# 第五章：大数定律与中心极限定理[1]

Xiao Yuan[1]

[1]Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China
xiaoyuan@pku.edu.cn

March 28, 2024

[1]讲义内容基于三本参考教材以及网络素材

# 目录

# 频率收敛于概率?





**频率**

频率：$n$ 次试验中事件发生 ($n_A$ 次) 的比例：$f_n(A) = n_A/n$

# 频率收敛于概率?

## 频率

频率: $n$ 次试验中事件发生 ($n_A$ 次) 的比例: $f_n(A) = n_A/n$

- 频率 $f_n(A)$ 随着 $n$ 增大逐渐趋于一个稳定值，也即是事件 $A$ 发生的概率

# 频率收敛于概率？





**频率**

频率：$n$ 次试验中事件发生 ($n_A$ 次) 的比例：$f_n(A) = n_A/n$

- 频率 $f_n(A)$ 随着 $n$ 增大逐渐趋于一个稳定值，也即是事件 $A$ 发生的概率
- 考虑 $n$ 重伯努利试验，也即是每次试验只有 $A, \bar{A}$ 两个事件可能发生，将试验独立重复 $n$ 次，设事件 $A$ 发生 $n_A$ 次，则上述结论意味着 $\lim_{n\to\infty} \frac{n_A}{n} = p$?

# 频率收敛于概率?





**频率**

频率: $n$ 次试验中事件发生 ($n_A$ 次) 的比例: $f_n(A) = n_A/n$

- 频率 $f_n(A)$ 随着 $n$ 增大逐渐趋于一个稳定值，也即是事件 $A$ 发生的概率
- 考虑 $n$ 重伯努利试验，也即是每次试验只有 $A, \bar{A}$ 两个事件可能发生，将试验独立重复 $n$ 次，设事件 $A$ 发生 $n_A$ 次，则上述结论意味着 $\lim_{n\to\infty} \frac{n_A}{n} = p$?
- 注意到尽管在实际进行 $n$ 重伯努利试验时，我们总是很大概率观察到 $\lim_{n\to\infty} \frac{n_A}{n} = p$ 成立，但是总是存在该式不成立的情况，例如 $n$ 个 $A$ 发生，且违背该式的事件有随 $n$ 增加有无穷多个

# 频率收敛于概率？



## 频率

频率：$n$ 次试验中事件发生 ($n_A$ 次) 的比例：$f_n(A) = n_A/n$

- 频率 $f_n(A)$ 随着 $n$ 增大逐渐趋于一个稳定值，也即是事件 $A$ 发生的概率
- 考虑 $n$ 重伯努利试验，也即是每次试验只有 $A, \bar{A}$ 两个事件可能发生，将试验独立重复 $n$ 次，设事件 $A$ 发生 $n_A$ 次，则上述结论意味着 $\lim_{n\to\infty} \frac{n_A}{n} = p$?
- 注意到尽管在实际进行 $n$ 重伯努利试验时，我们总是很大概率观察到 $\lim_{n\to\infty} \frac{n_A}{n} = p$ 成立，但是总是存在该式不成立的情况，例如 $n$ 个 $A$ 发生，且违背该式的事件有随 $n$ 增加有无穷多个
- 满足 $\lim_{n\to\infty} \frac{n_A}{n} = p$ 成立的事件个数相对总的事件个数是指数小的

# 依概率收敛





- 尽管如此，我们总是很大概率观察到 $\lim_{n\to\infty} \frac{n_A}{n} \approx p$ 成立

# 依概率收敛





- 尽管如此，我们总是很大概率观察到 $\lim_{n\to\infty}\frac{n_A}{n}\approx p$ 成立
- 也即是事件 $B=\{\frac{n_A}{n}\approx p\}$ 随 $n$ 的增加成立的概率也增加

- 尽管如此，我们总是很大概率观察到 $\lim_{n\to\infty} \frac{n_A}{n} \approx p$ 成立
- 也即是事件 $B = \{\frac{n_A}{n} \approx p\}$ 随 $n$ 的增加成立的概率也增加
- 定量来说，对于任意 $\varepsilon > 0$，我们考虑 $B = \{|\frac{n_A}{n} - p| < \varepsilon\}$，则有 $\lim_{n\to\infty} P(B) = 1$

# 依概率收敛



- 尽管如此，我们总是很大概率观察到 $\lim_{n\to\infty}\frac{n_A}{n}\approx p$ 成立
- 也即是事件 $B=\{\frac{n_A}{n}\approx p\}$ 随 $n$ 的增加成立的概率也增加
- 定量来说，对于任意 $\varepsilon>0$，我们考虑 $B=\{|\frac{n_A}{n}-p|<\varepsilon\}$，则有 $\lim_{n\to\infty}P(B)=1$
- 也即是，**频率依概率收敛域概率**

# 伯努利定理

PEKING UNIVERSITY



## 伯努利定理

设事件 $A$ 发生的概率为 $p \in (0,1)$，$n$ 重伯努利试验中 $A$ 事件发生的次数为 $n_A$，则对于任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

# 伯努利定理

PEKING UNIVERSITY

**伯努利定理**

设事件 $A$ 发生的概率为 $p \in (0,1)$, $n$ 重伯努利试验中 $A$ 事件发生的次数为 $n_A$，则对于任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

- 伯努利定理从理论上论证了频率“收敛”于概率的规律，也被称作伯努利大数定律
- 伯努利定理也是 $0-1$ 分布的大数定律

## 马尔科夫不等式

设随机变量 $X$ 的 $k$ 阶矩存在，也即是 $E(|X|^k) < \infty$，则对于任意 $\varepsilon > 0$，有

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|^k)}{\varepsilon^k}$$

# 马尔科夫不等式

PEKING UNIVERSITY

## 马尔科夫不等式

设随机变量 $X$ 的 $k$ 阶矩存在，也即是 $E(|X|^k) < \infty$，则对于任意 $\varepsilon > 0$，有

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|^k)}{\varepsilon^k}$$

证明：以连续型随机变量为例，有

$$\begin{aligned} P(|X| \geq \varepsilon) &= \int_{|x| \geq \varepsilon} f(x)dx \\ &\leq \int_{|x| \geq \varepsilon} \frac{|x|^k}{\varepsilon^k} f(x)dx \\ &\leq \frac{1}{\varepsilon^k} \int_{|x| \geq \varepsilon} |x|^k f(x)dx \\ &\leq \frac{E(|X|^k)}{\varepsilon^k} \end{aligned}$$

# 马尔科夫不等式

## 马尔科夫不等式

设随机变量 $X$ 的 $k$ 阶矩存在，也即是 $E(|X|^k) < \infty$，则对于任意 $\varepsilon > 0$，有

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|^k)}{\varepsilon^k}$$

# 马尔科夫不等式

## 马尔科夫不等式

设随机变量 $X$ 的 $k$ 阶矩存在，也即是 $E(|X|^k) < \infty$，则对于任意 $\varepsilon > 0$，有

$$P(|X| \geq \varepsilon) \leq \frac{E(|X|^k)}{\varepsilon^k}$$

另证：定义事件 $A = \{|X| \geq \varepsilon\}$，则随机变量 $1_A \leq \frac{|X|^k}{\varepsilon^k}$.

- 当 $1_A = 1$ 时，事件 $A$ 发生，则 $\frac{|X|^k}{\varepsilon^k} \geq 1$.
- 当 $1_A = 0$ 时，显然 $\frac{|X|^k}{\varepsilon^k} \geq 0$.

$$1_A \leq \frac{|X|^k}{\varepsilon^k} \Rightarrow P(|X| \geq \varepsilon) = E(1_A) \leq E\left(\frac{|X|^k}{\varepsilon^k}\right)$$

# 切比雪夫不等式





## 切比雪夫不等式

设随机变量 $X$ 的期望 $\mu = E(X)$ 方差 $\sigma^2 = D(X)$ 存在，则对于任意 $\varepsilon > 0$，有

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

# 切比雪夫不等式



**切比雪夫不等式**

设随机变量 $X$ 的期望 $\mu = E(X)$ 方差 $\sigma^2 = D(X)$ 存在，则对于任意 $\varepsilon > 0$，有

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- 切比雪夫不等式是马尔科夫不等式 $n = 2$, $X = X - E(X)$ 的特殊情况

# 切比雪夫不等式



## 切比雪夫不等式

设随机变量 $X$ 的期望 $\mu = E(X)$ 方差 $\sigma^2 = D(X)$ 存在，则对于任意 $\varepsilon > 0$, 有

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- 切比雪夫不等式是马尔科夫不等式 $n = 2$, $X = X - E(X)$ 的特殊情况
- 例如：考虑 $\varepsilon = 3\sigma$, 则有 $P(|X-\mu| \geq 3\sigma) \leq 1/9 \approx 0.11$

