第三节 · 常用分布

司继春

上海对外经贸大学统计与信息学院

1 离散分布

1.1 离散均匀分布

随机变量 X 服从离散均匀分布如果 $P(X=x)=\frac{1}{N}, x\in(a_1,a_2,...,a_N)$。

- 期望：$\mathbb{E}(X)=\frac{1}{N}\sum_{i=1}^{N}a_i$
- 方差：$\mathrm{Var}(X)=\frac{1}{N}\sum_{i=1}^{N}(a_i-\mathbb{E}X)^2$

1.2 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），独立重复 N 次的试验结果。一个随机变量 X 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**）$X\sim Ber(p)$ 即：

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

- 期望：$\mathbb{E}(X)=p$
- 方差：$\mathrm{Var}(X)=p(1-p)$

1.3 二项分布

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

$$P(Y=y|N,p)=\begin{pmatrix}N\\ y\end{pmatrix}p^y(1-p)^{N-y}, y=1,2,...,N$$

我们成 Y 服从**二项分布**（**Binomial Distribution**），记为 $Y\sim Bi(N,p)$。由于对于任意的实数 a,b 以及任意整数 $N\ge 0$，有：$(a+b)^N=\sum_{i=0}^{N}\begin{pmatrix}N\\ i\end{pmatrix}a^ib^{n-i}$，

因而:

$$\sum_{y=1}^{N} P\left(Y=y|N,p\right) = \sum_{y=0}^{N} \left(\begin{array}{c} N \\ y \end{array}\right) p^{y}\left(1-p\right)^{N-y} = \left(p+1-p\right)^{N} = 1$$

- 期望: $\mathbb{E}\left(Y\right) = Np$
- 方差: $\mathrm{Var}\left(Y\right) = Np\left(1-p\right)$

1.4 负二项分布

二项分布关心在 N 次伯努利实验中成功的次数，而有时我们会关心为了达到 r 次成功所必须要做的试验的次数。记 Z 为为了达到 r 次成功所需的试验次数，那么事件 $Z=z$ 即 z 次试验达到了 r 次成功，这一事件可以进一步分解为在前 $z-1$ 次试验中得到了 $r-1$ 次成功，而第 r 次也成功了。所以其概率质量函数为:

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

我们称随机变量 Z 服从**负二项分布**（**Negative Binomial Distribution**），记为 $Z \sim NB\left(r,p\right)$。

下面计算 Z 的期望:

$$\begin{aligned} E\left(Z\right) &= \sum_{z=r}^{\infty} z \cdot \left(\begin{array}{c} z-1 \\ r-1 \end{array}\right) p^{r}\left(1-p\right)^{z-r} \\ &= \sum_{z=r}^{\infty} z \cdot \frac{\left(z-1\right)!}{\left(r-1\right)!\left(z-r\right)!} p^{r}\left(1-p\right)^{z-r} \\ &= \frac{r}{p} \sum_{z=r}^{\infty} \frac{z!}{r!\left(z-r\right)!} p^{r+1}\left(1-p\right)^{z-r} \\ &\overset{z'=z+1,r'=r+1}{=} \frac{r}{p} \sum_{z'=r+1}^{\infty} \frac{\left(z'-1\right)!}{\left(r'-1\right)!\left(z'-r'\right)!} p^{r'}\left(1-p\right)^{z'-r'} \\ &\overset{NB\left(r+1,p\right)\text{的质量函数和为 } 1}{=} \frac{r}{p} \end{aligned}$$

同理可计算其方差。

- 期望: $\mathbb{E}\left(Z\right) = \frac{r}{p}$
- 方差: $\mathrm{Var}\left(Z\right) = \frac{r(1-p)}{p^{2}}$

练习 1. 求负二项分布的方差。

1.5 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 $V \sim NB(1,p)$，则随机变量 V 服从几何分布：

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

我们记 $V \sim G(p)$。几何分布具有无记忆性，即：

$$P(V>s|V>t)=P(V>s-t), s>t$$

即在给定我们为了等待成功已经等了 t 次的情况下，还需要等待的次数跟已经等待的次数是没有关系的。比如北京的车牌摇号，10 次没有中签的人跟 100 次没有中签的人，其需要继续等待时间的分布是一样的。因而此分布不适于建模带有生命长度的问题。

