概率论笔记

作者：肖程哲

时间：August 26, 2022

苟日新，日日新，又日新

目录

第 4 章 常见分布

考试重点

- ❑ 各分布的特征
- ❑ 各分布联系与转换
- ❑ 多维正态分布

4.1 离散分布

4.1.1 均匀分布

定义 4.1 (离散均匀分布)

若随机变量只能在 $a_1, \cdots, a_n$ 中取值，并且对应的概率相同，则称其遵循**均匀分布** (Uniform distribution)，记为 $X \sim U(a_1, \cdots, a_n)$。

注 此即古典概型。

离散均匀分布的特征：

参数 $a_i \in \mathbb{R}$

概率质量函数 $p(x) = \begin{cases} \frac{1}{m} & x = a_1, a_2, \cdots, a_n \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = \frac{\sum_{i=1}^m e^{a_i t}}{m}$

均值 $\mu = \frac{\sum_{i=1}^m a_i}{m} = \overline{a}$

方差 $\sigma^2 = \frac{\sum_{i=1}^m (a_i - \overline{a})^2}{m}$

实例 丢一个均匀的骰子

4.1.2 两点分布

定义 4.2 (两点分布)

若随机变量只能取 0 或 1，并且对应的概率分别为 $1-p$ 与 p，则称其遵循**两点分布** (Bernouli distribution)（也称 0-1 分布、两点分布分布），记为 $X \sim B(1, p)$。

两点分布的特征：

参数 $p \in [0, 1]$

概率质量函数 $p(x) = \begin{cases} p^x (1-p)^{1-x} & x \in \{0, 1\} \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = pe^t + 1 - p$

均值 $\mu = p$

方差 $\sigma^2 = p(1-p)$

实例 丢一次硬币，p 代表某一面出现的概率

注 若 A 是一个事件，出现概率为 p_A，则指示随机变量 I_A（若 A 出现记为 1，否则为 0）遵循 Bernoulli 分布，即 $I_A \sim B(p_A)$

4.1.3 二项分布

定义 4.3 (二项分布)

若进行 n 次独立的 Bernoulli 实验（定义??）$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} B(1,p)$，则这些随机变量之和 $Y = X_1 + \cdots + X_n$ 遵循**二项分布** (binomial distribution)，记为 $Y \sim B(n,p)$。♣

二项分布的特征：

参数 $p \in [0,1], n \in \mathbb{N}_+$

概率质量函数 $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \mathbb{N}^n \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = (pe^t + 1 - p)^n$，求解方法有：

- 独立变量 之和的矩母函数等于各变量矩母函数之积（定理3.15）
- 定义

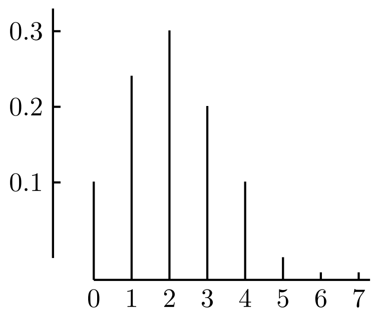
均值 $\mu = np$，求解方法有：

- 定义（例??）
- 矩母函数在 $t = 0$ 的一阶导
- 变量之和的均值等于各变量均值之和

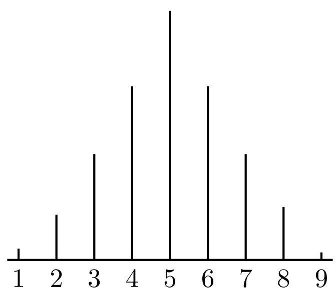
方差 $\sigma^2 = np(1-p)$，求解方法有：

- 定义（例??）
- 矩母函数在 $t = 0$ 的二阶导，$\mathrm{Var}(X) = E(X^2) - (E(X))^2$
- 独立变量 之和的方差等于各变量均值之和

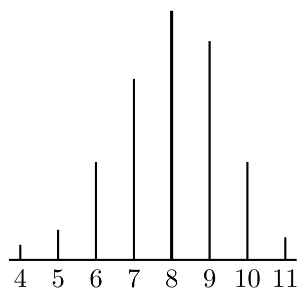
实例 丢 n 次硬币，p 代表某一面出现的概率，出现此面的次数



(a) $b(10, 0, 2)$ 的线条图 (右偏)



(b) $b(10, 0, 5)$ 的线条图 (对称)



(c) $b(10, 0, 8)$ 的线条图 (左偏)

图 4.1: 二项分布 $b(n,p)$ 的线条图

从上图可以看出：

- 位于均值 np 附近概率较大；
- 随着 p 的增加，分布的峰逐渐右移.

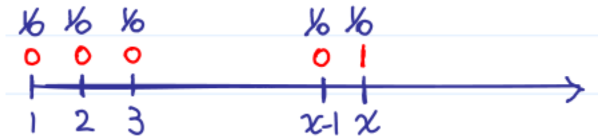
命题 4.1

二项分布之和仍是二项分布。若 $X_1+\cdots+X_k$ 独立且 $X_i \sim B(n_i,p)$，则 $Y = X_1+\cdots+X_k \sim B(n_1+\cdots+n_k,p)$

4.1.4 几何分布

定义 4.4

若进行无限次独立的 Bernouli 实验，则第一次出现“1”的实验次数遵循**几何分布** (geometric distribution)，记为 $X \sim G(p)$。



几何分布的特征：

参数 $p \in [0,1]$

概率质量函数 $p(x) = \begin{cases} (1-p)^{x-1}p & x \in \mathbb{N}_+ \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = \frac{pe^t}{1-(1-p)e^t}, t < -\ln(1-p)$，求解方法有：

- 独立变量 之和的矩母函数等于各变量矩母函数之积（参见定理3.15）
- 定义

均值 $\mu = \frac{1}{p}$，求解方法有：

- $E(X) = \sum_{k=1}^{\infty} P(X \geq k)$
- 矩母函数在 $t = 0$ 的一阶导
- 定义（下述微分法）

方差 $\sigma^2 = \frac{1-p}{p^2}$，求解方法有：

- 定义（下述微分法）
- 矩母函数在 $t = 0$ 的二阶导，$\mathrm{Var}(X) = E(X^2) - (E(X))^2$

实例 买彩票时，中奖所需购买张数。

笔记 微分法求解几何分布均值：

$$\begin{aligned} E(X) &= p\sum_{k=1}^{\infty} kq^{k-1} = p\frac{d}{dq}\sum_{k=1}^{\infty} q^k \\ &= p\frac{d}{dq}\frac{q}{1-q} = \frac{p}{(1-p)^2} \\ &= \frac{1}{p} \end{aligned}$$

