# Lecture 6 : Common Probability Distribution

## §1 Sampling

### 1. Sampling（抽样）

1° Sampling可以被用来 do simulation.

2° Sampling可以被用来估计人口数，因为样本数据可以近似真实模型.

### 2. To sample X :

1° 对于 Common R.V. (random variable) : Python有内置 function.

2° 对于 Uncommon R.V. : Need tricks

## §2 Common R.V. : Discrete

### 1. Bernoulli distribution（伯努利分布） Ber(p)

1° 伯努利分布（0-1分布）

仅进行一次实验，对于随机变量X，有概率p取1为值，有概率 $q=1-p$ 取0为值.

e.g. 抛一枚硬币正面朝上，一次游戏胜利/失败.

2° 期望与方差

Mean = p

Variance = p·q

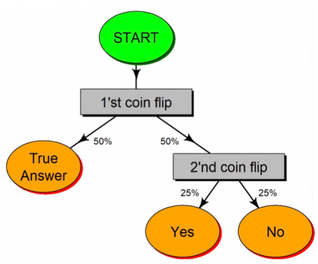
证: $E(X) = 1\cdot p + 0\cdot q = p$

$$Var(X) = E(X^2) - E(X)^2 = (1^2\cdot p + 0^2\cdot q) - p^2 = p(1-p) = pq$$

3° 应用

**1.Bernoulli distribution: Application**

- The platform knows whether a consumer has purchased a specific product.
- A seller wants to inquire this information.
- To **protect the privacy**, the platform cannot answer exactly.
- How to implement?




**Randomized response!!!**

### 2. Binomial distribution（二项分布） Bin(n,p)

1° 二项分布

① n次伯努利实验

② n次实验相互独立（有放回）

③ 含有系数 n（实验次数），p（每次成功概率）

④ 变量 X：成功/失败 次数

e.g. 抛一枚硬币3次，2次正面朝上

2° 概率（研究 n 次实验成功/失败 k 次的概率）

$Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ ，$\binom{n}{k}=\frac{n!}{k!(n-k)!}$ $(C_n^k)$ ，$k=0,1,2,\cdots\cdots n$

3° 期望与方差（研究 n 次实验成功/失败的次数）

Mean $= n\cdot p$

Variance $= n\cdot p\cdot(1-p)$

证1: 
$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\cdot C_n^k p^k q^{n-k} \\
&= \sum_{k=0}^{n} k\cdot\frac{n!}{k!(n-k)!}\cdot p^k\cdot q^{n-k} \\
&= \sum_{k=0}^{n} \frac{n!}{(k-1)!(n-k)!} p^k\cdot q^{n-k} \\
&= np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}\cdot q^{n-k} \quad (k=1:k=0\text{的那一项值为}0) \\
&= np\underbrace{\sum_{k-1=0}^{n} C_{n-1}^{k-1} p^{k-1} q^{n-k}}_{} \longrightarrow =1 \\
&= np
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= E(X^2)-E(X)^2 \\
&= \sum_{k=0}^{n} k^2\cdot C_n^k p^k q^{n-k} - n^2p^2 \\
&= np\sum_{k-1=0}^{n} k\cdot C_{n-1}^{k-1} p^{k-1} q^{n-k} - n^2p^2 \\
&= np\cdot\left(1+\sum_{k-1=0}^{n}(k-1)\cdot C_{n-1}^{k-1} p^{k-1} q^{n-k}\right) - n^2p^2 \\
&= np\cdot(1+(n-1)p) - n^2p^2 \\
&= npq + n^2p^2 - n^2p^2 \\
&= npq
\end{aligned}
$$

证2: 因为 n 次实验相互独立

$E(X)=E(X_1+X_2+\cdots+X_n)=EX_1+EX_2+\cdots+EX_n=np$

$D(X)=D(X_1+X_2+\cdots+X_n)=DX_1+DX_2+\cdots+DX_n=npq$

4° 应用

## 2.Binomial distribution: Application

- A buyer buys N products from a producer.
- Each product can be flawed (w.p. p) or not (w.p. 1-p)
- If k products are flawed, the producer has to refund $R_k$


- How much the producer will pay to the buyer on average?