- 期望：$\mathbb{E}(V)=\frac{1}{p}$
- 方差：$\mathrm{Var}(V)=\frac{1-p}{p^2}$

练习 2. 证明几何分布的无记忆性。

1.6 超几何分布

即在一个 N 个试验组成的总体中，有已知 K 次成功，从这 N 个总体中抽取出 n 个样本，抽到的成功次数 $M=k$ 的概率。我们称随机变量 M 服从**超几何分布**（**Hypergeometric Distribution**），其概率质量函数为：

$$P(M=k|N,K,n)=\frac{\begin{pmatrix}K\\k\end{pmatrix}\begin{pmatrix}N-K\\n-k\end{pmatrix}}{\begin{pmatrix}N\\n\end{pmatrix}}$$

- 期望：$\mathbb{E}(M)=\frac{n}{N}\cdot K$
- 方差：$\mathrm{Var}(M)=n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$

1.7 泊松分布

在第一节中我们研究了到达次数问题，相应的，如果随机变量 X 的概率质量函数为：

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

那么我们称随机变量 X 服从**泊松分布**（**Poisson Distribution**），记为 $X \sim P(\lambda)$。泊松分布经常被用来建模次数问题。

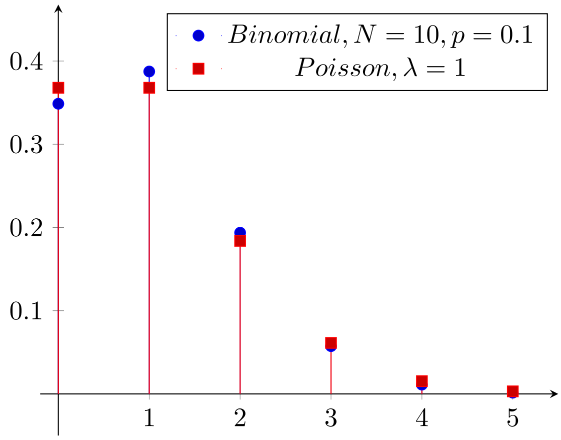

图 1: 泊松分布与二项分布

- 期望：$\mathbb{E}(X) = \lambda$
- 方差：$\mathrm{Var}(X) = \lambda$

练习 3. 计算泊松分布的方差。

根据第一节的例题中，我们已经知道如果一个随机变量 $Y \sim Bi(N, p)$，令 $\lambda = Np$，并令 $N \to \infty$，那么随机变量 Y 近似服从泊松分布，因而对于相对较大的 N，可以使用泊松分布去近似二项分布。

例 1. 如果一个人打字出错的概率为 $p = 0.001$，那么他写 1000 字出错的概率应该服从二项分布。记出错的字数为 Y，则 $Y \sim Bi(1000, 0.001)$。比如写 1000 字出错两个以内的概率为: :

$$P(Y \leq 2) = \sum_{y=0}^{2} \begin{pmatrix} 1000 \\ y \end{pmatrix} 0.001^y 0.999^{1000-y} \approx 0.91979$$

以上计算比较繁琐，如果使用泊松逼近，令 $X \sim P(1)$：

$$P(Y \leq 2) \approx P(X \leq 2) = \sum_{y=0}^{2} \frac{1^y}{y!} e^{-1} = 0.91970$$

两者非常接近。

2 连续分布

2.1 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：

$$f(x|a,b)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & otherwise\end{cases}$$

则称随机变量 X 服从区间 $[a,b]$ 上的**均匀分布**（**Uniform Distribution**），记为 $X\sim U(a,b)$。

- 分布函数：
$$F(x|a,b)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & x\in[a,b]\\ 1 & x>b\end{cases}$$
- 期望：$\mathbb{E}(X)=\frac{a+b}{2}$
- 方差：$\mathrm{Var}(X)=\frac{(b-a)^2}{12}$

2.2 指数分布

若随机变量 X 的概率密度函数为：

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

那么我们称 X 服从**指数分布**（**Exponential Distribution**），其密度函数如图(2) 所示，记为 $X\sim E(a,\beta)$。

- 分布函数：$F(x)=1-e^{-\frac{x-a}{\beta}},x>a$
- 期望：$\mathbb{E}(X)=\beta+a$
- 方差：$\mathrm{Var}(X)=\beta^2$

例 2.（指数分布）如果一个随机变量 $Y\sim P(\lambda)$，其中参数 λ 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 $P(X>t)=e^{-\frac{\lambda t}{T}}$，所以 $P(X\le t)=1-e^{-\frac{\lambda t}{T}}$，因而等待时间 X 服从指数分布。