微分法求解几何分布方差：

$$
\begin{aligned}
E(X(X-1)) &= p\sum_{k=1}^{\infty} k(k-1)q^{k-1} = p\frac{d}{dq}\sum_{k=1}^{\infty} q^{k} \\
&= p\frac{d^2}{dq^2}\frac{q}{1-q} = \frac{2p}{(1-q)^3} \\
&= \frac{2}{p^2} \\
\mathrm{Var}(X) &= E(X(X-1)) + E(X) - (E(X))^2 \\
&= \frac{2}{p^2} + \frac{1}{p} - \frac{1}{p^2}p^2 \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

命题 4.2

离散分布中有且只有几何分布具备无记忆性：

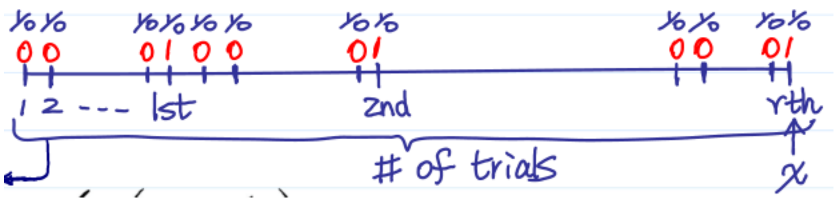
$$P\{X = s+t|X > t\} = P\{X = s\}, \quad \forall s > 0, t > 0$$

证明

4.1.5 负二项分布

定义 4.5

若进行无限次独立的 Bernouli 实验，则第 r 次出现“1”的实验次数遵循**负二项分布** (negative binomial distribution)，记为 $X \sim NB(r,p)$。



负二项分布的特征：

参数 $p \in [0,1]$

概率质量函数 $p(x) = \begin{cases} \binom{x-1}{r-1}p^r(1-p)^{x-r}p & x \in \mathbb{N}_r \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = \frac{p^r e^{rt}}{(1-(1-p)e^t)^r}, t < -\ln(1-p)$，求解方法有：

- 独立变量 之和的矩母函数等于各变量矩母函数之积（参见定理3.15）
- 定义凑一法
- 利用公式 $\sum_{x=0}^{\infty}\binom{n+x-1}{x}t^x = \frac{1}{(1-t)^n}, -1 < t < 1$

均值 $\mu = \frac{r}{p}$，求解方法有：

- 定义凑一法
- 矩母函数在 $t = 0$ 的一阶导
- 独立几何分布之和 (定理4.3)

方差 $\sigma^2=\frac{r(1-p)}{p^2}$，求解方法有：

- 定义（凑一法）
- 矩母函数在 $t=0$ 的二阶导，$\mathrm{Var}(X)=E(X^2)-(E(X))^2$
- 独立几何分布之和 (定理4.3)

实例 买彩票时，中奖 r 次所需购买张数。

命题 4.3

几何分布之和是负二项分布。若 $X_1+\cdots+X_r$ 独立且 $X_i\sim G(p)$，则 $Y=X_1+\cdots+X_r\sim NB(r,p)$

证明

命题 4.4

负二项分布之和仍是负二项分布。若 $X_1+\cdots+X_k$ 独立且 $X_i\sim NB(r_i,p)$，则 $Y=X_1+\cdots+X_k\sim B(r_1+\cdots+r_k,p)$

证明

4.1.6 多项分布

定义 4.6 (多项分布)

若进行 n 次独立的实验，每次实验有 r 种结果，每种结果对于概率分别为 $p_1,\cdots,p_r$。令 X_i 代表得出结果 i 的次数，则随机向量 $\mathbf{X}=(X_1,\cdots,X_r)$ 遵循**多项分布** (multinomial distribution)，记为 $\mathbf{X}\sim Multinomial(n,p_1,\cdots,p_r)$。

多项分布的特征：

参数 $p_i\in[0,1],\sum_{i=1}^r p_i=1$

概率质量函数 $p(x)=\begin{cases}\binom{n}{x_1\cdots x_r}p_1^{x_1}\cdots p_r^{x_r} & x\in\mathbb{N}^n \,\&\sum_{i=1}^r x_i=n\\ 0 & \text{其他}\end{cases}$

矩母函数 $M(t_1,\cdots,t_r)=(p_1e^{t_1}+\cdots+p_re^{t_r})^n$，求解方法有：

- 利用公式
- 凑一法

边缘分布 $X_i\sim B(n,p_i)$

均值 $E(X_i)=np_i$

方差 $\mathrm{Var}(X_i)=np_i(1-p_i)$

协方差 $\mathrm{Cov}(X_i,X_j)=-p_ip_j$，求解方法有：

- 矩母函数求解 $E(X_iX_j)$
- 凑一法求解 $E(X_iX_j)$

实例

注 多项分布是二项分布的泛化。

4.1.7 泊松分布

定理 4.1 (Poisson 逼近)

$$\lim_{np_n\to\lambda}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$

证明 记 $np_n=\lambda_n$，记 $p_n=\lambda_n/n$，我们可得

$$\begin{aligned}\binom{n}{k}p_n^k(1-p_n)^{n-k}&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}\\&=\frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n-k}.\end{aligned}$$

对固定的 k 有

$$\lim_{n\to+\infty}\lambda_n=\lambda;\qquad \lim_{n\to+\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=e^{-\lambda};\ \lim_{n\to+\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1.$$

从而

$$\lim_{n\to+\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$

对任意的 k（$k=0,1,2,\cdots$）成立

定义 4.7 (Poisson 分布)

若随机变量 X 的概率分布列满足以下形式：

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,2,\cdots,$$

则称其遵循 **Poisson 分布**记为 $X\sim P(\lambda)$.

注 由定理4.1可看出，Poisson 分布可作为二项分布的近似。λ 的涵义

Poisson 分布的特征：

参数 $\lambda>0$

概率质量函数 $p(k)=\begin{cases}\frac{\lambda^k}{k!}e^{-\lambda} & x\in\mathbb{N}^n\\ 0 & \text{其他}\end{cases}$

矩母函数 $M(t)=e^{\lambda(e^t-1)}$，求解方法有：

- 利用公式
- 凑一法

均值 $\mu=\lambda$

方差 $\sigma^2=\lambda$

实例 公共汽车站来到的乘客数

命题 4.5

Poisson 分布之和仍是 Poisson 项分布。若 $X_1+\cdots+X_k$ 独立且 $X_i\sim P(\lambda_i)$，则 $Y=X_1+\cdots+X_k\sim P(\lambda_1+\cdots+\lambda_k)$

证明

引理 4.1

若 $f(x)$ 是连续函数（或单调函数），且对一切 x,y（或一切 $x,y \geq 0$）成立：

$$f(x)f(y) = f(x+y)$$

则 $f(x)$ 为**指数函数**。

证明

$$\because f(1) = [f(\frac{1}{n})]^n$$
$$且 a = f(1) = [f(\frac{1}{2})]^2 \geq 0$$
$$\therefore f(\frac{1}{n}) = a^{\frac{1}{n}}$$
$$\therefore f(\frac{m}{n}) = [f(\frac{1}{n})]^m = a^{\frac{m}{n}}$$

