概率分布

Guangyao Zhao

2022-08-04

Contents

离散型随机变量

二项式分布 (Binomial distribution)

二项分布是一种最简单，最常用的分布。它描述的是一个事件发生的概率是 p，即不发生的概率就是 $1-p$。在 n 次中发生 k 次的概率是：

$$P_k = b(k, n, p) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

泊松分布 (Poisson distribution)

泊松分布是一种进阶的二项式分布，个人觉得在离散分布中最重要的一个分布。他描述的是当二项式分布的 n 很大，概率 p 特别小的时候，发生 k 次的概率。其中 $\lambda = np$:

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

证明过程如下：已知前提知识

$$\lim_{n\to\infty}\left(1+\frac{\lambda}{n}\right)^{\frac{n}{\lambda}} = e^{\lambda}$$

具体证明如下：

$$\begin{aligned}
\lim_{n\to\infty} P(X=k) &= \lim_{n\to\infty}\begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \\
&= \lim_{n\to\infty}\frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty}\underbrace{\left[\frac{n!}{n^k(n-k)!}\right]}_{F}\left(\frac{\lambda^k}{k!}\right)\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to\exp(-\lambda)}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&= \lim_{n\to\infty}\underbrace{\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)\right]}_{\to 1}\left(\frac{\lambda^k}{k!}\right)\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\exp(-\lambda)}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&= \left(\frac{\lambda^k}{k!}\right)\exp(-\lambda)
\end{aligned}$$

例子

计算机硬件公司制造某种特殊型号的微型芯片，次品率达到 0.1%，各芯片成为次品相互独立。请用两种方法，即二项分布和泊松分布求：在 1000 只产品中至少有 2 只次品的概率。以 X 表示产品中的次品数，$X \sim b(1000, 0.001)$

二项分布

$$\begin{aligned}
P\{X\geq 2\} &= 1 - P\{X=0\} - P\{X=1\} \\
&= 1 - 0.999^{1000} - 1000\times 0.001^1\times 0.999^{999} \\
&= 0.2642411
\end{aligned}$$

泊松分布

参数: $n=1000, p=0.001$, 即 $\lambda=np=1$

$$\begin{aligned}P\{X\geq 2\}&=1-P\{X=0\}-P\{X=1\}\\&=1-\frac{1^0}{0!}e^{-1}-\frac{1^1}{1!}e^{-1}\\&=0.2642411\end{aligned}$$

几何分布 (Geometric distribution)

几何分布比较简单: 一堆产品的次品率是 p, 抽到第 k 次抽到次品率的概率是多少

$$P(X=k)=(1-p)^{k-1}p$$

超几何分布 (Hypergeometric distribution)

超几何分布也不难: 一堆产品中良品和次品分别有 N 和 M。举行一次抽样, 抽出 n 个产品中, 有 k 个次品的概率

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

连续型随机变量

连续型随机变量的概率密度要满足以下条件:

- $f(x)\geq 0$
- $\int_{-\infty}^{+\infty}f(x)dx=1$
- $P(a\leq X\leq b)=F(b)-F(a)=\int_a^b f(x)dx$

正态分布 (Normal distribution)

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

指数分布 (Exponential distribution)

指数分布最大的特点是『无记忆性』。其中 λ 表示的是『每单位时间发生该事件的次数』，该值和 x 无关

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{其他}\end{cases}$$

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

例子

顾客到银行的窗口等待时间 X 服从指数分布，概率密度为：

$$f_X(x)=\begin{cases}\frac{1}{5}e^{-x/5} & ,x>0\\ 0, & \text{其他}\end{cases}$$

某顾客在窗口等待服务，若超过 10 他就离开，他一个月要到银行 5 次，以 Y 表示一个月内他未等到服务而离开窗口的次数。写出 Y 的分布律，并求出 $P\{Y\geq 1\}$

离开的概率计算如下：

$$P\{X>10\}=1-P\{X\leq 10\}=e^{-2}$$

二项分布计算最终结果：

$$P\{Y>1\}=1-P\{Y=1\}=1-\binom{5}{0}\left(e^{-2}\right)^0\left(1-e^{-2}\right)^5$$

均匀分布 (Uniform distribution)

$$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases}$$

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\leq x<b\\ 1, & x\geq b\end{cases}$$