同样，指数分布也具有无记忆性，即如果 $X\sim E(0,\beta)$，那么

$$P(X>s|X>t)=P(X>s-t),s>t$$

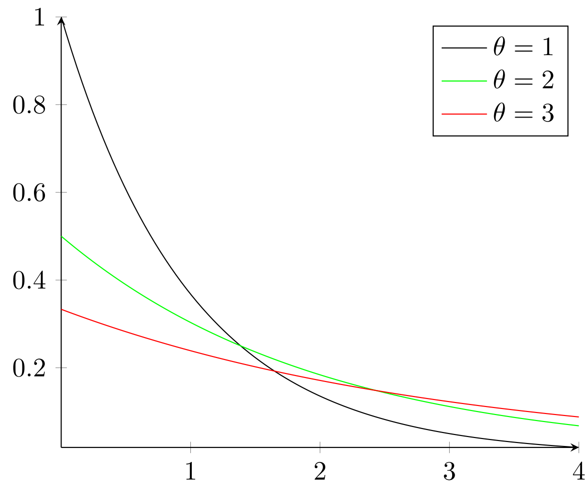

图 2: 指数分布密度函数

以为着为了等待第一个人到达，如果已经等了 t 时间，那么继续等待的时间仍然服从指数分布：给定已经等了五分钟的情况下，继续等待五分钟有到达的概率与第一个五分钟有到达的概率是一样的。

现在考虑，如果一件产品的使用寿命为 T，其分布函数为 $F(t)$，那么寿命大于 t 的概率为 $S(t) = 1 - F(t)$，我们称之为**生存函数**（**Survival function**）。很多时候，我们关心一件产品已经使用了时间 t，在 $(t, t + dt)$ 一小段时间内死亡的概率，如果令 $dt \to 0$，即瞬时的死亡风险。可以计算：

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt)}{dt \cdot S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t)$$

我们称 $\lambda(t)$ 为**风险函数**（**hazard function**）。指数分布的风险函数为：

$$\lambda(t) = -\frac{d}{dt} \ln S(t) = -\frac{d}{dt} \ln \left(e^{-\frac{x}{\beta}}\right) = \frac{1}{\beta}$$

因而指数函数的风险函数为常数，即其死亡（到达）的可能性不随着时间的增加而增加（或者减少）。

2.3 伽马分布

在介绍伽马分布之前，我们需要先介绍伽马函数（Γ Function）。我们定义 Γ 函数为：

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha - 1} e^{-t} dt$$

易知当 $\alpha > 0$ 时，$\Gamma(\alpha) < \infty$，而当 $\alpha \leq 0$ 时，该积分不一定有限。图 (3) 展示了 Γ 函数的逆函数。

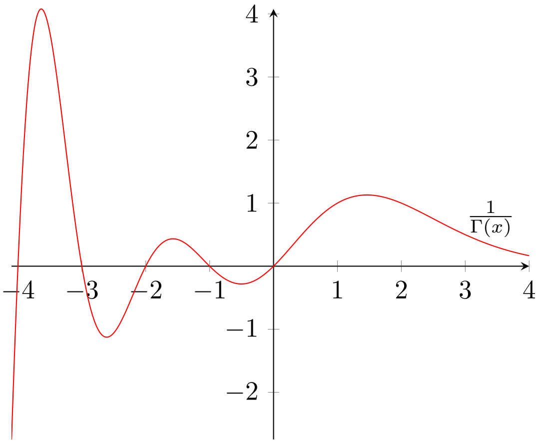

图 3: $\Gamma^{-1}(\alpha)$

根据 Γ 函数的定义，可知 $\Gamma(1)=1$，而当 $\alpha>0$ 时，有：

$$\begin{aligned}
\Gamma(\alpha+1) &= \int_0^\infty t^\alpha e^{-t} dt \\
&= -\int_0^\infty t^\alpha de^{-t} \\
&= -t^\alpha e^{-t}\Big|_0^\infty + \int_0^\infty e^{-t} dt^\alpha \\
&= \alpha \int_0^\infty t^{\alpha-1} e^{-t} dt \\
&= \alpha\Gamma(\alpha)
\end{aligned}$$

结合 $\Gamma(1)=1$，可以得到对于任意的正整数 n，$\Gamma(n)=(n-1)!$。此外，$\Gamma\left(\frac{1}{2}\right)=\int_0^\infty \frac{1}{\sqrt{t}} e^{-t} dt=\sqrt{\pi}$，因而：

$$\begin{aligned}
\int_0^\infty \frac{1}{\sqrt{t}} e^{-t} dt &\overset{t=\frac{x^2}{2}}{=} \int_0^\infty \frac{\sqrt{2}}{x} e^{-\frac{x^2}{2}} d\frac{x^2}{2} \\
&= \sqrt{2}\int_0^\infty e^{-\frac{x^2}{2}} dx \\
&= \sqrt{\pi}
\end{aligned}$$