即此命题对一切有理数成立。又实数具有连续性，故此命题对一切实数成立。

定义 4.8 (泊松过程)

若某一随机过程满足以下特征，则称其为**泊松过程**：

平稳性 随机事件 A 在 $[t_0, t_0+t)$ 中的发生次数数只与时间间隔长度 t 有关而与时间起点 t_0 无关。以 $P_k(t)$ 记在长度为 t 的时间区间中发生 K 次事件 A 的概率。

独立增量性 在 $[t_0, t_0+t)$ 中发生 K 次事件 A 的概率与时刻 t_0 以前发生的事件独立。

普通性 在充分小的时间间隔中，最多发生一次事件 A。即，若记 $\psi(t) = \sum_{k=2}^{\infty} P_k(t) = 1 - P_0(t) - P_1(t)$，则有 $\lim_{t\to 0}\psi(t)/t = 0$

对 $\Delta t > O$，考虑 $[O, t+\Delta t)$ 中发生 K 次事件 A 的概率 $P_k(t+\Delta t)$，由独立增量性及全概率公式可得：

$$P_k(t+\Delta t) = P_k(t)P_0(\Delta r) + P_{k-1}P_1(\Delta t) + \cdots + P_0(t)P_k(\Delta t), k \geq 0$$

（对 $n \geq 1$，假定 $P_{-n}(t) = 0$）

特别地

$$P_0(t+\Delta t) = P_0(t)P_0(\Delta t)$$

由引理4.1知

$$P_0(t) = a^t, a \geq 0$$

若 $a = O$，则 $P_0(t) \equiv 0$，说明在不管怎么短的时间间隔内事件 A 都发生，因此在有限时间间隔中将发生无穷多个次事件 A，这种情形不在我们的考虑之列。此外，因 $P_0(t)$ 是概率，故应有 $a \geq 1$，而当 $a = 1$ 时，$P_0(t) \equiv 1$，表明事件 A 永不发生，也不是我们感兴趣的情形，所以应有 $0 < a < 1$，从而存在 $\lambda > 0$，使

$$P_0(t) = e^{-\lambda t}$$

因此当 $\Delta t \to 0$ 时，有

$$P_0(\Delta t) = \mathrm{e}^{-\lambda\Delta t} = 1 - \lambda\Delta t + o(\Delta t)$$
$$P_1(\Delta t) = 1 - P_0(\Delta t) - \psi(\Delta t) = \lambda\Delta t + o(\Delta t)$$
$$\sum_{l=2}^{\infty} P_{k-l}(t)P_l(\Delta t) \leqslant \sum_{l=2}^{\infty} P_l(\Delta t) = \psi(\Delta t) = o(\Delta t)$$

所以：

$$P_k(t+\Delta t) = P_k(t)(1-\lambda\Delta t) + P_{k-1}\lambda\Delta t + o(\Delta t), k \geq 1$$

因此：

$$P_k'(t)=\frac{P_k(t+\Delta t)-P_k(t)}{\Delta t}=\lambda[P_{k-1}-P_k(t)]$$

由于 $P_0(t)=\mathrm{e}^{-\lambda t}$，故有 $P_1'(t)=\lambda[\mathrm{e}^{-\lambda t}-P_1(t)]$，可解得 $P_0(t)=\lambda t\mathrm{e}^{-\lambda t}$，依次可递推可解得：

$$P_k(t)=\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t},k\in\mathbb{N}$$

正是参数为 λt 的泊松分布。

4.1.8 超几何分布

定义 4.9

设有 N 个产品，其中有 M 个不合格品. 若从中不放回地随机抽取 n 个，则其中含有的不合格品的个数 X 服从超几何分布，记为 $X\sim h(n,N,M)$.

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\ k=0,1,\cdots,r, \tag{4.1}$$

其中

超几何分布的特征：

参数	$r=\min\{M,n\}$，且 $M\leq N,n\leq N,n,N,M$ 均为正整数.
概率质量函数	$p(k)=\begin{cases}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} k=0,1,\cdots,r\\ 0 & \text{其他}\end{cases}$
矩母函数	存在，但没有简单的表达
均值	$\mu=\frac{rn}{n+m}$
方差	$\sigma^2=\frac{rnm(n+m-r)}{(n+m)^2(n+m-1)}$
实例	从一个有限总体中进行不放回抽样

注 若 $m,n\to\infty$ 时有 $p_{m,n}=\frac{n}{m+n}\to p$，则超几何分布可近似为二项分布。$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\cong\binom{n}{k}p^k(1-p)^{n-k}$

4.2 连续分布

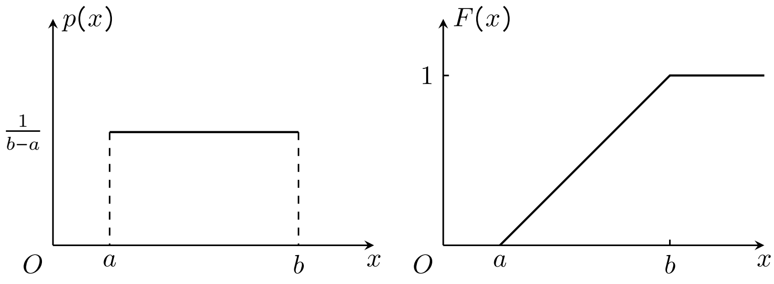
4.2.1 均匀分布

定义 4.10

若随机变量 X 在 $[a,b]$ 中任一区域的概率与其测度成正比，则称 X 服从区间 (a,b) 上的**均匀分布** (Uniform distribution)，记作 $X\sim U(a,b)$。

均匀分布的特征：

参数	$a<b,\quad a,b\in\mathbb{R}$
概率质量函数	$p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0 & \text{其他}\end{cases}$



(a) 密度函数 $p(x)$　　(b) 分布函数 $F(x)$

图 4.2: (a,b) 上的均匀分布

分布函数　$F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \le x < b; \\ 1, & x \ge b. \end{cases}$

矩母函数　$M(t) = \frac{e^{bt}-e^{at}}{t(b-a)}, \quad t \in \mathbb{R}$

均值　$\mu = \frac{a+b}{2}$

方差　$\sigma^2 = \frac{(b-a)^2}{12}$

注 常利用 $U(0,1)$ 生成特定分布的伪随机数（参见定理2.6）

4.2.2 指数分布

定义 4.11

若随机变量 X 的密度函数（见图 4.3）为

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0; \\ 0, & x < 0, \end{cases}$$

则称 X 服从**指数分布**，记作 $X \sim Exp(\lambda)$，其中参数 $\lambda > 0$. 指数分布的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0; \\ 0, & x < 0. \end{cases}$$