## 切比雪夫不等式

设随机变量 $X$ 的期望 $\mu = E(X)$ 方差 $\sigma^2 = D(X)$ 存在，则对于任意 $\varepsilon > 0$，有

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- 切比雪夫不等式是马尔科夫不等式 $n = 2$, $X = X - E(X)$ 的特殊情况
- 例如：考虑 $\varepsilon = 3\sigma$，则有 $P(|X - \mu| \geq 3\sigma) \leq 1/9 \approx 0.11$
- 另一方面，假设 $X \sim N(\mu, \sigma^2)$，则有 $P(|X - \mu| \geq 3\sigma) \approx 0.003$

# 切比雪夫不等式





**切比雪夫不等式**

设随机变量 $X$ 的期望 $\mu = E(X)$ 方差 $\sigma^2 = D(X)$ 存在，则对于任意 $\varepsilon > 0$，有

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

- 切比雪夫不等式是马尔科夫不等式 $n = 2$，$X = X - E(X)$ 的特殊情况
- 例如：考虑 $\varepsilon = 3\sigma$，则有 $P(|X-\mu| \geq 3\sigma) \leq 1/9 \approx 0.11$
- 另一方面，假设 $X \sim N(\mu, \sigma^2)$，则有 $P(|X-\mu| \geq 3\sigma) \approx 0.003$
- 尽管切比雪夫不等式给出更松的界，它不依赖于随机变量的分布

## Hoeffding 不等式

设随机变量 $X \in [a, b]$ 的期望存在，则对于任意 $\varepsilon > 0$，有

$$P(X - E(X) \geq \varepsilon) \leq e^{-\frac{2\varepsilon^2}{(b-a)^2}}$$

$$P(X - E(X) \leq -\varepsilon) \leq e^{-\frac{2\varepsilon^2}{(b-a)^2}}$$

# Hoeffding 不等式





## Hoeffding 不等式

设随机变量 $X \in [a, b]$ 的期望存在，则对于任意 $\varepsilon > 0$，有

$$P(X - E(X) \geq \varepsilon) \leq e^{-\frac{2\varepsilon^2}{(b-a)^2}}$$

$$P(X - E(X) \leq -\varepsilon) \leq e^{-\frac{2\varepsilon^2}{(b-a)^2}}$$

证明：$P(X - E(X) \geq \varepsilon) \leq e^{-t\varepsilon} E\left(e^{t(X-E(X))}\right) \leq e^{-t\varepsilon} e^{\frac{t^2(b-a)^2}{8}}$
（第一个不等号类似于马尔科夫不等式证明，第二个不等号用到 Hoeffding 不等式）

取 $t = 4\varepsilon/(b-a)^2$，因此有 $P(X - E(X) \geq \varepsilon) \leq e^{-\frac{2\varepsilon^2}{(b-a)^2}}$.
考虑 $-X$ 则得到第二个不等式。

# 例子





- 例：（Hoeffding 不等式）设 $X$ 是在区间 $[a,b]$ 中取值且 $\mathbb{E}X=0$ 的随机变量，则对每一个 $t>0$，有 $\mathbb{E}\exp(tX)\leqslant\exp\left(\frac{t^2(b-a)^2}{8}\right)$

# 例子

- 例：（Hoeffding 不等式）设 $X$ 是在区间 $[a,b]$ 中取值且 $\mathbb{E}X=0$ 的随机变量，则对每一个 $t>0$，有 $\mathbb{E}\exp(tX)\leqslant\exp\left(\frac{t^2(b-a)^2}{8}\right)$

证明：将 $\exp(tX)$ 视为 $X$ 的函数，利用凸性得到 $\exp(tX)\leqslant\frac{X-a}{b-a}\exp(tb)+\frac{b-X}{b-a}\exp(ta)$. 由于 $\mathbb{E}X=0$，于是

$$\mathbb{E}\exp(tX)\leqslant\frac{b}{b-a}\mathbb{E}\exp(ta)-\frac{a}{b-a}\mathbb{E}\exp(tb)$$

定义 $\varphi(t)=\log\left(\frac{b}{b-a}\exp(ta)-\frac{a}{b-a}\exp(tb)\right)$. 则 $\varphi(0)=0,\varphi'(0)=0$，而

$$\varphi''(\xi)=\frac{ab\left(ae^{\xi a}-be^{\xi b}\right)\left(be^{\xi a}-ae^{\xi b}\right)-a^2b^2\left(e^{\xi a}-e^{\xi b}\right)^2}{\left(be^{\xi a}-ae^{\xi b}\right)^2}$$
$$=\frac{-ab(b-a)^2}{-2ab+a^2e^{\xi(a-b)}+b^2e^{\xi(b-a)}}\leqslant\frac{(b-a)^2}{4}$$

利用带 Lagrange 余项的展开得，存在 $\xi\in(0,t)$

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\frac{\varphi''(\xi)}{2}t^2\leqslant\frac{t^2(b-a)^2}{8}.$$

故

$$\mathbb{E}\exp(tX)\leqslant\mathbb{E}e^{\varphi(t)}\leqslant\exp\left(\frac{t^2(b-a)^2}{8}\right)$$

# 伯努利定理证明

PEKING UNIVERSITY

## 伯努利定理

设事件 $A$ 发生的概率为 $p \in (0,1)$, $n$ 重伯努利试验中 $A$ 事件发生的次数为 $n_A$，则对于任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

## 伯努利定理

设事件 $A$ 发生的概率为 $p \in (0,1)$，$n$ 重伯努利试验中 $A$ 事件发生的次数为 $n_A$，则对于任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

证明：定义事件 $A_i = \{$事件 $A$ 第 $i$ 次试验中发生$\}$，则 $n_A = \sum_{i=1}^n 1_{A_i}$，且 $E(1_{A_i}) = p$，$D(1_{A_i}) = p(1-p)$，$E(n_A) = np$，$D(n_A) = np(1-p)$．因此利用切比雪夫不等式$\left[P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}\right]$ 有

$$P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) = P(|n_A - np| \geq n\varepsilon) \leq \frac{np(1-p)}{n^2\varepsilon^2} = \frac{pq}{n\varepsilon^2}$$

因此 $\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) = 0$

# 伯努利定理证明





**伯努利定理**

设事件 $A$ 发生的概率为 $p \in (0,1)$, $n$ 重伯努利试验中 $A$ 事件发生的次数为 $n_A$，则对于任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

## 伯努利定理

设事件 $A$ 发生的概率为 $p \in (0,1)$, $n$ 重伯努利试验中 $A$ 事件发生的次数为 $n_A$，则对于任意 $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$$

证明：定义事件 $A_i = \{$事件 $A$ 第 $i$ 次试验中发生$\}$, 则 $n_A = \sum_{i=1}^n 1_{A_i}$,
$P(n_A - np \geq \varepsilon) \leq e^{-t\varepsilon} E\left(e^{t(n_A - np)}\right) = e^{-t\varepsilon} \prod_{i=1}^n E\left(e^{t(1_{A_i} - p)}\right) \leq$
$e^{-t\varepsilon} e^{\frac{t^2 n(b-a)^2}{8}} = e^{-t\varepsilon + \frac{t^2 n}{8}}$，取 $t = \frac{4\varepsilon}{n}$，则有 $P(n_A - np \geq \varepsilon) \leq e^{-\frac{2\varepsilon^2}{n}}$
同理可得 $P(n_A - np \leq -\varepsilon) \leq e^{-\frac{2\varepsilon^2}{n}}$
因此有 $P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) \leq 2e^{-\frac{2\varepsilon^2}{n}}$，因此 $\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right) = 0$

# 例子



- 例：某天文机构想测量宇宙中两颗行星的距离，进行了 $n$ 次独立的观测，测量值分别为 $X_i$ (光年)，$i = 1, 2, \cdots, n$ 若 $E(X_i) = \mu($ 为两颗行星的真实距离，未知 $)$，$D(X_i) = 5$. 现取这 $n$ 次观测的平均作为真实距离 $\mu$ 的估计.

  (1) 若 $n = 100$, 那公估计值与真实值之间的误差在 $\pm 0.5$ 光年之内的概率至少有多大?

# 例子





- 例：某天文机构想测量宇宙中两颗行星的距离，进行了 $n$ 次独立的观测，测量值分别为 $X_i$ (光年), $i = 1, 2, \cdots, n$ 若 $E(X_i) = \mu($ 为两颗行星的真实距离，末知 ), $D(X_i) = 5$. 现取这 $n$ 次观测的平均作为真实距离 $\mu$ 的估计.