因而

$$\int_0^\infty e^{-\frac{x^2}{2}} dx = \sqrt{\frac{\pi}{2}} \tag{1}$$

根据 Γ 函数的定义，函数：

$$f(t|\alpha)=\frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}, t>0$$

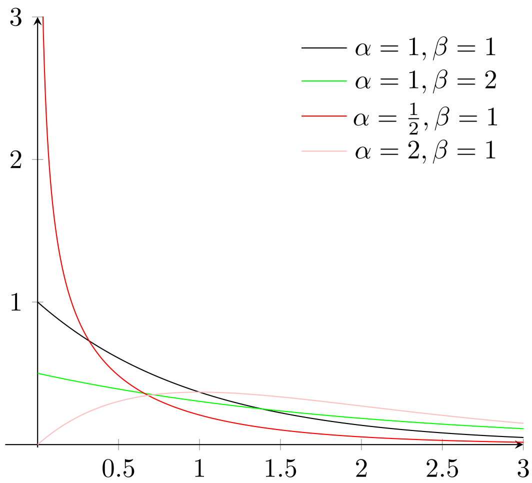

图 4: $\Gamma(\alpha, \beta)$ 分布的密度函数

为一个密度函数。如果令 $T \sim f(t|\alpha)$，那么对于任意的 $\beta > 0$，我们称 $X = \beta T$ 服从**伽马分布**（**Gamma Distribution**），记为 $X \sim \Gamma(\alpha, \beta)$，其密度函数为：

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, x > 0, \alpha > 0, \beta > 0$$

图 (4) 展示了不同参数下 $\Gamma(\alpha, \beta)$ 分布的密度函数。可知，指数分布为 $\alpha = 1$ 时 Γ 分布的特例。其期望：

$$\begin{aligned}
E(X) &= \int_0^{\infty} x f(x|\alpha, \beta)\, dx \\
&= \int_0^{\infty} x \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha} e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Gamma(\alpha+1)\beta^{\alpha+1} \\
&= \alpha\beta
\end{aligned}$$

类似地，可以计算其方差：

- 期望：$\mathbb{E}(X) = \alpha\beta$
- 方差：$\mathrm{Var}(X) = \alpha\beta^2$

练习 4. 计算 $\Gamma(\alpha, \beta)$ 分布的方差。

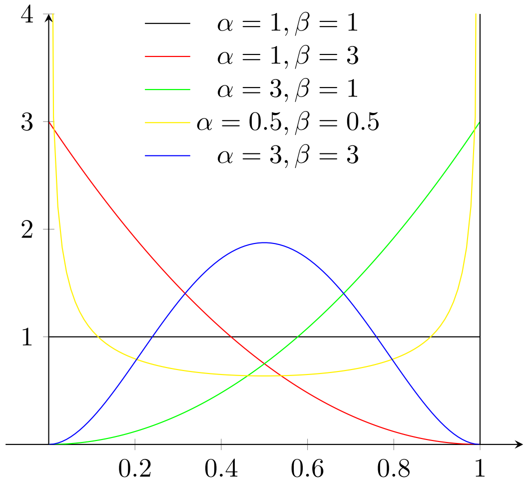

图 5: 指数分布密度函数

2.4 Beta 分布

如果一个随机变量 $X \in (0,1)$，其分布函数为：

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

那么我们称 X 服从 **Beta 分布**（**Beta Distribution**）。其中

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

图 (5) 给出了 Beta 分布的密度函数。当 $\alpha > 1, \beta = 1$ 时，Beta 分布的密度函数为单调递增的；当 $\alpha = 1, \beta > 1$ 时，Beta 分布的密度函数为单调递减的；当 $\alpha < 1, \beta < 1$ 时，Beta 分布的密度函数为 U 型的；当 $\alpha > 1, \beta > 1$ 时，Beta 分布的密度函数为钟型的；当 $\alpha = 1, \beta = 1$ 时，Beta 分布的密度函数退化为均匀分布。由于 Beta 分布的取值范围在 $(0,1)$ 范围内，因而 Beta 分布经常被用于研究比率等问题，或者在贝叶斯分析中作为概率的先验。