♣

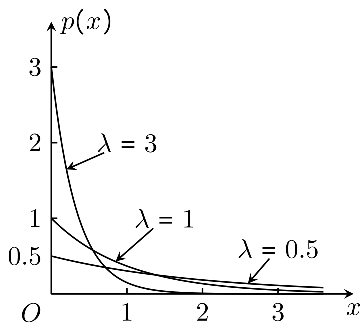


图 4.3: 参数为 λ 的指数分布密度函数

指数分布的特征：

矩母函数　$M(t) = \frac{\lambda}{\lambda-t}, \quad t < \lambda$

均值 $\mu = \frac{1}{\lambda}$

方差 $\sigma^2 = \frac{1}{\lambda^2}$

实例 物品使用寿命，等待时间

注 有其均值可看出，对于等待时间模型，$\frac{1}{\lambda}$ 代表平均等待时间，即（时间/次）；λ 代表平均发生频率，即（次/时间）。

定理 4.2 (指数分布的无记忆性)

连续分布中有且只有指数分布具备无记忆性：

$$P(X > s + t | X > s) = P(X > t), \quad \forall s > 0, t > 0$$

证明

命题 4.6

如果某设备在任何长为 t 的时间 $[0,t]$ 内发生故障的次数 $N9t)$ 服从参数为 λt 的泊松分布，则相继两次故障之间的时间间隔 T 服从参数为 λ 的指数分布.

证明 设 $N(t) \sim P(\lambda t)$，即

$$P\big(N(t) = k\big) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \cdots.$$

注意到两次故障之间的时间间隔 T 是非负随机变量，且事件 $\{T \geq t\}$ 说明此设备在 $[0,t]$ 内没有发生故障，即 $\{T \geq t\} = \{N(t) = 0\}$，由此我们得

当 $t < 0$ 时，有 $F_T(t) = P(T \leq t) = 0$；

当 $t \geq 0$ 时，有

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - P\big(N(t) = 0\big) = 1 - e^{-\lambda t},$$

所以 $T \sim Exp(\lambda)$，相继两次故障之间的时间间隔 T 服从参数为 λ 的指数分布.

4.2.3 伽马分布

定义 4.12

若随机变量 X 的密度函数为

$$p(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\lambda x}, & x \geq 0; \\ 0, & x < 0, \end{cases}$$

则称 X 服从**伽玛分布**，记作 $X \sim \Gamma(\alpha, \lambda)$，其中 $\alpha > 0$ 为形状参数，$\lambda > 0$ 为尺度参数.

图 4.4 给出若干条 λ 固定、α 不同的伽玛密度函数曲线，从图中可以看出：

- $0 < \alpha < 1$ 时，$p(x)$ 是严格下降函数，且在 $x = 0$ 处有奇异点；
- $\alpha = 1$ 时，$p(x)$ 是严格下降函数，且在 $x = 0$ 处 $p(0) = \lambda$；
- $1 < \alpha \leq 2$，$p(x)$ 是单峰函数，向上凸、后下凸；
- $2 < \alpha$ 时，$p(x)$ 是单峰函数，先下凸、中间上凸、后下凸. 且 α 越大，$p(x)$ 越近似于正态分布.

Gamma 分布的特征：

矩母函数 $M(t) = (\frac{\lambda}{\lambda - t})^\alpha, \quad t < \lambda$

均值 $\mu = \frac{\alpha}{\lambda}$

方差 $\sigma^2 = \frac{\alpha}{\lambda^2}$

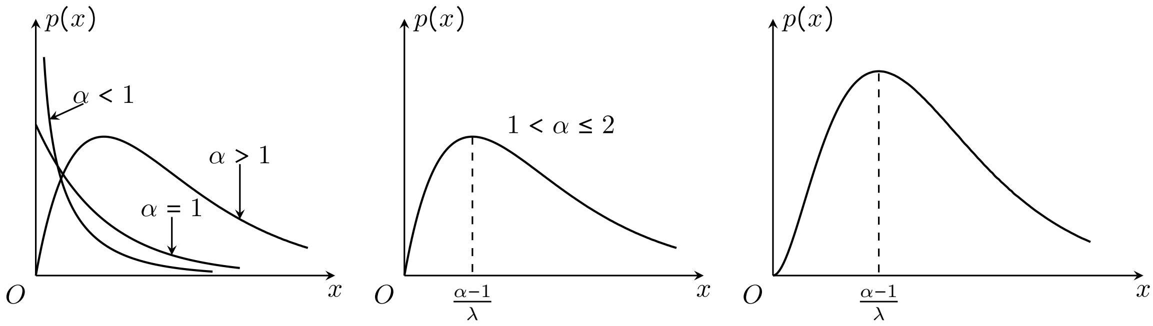


图 4.4: λ 固定、不同 α 的伽玛密度曲线

Gamma 函数 $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}\mathrm{e}^{-y}dy$ 的特征：

- $\Gamma(1) = 1$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$
- $\Gamma(n) = (n-1)!, \quad n \in \mathbb{Z}$
- $\Gamma(\frac{\alpha}{2}) = \frac{\sqrt{\pi}(\alpha-1)!}{2^{\alpha-1}(\frac{\alpha-1}{2})!}, \quad \alpha = 2k+1, k \in \mathbb{N}$

命题 4.7

Γ 分布是指数分布的泛化，即 $\Gamma(1,\lambda) = E(\lambda)$。进一步有：若随机变量 $X_1,\cdots,X_k,\mathrm{i.i.d.} \sim E(\lambda)$，则 $Y = X_1+\cdots+X_k \sim \Gamma(k,\lambda)$。更进一步有：若随机变量 $X_1',\cdots,X_k',\mathrm{i.i.d.} \sim \Gamma(\alpha_i,\lambda)$，则 $Y' = X_1'+\cdots+X_k' \sim \Gamma(\alpha_1+\cdots+\alpha_k,\lambda)$。

证明

$$M_Y(t) = \prod_{i=1}^k M_{X_i}(t_i) = \prod_{i=1}^k \frac{\lambda}{\lambda-t} = (\frac{\lambda}{\lambda-t})^k$$
$$M_{Y'}(t) = \prod_{i=1}^k M_{X_i'}(t_i) = \prod_{i=1}^k (\frac{\lambda}{\lambda-t})^{\alpha_i} = (\frac{\lambda}{\lambda-t})^{\alpha_1+\cdots+\alpha_k}$$

命题 4.8

若令随机变量 $X \sim \Gamma(\alpha,\lambda)$，则 $cX \sim \Gamma(\alpha,\frac{\lambda}{c}), \quad c > 0$

证明

$$M_{cX}(t) = M_X(ct) = (\frac{\lambda}{\lambda-ct})^\alpha = (\frac{\frac{\lambda}{c}}{\frac{\lambda}{c}-t})^\alpha$$