  (1) 若 $n = 100$, 那公估计值与真实值之间的误差在 $\pm 0.5$ 光年之内的概率至少有多大?

  $P(|\frac{1}{100}\sum_i X_i - \mu| \geq 0.5) = P(|\sum_i X_i - \mu| \geq 50) \leq \frac{500}{50^2} = 0.2$, 因此
  $P(|\frac{1}{100}\sum_i X_i - \mu| < 0.5) = 0.8$

# 例子

- 例：某天文机构想测量宇宙中两颗行星的距离，进行了 $n$ 次独立的观测，测量值分别为 $X_i$ (光年)，$i=1,2,\cdots,n$ 若 $E(X_i)=\mu($ 为两颗行星的真实距离，未知 $)$，$D(X_i)=5$. 现取这 $n$ 次观测的平均作为真实距离 $\mu$ 的估计.

  (1) 若 $n=100$，那么估计值与真实值之间的误差在 $\pm 0.5$ 光年之内的概率至少有多大?

  $P(|\frac{1}{100}\sum_i X_i-\mu|\geq 0.5)=P(|\sum_i X_i-\mu|\geq 50)\leq \frac{500}{50^2}=0.2$，因此
  $P(|\frac{1}{100}\sum_i X_i-\mu|<0.5)=0.8$

  (2) 若要以不低于 95% 的把握控制估计值与真实值之间的误差在 $\pm 0.5$ 光年之内，至少要观测多少次?

- 例：某天文机构想测量宇宙中两颗行星的距离，进行了 $n$ 次独立的观测，测量值分别为 $X_i$ (光年), $i=1,2,\cdots,n$ 若 $E(X_i)=\mu($ 为两颗行星的真实距离，未知 $)$, $D(X_i)=5$. 现取这 $n$ 次观测的平均作为真实距离 $\mu$ 的估计.

  (1) 若 $n=100$, 那么估计值与真实值之间的误差在 $\pm0.5$ 光年之内的概率至少有多大?

  $P(|\frac{1}{100}\sum_i X_i-\mu|\geq 0.5)=P(|\sum_i X_i-\mu|\geq 50)\leq\frac{500}{50^2}=0.2$, 因此 $P(|\frac{1}{100}\sum_i X_i-\mu|<0.5)=0.8$

  (2) 若要以不低于 95% 的把握控制估计值与真实值之间的误差在 $\pm0.5$ 光年之内，至少要观测多少次?

  $P(|\frac{1}{n}\sum_i X_i-\mu|\geq 0.5)=P(|\sum_i X_i-\mu|\geq 0.5n)\leq\frac{5n}{(0.5n)^2}\leq 0.05$, 因此 $n\geq 400$

## 依概率收敛

考虑随机变量 $X$ 和 $X_1, X_2, \ldots,$, 如果对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1,$$

则成随机变量序列 $\{X_n\}$ 依概率收敛于 $X$, 并记作 $\lim_{n\to\infty} X_n \stackrel{P}{=} X$ 或者 $X_n \xrightarrow{P} X\ (n \to \infty)$

## 依概率收敛

考虑随机变量 $X$ 和 $X_1, X_2, \ldots,$, 如果对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1,$$

则成随机变量序列 $\{X_n\}$ 依概率收敛于 $X$, 并记作 $\lim_{n\to\infty} X_n \overset{P}{=} X$ 或者 $X_n \xrightarrow{P} X\ (n\to\infty)$

- 若 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, 当 $n\to\infty$ 时. 函数 $g(x,y)$ 在点 $(a,b)$ 连续, 则 $g(X_n, Y_n) \xrightarrow{P} g(a,b)$, 当 $n\to\infty$ 时.
  如: 当 $n\to\infty$ 时
  $$X_n + Y_n \xrightarrow{P} a+b,$$
  $$X_n \times Y_n \xrightarrow{P} a\times b,$$
  $$X_n/Y_n \xrightarrow{P} a/b(b\neq 0).$$

## 随机变量序列的大数定律

考虑随机变量序列 $X_1, X_2, \ldots,$，若存在常数列 $a_1, a_2, \ldots,$ 使得对于任意 $\varepsilon > 0$ 满足

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - a_n\right| \leq \varepsilon\right) = 1$$

则成随机变量序列 $X_1, X_2, \ldots,$ 服从**大数定律**

## 随机变量序列的大数定律

考虑随机变量序列 $X_1, X_2, \ldots,$，若存在常数列 $a_1, a_2, \ldots$，使得对于任意 $\varepsilon > 0$ 满足

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - a_n\right| \leq \varepsilon\right) = 1$$

则成随机变量序列 $X_1, X_2, \ldots$，服从**大数定律**

- 伯努利定理是上述一般大数定律的特例

## 随机变量序列的大数定律

考虑随机变量序列 $X_1, X_2, \ldots,$，若存在常数列 $a_1, a_2, \ldots,$ 使得对于任意 $\varepsilon > 0$ 满足

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - a_n\right| \leq \varepsilon\right) = 1$$

则成随机变量序列 $X_1, X_2, \ldots,$ 服从**大数定律**

- 伯努利定理是上述一般大数定律的特例
- 一般我们取 $a_n = \sum_{i=1}^n E(X_i)/n$
- 另外三个代表性的大数定律有：切比雪夫大数定律、马尔科夫大数定律、辛钦大数定律

## 切比雪夫大数定律

设随机变量序列 $X_1, X_2, \ldots$, 两两不相关，且方差有界，也即是对于常数 $C$ 有 $D(X_i) \leq C$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \leq \varepsilon\right) = 1$$

## 切比雪夫大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 两两不相关，且方差有界，也即是对于常数 $C$ 有 $D(X_i) \leq C$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \leq \varepsilon\right) = 1$$

证明：根据切比雪夫不等式 $\left[P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}\right]$ 有

$$P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \geq \varepsilon\right) \leq \frac{D\left(\frac{\sum_{i=1}^n X_i}{n}\right)}{\varepsilon^2} = \frac{D(\sum_{i=1}^n X_i)}{n^2\varepsilon^2}$$

$$= \frac{\sum_{i=1}^n D(X_i)}{n^2\varepsilon^2} \leq \frac{C}{n\varepsilon^2}$$

## 马尔科夫大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 满足 $\lim_{\infty} \frac{1}{n^2} D(\sum_{i=1}^n X_i) = 0$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \leq \varepsilon\right) = 1$$

## 马尔科夫大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 满足 $\lim_{\infty} \frac{1}{n^2} D(\sum_{i=1}^n X_i) = 0$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \le \varepsilon\right) = 1$$

- 证明方法仅需利用切比雪夫不等式，也即是注意到 $P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \ge \varepsilon\right) \le \frac{D(\sum_{i=1}^n X_i)}{n^2\varepsilon^2}$

## 马尔科夫大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 满足 $\lim_{\infty} \frac{1}{n^2} D(\sum_{i=1}^n X_i) = 0$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \le \varepsilon\right) = 1$$

- 证明方法仅需利用切比雪夫不等式，也即是注意到 $P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n E(X_i)}{n}\right| \ge \varepsilon\right) \le \frac{D(\sum_{i=1}^n X_i)}{n^2\varepsilon^2}$
- 切比雪夫大数定律是马尔科夫大数定律的特例
- 马尔科夫大数定律不需要假设随机变量的独立性

# 例子



- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立，且它们的分布律为 $P\{X_i = \sqrt{i}\} = P\{X_i = -\sqrt{i}\} = \frac{1}{2i}, P\{X_i = 0\} = 1 - \frac{1}{i}, i = 1, 2, \cdots$ . 试判断 $\{X_i, i \geq 1\}$ 是否服从大数定律？

# 例子



- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立，且它们的分布律为 $P\{X_i = \sqrt{i}\} = P\{X_i = -\sqrt{i}\} = \frac{1}{2i}, P\{X_i = 0\} = 1 - \frac{1}{i}, i = 1, 2, \cdots$ . 试判断 $\{X_i, i \geq 1\}$ 是否服从大数定律？

解：由于对任意的 $i \geq 1$，有 $E(X_i) = 0$，

$$D(X_i) = E(X_i^2) = 0 + (\sqrt{i})^2 \cdot \frac{1}{2i} + (\sqrt{i})^2 \cdot \frac{1}{2i} = 1,$$

所以 $\{X_i, i \geq 1\}$ 相互独立，方差相同，由马尔科夫大数定律知满足大数定律，且

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} 0, \text{ 当} n \to +\infty.$$