Beta 分布的 r 阶矩可以计算为：

$$\begin{aligned} E(X^r) &= \frac{1}{B(\alpha,\beta)}\int_0^1 x^r x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{1}{B(\alpha,\beta)}\int_0^1 x^{r+\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{B(r+\alpha,\beta)}{B(\alpha,\beta)}\frac{1}{B(r+\alpha,\beta)}\int_0^1 x^{r+\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{B(r+\alpha,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+r)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)} \end{aligned}$$

从而：

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

2.5 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 $X\sim N\left(\mu,\sigma^2\right)$。

- 期望：$\mathbb{E}(X)=\mu$
- 方差：$\mathrm{Var}(X)=\sigma^2$

如果 $X\sim N\left(\mu,\sigma^2\right)$，如果令 $Z=\frac{X-\mu}{\sigma}$，则 $E(Z)=\frac{E(X)-\mu}{\sigma}=0$，$\mathrm{Var}(Z)=\frac{1}{\sigma^2}\cdot\mathrm{Var}(X)=1$，随机变量 Z 服从 $\mu=0,\sigma=1$ 的正态分布，即 $Z\sim N(0,1)$，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：

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

根据公式 (1) 可知，其密度函数在 $\mathbb{R}$ 上的积分为 1。而其分布函数为：

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

由于 $\Phi(x)$ 没有显示解，因而一般我们用 $\Phi(z)$ 来代表标准正态的分布函数。

图 (6) 列出了标准正态分布的密度函数以及分布函数。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 $X\sim N\left(\mu,\sigma^2\right)$，那么 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$，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：

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

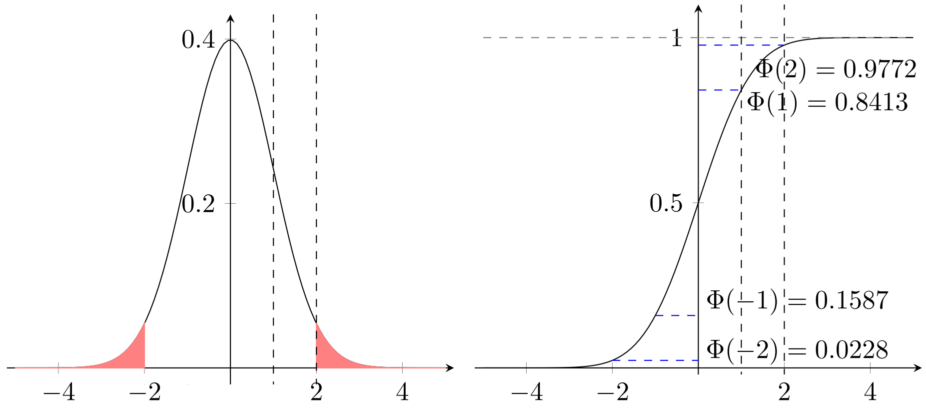

图 6: 标准正态分布密度函数与分布函数

同理，有:

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

正态分布有很多优良的性质，因而在建模中是最经常被使用的一种分布。比如，如果两个随机变量 X, Y 独立，且 $X \sim N\left(\mu_1, \sigma_1^2\right), Y \sim N\left(\mu_2, \sigma_2^2\right)$，那么 $V = X + Y \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$，即两个独立的正态分布的和仍然是正态分布。在下一节多元随机变量中，我们还将介绍更多的正态分布独有的性质。

练习 5. 求标准正态分布的偏度、峰度。

2.6 对数正态分布

若随机变量 $Y = e^X, X \sim N\left(\mu, \sigma^2\right)$，则我们称随机变量 Y 服从**对数正态分布**（**Lognormal Distribution**），记为 $Y \sim LN\left(\mu, \sigma^2\right)$。

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

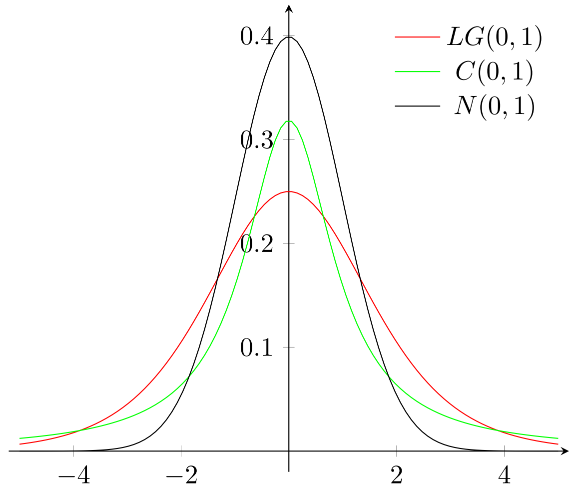

图 7: Logistic 分布、Cauchy 分布与正态分布密度函数

2.7 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: $X\sim LG\left(\mu,\sigma\right)$。Logistic 分布被广泛运用在**离散选择**（**Discrete Choice**）模型或者**分类器**（**Classifier**）中。