命题 4.9

若令随机变量 $X \sim \Gamma(\alpha,\lambda)$，则 $\mu_k = \frac{\Gamma(\alpha+k)}{\lambda^k\Gamma(\alpha)}, \quad 0 < k$，且则 $\mu_{-k} = \frac{\lambda^k\Gamma(\alpha-k)}{\Gamma(\alpha)}, \quad 0 < k < \alpha$

证明

$$\mu_k = M_X^{(k)}(0) = \frac{1}{\lambda^k}\alpha(\alpha+1)\cdots(\alpha+k-1) = \frac{\Gamma(\alpha+k)}{\lambda^k\Gamma(\alpha)}$$
$$\mu_{-k} = M_X^{(-k)}(0) = \lambda^k\frac{1}{\alpha-1}\cdots\frac{1}{\alpha-k} = \frac{\lambda^k\Gamma(\alpha-k)}{\Gamma(\alpha)}$$

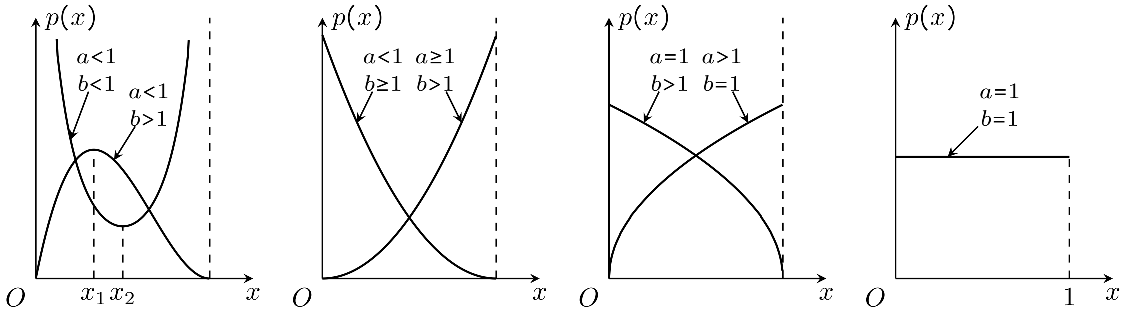


图 4.5: 贝塔密度函数曲线

命题 4.10

4.2.4 贝塔分布

定义 4.13

称以下函数

$$\beta(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$$

为**贝塔函数**，其中参数 $a>0, b>0$.

命题 4.11

贝塔函数具有如下性质：

- $\beta(a,b)=\beta(b,a)$
- $\beta(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

证明 在贝塔函数的积分中令 $y=1-x$，即得

$$\beta(a,b)=\int_1^0(1-y)^{a-1}y^{b-1}(-\mathrm{d}y)=\int_0^1(1-y)^{a-1}y^{b-1}\,\mathrm{d}y=\beta(b,a)$$

由伽玛函数的定义知

$$\Gamma(a)\Gamma(b)=\int_0^{+\infty}\int_0^{+\infty}x^{a-1}y^{b-1}\mathrm{e}^{-(x+y)}\,\mathrm{d}x\,\mathrm{d}y$$

作变量变换 $x=uv, y=u(1-v)$，其雅可比行列式 $J=-u$，故

$$\begin{aligned}\Gamma(a)\Gamma(b)&=\int_0^{+\infty}\int_0^1(uv)^{a-1}[u(1-v)]^{b-1}\mathrm{e}^{-u}u\,\mathrm{d}u\,\mathrm{d}v\\&=\int_0^{+\infty}u^{a+b-1}\mathrm{e}^{-u}\int_0^1 v^{a-1}(1-v)^{b-1}\,\mathrm{d}v=\Gamma(a+b)\end{aligned}$$

$beta(a,b)$,

定义 4.14

若随机变量 X 的密度函数为

$$p(x)=\begin{cases}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, & 0<x<1;\\ 0, & \text{其他},\end{cases}$$

则称 X 服从**贝塔分布**，记作 $X\sim Be(a,b)$，其中 $a>0, b>0$ 都是形状参数.

从图 4.5 可以看出：

- $a < 1, b < 1$ 时，$p(x)$ 是下凸的单峰函数.
- $a > 1, b > 1$ 时，$p(x)$ 是上凸的单峰函数.
- $a < 10, b \geq 1$ 时，$p(x)$ 是下凸的单调减函数.
- $a \geq 1, b < 1$ 时，$p(x)$ 是下凸的单调增函数.
- $a = 1, b = 1$ 时，$p(x)$ 是常函数，且 $Be(1,1) = U(0,1)$.

Beta 分布的特征：

矩母函数	$M(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1} \frac{a+r}{a+b+r}\right)\frac{t^k}{k!}, \quad t < \lambda$
均值	$\mu = \frac{a}{a+b}$
方差	$\sigma^2 = \frac{ab}{(a+b+1)(a+b)^2}$

注

$$\beta(1,1) = U(0,1)$$

命题 4.12

令独立随机变量 $X_1 \sim \Gamma(\alpha_1, \lambda), X_2 \sim \Gamma(\alpha_2, \lambda)$，则 $\frac{X_1}{X_1+X_2} \sim \beta(\alpha_1, \alpha_2)$

证明

4.3 正态分布及其导出分布

4.3.1 正态分布

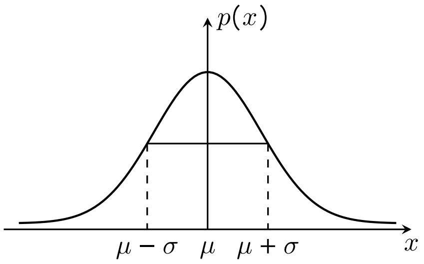
定义 4.15

若随机变量 X 的密度函数为

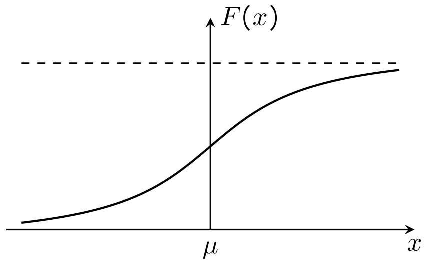
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R} \tag{4.2}$$

则称 X 服从**正态分布** (Normal distribution)，称 X 为**正态变量**，记作 $X \sim N(\mu, \sigma^2)$。其中参数 $\mu \in \mathbb{R}, \sigma > 0$。

正态分布的密度函数 $p(x)$ 的图形如图 4.6(a) 所示，是一条钟形曲线，左右关于 μ 对称，衰减速度由 σ 决定，$\mu \pm \sigma$ 是该曲线的拐点.