## 辛钦大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 满足独立同分布，且存在期望 $\mu = E(X_i) < \infty$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \leq \varepsilon\right) = 1$$

## 辛钦大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 满足独立同分布，且存在期望 $\mu = E(X_i) < \infty$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| \le \varepsilon\right) = 1$$

- 辛钦大数定律一方面需要更强的独立同分布假设，
- 一方面又对方差没有假设

# 例子

- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i$ 服从柯西分布 $f(x) = \frac{1}{\pi(x^2+1)}$，则 $\sum_{i=1}^{n} X_i/n$ 依概率收敛吗？如果依概率收敛，收敛于什么？

# 例子

- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i$ 服从柯西分布 $f(x) = \frac{1}{\pi(x^2+1)}$，则 $\sum_{i=1}^n X_i/n$ 依概率收敛吗？如果依概率收敛，收敛于什么？
  注意到柯西分布的期望发散，因此 $\sum_{i=1}^n X_i/n$ 的收敛性无法直接判断，其它证明方法？

- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i$ 服从柯西分布 $f(x) = \frac{1}{\pi(x^2+1)}$，则 $\sum_{i=1}^n X_i/n$ 依概率收敛吗？如果依概率收敛，收敛于什么？
  注意到柯西分布的期望发散，因此 $\sum_{i=1}^n X_i/n$ 的收敛性无法直接判断，其它证明方法？
- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i$ 服从分布 $f(x) = \frac{c}{(|x|^3+1)}$，其中 $c = 3\sqrt{3}/(4\pi)$ 为归一化常数，则 $\sum_{i=1}^n X_i/n$ 依概率收敛吗？如果依概率收敛，收敛于什么？

# 例子



- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i$ 服从柯西分布 $f(x) = \frac{1}{\pi(x^2+1)}$，则 $\sum_{i=1}^n X_i/n$ 依概率收敛吗？如果依概率收敛，收敛于什么？
  注意到柯西分布的期望发散，因此 $\sum_{i=1}^n X_i/n$ 的收敛性无法直接判断，其它证明方法？
- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i$ 服从分布 $f(x) = \frac{c}{(|x|^3+1)}$，其中 $c = 3\sqrt{3}/(4\pi)$ 为归一化常数，则 $\sum_{i=1}^n X_i/n$ 依概率收敛吗？如果依概率收敛，收敛于什么？
  注意到 $\int_{-\infty}^{\infty} |x| f(x) dx = 2c \int_0^{\infty} x/(x^3+1) dx$ 收敛，因此 $\sum_{i=1}^n X_i/n$ 依概率收敛于 $E(X_i) = 0$

# 例子

- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i \sim U(0,1)$，则 $\sqrt[n]{X_1 X_2 \ldots X_n}$ 依概率收敛吗？如果依概率收敛，收敛于什么？

# 例子





- 例：设随机变量 $X_1, \cdots, X_n, \cdots$，相互独立同分布，$X_i \sim U(0,1)$，则 $\sqrt[n]{X_1 X_2 \ldots X_n}$ 依概率收敛吗？如果依概率收敛，收敛于什么？

  解：记 $Y_n = \sqrt[n]{X_1 \ldots X_n}$，令
  $Z_n = \ln Y_n = \frac{1}{n}(\ln X_1 + \ldots + \ln X_n)$．则
  $\ln X_1, \cdots, \ln X_n, \cdots$，相互独立同分布，又
  $E(\ln X_1) = \int_0^1 \ln x dx = -1$，由辛钦大数定律知，
  $Z_n \xrightarrow{p} -1$，当 $n \to +\infty$．利用依概率收敛的性质，得
  $Y_n = e^{Z_n} \xrightarrow{p} e^{-1}$，当 $n \to +\infty$．

## 几乎必然收敛

考虑随机变量 $X$ 和 $X_1, X_2, \ldots,$ 若

$$P(\lim_{n\to\infty} X_n = X) = 1,$$

则称随机变量序列 $\{X_n\}$ 几乎必然收敛于 $X$, 并记作 $\lim_{n\to\infty} X_n \overset{\text{a.s}}{=} X$ 或者 $X_n \xrightarrow{\text{a.s}} X \ (n \to \infty)$

# 几乎必然收敛

## 几乎必然收敛

考虑随机变量 $X$ 和 $X_1, X_2, \ldots,$ 若

$$P(\lim_{n\to\infty} X_n = X) = 1,$$

则称随机变量序列 $\{X_n\}$ 几乎必然收敛于 $X$, 并记作 $\lim_{n\to\infty} X_n \stackrel{\text{a.s}}{=} X$ 或者 $X_n \xrightarrow{\text{a.s}} X \ (n \to \infty)$

## 几乎必然收敛

令 $A_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$, 则 $X_n \xrightarrow{\text{a.s}} X \ (n \to \infty)$ 等价于 $P(\cup_{m=n}^{\infty} A_m(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$

## 依概率收敛与几乎必然收敛

令 $A_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$,

则 $X_n \xrightarrow{\text{a.s}} X \ (n \to \infty)$ 等价于 $P(\cup_{m=n}^{\infty} A_m(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$

则 $X_n \xrightarrow{P} X \ (n \to \infty)$ 等价于 $P(A_n(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$

## 依概率收敛与几乎必然收敛

令 $A_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$,
则 $X_n \xrightarrow{\mathrm{a.s}} X\ (n \to \infty)$ 等价于 $P(\cup_{m=n}^{\infty} A_m(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$
则 $X_n \xrightarrow{P} X\ (n \to \infty)$ 等价于 $P(A_n(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$

- 例: $X_n \sim B(1, 1/n)$ 且相互独立，则有 $X_n \xrightarrow{P} B(1, 0)$

### 依概率收敛与几乎必然收敛

令 $A_n(\varepsilon) = \{|X_n - X| \geq \varepsilon\}$,
则 $X_n \xrightarrow{a.s} X\ (n \to \infty)$ 等价于 $P(\cup_{m=n}^{\infty} A_m(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$
则 $X_n \xrightarrow{P} X\ (n \to \infty)$ 等价于 $P(A_n(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall \varepsilon > 0$

- 例：$X_n \sim B(1, 1/n)$ 且相互独立，则有 $X_n \xrightarrow{P} B(1, 0)$
- 另一方面，对于 $\varepsilon > 0$, 我们有
  $P(A_n(\varepsilon)) = P(X_n = 1) = 1/n$, 因此
  $P(\cup_{m=n}^{\infty} A_m(\varepsilon)) = 1 - \prod_{m=n}^{\infty}(1 - 1/n) \xrightarrow{n\to\infty} 1$，因此
  $X_n$不几乎必然收敛于 $X$

# 依概率收敛与几乎必然收敛



## 依概率收敛与几乎必然收敛

令 $A_n(\varepsilon)=\{|X_n-X|\geq\varepsilon\}$,
则 $X_n \xrightarrow{\text{a.s}} X\ (n\to\infty)$ 等价于 $P(\cup_{m=n}^{\infty}A_m(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall\varepsilon>0$
则 $X_n \xrightarrow{P} X\ (n\to\infty)$ 等价于 $P(A_n(\varepsilon)) \xrightarrow{n\to\infty} 0, \forall\varepsilon>0$

- 例： $X_n\sim B(1,1/n)$ 且相互独立，则有 $X_n \xrightarrow{P} B(1,0)$
- 另一方面，对于 $\varepsilon>0$，我们有
  $P(A_n(\varepsilon))=P(X_n=1)=1/n$，因此
  $P(\cup_{m=n}^{\infty}A_m(\varepsilon))=1-\prod_{m=n}^{\infty}(1-1/n) \xrightarrow{n\to\infty} 1$，因此
  $X_n$ 不几乎必然收敛于 $X$

## 几乎必然收敛 $\to$ 依概率收敛

几乎必然收敛 $\to$ 依概率收敛，反之不成立

**依分布收敛**

考虑随机变量 $X$ 和 $X_1, X_2, \ldots,$ 若对于 $F_X(x)$ 的每个连续点 $x$ 有

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

则称 $\{F_{X_n}(x)\}$ 弱收敛于 $F(x)$, 随机变量序列 $\{X_n\}$ 依分布收敛于 $X$, 并记作 $\lim_{n\to\infty} X_n \xrightarrow{d} X$ 或者 $X_n \xrightarrow{d} X$ $(n \to \infty)$.

## 依分布收敛

考虑随机变量 $X$ 和 $X_1, X_2, \ldots$, 若对于 $F_X(x)$ 的每个连续点 $x$ 有

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

则称 $\{F_{X_n}(x)\}$ 弱收敛于 $F(x)$, 随机变量序列 $\{X_n\}$ 依分布收敛于 $X$, 并记作 $\lim_{n\to\infty} X_n \xrightarrow{d} X$ 或者 $X_n \xrightarrow{d} X$ $(n \to \infty)$.

