常用分布

司继春

[1] 上海对外经贸大学

2026年10月

概览

1. 离散型随机变量
2. 连续型随机变量
3. 随机变量的变换
4. 分布族
5. 单参数指数分布族

伯努利分布

伯努利分布即伯努利试验（Bernoulli trial）的结果. 伯努利试验即试验结果只有两种可能性（成功，失败），定义为$X = 1\ if$ 成功, $= 0\ if$ 失败. 如果成功的概率为p，那么**伯努利分布（Bernoulli distribution）**$X \sim \mathcal{B}er(p)$即

$$X = \begin{cases} 1 & with\ probability\ p \\ 0 & with\ probability\ 1-p \end{cases} \quad 0 \leq p \leq 1$$

- 期望：$\mathbb{E}(X) = p$
- 方差：$\mathbb{V}(X) = p(1-p)$
- 矩母函数：$1 - p + pe^t$

二项分布

二项分布即独立重复N次伯努利实验，成功次数$Y=\sum_{i=1}^{N}X_i, X_i\sim\mathcal{B}er\left(p\right)$的分布. 简单计算可以得到，随机变量$Y$的概率质量函数:

$$\mathrm{P}\left(Y=y|N,p\right)=\left(\begin{array}{c}N\\y\end{array}\right)p^{y}\left(1-p\right)^{N-y},y=1,2,...,N$$

我们成Y服从**二项分布（binomial distribution）**，记为$Y\sim\mathcal{B}i\left(N,p\right)$.

- 期望：$\mathbb{E}\left(Y\right)=Np$
- 方差：$\mathbb{V}\left(Y\right)=Np\left(1-p\right)$
- 矩母函数：$\left(1-p+pe^{t}\right)^{N}$

负二项分布

负二项分布（negative binomial distribution）用于建模为了达到r次成功所需要的试验次数，如果记为Z，那么

$$\begin{aligned}\mathrm{P}(Z=z|r,p) &= \left(\begin{array}{c} z-1 \\ r-1 \end{array}\right) p^{r-1}(1-p)^{z-r}\cdot p \\ &= \left(\begin{array}{c} z-1 \\ r-1 \end{array}\right) p^{r}(1-p)^{z-r}, z=r,r+1,...\end{aligned}$$

进一步，为了达到r次成功，需要经过的失败次数$X=Z-r$，从而

$$\begin{aligned}\mathrm{P}(X=x|r,p) &= \mathrm{P}(Z-r=x|r,p)=\mathrm{P}(Z=r+x|r,p) \\ &= \left(\begin{array}{c} r+x-1 \\ r-1 \end{array}\right) p^{r}(1-p)^{x} = \left(\begin{array}{c} r+x-1 \\ x \end{array}\right) p^{r}(1-p)^{x}, x=0,1,...\end{aligned}$$

记为$X\sim\mathcal{NB}(r,p)$.

- 期望：$\mathbb{E}(Z)=\frac{r(1-p)}{p}$
- 方差：$\mathbb{V}(Z)=\frac{r(1-p)}{p^2}$
- 矩母函数：$\left(\frac{p}{1-e^t+pe^t}\right)^r, t<-\ln(1-p)$

几何分布

- **几何分布**（geometric distribution）是最简单形式的负二项分布，如果一个随机变量$(V-1)\sim\mathcal{NB}(1,p)$，则随机变量$V$服从几何分布：

$$\mathrm{P}(V=v|p)=p(1-p)^{v-1}$$

  我们记$V\sim\mathcal{G}(p)$.
- 几何分布具有无记忆性，即

$$\mathrm{P}(V>s|V>t)=\mathrm{P}(V>s-t),s>t$$

泊松分布

如果随机变量X的概率质量函数为

$$\mathrm{P}(X=x)=\frac{\lambda^x}{x!}e^{-\lambda}, x=0,1,2,...$$

那么我们称随机变量X服从**泊松分布（Poisson distribution）**，记为$X\sim\mathcal{P}(\lambda)$. 泊松分布经常被用来建模次数、个数等相关问题.

- 期望：$\mathbb{E}(X)=\lambda$
- 方差：$\mathbb{V}(X)=\lambda$
- 矩母函数：$e^{\lambda\left(e^t-1\right)}$

均匀分布

如果随机变量X在区间$[a,b]$内的概率密度函数为常数，即

$$f\left(x|a,b\right)=\mathbb{1}\left\{a\leq x\leq b\right\}\frac{1}{b-a}$$

则称随机变量X服从区间$[a,b]$上的**均匀分布（uniform distribution）**，记为$X\sim\mathcal{U}\left(a,b\right)$.

- 期望：$\mathbb{E}\left(X\right)=\frac{a+b}{2}$
- 方差：$\mathbb{V}\left(X\right)=\frac{(b-a)^2}{12}$
- 矩母函数：$\frac{e^{tb}-e^{ta}}{t(b-a)}$

指数分布

若随机变量X的概率密度函数为

$$f\left(x|a,\beta\right)=\mathbb{1}\left\{x>a\right\}\frac{1}{\beta}e^{-\frac{x-a}{\beta}}$$

那么我们称X服从**指数分布**（**exponential distribution**），记为$X\sim\mathcal{E}\left(a,\beta\right)$. 如果$a=0$，可以简记为$X\sim\mathcal{E}\left(\beta\right)$.