(a) 密度函数 $p(x)$



(b) 分布函数 $F(x)$

图 4.6: 正态分布

正态分布的特征：

矩母函数	$M(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}), \quad t \in \mathbb{R}$

均值 $\mu = \mu$

方差 $\sigma^2 = \sigma^2$

命题 4.13

若随机变量 $X \sim N(\mu, \sigma^2)$，则其线性变换 $a, b \in \mathbb{R}, aX + b \sim N(a\mu + b, a^2\sigma^2)$。特别的，常对正态随机变量作标准化：$\frac{X-\mu}{\sigma} \sim N(0,1)$

证明

$$M_{aX+b}(t) = e^{bt}M_X(at) = \exp(bt + a\mu t + \frac{a^2\sigma^2t^2}{2})$$

命题 4.14

若独立随机变量 $X_1, \cdots, X_k$ 满足 $X_i \sim N(\mu_i, \sigma_i^2)$，则其和 $Y = X_1 + \cdots + X_k \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2)$

证明

$$M_{X_1+\cdots+X_n} = \prod_{i=1}^n M_{X_i}(t) = \exp(\sum_{i=1}^k \mu_i t + \sum_{i=1}^k \frac{\sigma_i^2 t^2}{2})$$

推论 4.1

若随机变量 $X_1, \cdots, X_k$ i. i. d. $\sim N(\mu, \sigma^2)$，则其平均 $\overline{X} = (X_1 + \cdots + X_k)/k \sim N(\mu, \frac{\sigma^2}{k})$

定理 4.3

若随机变量 $X_1, \cdots, X_k$ i. i. d $\sim N(\mu, \sigma^2)$，并且定义其样本均值为 $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$，其样本方差为 $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$，则有以下关系：

1. $\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$，即 $\sqrt{n}(\overline{X_n} - \mu)/\sigma \sim N(0,1)$
2. 随机变量 $\overline{X_n}$ 与随机向量 $(X_1 - \overline{X_n}, \cdots, X_n - \overline{X_n})$ 独立
3. $\overline{X_n}, S_n^2$ 独立
4. $\frac{(n-1)S_n^2}{\sigma} \sim \chi_n$
5. $\frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n} \sim t_{n-1}$

证明

4.3.2 卡方分布

定义 4.16

设 $Z_1, \ldots, Z_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $X = Z_1^2 + \ldots + Z_n^2$ 的分布称为 n **自由度的卡方分布** (Chi-square distribution)，记为 $X \sim \chi_n^2$.

卡方分布的特征：

参数 $n \in \mathbb{N}_+$

概率密度函数 $p(x) = \begin{cases} \frac{(1/2)^{\frac{n}{2}}}{\Gamma(n/2)} y^{\frac{n}{2}-1} \mathrm{e}^{-\frac{y}{2}} & y > 0 \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = (\frac{1}{1-2t})^{\frac{n}{2}}$

均值 $\mu = n$

方差 $\sigma^2 = 2n$

实例

该密度函数的图像是一个只取非负值的偏态分布, 见图 4.7。

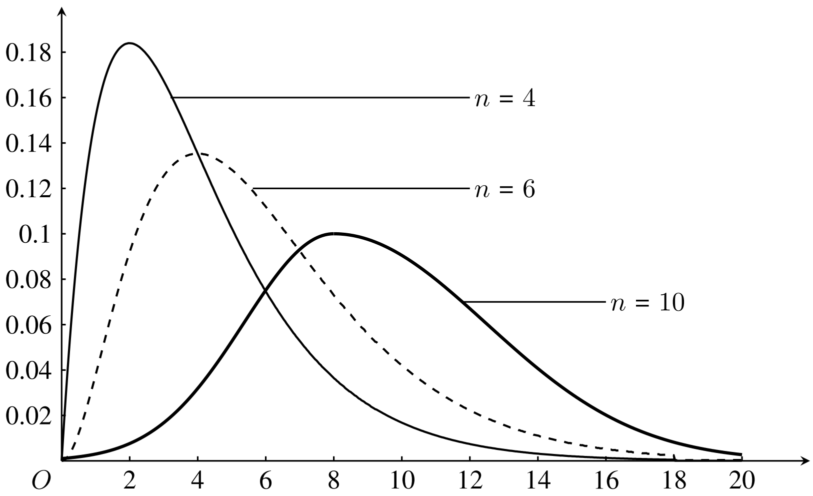


图 4.7: $\chi^2(n)$ 分布的密度函数

命题 4.15

若 $X \sim N(0,1)$，则 $X^2 \sim Ga(1/2, 1/2), Y \sim Ga(n/2, 1/2) = \chi^2(n)$

证明

推论 4.2

若随机变量 $X_1, \cdots, X_k, \text{i.i.d.} \sim \chi^2_{n_i}$，则 $Y = X_1 + \cdots + X_k \sim \chi^2_{n_1+\cdots+n_k}$。

4.3.3 F 分布

定义 4.17

设独立随机变量 $X_1 \sim \chi^2(m), X_2 \sim \chi^2(n)$，则称 $F = \frac{X_1/m}{X_2/n}$ 的分布是自由度为 m 与 n 的 **F 分布**，记为 $F \sim F(m,n)$, 其中 m 称为分子自由度，n 称为分母自由度。

F 分布的特征：

参数 $m, n \in \mathbb{N}_+$

概率密度函数

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}$$

矩母函数 不存在

均值 $\mu = \frac{n}{n-2}$

方差 $\sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

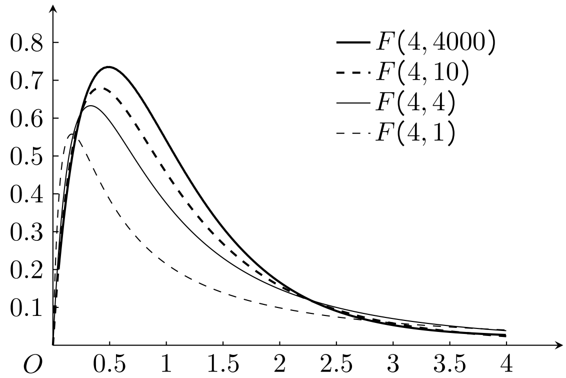


图 4.8: F 分布的密度函数，是一个只取非负值的偏态分布

实例

命题 4.16

F 分布的密度函数为

证明 首先我们导出 $Z = \frac{X_1}{X_2}$ 的密度函数, 若记 $p_1(x)$ 和 $p_2(x)$ 分别为 $\chi^2(m)$ 和 $\chi^2(n)$ 的密度函数, 根据独立随机变量商的分布的密度函数公式2.4。Z 的密度函数为

$$p_Z(z) = \int_0^{=\infty} x_2 p_1(zx_2) p_2(x_2)\,\mathrm{d}x_2$$
$$= \frac{z^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{\frac{m+n}{2}}} \int_0^{+\infty} x_2^{\frac{m+n}{2}-1} \mathrm{e}^{-\frac{x_2}{2}(1+z)}\,\mathrm{d}x_2.$$