## 依概率收敛 $\to$ 依分布收敛

依概率收敛 $\to$ 依分布收敛, 反之不成立

# 依分布收敛

**依概率收敛 $\rightarrow$ 依分布收敛**

依概率收敛 $\rightarrow$ 依分布收敛，反之不成立

# 依分布收敛

**依概率收敛 $\rightarrow$ 依分布收敛**

依概率收敛 $\rightarrow$ 依分布收敛，反之不成立

- 关键点在于分布函数 $F(x)$（实数函数）相同不意味着随机变量（样本空间到实数的函数）相同

# 依分布收敛

PEKING UNIVERSITY

## 依概率收敛 $\rightarrow$ 依分布收敛

依概率收敛 $\rightarrow$ 依分布收敛，反之不成立

- 关键点在于分布函数 $F(x)$（实数函数）相同不意味着随机变量（样本空间到实数的函数）相同
- 例：考虑样本空间 $S = \{\omega_1, \omega_2\}$，随机变量 $X(\omega_\pm) = \pm 1$，$Y(\omega_\pm) = -\pm 1$，且 $P_X(X = \pm 1) = P_Y(Y = \pm 1) = 1/2$，则随机变量 $X$ 与 $Y$ 不同，而分布函数 $F_X(x)$ 与 $F_X(x)$ 相同

# 依分布收敛





**依概率收敛 → 依分布收敛**

依概率收敛 → 依分布收敛，反之不成立

- 关键点在于分布函数 $F(x)$（实数函数）相同不意味着随机变量（样本空间到实数的函数）相同
- 例：考虑样本空间 $S=\{\omega_1,\omega_2\}$，随机变量 $X(\omega_\pm)=\pm 1$，$Y(\omega_\pm)=-\pm 1$，且 $P_X(X=\pm 1)=P_Y(Y=\pm 1)=1/2$，则随机变量 $X$ 与 $Y$ 不同，而分布函数 $F_X(x)$ 与 $F_X(x)$ 相同

**(常数分布) 依概率收敛 ≡ 依分布收敛**

对于常数 $c$，依概率收敛 $X_n \xrightarrow{P} X \equiv c\ (n\to\infty)$ 等价于

依分布收敛 $X_n \xrightarrow{d} X \equiv c\ (n\to\infty)$

## 三种收敛方式

考虑样本空间 $S=\{\omega\}$, 随机变量 $X$ 和 $X_1, X_2, \ldots,$ 分布函数 $F_X(x), F_{X_n}(x)$, $A=\{\omega\in S: \lim_{n\to\infty} X_n(\omega)=X(\omega)\}$, $A_n(\varepsilon)=\{\omega\in S: |X_n(\omega)-X(\omega)|\geq\varepsilon\}$

- 几乎必然收敛: $P(\lim_{n\to\infty} X_n = X)=1$
  $\Leftrightarrow P(A)=1 \Leftrightarrow \lim_{n\to\infty} P(\cup_{m=n}^{\infty} A_m(\varepsilon))=0, \forall\varepsilon>0$
- 依概率收敛: $\lim_{n\to\infty} P(|X_n-X|\geq\varepsilon)=1, \forall\varepsilon>0$
  $\Leftrightarrow \lim_{n\to\infty} P(A_n(\varepsilon))=0, \forall\varepsilon>0$
- 依分布收敛: $\lim_{n\to\infty} F_{X_n}(x)=F_X(x)$

几乎必然收敛 $\to$ 依概率收敛 $\xrightarrow{\text{常数分布等价}}$ 依分布收敛

# 总结

PEKING UNIVERSITY

## 三种收敛方式

考虑样本空间 $S=\{\omega\}$，随机变量 $X$ 和 $X_1, X_2, \dots$，分布函数 $F_X(x), F_{X_n}(x)$，$A=\{\omega\in S: \lim_{n\to\infty} X_n(\omega)=X(\omega)\}$，$A_n(\varepsilon)=\{\omega\in S: |X_n(\omega)-X(\omega)|\geq\varepsilon\}$

- 几乎必然收敛：$P(\lim_{n\to\infty} X_n = X)=1$
  $\Leftrightarrow P(A)=1 \Leftrightarrow \lim_{n\to\infty} P(\cup_{m=n}^{\infty} A_m(\varepsilon))=0, \forall\varepsilon>0$
- 依概率收敛：$\lim_{n\to\infty} P(|X_n-X|\geq\varepsilon)=1, \forall\varepsilon>0$
  $\Leftrightarrow \lim_{n\to\infty} P(A_n(\varepsilon))=0, \forall\varepsilon>0$
- 依分布收敛：$\lim_{n\to\infty} F_{X_n}(x)=F_X(x)$

几乎必然收敛 $\rightarrow$ 依概率收敛 $\xrightarrow{\text{常数分布等价}}$ 依分布收敛

- 依概率收敛下的大数定理为弱大数定律

# 总结

PEKING UNIVERSITY



## 三种收敛方式

考虑样本空间 $S=\{\omega\}$, 随机变量 $X$ 和 $X_1, X_2, \ldots$, 分布函数 $F_X(x), F_{X_n}(x)$, $A=\{\omega\in S: \lim_{n\to\infty} X_n(\omega)=X(\omega)\}$, $A_n(\varepsilon)=\{\omega\in S: |X_n(\omega)-X(\omega)|\geq\varepsilon\}$

- 几乎必然收敛: $P(\lim_{n\to\infty} X_n=X)=1$ $\Leftrightarrow P(A)=1\Leftrightarrow \lim_{n\to\infty} P(\cup_{m=n}^{\infty}A_m(\varepsilon))=0, \forall\varepsilon>0$
- 依概率收敛: $\lim_{n\to\infty} P(|X_n-X|\geq\varepsilon)=1, \forall\varepsilon>0$ $\Leftrightarrow \lim_{n\to\infty} P(A_n(\varepsilon))=0, \forall\varepsilon>0$
- 依分布收敛: $\lim_{n\to\infty} F_{X_n}(x)=F_X(x)$

几乎必然收敛 $\rightarrow$ 依概率收敛 $\xrightarrow{\text{常数分布等价}}$ 依分布收敛

- 依概率收敛下的大数定理为弱大数定律
- 几乎必然收敛下的大数定理为强大数定律

## 特征函数

随机变量 $X$ 的特征函数为

$$\psi_X(t) = E(e^{itX}), \quad t \in (-\infty, \infty)$$

## 特征函数

随机变量 $X$ 的特征函数为

$$\psi_X(t) = E(e^{itX}), \quad t \in (-\infty, \infty)$$

- 对于连续型随机变量，特征函数也即是密度函数 $f_X(x)$ 的傅里叶变换 $\psi_X(t) = \int e^{itx} f(x) dx$

## 特征函数

**特征函数**

随机变量 $X$ 的特征函数为

$$\psi_X(t) = E(e^{itX}), \ \ t \in (-\infty, \infty)$$

- 对于连续型随机变量，特征函数也即是密度函数 $f_X(x)$ 的傅里叶变换 $\psi_X(t) = \int e^{itx} f(x) dx$
- 退化分布 $P(X = x) = 1$, $\psi_X(t) = e^{itx}$

## 特征函数

随机变量 $X$ 的特征函数为

$$\psi_X(t) = E(e^{itX}),\ \ t \in (-\infty, \infty)$$

- 对于连续型随机变量，特征函数也即是密度函数 $f_X(x)$ 的傅里叶变换 $\psi_X(t) = \int e^{itx} f(x) dx$
- 退化分布 $P(X = x) = 1$, $\psi_X(t) = e^{itx}$
- 泊松分布 $\pi(\lambda)$ 的特征函数为 $\psi_X(t) = \sum_k e^{itk} e^{-\lambda} \lambda^k / k! = e^{\lambda(e^{it}-1)}$

## 特征函数

随机变量 $X$ 的特征函数为

$$\psi_X(t) = E(e^{itX}),\ \ t \in (-\infty, \infty)$$

- 对于连续型随机变量，特征函数也即是密度函数 $f_X(x)$ 的傅里叶变换 $\psi_X(t) = \int e^{itx} f(x)dx$
- 退化分布 $P(X = x) = 1$, $\psi_X(t) = e^{itx}$
- 泊松分布 $\pi(\lambda)$ 的特征函数为 $\psi_X(t) = \sum_k e^{itk} e^{-\lambda}\lambda^k/k! = e^{\lambda(e^{it}-1)}$
- 正态分布的特征函数为 $\psi_X(t) = \int e^{itx} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} = e^{i\mu t - \sigma^2 t^2/2}$