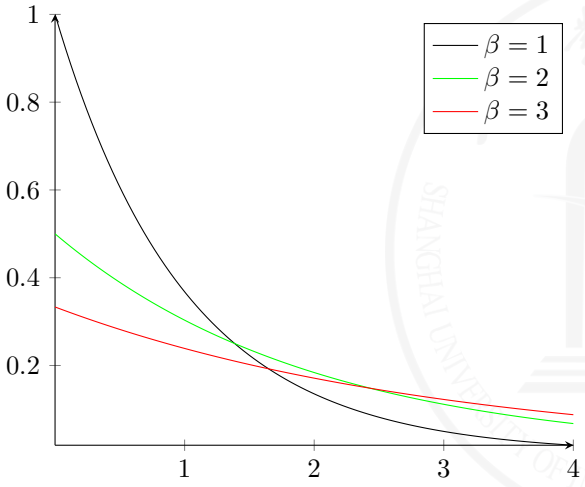
- 分布函数：$F\left(x\right)=\mathbb{1}\left\{x>a\right\}\left(1-e^{-\frac{x-a}{\beta}}\right)$
- 期望：$\mathbb{E}\left(X\right)=\beta+a$
- 方差：$\mathbb{V}\left(X\right)=\beta^{2}$
- 矩母函数：$\left(1-\beta t\right)^{-1}e^{at}$

指数分布

- 如果一个随机变量$Y \sim \mathcal{P}(\lambda)$，其中参数$\lambda$代表一段时间$T$内到达的平均次数，所以对于时间$t$内，平均到达的人数即$\lambda\frac{t}{T}$，因而在$t$时间内只有0次到达的概率为$e^{-\frac{\lambda t}{T}}$.
- 如果在时间t内有0次到达，意味着等待时间大于t，因而等待时间X大于t的概率$\mathrm{P}(X > t) = e^{-\frac{\lambda t}{T}}$，所以$\mathrm{P}(X \leq t) = 1 - e^{-\frac{\lambda t}{T}}$，因而等待时间$X$服从指数分布，即$X \sim \mathcal{E}\left(0, \frac{T}{\lambda}\right)$.
- 同样，指数分布也具有无记忆性，即如果令$\beta = \frac{T}{\lambda}$，那么$X \sim \mathcal{E}(0, \beta)$，那么

$$\mathrm{P}(X > s | X > t) = \mathrm{P}(X > s - t), s > t$$

指数分布



Beta分布

如果一个随机变量$X \in (0,1)$，且其分布函数为

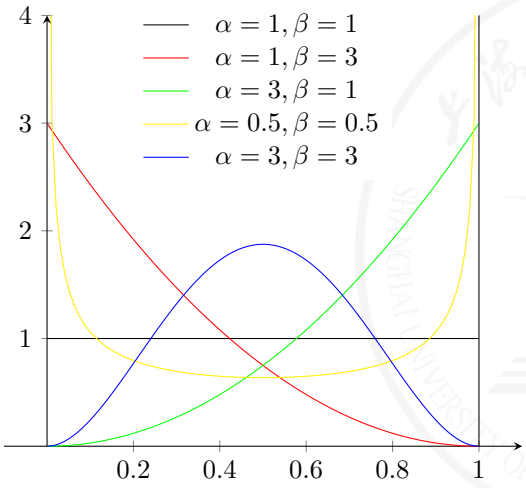
$$f(x|\alpha,\beta) = \mathbb{1}\{0 < x < 1\} \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \alpha > 0, \beta > 0$$

那么我们称X服从**beta分布（beta distribution）**，记为$X \sim \mathcal{B}eta(\alpha,\beta)$. 其中

$$\mathrm{B}(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- 期望：$\mathbb{E}(X) = \frac{\alpha}{\beta+a}$
- 方差：$\mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Beta分布


4
3
2
1
0.2
0.4
0.6
0.8
1
$\alpha = 1, \beta = 1$
$\alpha = 1, \beta = 3$
$\alpha = 3, \beta = 1$
$\alpha = 0.5, \beta = 0.5$
$\alpha = 3, \beta = 3$

逻辑斯蒂分布

若随机变量X的分布函数为

$$F\left(x|\mu,\sigma\right)=\frac{e^{\frac{x-\mu}{\sigma}}}{1+e^{\frac{x-\mu}{\sigma}}}$$

那么我们称X服从**逻辑斯蒂分布（logistic distribution）**，其密度函数为

$$f\left(x|\mu,\sigma\right)=\frac{1}{\sigma}\frac{e^{-\frac{x-\mu}{\sigma}}}{\left[1+e^{-\frac{x-\mu}{\sigma}}\right]^2}$$

记为：$X\sim\mathcal{LG}\left(\mu,\sigma\right)$. 标准Logistic分布的分布函数

$$F\left(x|0,1\right)=\frac{e^{x}}{1+e^{x}}$$

经常被应用在离散选择（discrete choice）模型、分类器（classifier）等模型中.

- 期望：$\mathbb{E}\left(X\right)=\mu$
- 方差：$\mathbb{V}\left(X\right)=\sigma^2\frac{\pi^2}{3}$
- 矩母函数：$e^{\mu t}\mathrm{B}\left(1-\sigma t,1+\sigma t\right),t\in\left(-\frac{1}{\sigma},\frac{1}{\sigma}\right)$

正态分布

如果一个随机变量X潜在受到非常多的独立因素的影响，即$X = f_1 + f_2 + \cdots$，而每个f_i又不能单独对X有非常大的影响，那么一般来说X将会服从**正态分布**（**normal distribution**）或者**高斯分布**（**Gaussian distribution**）. 正态分布的密度函数为

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

记为$X \sim \mathcal{N}(\mu, \sigma^2)$.

