二项随机变量

zcl.space

目录

1 伯努利随机变量和二项随机变量

假定一个实验，其结果可以分为成功或者失败。如果我们在试验的机故宫是成功时令$X=1$，而在试验的结果是失败时令$X=0$，那么X的概率质量函数是：

$$
\begin{aligned}
p(0) &= P(X=0) = 1-p && (1.1)\\
p(1) &= P(X=1) = p && (1.2)
\end{aligned}
$$

其中$p, 0 \le p \le 1$是试验的结果为成功的概率。

随机变量X成为伯努利随机变量，如果其概率密度函数由式(1.1)给出. 现在我们推广伯努利随机变量。

假定做了n 次试验，其中每次结果成功的概率为p，失败的概率为$1-p$，如果以X代表出现在n次试验中成功的次数，那么X称为具有参数(n,p)的二项随机变量，其概率质量函数为：

$$
p(i) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, \dots, n \tag{1.3}
$$

可以通过二项式定理验证，这些概率加起来是1：

$$
\sum_{i=0}^{n} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1 \tag{1.4}
$$

2 性质

接下来我们讨论二项分布的性质，先看期望和方差。

首先我们注意到：

$$E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \tag{2.1}$$

利用恒等式：

$$i\binom{n}{i} = n\binom{n-1}{i-1} \tag{2.2}$$

可得：

$$\begin{aligned} E[X^k] &= np\sum_{i=1}^{n} i^{k-1}\binom{n-1}{k-1}p^{i-1}(1-p)^{n-i} && (2.3)\\ &= np\sum_{i=1}^{n}(j+1)^{k-1}\binom{n-1}{j}p^j(1-p)^{n-1-j} && (2.4)\\ &= npE[(Y+1)^{k-1}] && (2.5)\end{aligned}$$

其中Y是一个$(n-1,p)$的二项随机变量。在上面的式子中令$k=1$，可得：

$$E[X] = np \tag{2.6}$$

即如果每次试验成功的概率为p，那么n次独立重复试验的成功次数的期望等于np. 令式 (2.3)中的$k=2$，结合二项随机变量的期望公式，可得：

$$E[X^2] = npE[Y+1] = np[(n-1)p+1] \tag{2.7}$$

结合式 (2.6)，有：

$$\mathrm{Var}(X) = E[X^2] - (E[x])^2 = np[(n-1)p+1] - (np)^2 = np(1-p) \tag{2.8}$$

综上可得结论：如果X是一个参数为n,p的二项随机变量，那么：

$$E[X] = np \qquad \mathrm{Var}(X) = np(1-p) \tag{2.9}$$

关于二项分布还有一个很重要的结论：

定理 2.1

如果X是一个参数为n,p的二项随机变量，其中$0 < p < 1$，那么当k从0到n时，$P\{X=k\}$一开始单调递增，然后一直单调递减，当$k=\lceil (n+1)p \rceil$时取的最大值。

证明 2.1

为证明这个命题，我们考虑$P\{X=k\}/P\{X=k-1\}$，对于给定的k，判定其与1的大小关系。

$$\frac{P\{X=k\}}{P\{X=k-1\}} = \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\binom{n}{k-1}p^{k-1}(1-p)^{n-k+1}} \tag{2.10}$$

$$= \frac{(n-k+1)p}{k(1-p)} \tag{2.11}$$

因此$P\{X=k\} \geq P\{X=k-1\}$，当且仅当:

$$(n-k+1)p \geq k(1-p) \tag{2.12}$$

等价于$k \leq (n+1)p$

注意上面的证明过程告诉我们了一种递归的计算二项分布的方法。

3 例子

针对上面的例子。使用python画出二项分布的pmf图。我使用 scipy.stats 提供的 binom 函数。

```
from scipy.stats import binom
import numpy as np
import matplotlib.pyplot as plt
fig,ax = plt.subplots(1,1)
n,p = 5,0.4
x = np.arange(binom.ppf(0,n,p),binom.ppf(1,n,p))
ax.plot(x, binom.pmf(x, n, p), 'bo', ms=8, label='binom pmf')
ax.vlines(x, 0, binom.pmf(x, n, p), colors='b', lw=5, alpha=0.5)
rv = binom(n, p)
ax.vlines(x, 0, rv.pmf(x), colors='k', linestyles='-', lw=1,label='froze
ax.legend(loc='best', frameon=False)
plt.show()
```

其概率质量函数如图1 所示。

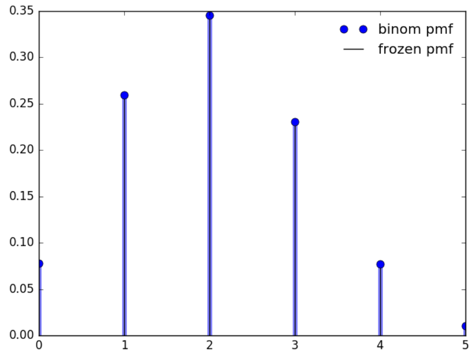

图 1: (5,0.4)的二项分布