# 特征函数的性质

- $aX+b$ 的特征函数为
  $\psi_{(aX+b)}(t) = E(e^{it(aX+b)}) = e^{itb}E(e^{iatX}) = e^{itb}\psi_X(at)$

- $aX+b$ 的特征函数为
  $\psi_{(aX+b)}(t) = E(e^{it(aX+b)}) = e^{itb}E(e^{iatX}) = e^{itb}\psi_X(at)$
- 假设 $\{X_n\}$ 相互独立且 $X = \sum X_n$，则
  $\psi_X(t) = E(e^{it\sum X_n}) = \prod E(e^{itX_n}) = \prod \psi_{X_n}(t)$

# 特征函数的性质

- $aX+b$ 的特征函数为
$\psi_{(aX+b)}(t)=E(e^{it(aX+b)})=e^{itb}E(e^{iatX})=e^{itb}\psi_X(at)$
- 假设 $\{X_n\}$ 相互独立且 $X=\sum X_n$，则
$\psi_X(t)=E(e^{it\sum X_n})=\prod E(e^{itX_n})=\prod\psi_{X_n}(t)$
- 若 $X$ 的 $k$ 阶矩存在，则 $E(X^k)=(-i)^k\psi_X^{(k)}(0)$

# 特征函数的性质



- $aX+b$ 的特征函数为
  $\psi_{(aX+b)}(t) = E(e^{it(aX+b)}) = e^{itb}E(e^{iatX}) = e^{itb}\psi_X(at)$
- 假设 $\{X_n\}$ 相互独立且 $X=\sum X_n$，则
  $\psi_X(t) = E(e^{it\sum X_n}) = \prod E(e^{itX_n}) = \prod \psi_{X_n}(t)$
- 若 $X$ 的 $k$ 阶矩存在，则 $E(X^k) = (-i)^k \psi_X^{(k)}(0)$
- 以连续型随机变量为例，$\psi_X^{(k)}(t) = i^k \int x^k e^{itx} f(x)dx$，因此
  $(-i)^k \psi_X^{(k)}(0) = \int x^k f(x)dx = E(X^k)$

# 特征函数的性质



- $aX+b$ 的特征函数为
  $\psi_{(aX+b)}(t) = E(e^{it(aX+b)}) = e^{itb}E(e^{iatX}) = e^{itb}\psi_X(at)$
- 假设 $\{X_n\}$ 相互独立且 $X = \sum X_n$，则
  $\psi_X(t) = E(e^{it\sum X_n}) = \prod E(e^{itX_n}) = \prod \psi_{X_n}(t)$
- 若 $X$ 的 $k$ 阶矩存在，则 $E(X^k) = (-i)^k \psi_X^{(k)}(0)$
- 以连续型随机变量为例，$\psi_X^{(k)}(t) = i^k \int x^k e^{itx} f(x)dx$，因此
  $(-i)^k \psi_X^{(k)}(0) = \int x^k f(x)dx = E(X^k)$
- 例：已知正态分布的特征函数为 $\psi_X(t) = e^{i\mu t - \sigma^2 t^2/2}$，则
  $E(X) = -i\psi_X'(0) = \mu$，$E(X^2) = -\psi_X''(0) = \mu^2 + \sigma^2$

# 唯一性定理

**唯一性定理**

随机变量的分布函数由特征函数唯一决定

# 唯一性定理



**唯一性定理**

随机变量的分布函数由特征函数唯一决定

- 给定相互独立 $X_n \sim N(\mu_n, \sigma_n^2)$，令 $X = \sum X_n$，证明 $X \sim N(\sum \mu_n, \sum \sigma_n^2)$

# 唯一性定理

## 唯一性定理

随机变量的分布函数由特征函数唯一决定

- 给定相互独立 $X_n \sim N(\mu_n, \sigma_n^2)$，令 $X = \sum X_n$，证明 $X \sim N(\sum \mu_n, \sum \sigma_n^2)$
  $\psi_X(t) = \prod \psi_{X_n}(t) = \prod e^{i\mu_n t - \sigma_n^2 t^2/2} = e^{i\sum \mu_n t - \sum \sigma_n^2 t^2/2}$，结论根据独立性定理得证

## 唯一性定理

随机变量的分布函数由特征函数唯一决定

- 给定相互独立 $X_n \sim N(\mu_n, \sigma_n^2)$，令 $X = \sum X_n$，证明 $X \sim N(\sum \mu_n, \sum \sigma_n^2)$
  $\psi_X(t) = \prod \psi_{X_n}(t) = \prod e^{i\mu_n t - \sigma_n^2 t^2/2} = e^{i\sum \mu_n t - \sum \sigma_n^2 t^2/2}$，结论根据独立性定理得证
- 类似的方法可以证明其它分布的可加性

# 连续性定理



## 连续性定理

分布函数 $\{F_{X_n}(x)\}$ 弱收敛于 $F_X(x)$ 的充要条件为相应的特征函数 $\{\psi_{X_n}(t)\}$ 收敛于 $F_X(x)$ 的特征函数 $\psi_X(t)$

## 连续性定理

分布函数 $\{F_{X_n}(x)\}$ 弱收敛于 $F_X(x)$ 的充要条件为相应的特征函数 $\{\psi_{X_n}(t)\}$ 收敛于 $F_X(x)$ 的特征函数 $\psi_X(t)$

- 例： 已知 $X \sim \pi(\lambda)$, $Y = (X-\lambda)/\sqrt{\lambda}$, 证明 $\lim_{\lambda\to\infty} Y \xrightarrow{d} Z \sim N(0,1)$

**连续性定理**

分布函数 $\{F_{X_n}(x)\}$ 弱收敛于 $F_X(x)$ 的充要条件为相应的特征函数 $\{\psi_{X_n}(t)\}$ 收敛于 $F_X(x)$ 的特征函数 $\psi_X(t)$

- 例： 已知 $X\sim\pi(\lambda)$, $Y=(X-\lambda)/\sqrt{\lambda}$, 证明 $\lim_{\lambda\to\infty} Y\xrightarrow{d} Z\sim N(0,1)$

  $X$ 的特征函数为 $\psi_X(t)=e^{\lambda(e^{it}-1)}$，因此 $Y$ 的特征函数为

  $\psi_Y(t)=\psi_X(t/\sqrt{\lambda})e^{-i\sqrt{\lambda}t}=e^{\lambda(e^{it/\sqrt{\lambda}}-1)-i\sqrt{\lambda}t}$

  注意 $e^{it/\sqrt{\lambda}}-1=it/\sqrt{\lambda}-t^2/(2\lambda)+o(1/\lambda)$

  因此 $\lambda(e^{it/\sqrt{\lambda}}-1)-i\sqrt{\lambda}t=-t^2/(2\lambda)+\lambda\cdot o(1/\lambda)$

  也即是 $\lim_{\lambda\to\infty}\psi_Y(t)=e^{\lim_{\lambda\to\infty}\lambda(e^{it/\sqrt{\lambda}}-1)-i\sqrt{\lambda}t}=e^{-t^2/2}$

  因此 $Y$ 的特征函数的极限为正态分布的特征函数，根据唯一性定理，则 $Y$ 依分布收敛于标准正态分布 $N(0,1)$

# 辛钦大数定律证明



**辛钦大数定律**

设随机变量序列 $X_1, X_2, \ldots,$ 满足独立同分布，且存在期望 $\mu = E(X_i) < \infty$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \leq \varepsilon\right) = 1$$

# 辛钦大数定律证明

## 辛钦大数定律

设随机变量序列 $X_1, X_2, \ldots,$ 满足独立同分布，且存在期望 $\mu = E(X_i) < \infty$，则对于任意 $\varepsilon > 0$ 有

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| \le \varepsilon\right) = 1$$

证明：弱大数定律（依概率收敛）$\overset{\text{常数分布}}{\equiv}$ 依分布收敛 $\overset{\text{定义}}{\equiv}$ 分布函数弱收敛 $\overset{\text{连续型定理}}{\equiv}$ 特征函数收敛

因为 $X_k$ 期望存在，因此其特征函数存在二阶导数且

$\psi_{X_k}(t) = \psi_{X_k}(0) + \psi'(0)t + o(t) = 1 + i\mu t + o(t)$

令 $\bar{X}_n = \frac{\sum_{k=1}^{n} X_k}{n}$，则有 $\psi_{\bar{X}_n}(t) = \prod_{i=k}^{n} \psi_{X_k}(t/n) = \left[1 + i\mu\frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n$