- 期望：$\mathbb{E}(X) = \mu$
- 方差：$\mathbb{V}(X) = \sigma^2$
- 矩母函数：$e^{\mu t + \frac{\sigma^2 t^2}{2}}$

正态分布

- 如果$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$，如果令$Z = \frac{X-\mu}{\sigma}$，
  则$\mathbb{E}(Z) = \frac{\mathbb{E}(X)-\mu}{\sigma} = 0$，$\mathbb{V}(Z) = \frac{1}{\sigma^2} \cdot \mathbb{V}(X) = 1$，随机变量$Z$服从$\mu = 0, \sigma = 1$的正态分布，即$Z \sim \mathcal{N}(0,1)$，我们称之为**标准正态分布**（**standard normal distribution**）.
- 标准正态分布的密度函数为

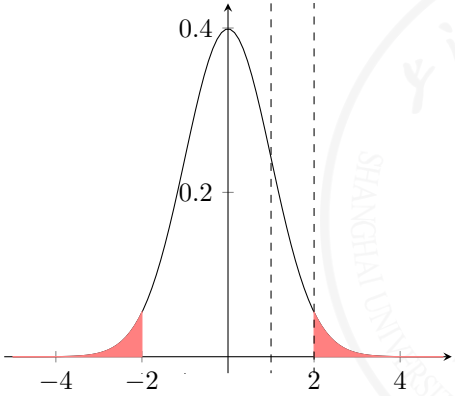
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- 而其分布函数为

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t$$

由于$\Phi(x)$没有初等函数表示，因而一般我们用$\Phi(x)$来代表标准正态的分布函数.

正态分布



正态分布

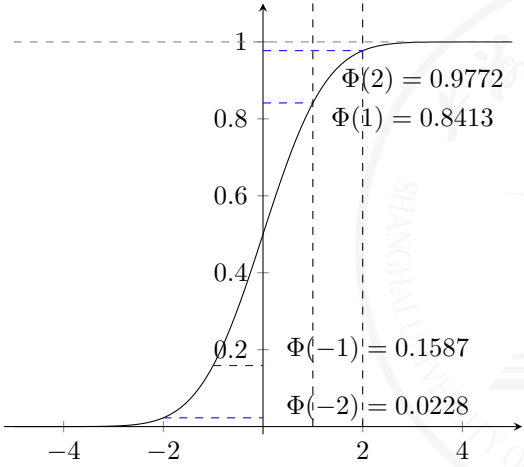
- 由于正态分布为对称分布，即$\phi(x)=\phi(-x)$，因而分布函数$\Phi(x)=1-\Phi(-x)$.
- 如果$X\sim\mathcal{N}(\mu,\sigma^2)$，那么$Z=\frac{X-\mu}{\sigma}\sim\mathcal{N}(0,1)$，根据标准正态的分布函数$\Phi(x)$可以计算$X$在区间内取值的概率，比如

$$\mathrm{P}(|X-\mu|\leq\sigma)=\mathrm{P}(|Z|<1)=\Phi(1)-\Phi(-1)=2\Phi(1)-1=0.6827$$

同理，有

$$\begin{aligned}
\mathrm{P}(|X-\mu|\leq 1.65\sigma)&=P(|Z|<1.65)\approx 0.90\\
\mathrm{P}(|X-\mu|\leq 1.96\sigma)&=P(|Z|<1.96)\approx 0.95\\
\mathrm{P}(|X-\mu|\leq 2\sigma)&=P(|Z|<2)=0.9545\\
\mathrm{P}(|X-\mu|\leq 2.58\sigma)&=P(|Z|<2.58)\approx 0.99\\
\mathrm{P}(|X-\mu|\leq 3\sigma)&=P(|Z|<3)=0.9973\\
\mathrm{P}(|X-\mu|\leq 5\sigma)&=P(|Z|<5)\geq 1-10^{-6}\\
\mathrm{P}(|X-\mu|\leq 6\sigma)&=P(|Z|<6)\geq 1-10^{-8}
\end{aligned}$$

正态分布



对数正态分布

若随机变量$Y = e^X, X \sim \mathcal{N}\left(\mu, \sigma^2\right)$，则我们称随机变量$Y$服从**对数正态分布（lognormal distribution）**，记为$Y \sim \mathcal{LN}\left(\mu, \sigma^2\right)$.

- 期望：$\mathbb{E}\left(Y\right) = e^{\mu + \frac{\sigma^2}{2}}$
- 方差：$\mathbb{V}\left(Y\right) = e^{2\left(\mu+\sigma^2\right)} - e^{2\mu+\sigma^2}$

对数正态分布经常用于建模城市人口大小、企业规模等变量，Gibrat（1931）指出了这一现象，因而被称为**Gibrat法则（Gibrat's law）**.

- 之所以会出现对数正态分布，与城市、企业的随机比例增长（proportional rate of growth）有关系：如果城市或者企业规模的增长率的期望和方差相同，那么经过一段时间的增长，其大小就会近似服从对数正态分布.

幂律

- 所谓幂律，即两个变量X, Y之间存在着类似于

$$Y = cX^{\beta}$$

  的关系，其中：
  - c是一个一般不太重要的常数
  - β则是**幂律指数**（power law exponent）
- 如果将其取对数，则幂律关系会变成一个线性关系：

$$\ln Y = \ln c + \beta \ln X$$

- c仅仅反映尺度，而幂律具有**尺度不变**（scale invariance）的性质：
  - 比如，如果X是一个城市的人口，以“人数”为单位，那么以“百万人”为单位的人数$\tilde{X} = {}^{X}/_{10^6} \triangleq kX$，那么

$$Y = cX^{\beta} = c\left(\frac{\tilde{X}}{k}\right)^{\beta} = \tilde{c}\tilde{X}^{\beta}$$

    同样满足幂律，仅仅是前面的常数变化了.

齐夫定律

- 具体到城市大小，如果令Y是城市从大到小的排名（Rank），X为城市的人口数（Size），大量研究发现这两个变量几乎满足幂律的关系，比如Gabaix（2016）对美国25万人以上城市的估计，这两者的关系大约是

$$\ln(\text{Rank}) = 7.88 - 1.03 \ln(\text{Size})$$

  其中Rank作为排序，数值越小代表城市越大，如果有N个城市，那么$\frac{\text{Rank}}{N}$应该就是城市人口大于等于这个城市的比例.