- 期望: $\mathbb{E}\left(X\right)=\mu$
- 方差: $\mathrm{Var}\left(X\right)=\sigma^2\frac{\pi^2}{3}$

2.8 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 $X\sim C\left(\mu,\sigma\right)$。

- 期望: 不存在

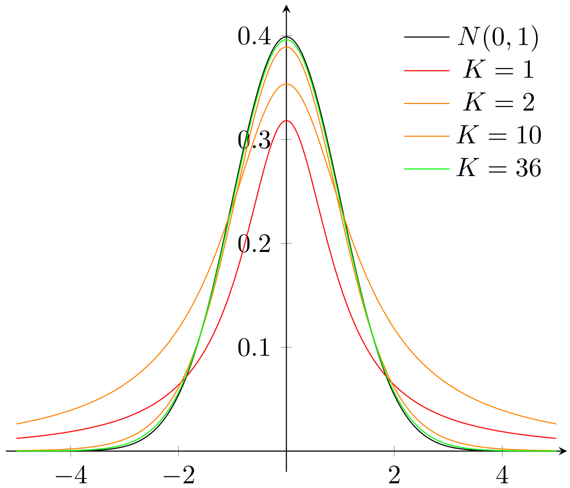

图 8: t 分布与正态分布密度函数

- 方差：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。因而在实践中计算两个数的比例时需要特别小心。

2.9 χ^2 分布

K 个独立的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 $X_1,X_2,...,X_K$ 为 K 个**独立**的**标准**正态分布，那么

$$X=\sum_{i=1}^{K}X_i^2\sim\chi_K^2$$

其中参数 K 为卡方分布的**自由度**（**degrees of freedom**）。χ^2 分布的密度函数为：

$$f\left(x|K\right)=\frac{1}{\Gamma\left(K/2\right)2^{K/2}}x^{K/2-1}e^{-x/2},x>0$$

因而，χ^2 分布实际是 Γ 分布的特殊形式（$\alpha=K/2,\beta=2$）。

- 期望：$\mathbb{E}\left(X\right)=K$
- 方差：$\text{Var}\left(X\right)=2K$

χ^2 分布在假设检验中非常常用。

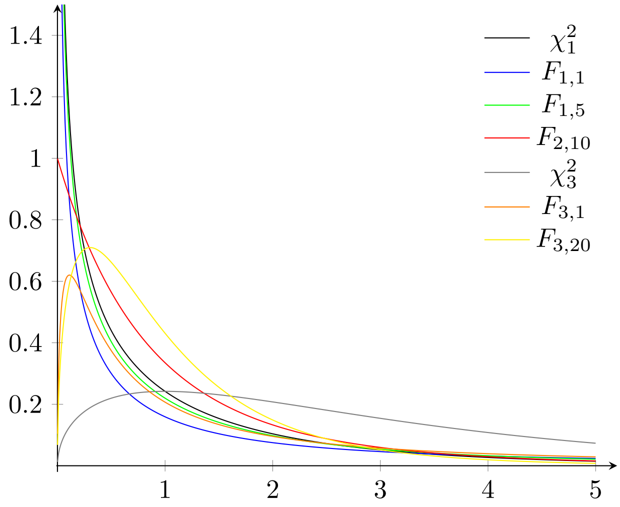

图 9: F 分布与 χ^2 分布密度函数

2.10 学生氏 t 分布

如果存在一个标准正态分布 $Z \sim N(0,1)$，以及一个 χ^2 分布 $X \sim \chi_K^2$，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 $T \sim t_K$，参数 K 为**自由度**。t 分布的密度函数为：

$$f(x|K) = \frac{\Gamma[(K+1)/2]}{\sqrt{K\pi}\Gamma(K/2)}\left(1+\frac{x^2}{K}\right)^{-\frac{K+1}{2}} = \frac{1}{\sqrt{K}B(1/2,K/2)}\left(1+\frac{x^2}{K}\right)^{-\frac{K+1}{2}}$$

当 $K=1$ 时，t 分布即退化为柯西分布，而当 $K \to \infty$ 时，t 分布趋向于正态分布。图 (8) 显示了不同自由度下的 t 分布与正态分布的比较。