R(N,p) = $\sum_{k=1}^{N}\Pr(X=k)\,R_k$

- Suppose the producer can incur cost C(p) to ensure that the product is flawed w.p. p.
- How much to invest?
- Choose p such that R(N,p) + C(p) is minimized

*5° 最可能值

① $(n+1)p$ 不为整数，则 $x=[(n+1)\cdot p]$ 是最可能值

② $(n+1)p$ 为整数，则 $x=(n+1)p$，$x=(n+1)p-1$ 是最可能值

例：彩票每周开一次，中奖率 $10^{-5}$，十年买了520次，求从未中奖的概率

$$P(X=0)=\binom{520}{0}\cdot(10^{-5})^0\cdot(1-10^{-5})^{520}$$
$$=0.99999^{520}$$
$$\approx 0.9948$$

*可用泊松分布近似计算.

例：报警器，报0.8，要以99%报警，求要装多少台

设安装台数为 $n$，报警台数为 $X$，$X\sim B(n,0.8)$

$$P\{X\geq 1\}=1-P\{X=0\}=1-C_n^0\,0.8^0\cdot 0.2^n=1-0.2^n\geq 0.99$$

$$n\geq 3$$

3. Geometric distribution（几何分布） $Geo(p)$

1° 几何分布

① 连续进行伯努利实验，直至成功

② 实验间相互独立（有放回）

③ 含有参数 $p$（每次成功概率）

④ 变量 $X$：第 $X$ 次实验首次成功，前 $X-1$ 次均未成功

e.g. 抛一枚硬币直至正面朝上

2° 概率（研究实验首次成功需要 $k$ 次的概率）

$$Pr(X=k)=(1-p)^{k-1}p,\quad k=1,2,3,\cdots\cdots$$

3° 期望与方差（研究实验首次成功需要的次数）

$$Mean=1/p$$

$$Variance=(1-p)/p^2$$

证：引理：$\sum\limits_{k=1}^{\infty}kx^{k-1}=\left(\sum\limits_{k=1}^{\infty}x^k\right)'=\left(\frac{x}{1-x}\right)'=\frac{1}{(1-x)^2}$

$$E(X)=\sum_{k=1}^{\infty}k(1-p)^{k-1}\cdot p$$
$$=\left(\sum_{k=1}^{\infty}k(1-p)^{k-1}\right)\cdot p$$
$$=\frac{1}{p^2}\cdot p$$
$$=1/p$$

引理：$\sum\limits_{k=1}^{\infty}k^2x^{k-1}=\sum\limits_{k=1}^{\infty}k\cdot kx^{k-1}=\left(\sum\limits_{k=1}^{\infty}k\cdot x^k\right)'=\left(x\sum\limits_{k=1}^{\infty}k\cdot x^{k-1}\right)'=\left(\frac{x}{(1-x)^2}\right)'=\frac{1+x}{(1-x)^3}$

$$Var(X)=E(X^2)-E(X)^2$$
$$=\left(\sum_{k=0}^{\infty}k^2\cdot(1-p)^{k-1}\right)\cdot p-\frac{1}{p^2}$$
$$=\frac{2-p}{p^3}\cdot p-\frac{1}{p^2}$$
$$=(1-p)/p^2$$

## * Connections

- A machine breaks down w.p. p on each date.

| Bernoulli distribution | Binomial distribution | Geometric Distribution |
|---|---|---|
| X: Whether the machine breaks down on a spefic date. | X: the number of breadowns during the first N dates. | X: the first date the machine breaks down |

## 4. Poisson distribution (泊松分布) $Poi(\lambda)$

### 1° 泊松分布

① 泊松分布是二项分布在 n 很大 p 很小时的一种极限形式

② 用于估计在某段连续的时间内，某件事发生的次数

③ 含有系数 $\lambda$（单位时间内随机事件的平均发生次数）（历史测算数值）

④ 变量 X：某段时间内随机事件的发生次数（未来估算数值）

e.g. 在 11:00 ~ 12:00 到达某一车站的人数

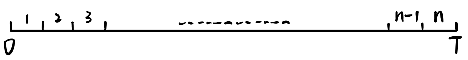
在显微镜某一特定面积内的红细胞数

### 2° 概率（研究某段时间内随机事件发生 k 次的概率）

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

证：求单位时间 T 内事件 A 发生 k 次的概率

- 将单位时间 T 等分为 n 段，使 $n\to\infty$




- 理想化假设（化为二项分布）：

① 独立性假设：各小时段内发生事件 A 的概率不互相影响.