- 如果将其写为幂律的形式，那么这个关系意味着

$$\mathrm{P}(X \geq x) = \frac{\text{Rank}}{N} = e^{\ln(\text{Rank}) - \ln N} = \frac{a}{x^{\zeta}}$$

  其中$\zeta = 1.03 \approx 1$，而$a = \frac{e^{7.88}}{N}$.

齐夫定律

根据如上幂律，如果X满足

$$\mathrm{P}\left(X \geq x\right) \propto \frac{1}{x^{\zeta}}$$

那么X的密度函数为

$$f_X\left(x\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left[1 - \mathrm{P}\left(X \geq x\right)\right] \propto \frac{1}{x^{\zeta+1}}$$

其中$\propto$是“成比例”的意思，即

$$f_X\left(x\right) = C \cdot \frac{1}{x^{\zeta+1}}$$

其中C使得密度函数线下面积为1.

齐夫定律

- 如果令$\zeta = 1$，那么以上关系可以写为

$$\mathrm{P}\left(\mathrm{Size} \geq x\right) \propto \frac{1}{x} \Longleftrightarrow x \propto \frac{1}{\mathrm{Rank}}$$

即城市人口数与其排名的倒数成比例.

- 实际上，以上关系不仅仅对于城市的大小成立，在自然和社会中的很多数量都满足这一规律，比如语言中单词的频率、一个社会中姓氏人数的分布、企业的规模等等，这一规律以以语言学家Zipf命名，被称作**齐夫定律（Zipf’s law）**.
- 齐夫定律要求$\zeta = 1$，当然也有很多现象不满足这一要求，但是同样满足幂律. 比如，在股票市场中，股票收益率满足“三次法则”（cubic law），即

$$\mathrm{P}\left(|r_t| > x\right) \propto \frac{1}{x^3}$$

意味着收益率的密度函数在尾部收敛到0的速度大约是x^{-3}级别，远慢于正态分布的指数级，从而极端涨跌的概率比正态分布所预言的要大.

二八定律

- 除齐夫定律外，另一个与幂律有关的定律是**二八定律**（80/20 rule，或**帕累托法则**，Pareto principle）
- 该定律来自于意大利经济社会学家帕累托. 帕累托发现，在意大利，大约80%的土地集中在20%的人手上
- 此后发现很多与分配、质量控制等问题有关的现象都可以适用该定律.

帕累托分布

与帕累托法则对应的有**帕累托分布**（Pareto distribution），同样符合幂律.

帕累托分布

如果令随机变量X为一个家庭的财富，那么财富高于x的人的比例可以用生存函数$S(x) = 1 - F(x)$来描述，如果假设该比例满足幂律，即

$$S(x) = \mathrm{P}\{X > x\} = \begin{cases} \left(\frac{x_m}{x}\right)^{\alpha} & x \geq x_m \\ 1 & x < x_m \end{cases}$$

其中x_m为X能取到的最小值，那么我们称X服从形状参数为α的帕累托分布，记为$X \sim \mathcal{P}a(x_m, \alpha)$.

帕累托分布

- 其分布函数为

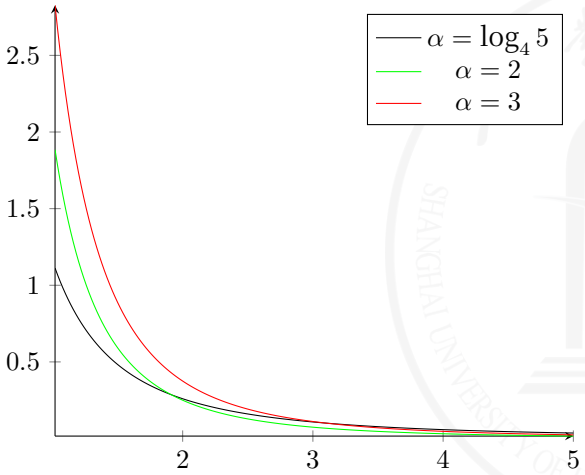
$$F(x) = \mathbb{1}\{x \geq x_m\}\left[1 - \left(\frac{x_m}{x}\right)^{\alpha}\right]$$

- 对应的密度函数为

$$f(x) = \mathbb{1}\{x \geq x_m\}\frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$$

- 根据生存函数，如果α越小，那么大于给定财富x的人数比例收敛到0的速度越慢，即右侧尾部收敛到0的速度更慢，意味着更大的分配不公平，故α也被称为尾部指数（tail index）或者帕累托指数（Pareto index）.
- 期望：$\frac{\alpha}{\alpha-1}x_m, \alpha > 1$，否则为$\infty$
- 方差：$\frac{\alpha}{(\alpha-1)^2(\alpha-2)}x_m^2, \alpha > 2$，否则为$\infty$

帕累托分布



帕累托分布与二八定律

现在考虑当α取何值时整个社会的财富分布会满足二八定律.

- 首先，20%的最富有人对应于$S\left(x^*\right)=0.2$，从而

$$\ln\left(\frac{x_m}{x^*}\right)=\frac{-\ln 5}{\alpha} \tag{1}$$

- 为了计算所有$X>x^*$的人占有全社会财富的比例，考虑一个划分$\Pi: x_m=t_0<t_1<\cdots<t_n=x^m$，那么财富在$[x_m,x^m]$区间内所占有的财富可以近似计算为

$$I=\sum_{k=1}^{n}\frac{t_k+t_{k-1}}{2}\left[F\left(t_k\right)-F\left(t_{k-1}\right)\right]$$