- 期望：$\mathbb{E}(X) = 0, K > 1$
- 方差：$\mathrm{Var}(X) = \frac{K}{K-2}, K > 2$

t 分布在假设检验中也扮演着至关重要的地位。

2.11 F 分布

如果存在两个**独立**的 χ^2 分布 $X_1 \sim \chi_n^2, X_2 \sim \chi_m^2$，那么随机变量 $F = \frac{X_1/n}{X_2/m}$ 即服从 F **分布**（**F-distribution**），记为 $T \sim F_{n,m}$，参数 n,m 为**自由度**。F 分布的密度函数为：

$$f(x|K) = \frac{n^{n/2}m^{m/2}\Gamma[(n+m)/2]}{\Gamma(n/2)\Gamma(m/2)(m+nx)^{(n+m)/2}}x^{\frac{n}{2}-1}, x > 0$$

图 (9) 显示了不同自由度下的 χ^2 分布与 F 分布的密度函数。

- 期望：$\mathbb{E}(X)=\frac{m}{m-2},m>2$
- 方差：$\mathrm{Var}(X)=\frac{2m^2(m+n-2)}{n(m-2)^2(m-4)},m>4$

F 分布与其他多种分布有关系：

- 如果 $X_k\sim\Gamma(\alpha_k,\beta_k)$ 且独立，那么 $\frac{\alpha_2\beta_1X_1}{\alpha_1\beta_2X_2}\sim F_{2\alpha_1,2\alpha_2}$
- 如果 $X\sim B(n/2,m/2)$，那么 $\frac{mX}{n(1-X)}\sim F_{n,m}$，反之亦成立
- 如果 $X\sim F_{n,m}$，那么 $\lim_{m\to\infty}nX\sim\chi_n^2$
- 如果 $T\sim t_K$，那么 $T^2\sim F_{1,K}$

3 分布族

给定任意的一个 $(\Omega,\mathcal{F})$，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，因而我们经常把研究的重点放在一个概率函数的可能集合内，并在此集合内部对总体的概率函数做出推断。为此我们可以定义参数族的概念。

定义 1. 如果对于每一个已知的 $\theta\in\Theta,\Theta\subset\mathbb{R}^d$，$P_\theta$ 为在 $(\Omega,\mathcal{F})$ 上的一个已知的概率函数，那么 $\{P_\theta,\theta\in\Theta\}$ 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。

其中**位置尺度族**（**Location-scale families**）和**指数分布族**（**Exponent families**）是两类最为特殊且重要的参数族，我们这节将分别介绍这两类分布族。

3.1 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：

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

所以其密度函数满足：

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

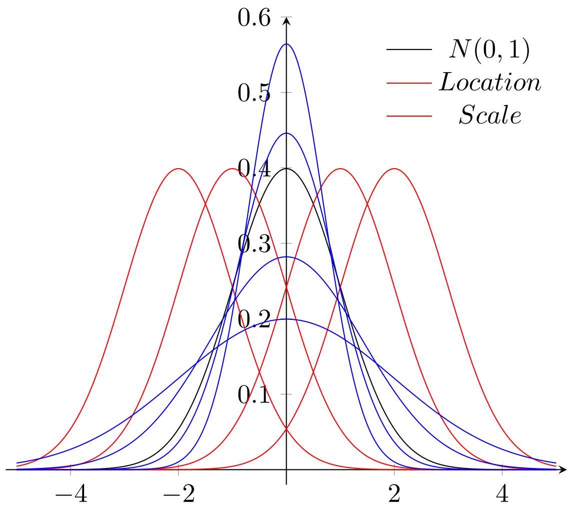

图 10: 正态分布的位置尺度族的密度函数

同时，其期望和方差满足：$\mathbb{E}(Y)=\sigma\mathbb{E}(X)+\mu$，$\mathrm{Var}(Y)=\sigma^2\mathrm{Var}(X)$。

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。位置尺度族即对于任意的随机变量做位移和数乘所得到的随机变量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**）。

例 3. 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 $\{P_{\mu,\sigma},\mu\in\mathbb{R},\sigma\in\mathbb{R}^+\}$ 为一个位置尺度族。图 (10) 展示了正态分布的位置族和尺度族。

上面介绍的包括指数分布、Γ 分布等都是各种类型的位置尺度族。正如正态分布一样，很多时候我们会将一个随机变量变换为期望为 0、方差为 1 的随机变量，即令随机变量 $Z=\frac{X-\mu}{\sigma}$，易得 $\mathbb{E}(Z)=0,\mathrm{Var}(Z)=1$，我们称之为**标准化**（**standardized**）。由于 $P(X\leq x)=P(Z\leq\sigma x+\mu)$，因而应用中对于位置尺度族，经常将先研究标准化之后的随机变量，进而推广到其位置尺度族。