② 各小时段内事件 A 最多发生 1 次（n 足够大时）且概率为 p

- $P\{X=k\} = C_n^k \cdot p^k \cdot (1-p)^{n-k}$

令 $\lambda = E(X) = np$，则 $p = \frac{\lambda}{n}$（$\lambda$ 为历史数据，已知）

- 则 $P\{X=k\} = \lim\limits_{n\to\infty} C_n^k \cdot (\frac{\lambda}{n})^k \cdot (1-\frac{\lambda}{n})^{n-k}$

$\lim\limits_{n\to\infty} \frac{C_n^k}{n^k} = \lim\limits_{n\to\infty} \frac{n!}{k!\,(n-k)!\,n^k} = \frac{1}{k!}$

$\lim\limits_{n\to\infty} \lambda^k = \lambda^k$

$\lim\limits_{n\to\infty}(1-\frac{\lambda}{n})^n = \lim\limits_{n\to\infty}(1+\frac{1}{n/-\lambda})^{\frac{n}{-\lambda}\cdot(-\lambda)} = e^{-\lambda}$

$\lim\limits_{n\to\infty}(1-\frac{\lambda}{n})^{-k} = 1$

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

### 3° 期望与方差

Mean $= \lambda$

Variance $= \lambda$

# §3 Common R.V.: Continuous

## 1. Uniform distribution (平均分布)  Unif [a, b]

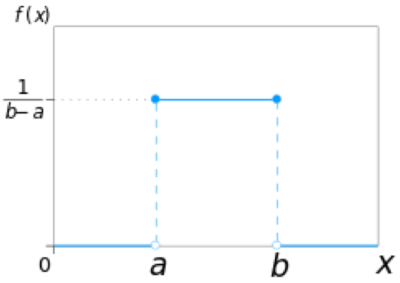
1° 平均分布

① X 在 [a, b] 上取值有着相同的概率

② 系数: a (区间下界), b (区间上界)

2° 概率

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



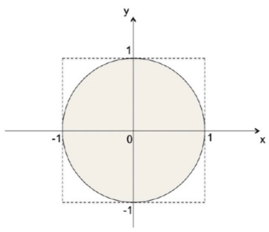

3° Cumulative distribution function

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

4° 应用

Estimation the value of $\pi$

- $\pi$ is the ratio of the circumference of a circle to its diameter
- $\pi$ is also the area of a unit disc
- Can write it as an integral: $\pi = \int_{-1}^{1}\int_{-1}^{1} 1\{x^2 + y^2 \le 1\}dxdy$




## 2. Normal distribution (正态分布)  $N(\mu, \sigma^2)$

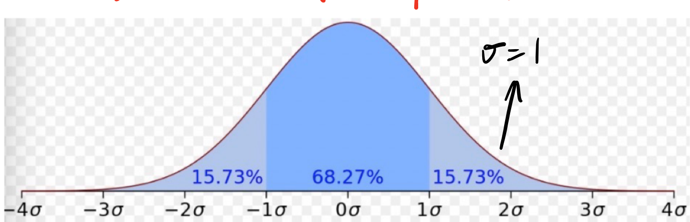
1° 正态分布

① 系数: $\mu$ (均值), $\sigma$ (标准差)