对于任意 $t$，则有 $\lim_{n\to\infty} \psi_{\bar{X}_n}(t) = \lim_{n\to\infty} \left[1 + i\mu\frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n = e^{it\mu}$

注意 $e^{it\mu}$ 是退化分布 $X = \mu$ 的特征函数，因此有 $F_{\bar{X}_n}(x)$ 弱收敛于 $F_{X=\mu}(x)$，也即是 $\bar{X}_n$ 依分布收敛于 $X = \mu$，也即是 $\bar{X}_n$ 依概率收敛于 $X = \mu$

# 标准化随机变量的规律





- 大数定律告诉我们 $\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} \frac{\sum_{i=1}^{n} E(X_i)}{n}$ 或者说 $\frac{\sum_{i=1}^{n} X_i - E(X_i)}{n} \xrightarrow{P} 0$

# 标准化随机变量的规律



- 大数定律告诉我们 $\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} \frac{\sum_{i=1}^{n} E(X_i)}{n}$ 或者说 $\frac{\sum_{i=1}^{n} X_i - E(X_i)}{n} \xrightarrow{P} 0$
- 考虑独立同分布的情况，大数定律为 $\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} \mu$ 或者说 $\frac{\sum_{i=1}^{n} X_i - \mu}{n} \xrightarrow{P} 0$，其中 $\mu = E(X_i)$

# 标准化随机变量的规律





- 大数定律告诉我们 $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \frac{\sum_{i=1}^n E(X_i)}{n}$ 或者说 $\frac{\sum_{i=1}^n X_i - E(X_i)}{n} \xrightarrow{P} 0$
- 考虑独立同分布的情况，大数定律为 $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$ 或者说 $\frac{\sum_{i=1}^n X_i - \mu}{n} \xrightarrow{P} 0$，其中 $\mu = E(X_i)$
- 考虑 $\sum_{i=1}^n X_i$ 的标准化，也即是 $\tilde{S}_n = \frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma}$

- 大数定律告诉我们 $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \frac{\sum_{i=1}^n E(X_i)}{n}$ 或者说 $\frac{\sum_{i=1}^n X_i - E(X_i)}{n} \xrightarrow{P} 0$
- 考虑独立同分布的情况，大数定律为 $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$ 或者说 $\frac{\sum_{i=1}^n X_i - \mu}{n} \xrightarrow{P} 0$，其中 $\mu = E(X_i)$
- 考虑 $\sum_{i=1}^n X_i$ 的标准化，也即是 $\tilde{S}_n = \frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma}$
- 对于随机变量 $\tilde{S}_n$，我们有 $E(\tilde{S}_n) = 0$，$D(\tilde{S}_n) = 1$，因此我们不再有 $\tilde{S}_n \xrightarrow{P} 0$

- 大数定律告诉我们 $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \frac{\sum_{i=1}^n E(X_i)}{n}$ 或者说 $\frac{\sum_{i=1}^n X_i - E(X_i)}{n} \xrightarrow{P} 0$
- 考虑独立同分布的情况，大数定律为 $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$ 或者说 $\frac{\sum_{i=1}^n X_i - \mu}{n} \xrightarrow{P} 0$，其中 $\mu = E(X_i)$
- 考虑 $\sum_{i=1}^n X_i$ 的标准化，也即是 $\tilde{S}_n = \frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma}$
- 对于随机变量 $\tilde{S}_n$，我们有 $E(\tilde{S}_n) = 0$，$D(\tilde{S}_n) = 1$，因此我们不再有 $\tilde{S}_n \xrightarrow{P} 0$
- 那么 $\tilde{S}_n$ 随 $n$ 的增加是否仍然服从一定的规律呢？

# De Moivre-Laplace 定理

## De Moivre-Laplace 定理

$X_i \sim B(1,p)$ 独立同分布，$\tilde{S}_n = \frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma}$，其中 $\sigma = \sqrt{pq}$，则有 $\tilde{S}_n \xrightarrow{d} Z \sim N(0,1)$

# De Moivre-Laplace 定理

**De Moivre-Laplace 定理**

$X_i \sim B(1,p)$ 独立同分布, $\tilde{S}_n = \frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma}$, 其中 $\sigma = \sqrt{pq}$, 则有 $\tilde{S}_n \xrightarrow{d} Z \sim N(0,1)$

证明：首先我们注意到 $S_n = \sum_{i=1}^n X_i \sim B(n,p)$.
另一方面，我们知道 $n$ 很大时，我们有 $B(n,p) \approx \pi(\lambda)$, 其中 $\lambda = np$. 因此 $S_n \sim \pi(\lambda)$.
在上面的例题中我们知道，当 $X \sim \pi(\lambda)$, $Y = (X-\lambda)/\sqrt{\lambda}$ 时我们有 $\lim_{\lambda\to\infty} Y \xrightarrow{d} Z \sim N(0,1)$.
注意到 $Y$ 也即是 $X$ 的标准化，因此我们有
$\tilde{S}_n \xrightarrow{d} Z \sim N(0,1)$

**Lindeberg-Lévy 定理**

$X_i$ 独立同分布，$E(X_i) = \mu$，$D(X_i) = \sigma^2$，$\tilde{S}_n = \frac{\sum_{i=1}^n X_i - \mu}{\sqrt{n}\sigma}$，则有 $\tilde{S}_n \xrightarrow{d} Z \sim N(0,1)$

**Lindeberg-Lévy 定理**

$X_i$ 独立同分布，$E(X_i)=\mu$，$D(X_i)=\sigma^2$，$\tilde{S}_n=\frac{\sum_{i=1}^n X_i-\mu}{\sqrt{n}\sigma}$，则有 $\tilde{S}_n \xrightarrow{d} Z\sim N(0,1)$

证明：假设 $X_i-\mu$ 的特征函数为 $\psi_{X_i}(t)$，则有

$$\begin{aligned}\psi_{X_i}(t) &= \psi_{X_i}(0)+\psi'_{X_i}(0)t+\psi''_{X_i}(0)t^2/2+o(t^2)\\ &= 1-\sigma^2t^2/2+o(t^2)\end{aligned}$$

因此 $\tilde{S}_n$ 的特征函数为

$$\left[\psi_{X_i}(t/\sqrt{n}\sigma)\right]^n=\left[1-\frac{t^2}{2n}+o(t^2/n)\right]^n \xrightarrow{n\to\infty} e^{-t^2/2}$$

也即是 $\tilde{S}_n \xrightarrow{d} Z\sim N(0,1)$

# 应用





- 中心极限定理告诉我们，当 $n$ 很大，$X_i$ 为独立同分布时，$E(X_i) = \mu$，$D(X_i) = \sigma^2$，我们有

$$\sum_{i=1}^{n} X_i \approx Z \sim N(n\mu, n\sigma^2)$$

# 应用





- 中心极限定理告诉我们，当 $n$ 很大，$X_i$ 为独立同分布时，$E(X_i)=\mu$，$D(X_i)=\sigma^2$，我们有

$$\sum_{i=1}^{n} X_i \approx Z \sim N(n\mu, n\sigma^2)$$

- 因此对于大数定律的情况，我们有

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx Z \sim N(\mu, \sigma^2/n)$$

# 应用



- 中心极限定理告诉我们，当 $n$ 很大，$X_i$ 为独立同分布时，$E(X_i)=\mu$，$D(X_i)=\sigma^2$，我们有

$$\sum_{i=1}^{n} X_i \approx Z \sim N(n\mu, n\sigma^2)$$

- 因此对于大数定律的情况，我们有

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx Z \sim N(\mu, \sigma^2/n)$$

- 因此 $P(a < \sum_{i=1}^{n} X_i \leq b) \approx \Phi\left(\frac{b-n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{n}\sigma}\right)$

# 应用

- 中心极限定理告诉我们，当 $n$ 很大，$X_i$ 为独立同分布时，$E(X_i)=\mu$，$D(X_i)=\sigma^2$，我们有

$$\sum_{i=1}^{n} X_i \approx Z \sim N(n\mu, n\sigma^2)$$

- 因此对于大数定律的情况，我们有

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx Z \sim N(\mu, \sigma^2/n)$$

- 因此 $P(a<\sum_{i=1}^{n} X_i \le b) \approx \Phi\left(\frac{b-n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{n}\sigma}\right)$
- 特别地，对于 $X_i \sim B(1,p)$，
$P(a<\sum_{i=1}^{n} X_i \le b) \approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{np(1-p)}}\right)$