其中$F\left(t_k\right)-F\left(t_{k-1}\right)$为财富在$(t_{k-1},t_k]$区间内的人数比例，$\frac{t_k+t_{k-1}}{2}$为每人财富的近似计算，从而在$[x_m,x]$区间内所占有的财富可以使用积分

$$\int_{x_m}^{x}x\mathrm{d}F\left(x\right)=\int_{x_m}^{x}t\frac{\alpha x_m^{\alpha}}{t^{\alpha+1}}\mathrm{d}t=\alpha x_m^{\alpha}\int_{x_m}^{t}\frac{1}{t^{\alpha}}\mathrm{d}t=\alpha x_m^{\alpha}\left[\frac{1}{1-\alpha}t^{-\alpha+1}\right]_{x_m}^{x}=\frac{\alpha}{\alpha-1}\left(x_m-x^{1-\alpha}x_m^{\alpha}\right)$$

帕累托分布与二八定律

对于$\alpha > 1$，全社会的总财富量可以通过

$$\int_{x_m}^{\infty} x \mathrm{d}F(x) = \frac{\alpha}{\alpha - 1} \left(x_m - x^{1-\alpha} x_m^{\alpha} \right)_{x_m}^{\infty} = \frac{\alpha}{\alpha - 1} x_m$$

计算. 而大于x^*的总财富可以通过

$$\int_{x^*}^{\infty} x \mathrm{d}F(x) = \frac{\alpha}{\alpha - 1} \left(x_m - x^{1-\alpha} x_m^{\alpha} \right)_{x_m}^{\infty} = \frac{\alpha}{\alpha - 1} x_m^{\alpha} \left(x^* \right)^{1-\alpha}$$

计算，二八定律意味着两者比例为

$$\frac{\frac{\alpha}{\alpha-1} x_m^{\alpha} \left(x^* \right)^{1-\alpha}}{\frac{\alpha}{\alpha-1} x_m} = \left(\frac{x_m}{x^*} \right)^{\alpha-1} = 0.8 \tag{2}$$

帕累托分布与二八定律

结合式(1)与式(2)，可以得到

$$\ln\left(\frac{x_m}{x^*}\right)=\frac{-\ln 5}{\alpha}=\frac{1}{\alpha-1}\ln\frac{4}{5}$$

从而得到

$$\alpha=\frac{\ln 5}{\ln 4}=\log_4 5$$

即，二八法则可以对应$\alpha=\log_4 5$的帕累托分布.

χ^2分布

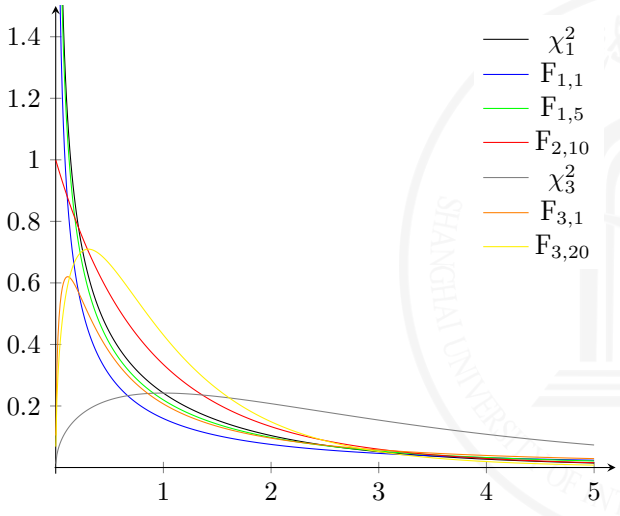
K个独立的标准正态分布的平方和的分布被称为**卡方分布**（Chi-square distribution）或者χ^2分布. 即，如果$Z_1, Z_2, ..., Z_K$为K个独立的标准正态分布，那么

$$X = \sum_{i=1}^{K} Z_i^2 \sim \chi^2(K)$$

其中参数K为卡方分布的**自由度**（degrees of freedom）.

- 期望：$\mathbb{E}(X) = K$
- 方差：$\mathbb{V}(X) = 2K$
- 矩母函数：$(1-2t)^{-\frac{K}{2}}$

χ^2分布与F分布



t分布

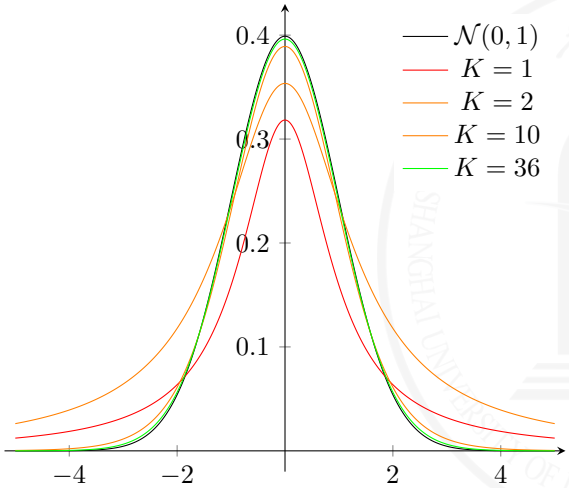
如果存在一个标准正态分布$Z \sim \mathcal{N}(0,1)$，以及一个χ^2分布$X \sim \chi^2(K)$，且Z和X**独立**，那么随机变量

$$T = \frac{Z}{\sqrt{X/K}}$$

即服从**学生氏t分布**（**students' t-distribution**）或者简称**t分布**（**t-distribution**），记为$T \sim \mathrm{t}(K)$，参数K为自由度.