e.g. 人群中的身高, 人群的财富分布

2° 概率

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < +\infty$$



*3° 性质

① $y=\varphi(x)$ 以 $x=\mu$ 为对称轴，呈钟形

$x=\mu$ 时，$\varphi(x)$ 取最大值 $\frac{1}{\sqrt{2\pi}\sigma}$

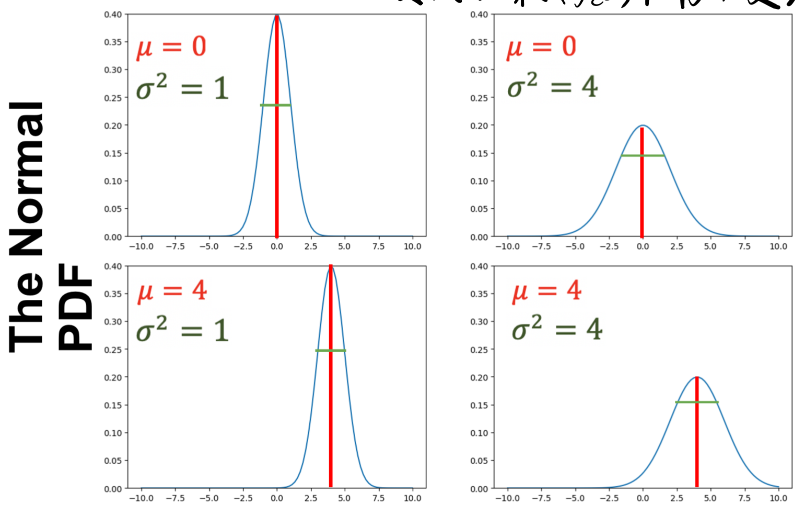
② $y=\varphi(x)$ 以 $x$ 轴为渐近线

$x=\mu\pm\sigma$ 为拐点

③ $\sigma$ 固定，$\mu$ 变化．左右移动

$\mu$ 固定，$\sigma$ 变化．$\sigma$ 变小，最高点上移，变陡

$\sigma$ 变大，最高点下移，变缓




*4° Cumulative distribution function

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

*5° 标准正态分布

$\mu=0$，$\sigma=1$

$\varphi_0(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，$-\infty<x<+\infty$

$\Phi_0(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$

性质：① 以 $y$ 轴为对称轴

② $\varphi_0(x)=\varphi_0(-x)$

③ $\Phi_0(-x)=1-\Phi_0(x)$

*6° 将一般正态分布化为标准正态分布

$\varphi(x)=\frac{1}{\sigma}\varphi_0\left(\frac{x-\mu}{\sigma}\right)$

$\Phi(x)=\Phi_0\left(\frac{x-\mu}{\sigma}\right)$

例：$X\sim N(1,4)$，$\mu=1$，$\sigma=2$，求 $P\{0<X<1.6\}$

$$\begin{aligned}P\{0<X<1.6\}&=\Phi(1.6)-\Phi(0)\\&=\Phi_0\left(\frac{1.6-1}{2}\right)-\Phi_0\left(\frac{0-1}{2}\right)\\&=\Phi_0(0.3)-\Phi_0(-0.5)\\&=\Phi_0(0.3)-1+\Phi_0(0.5)\end{aligned}$$

例：零件长度 $X \sim N(50,1)$，$\mu=50$，$\sigma=1$，$50\pm1$ 为合格，求：

(1) 合格率

(2) 抽3个，至少1个合格的概率

(1)
$$
\begin{aligned}
P(49 \le X \le 51) &= \Phi(51) - \Phi(49) \\
&= \Phi_0\left(\frac{51-50}{1}\right) - \Phi_0\left(\frac{49-50}{1}\right) \\
&= \Phi_0(1) - \Phi_0(-1) \\
&= 2\Phi_0(1) - 1
\end{aligned}
$$

(2) $P\{Y \ge 1\} = 1 - P\{Y=0\} = 1-(1-0.6826)^3 \approx 0.968$

## 3. Exponential distribution（指数分布） $Exp(\lambda)$

### 1° 指数分布

① 与泊松分布相对

② 用于估计同一事件相邻两次发生的时间间隔

- Exponential distribution models the interarrival time between two consecutive events.
- Poisson distribution models the number of arrivals in a unit time interval.