3.2 单参数指数分布族

指数分布族是非常常用的分布族，其包含了我们上面介绍的多数分布。在此我们首先讨论单参数的指数分布族。其定义如下：

定义 2. （指数分布族）对于一个参数族 $\{P_\theta, \theta \in \Theta\}$，如果其概率密度（质量）函数可以写成如下形式：

$$f(x|\theta) = h(x) \cdot \exp\{\eta(\theta) \cdot T(x) - B(\theta)\} \tag{2}$$

那么我们称 $\{P_\theta, \theta \in \Theta\}$ 为**单参数指数分布族**（**One-parameter exponential family**）。

由于很多随机变量的取值范围不是 $\mathbb{R}$，然而随机变量的值域为 $\mathbb{R}$，因而我们一般使用指示函数（indicator function）来表示随机变量的取值范围，即定义

$$1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

例如 $1_{(-\infty,0]}(x)$ 即当 $x <= 0$ 时，$1_{(-\infty,0]}(x) = 1$；当 $x > 0$ 时，$1_{(-\infty,0]}(x) = 0$。

我们之前遇到的很多分布都属于指数分布族。例如：

例 4. 令 N 为已知的正整数，如果随机变量 $X \sim Bi(N,p)$，参数 $\theta = p \in \Theta = (0,1)$，其质量函数：

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

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

例 5. 泊松分布参数为 $\theta = \lambda \in \Theta = (0,\infty)$，其质量函数为：

$$\begin{aligned} P(x|\lambda) &= \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \frac{1}{x!} \exp\{x\ln(\lambda) - \lambda\} \end{aligned}$$

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。

例 6. 指数分布的密度函数为:

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

因而令 $h(x)=1_{(0,\infty)}(x)$，$\eta(\theta)=-\frac{1}{\beta}$，$T(x)=x$，$B(\theta)=\ln\beta$，故指数分布属于指数分布族。

然而并非所有带指数的密度函数都属于指数分布族，比如:

例 7. 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:

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

由于 $1_{(\beta,\infty)}(x)$ 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。

练习 6. Pareto 分布的密度函数为:

$$f(x|a,\beta)=\beta\cdot a^{\beta}x^{-(\beta+1)},x>a$$

那么 Pareto 分布是否属于指数分布族?

很多时候为了方便起见，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:

$$f(x|\theta)=h(x)\cdot\exp\{\lambda\cdot T(x)-C(\lambda)\} \tag{3}$$

我们将指数分布族重新参数化为式 (3) 的形式，并将这种形式成为**规范形式**（**Canonical form**）。

例 8. 在例 (4) 中，如果令 $\lambda = \ln\left(\frac{p}{1-p}\right)$，那么

$$P(x|\lambda) = \left(\begin{array}{c} N \\ x \end{array}\right) \exp\left\{\lambda x - N\ln\left(1+e^{\lambda}\right)\right\}$$

即为二项分布的质量函数的规范形式。

对于指数分布族的规范形式，我们有如下定理：

定理 1. 如果随机变量 X 的概率密度（质量）函数为式 *(3)* 的形式，那么 $\mathbb{E}[T(X)] = C'(\lambda)$，$Var[T(X)] = C''(\lambda)$。

例 9. 在例 (8) 中，$T(x) = x$，因而 $\mathbb{E}[T(X)] = \mathbb{E}(X) = C'(\lambda) = \frac{Ne^{\lambda}}{1+e^{\lambda}} = Np$，$\mathrm{Var}[T(X)] = \mathrm{Var}(X) = C''(\lambda) = \frac{Ne^{\lambda}}{(1+e^{\lambda})^2} = Np(1-p)$。

练习 7. 请用定理 (1) 计算指数分布、泊松分布的期望和方差。

参考文献

[1] Bickel, P.J., Doksum, K.A., 2001. Mathematical Statistics: Basic Ideas and Selected Topics. Prentice-Hall, Inc, New Jersey.

[2] Casella, G., Berger, R.L., 2002. Statistical inference. Duxbury Pacific Grove, CA.

[3] Shao, J., 2007. Mathematical Statistics, 2nd ed. Springer, New York.

[4] Schervish, M.J., 1995. Theory of Statistics. Springer-Verlag, New York.