- 期望：$\mathbb{E}(X) = 0, K > 1$
- 方差：$\mathbb{V}(X) = \frac{K}{K-2}, K > 2$
- 矩母函数：不存在

标准正态分布与t分布



F分布

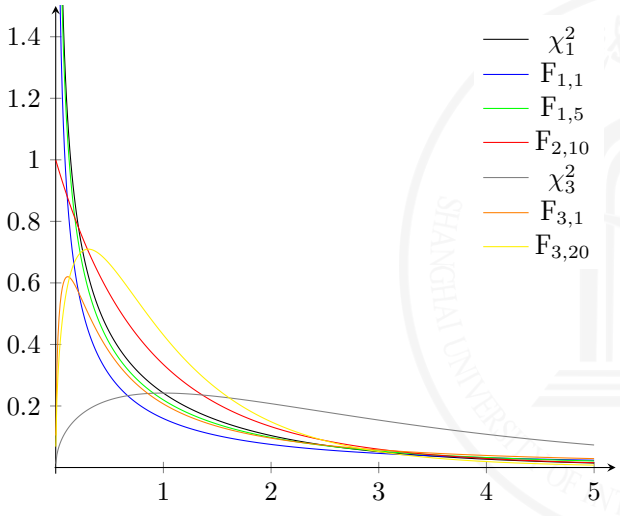
如果存在两个**独立的**χ^2分布$X_1 \sim \chi^2(n), X_2 \sim \chi^2(m)$，那么随机变量

$$F = \frac{X_1/n}{X_2/m}$$

即服从F**分布**（F-**distribution**），记为$F \sim \mathrm{F}(n,m)$，参数n,m为自由度.

- 期望：$\mathbb{E}(X) = \frac{m}{m-2}, m > 2$
- 方差：$\mathbb{V}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, m > 4$
- 矩母函数：不存在

χ^2分布与F分布



F分布的性质

- 如果$X_k \sim \Gamma(\alpha_k, \beta_k)$且独立，那么

$$\frac{\alpha_2\beta_1 X_1}{\alpha_1\beta_2 X_2} \sim \mathrm{F}(2\alpha_1, 2\alpha_2)$$

- 如果$X \sim \mathcal{B}eta(n/2, m/2)$，那么

$$\frac{mX}{n(1-X)} \sim \mathrm{F}(n, m)$$

反之亦成立

- 如果$X \sim \mathrm{F}(n, m)$，那么

$$\lim_{m\to\infty} nX \sim \chi^2(n)$$

- 如果$T \sim \mathrm{t}(K)$，那么

$$T^2 \sim \mathrm{F}(1, K)$$

柯西分布

若随机变量X的概率密度函数为

$$f\left(x|\mu,\sigma\right)=\frac{1}{\pi\sigma}\left[1+\left(\frac{x-\mu}{\sigma}\right)^{2}\right]^{-1}$$

那么我们称X服从**柯西分布**（**Cauchy distribution**），记为$X\sim\mathcal{C}\left(\mu,\sigma\right)$.

- 期望：不存在
- 方差：不存在
- 矩母函数：不存在

由于柯西分布的一阶矩不存在，因而通常作为不可积的分布的例子. 此外，可以得到，两个独立的标准正态分布X,Y，其比例$U=\frac{Y}{X}$刚好服从柯西分布. 此外，$\mathcal{C}\left(0,1\right)\sim\mathrm{t}\left(1\right)$.

随机变量的变换

- 对于一个可测函数$g(\cdot)\,\mathbb{R}\to\mathbb{R}$，$Y=g(X)$也是一个随机变量. 如果我们已知$X$的分布，如何获得新的随机变量$Y$的分布呢?
- 首先仿照随机变量的定义，我们定义函数$g(\cdot)$对于一个集合的逆为:

$$g^{-1}(A)=\{x\in\mathbb{R}:g(x)\in A\}$$

因而对于单点集

$$g^{-1}(\{y\})=\{x\in\mathbb{R}:g(x)=y\}$$

离散随机变量的变换

对于离散型的随机变量X，$g(X)$也是离散型的，因而随机变量Y的p.m.f.为:

$$\begin{aligned} f_Y(y) &= \mathrm{P}(Y=y) \\ &= \sum_{x\in g^{-1}(\{y\})} \mathrm{P}(X=x) \\ &= \sum_{x\in g^{-1}(\{y\})} f_X(x) \end{aligned}$$

由此可以确定随机变量$Y=g(X)$的概率质量函数.

离散随机变量的变换

抛骰子

若假设X为一次抛骰子的随机试验的结果，其概率质量函数为：

X	1	2	3	4	5	6
$\mathrm{P}(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

如果我们希望得到$Y=g(X)=(X-3.5)^2$的概率质量函数，可以通过以上步骤，比如：

$$g^{-1}(\{6.25\})=\left\{x\in\mathbb{R}:(x-3.5)^2=6.25\right\}=\{1,6\}$$

因而$\mathrm{P}(Y=6.25)=\mathrm{P}(X\in\{1,6\})=\frac{1}{3}$. 以此类推，我们得到$Y$的概率质量函数

Y	6.25	2.25	0.25
$\mathrm{P}(X=x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

连续随机变量的变换

对于一个一般的随机变量X，随机变量$Y = g(X)$的累积分布函数可以使用如下定义计算：

$$\begin{aligned} F_Y(y) &= \mathrm{P}(Y \leq y) \\ &= \mathrm{P}(g(X) \leq y) \\ &= \mathrm{P}(\{x : g(x) \leq y\}) \\ &= \int_{\{x:g(x)\leq y\}} \mathrm{d}F_X \end{aligned}$$

特别的，如果$g(\cdot)$严格单调递增，则：$F_Y(y) = F_X\left(g^{-1}(y)\right)$；
如果$g(\cdot)$严格单调递减，则：$F_Y(y) = 1 - F_X\left(g^{-1}(y)\right)$

连续随机变量的变换

均匀分布的变换

如果$U \sim \mathcal{U}(-1,1)$，欲求$Y = U^2$的密度函数，可以首先求其累计分布函数，对于$y \geq 0$，有：

$$\begin{aligned} F_Y(y) &= \mathrm{P}(Y \leq y) = \mathrm{P}\left(U^2 \leq y\right) \\ &= \mathrm{P}\left(\{-\sqrt{y} \leq U \leq \sqrt{y}\}\right) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} \mathrm{d}u = \sqrt{y} \end{aligned}$$

因而其密度函数为：

$$f_Y(y) = \mathbb{1}\{0 < y < 1\} \frac{1}{2\sqrt{y}}$$

分布族

- 给定任意的一个$(\Omega, \mathscr{F})$，在这个空间里面我们可以定义各种不同的概率函数，统计学需要解决的问题就是使用样本数据去推断总体的概率函数$\mathscr{P}$.
- 然而一般情况下，概率函数$\mathscr{P}$的可能性太多，因而我们经常将概率函数限制在一个可能集合内，为此我们可以定义参数族的概念.