③ 系数：$\lambda$（单位时间内随机事件的平均发生次数）（历史测算数值）

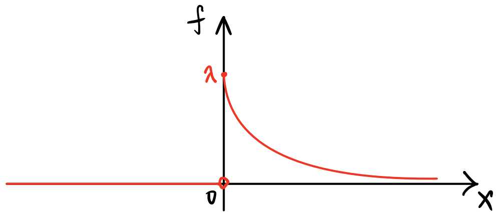
④ 变量 $X$：随机事件相邻两次发生的时间间隔（未来估算数值）

e.g. 乘客进入机场的时间间隔

相邻两次呼叫服务台的时间间隔

### 2° 概率

$$
f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}
$$




### 3° Cumulative distribution function

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

证：

- Suppose a T-length duration is split into n periods.
- At each period, an event occurs with probability $\frac{\lambda T}{n}$
- What is the probability that at least one event happens during the duration?

$$
1-\left(1-\frac{\lambda T}{n}\right)^n \xrightarrow{n\to\infty} 1-e^{-\lambda T}
$$

Continuous version of geometric distribution

X: time until the first event

F(T)=P(X≤T)
=P(at least one event during T)

*4° 无记忆性

X服从指数分布, $s>0$, $t>0$, 则 $P\{X>s+t \mid X>s\} = P\{X>t\}$

e.g. 若灯泡寿命服从指数分布，则其接着存活n年的概率与其现有寿命无关

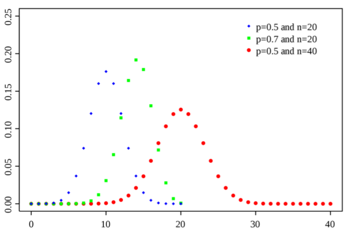
证: 
$$P\{X>s+t \mid X>s\} = \frac{P\{\{X>s+t\}\cap\{X>s\}\}}{P\{X>s\}}$$
$$= \frac{P\{X>s+t\}}{P\{X>s\}}$$
$$= \frac{\int_{s+t}^{+\infty} \lambda e^{-\lambda x} dx}{\int_{s}^{+\infty} \lambda e^{-\lambda x} dx}$$
$$= \frac{-e^{-\lambda x}\Big|_{s+t}^{+\infty}}{-e^{-\lambda x}\Big|_{s}^{+\infty}}$$
$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}$$
$$= e^{-\lambda t}$$

# Discrete

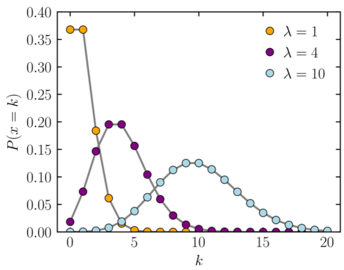
| | p.m.f | Mean | Variance |
|---|---|---|---|
| Bernoulli; Ber(p) | $P(X=1)=p$ <br> $P(X=0)=1-p$ | $p$ | $p(1-p)$ |
| Binomial; Bin(n,p) | $\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ | $np$ | $np(1-p)$ |
| Geometric; Geo(p) | $\Pr(X=k)=(1-p)^{k-1}p$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |
| Poisson; Poi($\lambda$) | $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\cdots$ | $\lambda$ | $\lambda$ |

| Bernoulli distribution | Binomial distribution | Geometric Distribution |
|---|---|---|
| X: Whether the machine breaks down on a spefic day. | X: the number of breadowns during the first N days. | X: the first day the machine breaks down |

# Discrete - visualization



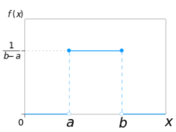

Comparison of different binomial

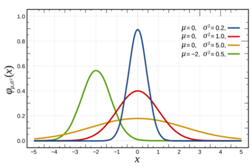



Comparison of different Poisson

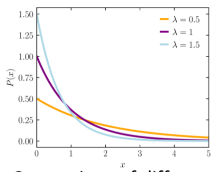
# Continuous

| | p.d.f | Mean | Variance |
|---|---|---|---|
| Uniform; Unif[a,b] | $f(x)=\frac{1}{b-a}, a\le x\le b$ | $\frac{a+b}{2}$ | $\frac{1}{12}(b-a)^2$ |
| Normal; $N(\mu,\sigma^2)$ | $f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ | $\mu$ | $\sigma^2$ |
| Exponential; Exp($\lambda$) | $f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0,\\ 0 & , x<0.\end{cases}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |



Comparison of different normal




Comparison of different exp