运用变换 $u = \frac{x_2}{2}(1+z)$, 可得

$$p_Z(z) = \frac{z^{\frac{m}{2}-1}(1+z)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} u^{\frac{m+n}{2}-1} \mathrm{e}^{-u}\,\mathrm{d}u.$$

最后的定积分为伽马函数 $\Gamma\left(\frac{m+n}{2}\right)$, 从而

$$p_Z(z) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}, \quad z > 0.$$

第二步, 我们导出 $F = \frac{n}{m}Z$ 的密度函数, 对 $y > 0$, 有

$$p_F(y) = p_Z\left(\frac{m}{n}y\right) \cdot \frac{m}{n}$$
$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}y\right)^{\frac{m}{2}-1} \left(1+\frac{m}{n}y\right)^{-\frac{m+n}{2}} \cdot \frac{m}{n}$$
$$= \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} y^{\frac{m}{2}-1} \left(1+\frac{m}{n}y\right)^{-\frac{m+n}{2}}.$$

注 由 F 分布的构造知, 若 $F \sim F(m,n)$, 则有 $1/F \sim F(n,m)$

4.3.4 t 分布

定义 4.18

设随机变量 $Z \sim N(0,1)$ 与 $U \sim \chi_n^2$ 独立，则称 $t = \frac{Z}{\sqrt{U/n}}$ 的分布为自由度为 n 的 **t 分布**, 记为 $t \sim t_n$.

t 分布的密度函数的图像是一个关于纵轴对称的分布 (图 4.9), 与标准正态分布的密度函数形状类似, 只是峰

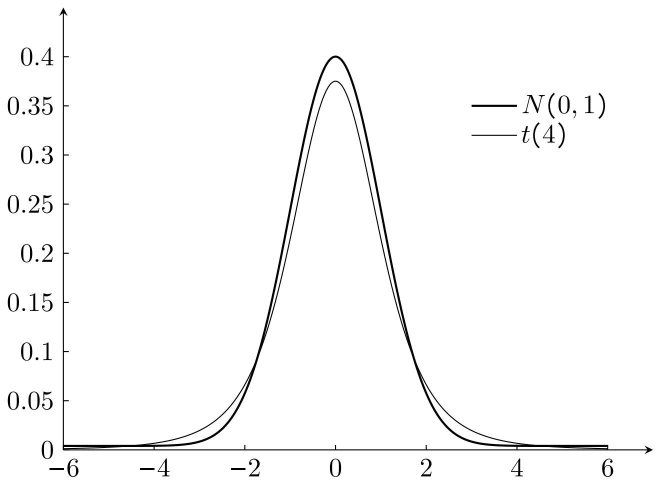


图 4.9: t 分布与 $N(0,1)$ 的密度函数，t 分布尾更重

比标准正态分布低一些, 尾部的概率比标准正态分布的大一些.

t 分布的特征：

参数	$n \in \mathbb{N}_+$
概率密度函数	$p(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}}$
矩母函数	不存在
均值	$\mu = 0, n > 1$
方差	$\sigma^2 = \frac{n}{n-2}, n > 2$
实例	

注

- 自由度 $n = 1$ 的 t 分布就是标准柯西分布，它的均值不存在;
- 自由度 $n \to \infty$ 时，t_n 趋近于 $N(0,1)$

命题 4.17

t 分布的密度函数为

证明 由标准正态密度函数的对称性知, X_1 与 $-X_1$ 有相同分布, 从而 t 与 $-t$ 有相同分布. 这意味着: 对任意实数 y 有

$$P(0 < t < y) = P(-y < -t < y) = P(-y < -t < 0).$$

于是

$$P(0 < t < y) = \frac{1}{2}P(t^2 < y^2).$$

由 F 变量构造可知, $t^2 = \frac{X_1^2}{X_2^2/n} \sim F(1,n)$, 将上式两边关于 y 求导可微 t 分布的密度函数为

$$\begin{aligned} p_t(y) = yp_F(y^2) &= \frac{\Gamma\left(\frac{1+n}{2}\right)\left(\frac{1}{n}\right)^{\frac{1}{n}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(y^2\right)^{\frac{1}{2}-1}\left(1+\frac{1}{n}y^2\right)^{-\frac{1+n}{2}} y \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < y < +\infty. \end{aligned}$$

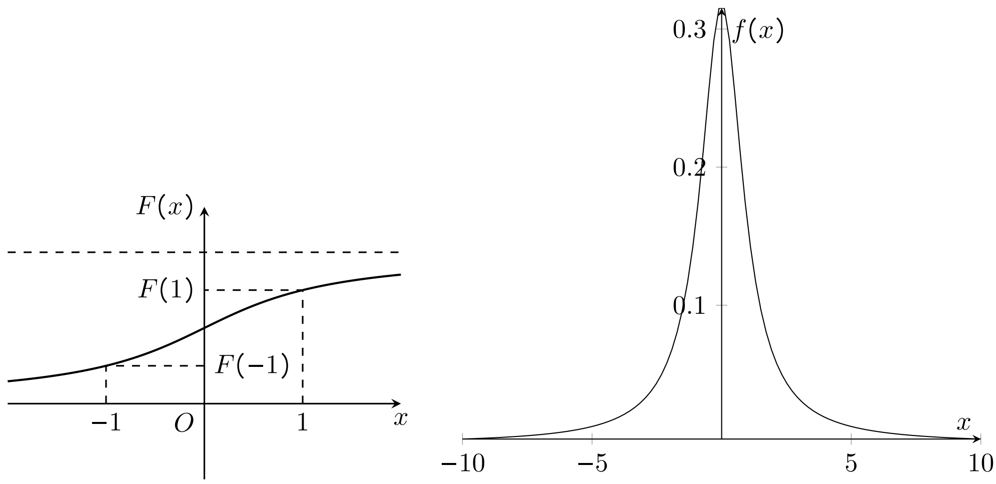
4.3.5 柯西分布

定义 4.19

若随机变量 X 的密度函数为

$$f(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2 + (x-\mu)^2}, x \in \mathbb{R}$$

则称 X 服从**柯西分布** (Cauchy distribution)，记作 $X \sim C(\mu, \sigma)$。其中参数 $\mu \in \mathbb{R}, \sigma > 0$。♣



(a) 柯西分布累积函数　(b) 柯西分布密度函数

图 4.10: 柯西分布

柯西分布的特征：

累积函数	$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(\mu + \sigma x)$
矩母函数	除 $t = 0$ 外不存在
特征函数	$\phi_X(t) = \exp(i\mu t - \sigma\lvert t\rvert)$
均值	不存在
方差	不存在

注 此类因极端值概率密度较高而导致均值、方差不存在的分布称为重尾分布。

命题 4.18

$C(0,1) = t_1$ ♠

命题 4.19

若随机变量 X, Y i.i.d. $\sim N(0,1)$，则 $\frac{X}{Y} \sim C(0,1)$

证明

命题 4.20

若随机变量 $X \sim C(\mu,\sigma)$，则其线性变换 $a, b \in \mathbb{R}, aX + b \sim C(a\mu + b, |a|\sigma)$。