参数族

如果对于每一个已知的$\theta \in \Theta, \Theta \subset \mathbb{R}^d$，$\mathrm{P}_\theta$为在$(\Omega, \mathcal{F})$上的一个已知的概率函数，那么$\{\mathrm{P}_\theta, \theta \in \Theta\}$即被称为**参数族**（parametric family），其中Θ为**参数空间**（parametric space），正整数d为参数空间的维数.

位置尺度族

对于一个随机变量X，我们令$Y=\sigma X+\mu, \sigma>0$，那么其分布函数：

$$\begin{aligned} F_Y(y) &= \mathrm{P}(Y \leq y) = \mathrm{P}(\sigma X+\mu \leq y) \\ &= \mathrm{P}\left(X \leq \frac{y-\mu}{\sigma}\right) \\ &= F_X\left(\frac{y-\mu}{\sigma}\right) \end{aligned}$$

所以其密度函数满足

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\right)$$

同时，其期望和方差满足

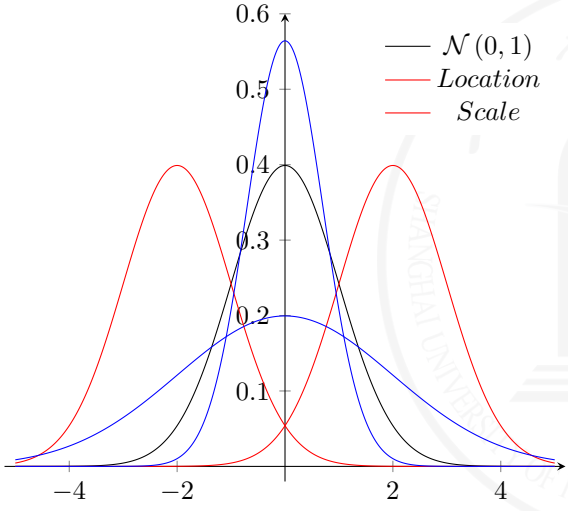
$$\begin{aligned} \mathbb{E}(Y) &= \sigma \mathbb{E}(X) + \mu \\ \mathbb{V}(Y) &= \sigma^2 \mathbb{V}(X) \end{aligned}$$

位置尺度族

位置尺度族

- 令$f(x)$为任意的密度函数，那么形如$\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$的密度函数就形成了以$\theta=[\mu,\sigma]'$为参数的参数族，我们称之为**位置尺度族**（location-scale family）.
- 位置尺度族即对于任意的“标准分布”做位移和数乘所得到的分布组成的分布族.
- 其中参数μ一般称为**位置参数**（location parameter），而σ一般称为**尺度参数**（scale parameter）.

位置尺度族



位置尺度族

正态分布

标准正态分布的密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

对于任意的$\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$，其位置尺度族的密度函数可以写为:

$$f(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma}}$$

即正态分布的密度函数. 因而正态分布族$\{\mathrm{P}_{\mu,\sigma}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$ 为一个位置尺度族.

标准分布

- 上面介绍的包括指数分布、Logistic分布等都是由“标准分布”生成的位置尺度族.
- 注意虽然多数情况下“标准分布”的期望选取为0、方差选取为1（如上例中的标准正态分布），但是也有例外
  - 比如“标准的”Logistic分布（$X \sim \mathcal{LG}(0,1)$）的期望为0，但是方差并不为1.

标准化

- 如果X为“标准分布”，对于$Y=\sigma X+\mu$，由于

$$\mathrm{P}(Y\leq y)=\mathrm{P}(\sigma X+\mu\leq y)=\mathrm{P}\left(X\leq\frac{y-\mu}{\sigma}\right)$$

  因而计算位置尺度组中任意分布的概率都可以转化为“标准分布”的概率
- 如此在计算时（统计表、计算机函数等）只需要将任意的分布变换为“标准分布”：

$$X=\frac{Y-\mu}{\sigma}$$

  即可，这一过程被称为**标准化（standardize）**.
- 多数情况下，可以选取$\mu=\mathbb{E}(Y),\sigma=\sqrt{\mathbb{V}(Y)}$，如此$\mathbb{E}(X)=0,\mathbb{V}(X)=1$，即将$Y$标准化为期望为0、方差为1的分布.

单参数指数分布族

指数分布族

对于一个参数族$\{\mathrm{P}_\theta, \theta \in \Theta\}$，如果其概率密度（质量）函数可以写成如下形式:

$$f(x|\theta) = h(x) \cdot \exp\{\eta(\theta) \cdot T(x) - B(\theta)\}$$

那么我们称$\{\mathrm{P}_\theta, \theta \in \Theta\}$为**单参数指数分布族**（one-parameter exponential family）.

