离散型

二项分布（binomial distribution）

如果离散型随机变量X服从二项分布，一般记作$X \sim B(n,p)$。

$$B(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x}, x=0,1,\dots$$
$$\mathrm{E}[X]=np$$
$$\mathrm{Var}[X]=np(1-p)$$

二项分布可以帮助纠正一个生活中很常见的谬误，比如说身高高于两米的人占人类总体的1%，那么是否说明随机选取的100个人中一定至少有1个人高于两米呢？记X为100个人中身高高于两米的人数，显然$X \sim B(100, 0.01)$，经计算可得$P(X=0) \approx 0.366$。其实也就意味着，100个人中，能至少看到1个身高高于两米的人的概率其实大约是$1-0.366=63.4\%$。

泊松分布（Poisson distribution）

泊松分布产生于表示在一定时间或空间内出现的事件个数的场景中。泊松分布有一些基本假设，设观察的这一单位时间或空间为$[0,1)$，取一个很大的自然数n，将$[0,1)$平分为n段窗口：$l_1=[0,\frac{1}{n}), l_2=[\frac{1}{n},\frac{2}{n}),\dots,l_n=[\frac{n-1}{n},1)$，则：

1. 在每段l_i内，恰发生一个事件的概率正比于这段的长度$\frac{1}{n}$，即可取为$\frac{\lambda}{n}$；又假定n很大故$\frac{1}{n}$很小时，不可能发生两次以上事件；
2. $l_1,l_2,\dots,l_n$中是否发生事件是相互独立的；

这样的基本假设下，单位窗口内发生事件的总数记为随机变量X。此时X应当服从二项分布，而当$n \to \infty$时，X则服从泊松分布，故泊松分布也可以看作是某种形式的二项分布取极限而得到：

$$P(X=i;\lambda)=\lim_{n\to\infty}\binom{n}{i}(\frac{\lambda}{n})^i(1-\frac{\lambda}{n})^{n-i}$$

将$\lim_{n\to\infty}\binom{n}{i}/n^i=1/i!$、$\lim_{n\to\infty}(1-\frac{\lambda}{n})^{n-i}=e^{-\lambda}$代入即可得到泊松分布的分布律。

一般如果$X \sim B(n,p)$且n较大、p较小、$np=\lambda$不太大时，X的分布接近于泊松分布$P(\lambda)$。

$$P(x;\lambda)=e^{-\lambda}\frac{\lambda^x}{x!}, x=0,1,\dots$$
$$\mathrm{E}[X]=\lambda$$
$$\mathrm{Var}[X]=\lambda$$

伯努利分布（Bernoulli distribution）

伯努利分布$B(1,p)$实际上是二项分布中$n=1$的一个特例：

$$B(1;1,p)=p, B(0;1,p)=1-p$$
$$\mathrm{E}[X]=p$$
$$\mathrm{Var}[X]=p(1-p)$$

多项分布（multinomial distribution）

多项分布其实就是二项分布的推广，不像二项分布，多项分布的取值的是多值而不是二值的（binary）。假设有k种结果，且这k种结果互相对立、完备穷举（mutually exclusive and collectively exhaustive），此时它们的概率之和为1，即$p_1+\dots+p_k=1$，多项分布计算的则是这k种结果分别发生$n_1,\dots,n_k$次时的概率。令$N=n_1+\dots+n_k, \vec{p}=[p_1,\dots,p_k], \vec{n}=[n_1,\dots,n_k]$，则：

$$P(\vec{n};\vec{p})=\frac{N!}{n_1!\dots n_k!}p_1^{n_1}\dots p_k^{n_k}$$

多项分布可以拓展到连续情况，此时$n_1,\dots,n_k\in\mathbb{R}_+$，而概率质量函数变为

$$p(\vec{n};\vec{p})=\frac{\Gamma(N+1)}{\Gamma(n_1+1)\dots\Gamma(n_k+1)}p_1^{n_1}\dots p_k^{n_k}$$

连续情况下的多项分布也是`sklearn`中能将TFIDF特征应用到`MultinomialNB`的基本原理。

分类分布（categorical distribution）

类似伯努利分布是二项分布$n=1$时的特例，分类分布则是多项分布$N=1$时的特例：

$$\begin{aligned}P(\vec{n};\vec{p},1)=\prod_{i=1}^{k}p_i^{n_i}\\ \mathrm{E}[X]=\vec{p}\\ \mathrm{Var}[X]=\vec{p}(1-\vec{p})\end{aligned}$$

连续型

指数分布（exponential distribution）

指数分布最常见的一个场景是寿命估计。设想一种大批生产的电器元件，其元件寿命X是随机变量，在"无老化"的假定下——即"若元件在时刻x尚正常工作，则其失效率总为某个与x无关的常数$\lambda>0$"，那么X服从参数为λ的指数分布。

上述假设用概率语言描述则是

$$\lim_{h\to 0}P(x\le X\le x+h|X>x)/h=\lambda$$

注意到

$$P(x\le X\le x+h|X>x)=\frac{P(\{x\le X\le x+h\}\cap\{X>x\})}{P(X>x)}=\frac{P(x<X\le x+h)}{P(X>x)}$$

所以

$$\begin{aligned}\lim_{h\to 0}\frac{P(x<X\le x+h)}{hP(x<X))}=\lambda\\ \lim_{h\to 0}\frac{F(x+h)-F(x)}{h(1-F(x))}=\lambda\\ \frac{F'(x)}{1-F(x)}=\lambda\end{aligned}$$

上述微分方程的通解为$F(x)=1-Ce^{-\lambda x}$，而$F(0)=0$，故$C=1$。

$$\begin{aligned}p(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}\\ \mathrm{E}[X]=\lambda^{-1}\\ \mathrm{Var}[X]=\lambda^{-2}\end{aligned}$$

正态分布（normal distribution）

正态分布也叫作高斯分布（Gaussian distribution），一维情况下：

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

二维情况下：

$$p\Big((x,y);\mu_X,\mu_Y,\sigma_X,\sigma_Y,\sigma_{XY}\Big)=\frac{1}{2\pi\sqrt{(\sigma_X^2\sigma_Y^2-\sigma_{XY}^2)}}e^{-\frac{1}{2(1-\sigma_{XY}^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}-\frac{2\sigma_{XY}(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}$$

n维情况下：

$$p(\mathrm{x};\mu,\Sigma)=\frac{1}{\sqrt{|2\pi\Sigma|}}e^{-\frac{1}{2}(\mathrm{x}-\mu)^T\Sigma^{-1}(\mathrm{x}-\mu)}$$