# 例子



- 在 $n$ 重贝努里试验中，若已知每次试验事件 $A$ 出现的概率为 0.75，试利用中心极限定理，(1) 若 $n=7500$，估计 $A$ 出现的频率在 0.74 至 0.76 之间的概率近似值；(2) 估计 $n$，使 $A$ 出现的频率在 0.74 至 0.76 之间的概率不小于 0.90

# 例子

PEKING UNIVERSITY

- 在 $n$ 重贝努里试验中，若已知每次试验事件 $A$ 出现的概率为 0.75，试利用中心极限定理，(1) 若 $n=7500$，估计 $A$ 出现的频率在 0.74 至 0.76 之间的概率近似值；(2) 估计 $n$，使 $A$ 出现的频率在 0.74 至 0.76 之间的概率不小于 0.90

(1) 记 $A$ 出现的次数为 $X$，则 $E(X)=np=0.75n$，
$D(X)=npq=0.1875n$，因此 $P(0.74n<X\le 0.76n)\approx$
$\Phi\left(\frac{(0.76-0.75)n}{\sqrt{0.1875n}}\right)-\Phi\left(\frac{(0.74-0.75)n}{\sqrt{0.1875n}}\right)\approx 0.9544$
(2) $\Phi\left(\frac{(0.76-0.75)n}{\sqrt{0.1875n}}\right)-\Phi\left(\frac{(0.74-0.75)n}{\sqrt{0.1875n}}\right)\ge 0.95$，得 $n\ge 5074$ 注意和利用切比雪夫不等式的结果比较

- 在 $n$ 重贝努里试验中，若已知每次试验事件 $A$ 出现的概率为 0.75，试利用中心极限定理，(1) 若 $n=7500$，估计 $A$ 出现的频率在 0.74 至 0.76 之间的概率近似值；（2）估计 $n$，使 $A$ 出现的频率在 0.74 至 0.76 之间的概率不小于 0.90

  (1) 记 $A$ 出现的次数为 $X$，则 $E(X)=np=0.75n$，$D(X)=npq=0.1875n$，因此 $P(0.74n<X\leq 0.76n)\approx \Phi\left(\frac{(0.76-0.75)n}{\sqrt{0.1875n}}\right)-\Phi\left(\frac{(0.74-0.75)n}{\sqrt{0.1875n}}\right)\approx 0.9544$

  (2) $\Phi\left(\frac{(0.76-0.75)n}{\sqrt{0.1875n}}\right)-\Phi\left(\frac{(0.74-0.75)n}{\sqrt{0.1875n}}\right)\geq 0.95$，得 $n\geq 5074$ 注意和利用切比雪夫不等式的结果比较

- 设随机变量 $X_1,\cdots,X_{20}$，相互独立同分布，$X_i\sim U(-1,1)$，分别求 (1) $\frac{1}{20}\sum_{k=1}^{20}X_k$，(2) $\frac{1}{20}\sum_{k=1}^{20}|X_k|$ (3) $\frac{1}{20}\sum_{k=1}^{20}X_k^2$ 的近似分布

- 在 $n$ 重贝努里试验中，若已知每次试验事件 $A$ 出现的概率为 0.75，试利用中心极限定理，(1) 若 $n=7500$，估计 $A$ 出现的频率在 0.74 至 0.76 之间的概率近似值；（2）估计 $n$，使 $A$ 出现的频率在 0.74 至 0.76 之间的概率不小于 0.90

  (1) 记 $A$ 出现的次数为 $X$，则 $E(X)=np=0.75n$，$D(X)=npq=0.1875n$，因此 $P(0.74n<X\leq 0.76n)\approx \Phi\left(\frac{(0.76-0.75)n}{\sqrt{0.1875n}}\right)-\Phi\left(\frac{(0.74-0.75)n}{\sqrt{0.1875n}}\right)\approx 0.9544$

  (2) $\Phi\left(\frac{(0.76-0.75)n}{\sqrt{0.1875n}}\right)-\Phi\left(\frac{(0.74-0.75)n}{\sqrt{0.1875n}}\right)\geq 0.95$，得 $n\geq 5074$ 注意和利用切比雪夫不等式的结果比较

- 设随机变量 $X_1,\cdots,X_{20}$，相互独立同分布，$X_i\sim U(-1,1)$，分别求 (1) $\frac{1}{20}\sum_{k=1}^{20}X_k$，(2) $\frac{1}{20}\sum_{k=1}^{20}|X_k|$ (3) $\frac{1}{20}\sum_{k=1}^{20}X_k^2$ 的近似分布

  三个分布均近似服从正态分布

  (1) $E(X)=0, D(X)=1/3$，$\frac{1}{200}\sum_{k=1}^{20}X_k\approx Z\sim N(0,1/60)$

  (2) $E(|X|)=1/2, D(|X|)=1/12$，$\frac{1}{200}\sum_{k=1}^{20}|X_k|\approx Z\sim N(1/2,1/240)$

  (3) $E(X^2)=1/3, D(X)=4/45$，$\frac{1}{200}\sum_{k=1}^{20}X_k^2\approx Z\sim N(1/3,1/225)$

# 例子



- 设某工厂有 400 台同类机器，各台机器发生故障的概率都是 0.01，各台机器工作是相互独立的，试用三种方法求机器出故障的台数不小于 2 的概率。

- 设某工厂有 400 台同类机器，各台机器发生故障的概率都是 0.01，各台机器工作是相互独立的，试用三种方法求机器出故障的台数不小于 2 的概率。

记 $X$ 为故障机器数量

(1) 根据二项分布有 $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0.99^{400} - 400 \times 0.01 \times 0.99^{399} \approx 0.9095$

(2) 用泊松分布近似计算 $\lambda = np = 400 \times 0.01 = 4$,
$P(X \geq 2) = 1 - e^{-4} - 4e^{-4} \approx 0.9084$

(3) 用正态分布近似计算
$\sqrt{npq} = \sqrt{400 \times 0.01 \times 0.99} = 1.99$
$P(X \geq 2) = 1 - P(X \leq 1) \approx 1 - \Phi\left(\frac{1-np}{\sqrt{npq}}\right) = \Phi\left(\frac{3}{1.99}\right) = 0.9341$

- 设某工厂有 400 台同类机器，各台机器发生故障的概率都是 0.02，各台机器工作是相互独立的，试用三种方法求机器出故障的台数不小于 2 的概率。

# 例子





- 设某工厂有 400 台同类机器，各台机器发生故障的概率都是 0.02，各台机器工作是相互独立的，试用三种方法求机器出故障的台数不小于 2 的概率。

记 $X$ 为故障机器数量

(1) 根据二项分布有 $P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - 0.98^{400} - 400 \times 0.02 \times 0.98^{399} \approx 0.9972$

(2) 用泊松分布近似计算 $\lambda = np = 400 \times 0.02 = 8$,
$P(X \geq 2) = 1 - e^{-8} - 8e^{-8} \approx 0.9969$

(3) 用正态分布近似计算
$\sqrt{npq} = \sqrt{400 \times 0.02 \times 0.98} = 2.8$
$P(X \geq 2) = 1 - P(X \leq 1) \approx 1 - \Phi\left(\frac{1-np}{\sqrt{npq}}\right) = \Phi\left(\frac{7}{2.8}\right) = 0.9938$

# 例子





- 某校 1500 名学生选修"概率统计"课程，共有 10 名教师主讲此课，假定每位学生可以随意地选择一位教师 (即，选择任意一位教师的可能性均为 1/10)，而且学生之间的选择是相互独立的．问：每位教师的上课教室应该设有多少座位才能保证该班因没有座位而使学生离开的概率小于 5%．

# 例子

- 某校 1500 名学生选修“概率统计”课程，共有 10 名教师主讲此课，假定每位学生可以随意地选择一位教师 (即，选择任意一位教师的可能性均为 1/10)，而且学生之间的选择是相互独立的. 问：每位教师的上课教室应该设有多少座位才能保证该班因没有座位而使学生离开的概率小于 5%.

考虑任意一位老师甲，且记

$$X_i = \begin{cases} 1, & \text{第}\, i\, \text{个学生选择教师甲} \\ 0, & \text{其它} \end{cases},\quad X_i \sim B(1, 1/10)$$ 且独立同分布. 记 $Y = \sum_{i=1}^{1500} X_i$，则 $Y \sim B(1500, 1/10)$.

设教室有 $a$ 个座位，因此

$$P(Y \le a) \approx \Phi\left(\frac{a - 1500 * 1/10}{\sqrt{1500 * (1/10) * (9/10)}}\right)$$ 注意

$\Phi(1.645) = 95\%$，因此有 $a \ge 169.11$