证明

$$\phi_{aX+b}(t) = e^{ibt}\phi_X(at) = \exp(ibt + ia\mu t - \sigma|at|)$$

命题 4.21

若独立随机变量 $X_1, \cdots, X_k$ 满足 $X_i \sim C(\mu_i, \sigma_i)$，则其和 $Y = X_1 + \cdots + X_k \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i)$

证明

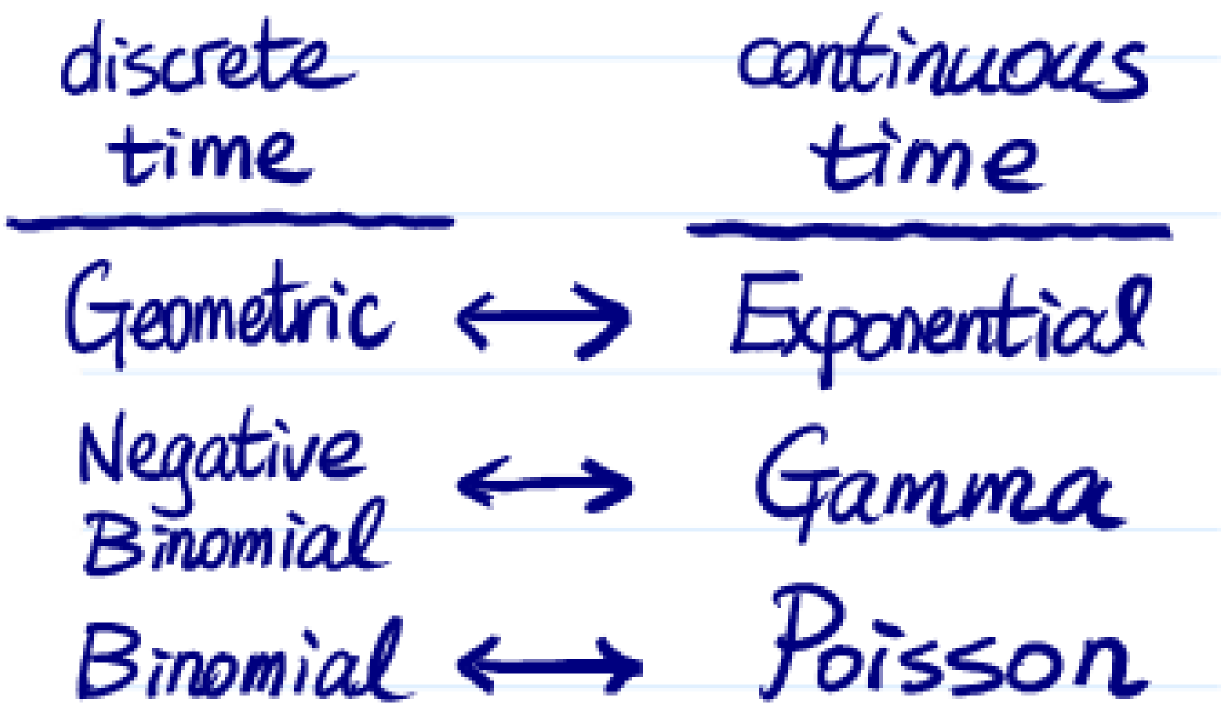
$$\phi_{X_1+\cdots+X_n} = \prod_{i=1}^n \phi_{X_i}(t) = \exp(it\sum_{i=1}^k \mu_i - |t|\sum_{i=1}^k \sigma_i)$$

推论 4.3

若随机变量 $X_1, \cdots, X_k$ i.i.d. $\sim N(\mu, \sigma^2)$，则其平均 $\overline{X} = (X_1 + \cdots + X_k)/k \sim C(\mu, \sigma)$

注 此处的 σ 不代表方差，所以不遵守 $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{k}$ 的关系。

4.4 各分布间关系



可参考网站http://www.math.wm.edu/~leemis/chart/UDR/UDR.html

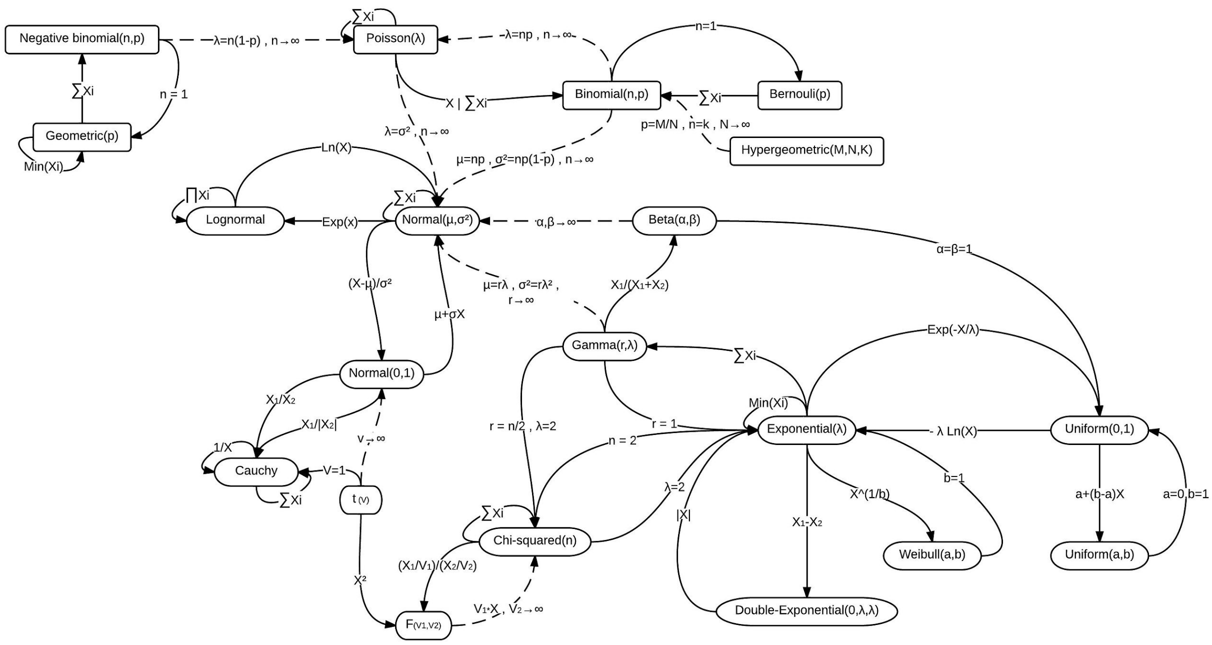


图 4.11: 各分布间的联系