单参数指数分布族

二项分布

令N为已知的正整数，如果随机变量$X \sim \mathcal{B}i(N,p)$，参数$\theta = p \in \Theta = (0,1)$，其质量函数:

$$\begin{aligned}
\mathrm{P}(x|p) &= \binom{N}{x} p^x (1-p)^{N-x} \\
&= \binom{N}{x} \exp\{x \ln(p) + (N-x)\ln(1-p)\} \\
&= \binom{N}{x} \exp\{x[\ln(p) - \ln(1-p)] + N\ln(1-p)\} \\
&= \binom{N}{x} \exp\left\{x\left[\ln\left(\frac{p}{1-p}\right)\right] + N\ln(1-p)\right\}
\end{aligned}$$

故令$h(x) = \binom{N}{x}$，$\eta(\theta) = \ln\left(\frac{p}{1-p}\right)$，$T(x) = x$，$B(\theta) = -N\ln(1-p)$，所以二项分布属于指数分布族.

单参数指数分布族

泊松分布

泊松分布参数为$\theta = \lambda \in \Theta = (0, \infty)$，其质量函数为：

$$\mathrm{P}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} = \frac{1}{x!} \exp\{x \ln(\lambda) - \lambda\}$$

故令$h(x) = \frac{1}{x!}$，$\eta(\theta) = \ln(\lambda)$，$T(x) = x$，$B(\theta) = \lambda$，所以泊松分布属于指数分布族.

单参数指数分布族

指数分布

指数分布的密度函数为:

$$
\begin{aligned}
f\left(x|\beta\right) &= \mathbb{1}\left\{x>0\right\}\frac{1}{\beta}e^{-\frac{x}{\beta}} \\
&= \mathbb{1}\left\{x>0\right\}\exp\left\{-x\cdot\frac{1}{\beta}-\ln\beta\right\}
\end{aligned}
$$

因而令$h\left(x\right)=\mathbb{1}\left\{x>0\right\}$，$\eta\left(\theta\right)=-\frac{1}{\beta}$，$T\left(x\right)=x$，$B\left(\theta\right)=\ln\beta$，故指数分布属于指数分布族.

单参数指数分布族

非指数分布族

若某一密度函数为:

$$f(x|\beta) = \frac{1}{\beta}\exp\left\{1 - \frac{x}{\beta}\right\}, x > \beta > 0$$

可知$f(x|\beta)$为密度函数. 然而:

$$\begin{aligned} f(x|\beta) &= \mathbb{1}\{x > \beta\}\frac{1}{\beta}\exp\left\{1 - \frac{x}{\beta}\right\} \\ &= \mathbb{1}\{x > \beta\} \cdot \exp\left\{-\frac{x}{\beta} - \ln\beta + 1\right\} \end{aligned}$$

由于$\mathbb{1}\{x > \beta\}$不仅仅依赖于x，而且依赖于β，因而这一分布不属于指数分布族.

指数分布族的规范形式

指数分布的规范形式

为了方便起见，我们会把密度函数重新参数化，即对于指数分布族

$$f(x|\theta) = h(x) \cdot \exp\{\eta(\theta) \cdot T(x) - B(\theta)\}$$

我们令$\lambda = \eta(\theta)$，那么指数分布族可以写为：

$$f(x|\lambda) = h(x) \cdot \exp\{\lambda \cdot T(x) - C(\lambda)\}$$

并将这种形式成为**规范形式（canonical form）**.

指数分布族的矩

如果随机变量X的概率密度（质量）函数为规范形式，那么

$$\mathbb{E}[T(X)] = C'(\lambda)$$
$$\mathbb{V}[T(X)] = C''(\lambda)$$

指数分布族矩的证明

对于规范形式$f(x|\lambda)=h(x)\cdot\exp\{\lambda\cdot T(x)-C(\lambda)\}$，由于其为密度函数，因而：

$$1=\int_{\mathbb{R}}f(x|\lambda)\,\mathrm{d}x=\int_{\mathbb{R}}h(x)\cdot\exp\{\lambda\cdot T(x)-C(\lambda)\}\,\mathrm{d}x$$

从而：

$$e^{C(\lambda)}=\int_{\mathbb{R}}h(x)\cdot\exp\{\lambda\cdot T(x)\}\,\mathrm{d}x$$

两边对λ求导可得：

$$\begin{aligned}e^{C(\lambda)}C'(\lambda)&=\frac{\partial}{\partial\lambda}\int_{\mathbb{R}}h(x)\cdot\exp\{\lambda\cdot T(x)\}\,\mathrm{d}x\\&=\int_{\mathbb{R}}h(x)\cdot\frac{\partial}{\partial\lambda}\exp\{\lambda\cdot T(x)\}\,\mathrm{d}x\\&=\int_{\mathbb{R}}h(x)\cdot T(x)\exp\{\lambda\cdot T(x)\}\,\mathrm{d}x\end{aligned}$$

整理可得：

$$C'(\lambda)=\int_{\mathbb{R}}T(x)\cdot h(x)\cdot\exp\{\lambda\cdot T(x)-C(\lambda)\}\,\mathrm{d}x=\mathbb{E}[T(X)]$$

指数分布族的规范形式

二项分布的规范形式

对于二项分布：

$$\mathrm{P}(x|p) = \binom{N}{x} \exp\left\{x\left[\ln\left(\frac{p}{1-p}\right)\right] + N\ln(1-p)\right\}$$

如果令$\lambda = \ln\left(\frac{p}{1-p}\right)$，那么

$$\mathrm{P}(x|\lambda) = \binom{N}{x} \exp\left\{\lambda \cdot x - N\ln\left(1+e^{\lambda}\right)\right\}$$

即为二项分布的质量函数的规范形式. 其中，$T(x) = x$，因而

$$\mathbb{E}[T(X)] = \mathbb{E}(X) = C'(\lambda) = \frac{Ne^{\lambda}}{1+e^{\lambda}} = Np$$

$$\mathbb{V}[T(X)] = \mathbb{V}(X) = C''(\lambda) = \frac{Ne^{\lambda}}{(1+e^{\lambda})^2} = Np(1-p)$$

作业

- 2.2, 2.4, 2.6, 2.9, 2.10