率密度函数.

例题 2.5 已知随机变量 X_1, X_2 的联合概率密度函数 $f(x_1, x_2)$, 求 $Y_1 = \frac{X_2}{X_1}$ 的分布

解 令 $Y_2 = X_1$, 则:

$$x_1 = y_2 \equiv w_1(y_1, y_2)$$

$$x_2 = y_1 y_2 \equiv w_2(y_1, y_2)$$

$$\frac{\partial w_1}{\partial y_1} = 0, \frac{\partial w_1}{\partial y_2} = 1, \frac{\partial w_2}{\partial y_1} = y_2, \frac{\partial w_2}{\partial y_2} = y_1$$

$$J = \begin{vmatrix} 0 & 1 \\ y_2 & y_1 \end{vmatrix} = -y_2$$

所以:

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(y_2, y_1 y_2) \, |y_2|$$

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1 Y_2}(y_1, y_2) \, dy_2 = \int_{-\infty}^{\infty} f_{X_1 X_2}(y_2, y_1 y_2) \, |y_2| \, dy_2 \tag{2.4}$$

命题 2.4

若随机向量 $\mathbf{X}=(X_1,\cdots,X_n)$ 各分类相互独立, 则其各分量的函数 $\mathbf{Y}=(g_1(X_1),\cdots,g_n(X_n))$ 也相互独立

证明

2.3.4 矩母函数法

2.3.5 次序统计量

定义 2.17

设 $X_1,X_2,\ldots,X_n$ 为随机变量, 将其按大小排序后记为 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$, 则将 $x_{(i)}$ 称为该样本的第 i 个**次序统计量**. 其中

- **最小次序统计量**定义为:$X_{(1)}=\min\{x_1,\ldots,x_n\}$
- **最大次序统计量**定义为:$X_{(n)}=\max\{x_1,\ldots,x_n\}$
- **极差**定义为:$R=X_{(n)}-X_{(1)}$
- 第 i 个**间差**定义为:$S_i=X_{(i)}-X_{(i-1)}$

注 虽然 $X_1,X_2,\ldots,X_n$ 独立, 但其次序统计量一般不独立

例题 2.6 若 $X_1,X_2,\ldots,X_n$ 独立同分布, 求 $X_{(1)}$ 与 $X_{(n)}$ 的分布

解

$$
\begin{aligned}
F_{X_{(n)}}(x) &= P(X_{(n)}\le x)=P(X_1\le x,\ldots,X_n\le x)\\
&= P(X_1\le x)\cdots P(X_n\le x)\\
&= [F(x)]^n\\
f_{X_{(n)}}(x) &= \frac{d}{dx}F_{X_{(n)}}(x)\\
&= nf(x)[F(x)]^{n-1}
\end{aligned}
$$

$$
\begin{aligned}
1-F_{X_{(1)}}(x) &= P(X_{(1)}>x)=P(X_1>x,\ldots,X_n>x)\\
&= P(X_1>x)\cdots P(X_n>x)\\
&= [1-F(x)]^n\\
f_{X_{(1)}}(x) &= \frac{d}{dx}F_{X_{(1)}}(x)\\
&= nf(x)[1-F(x)]^{n-1}
\end{aligned}
$$

定理 2.10

第 i 个**次序统计量**的概率密度函数为:

$$f_{X_{(k)}}=C(n;1,k-1,n-k)f(x)[F(x)]^{k-1}[1-F(x)]^{n-k}$$

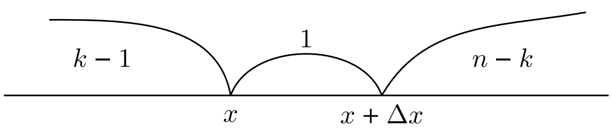


图 2.2: $X_{(k)}$ 取值的示意图

定理 2.11

次序统计量 $(x_{(i)}, x_{(j)})(i < j)$ 的联合分布密度函数为

$$p_{ij}(y,z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y)]^{i-1}[F(z)-F(y)]^{j-i-1} \cdot [1-F(z)]^{n-j}p(y)p(z), \quad y \le z,$$

证明 对正 $\Delta y, \Delta z$ 以及 $y < z$, 事件 “$x_{(i)} \in (y, y+\Delta y], x_{(j)} \in (z, z+\Delta z]$” 可以表示为 “容量为 n 的样本 $x_1, \dots, x_n$ 中有 $i-1$ 个观测值小于等于 y, 一个落入区间 $(y, y+\Delta y]$, $j-i-1$ 个落入区间 $(y+\Delta y, z]$, 一个落入区间 $(z, z+\Delta z]$, 而余下 $n-j$ 个大于 $z+\Delta z$”(见图 1.3).

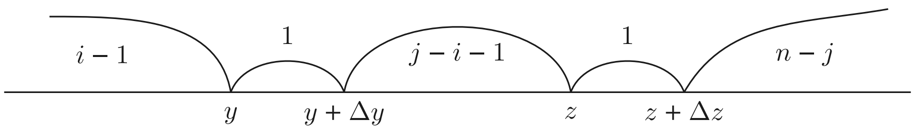


图 2.3: $x_{(i)}$ 与 $x_{(j)}$ 取值的示意图

例题 2.7 若 $X_1, X_2, \dots, X_n$ 独立同分布, 求 $R = X_{(n)} - X_{(1)}$ 的分布

解

$$f_{X_{(1)}X_{(n)}}(s,t) = n(n-1)f(s)f(t)[F(t)-F(s)]^{n-2}\mathbb{I}(s \le t)$$

$$f_R(r) = \mathbb{I}(r > 0)\int_{-\infty}^{\infty} f_{X_{(1)}X_{(n)}}(s, s+r)ds$$