# 1 公式

$\sum_{k=1}^{n} k^3 = \frac{1}{4}n^2(n+1)^2$
$\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}, r \neq 1$
$\sum_{k=1}^{n} kr^k = r\frac{1-(n+1)r^n+nr^{n+1}}{(1-r)^2}, r \neq 1$
$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$
$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots$
$\ln x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n = (x-1) - \frac{(x-1)^2}{2} - \cdots$
$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \cdots$
$\frac{\mathrm{d}}{\mathrm{d}x} \tan x = \sec^2 x$
$\frac{\mathrm{d}}{\mathrm{d}x} \cot x = -\csc^2 x$
$\frac{\mathrm{d}}{\mathrm{d}x} \sec x = \sec x \tan x$
$\frac{\mathrm{d}}{\mathrm{d}x} \csc x = -\csc x \cot x$
$\frac{\mathrm{d}}{\mathrm{d}x} \arcsin x = \frac{1}{\sqrt{1-x^2}}$
$\frac{\mathrm{d}}{\mathrm{d}x} \arccos x = -\frac{1}{\sqrt{1-x^2}}$
$\frac{\mathrm{d}}{\mathrm{d}x} \arctan x = \frac{1}{1+x^2}$
$\int \frac{1}{x^2+\alpha^2}\mathrm{d}x = \frac{\arctan \frac{x}{\alpha}}{\alpha}$
$\int \frac{1}{\pm x^2 \mp \alpha^2}\mathrm{d}x = \frac{\ln\left(\frac{x\mp\alpha}{\pm x+\alpha}\right)}{2\alpha}$
$\int \frac{1}{ax^2+b}\mathrm{d}x = \frac{1}{\sqrt{ab}} \arctan \frac{\sqrt{a}x}{\sqrt{b}}$
$\int \frac{1}{\sqrt{a^2-x^2}}\mathrm{d}x = \arcsin \frac{x}{a} = -\arccos \frac{x}{a}$
$\int \sec^2 x \mathrm{d}x = \tan x$
$\int \csc^2 x \mathrm{d}x = -\cot x$
$\int \sec x \tan x \mathrm{d}x = \sec x$
$\int \csc x \cot x \mathrm{d}x = -\csc x$
$\int \tan x \mathrm{d}x = -\ln|\cos x| = \ln|\sec x|$
$\int \cot x \mathrm{d}x = \ln|\sin x|$
$\int \sec x \mathrm{d}x = \ln|\sec x + \tan x|$
$\int \csc x \mathrm{d}x = \ln|\csc x - \cot x| = \ln\left|\frac{\tan x - \sin x}{\sin x \tan x}\right|$
$\int x^n e^{ax}\mathrm{d}x = \frac{1}{a}x^n e^{ax} - \frac{n}{a}\int x^{n-1}e^{ax}\mathrm{d}x$
$\int_{-\infty}^{\infty} e^{-ax^2+bx+c}\,\mathrm{d}x = \sqrt{\frac{\pi}{a}}\,e^{\frac{b^2}{4a}+c}$

# 2 Probability

## 2.1 Distribution

### 2.1.1 Poisson Distribution

本质上是一个 $n \to \infty$ 的二项分布，$\lambda = np$。
性质：$\mathbb{E}(X) = \lambda, \mathrm{Var}(X) = \lambda$
Approximate Bin: $n$ large, $p$ small $(n \geq 50, np \leq 5)$

### 2.1.2 Hypergeometric Distribution

记号：$X \sim \text{Hypergeomet}(n, N, m)$
概率：$p(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$
$N$ 个球，$m$ 个红球，不放回取出 $n$ 个，有 $k$ 红球。
$\mathbb{E}(X) = n \cdot \frac{m}{N}, \mathrm{Var}(X) = n \cdot \frac{m}{N}\left(1 - \frac{m}{N}\right)\left(1 - \frac{n-1}{N-1}\right)$

### 2.1.3 Normal Distribution

Approximate Bin: $np(1-p) \geq 10$
$Z \sim N(0,1), \mathbb{E}(g'(Z)) = \mathbb{E}(Zg(Z))$, assuming that $\lim_{x\to\infty} \frac{g(x)}{e^{\frac{x^2}{2}}} = 0$. So $\mathbb{E}(Z^{n+1}) = n\mathbb{E}(Z^{n-1})$.

### 2.1.4 Exponential Distribution

CDF：$F(X) = 1 - e^{-\lambda x}, x \geq 0$
$\mathbb{P}r(X > x) = e^{-\lambda x}, x \geq 0$
$\mathbb{E}(X^n) = \frac{n}{\lambda}\mathbb{E}(X^{n-1}) = \frac{n!}{\lambda^n}$
$\mathbb{E}(X) = \frac{1}{\lambda}, \mathrm{Var}(X) = \frac{1}{\lambda^2}$

### 2.1.5 Gamma Distribution

考试都用 $\Gamma(\alpha, \beta)$ 的形式
$\Gamma(x) = \int_0^{\infty} u^{x-1}e^{-u}\mathrm{d}u, x > 0$
$\Gamma(x) = (x-1)\Gamma(x-1)$
$\Gamma(n) = (n-1)!$
PDF：$f(x) = \frac{\lambda^\alpha}{\Gamma(a)}x^{\alpha-1}e^{-\lambda x}, x > 0$
$\alpha$：发生次数
$\mathbb{E}(X) = \frac{\alpha}{\lambda}, \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$
$\mathbb{E}(X^n) = \frac{n+\alpha-1}{\lambda}\mathbb{E}(X^{n-1}) = \frac{\alpha^{\overline{n}}}{\lambda^n}$

### 2.1.6 Chi-Squared Distribution

$X \sim \chi^2(k)$
$\mathbb{E}(X) = k$
$\mathcal{N}(0,1)^2 \sim \chi_1^2$
$\chi_n^2 \sim \Gamma\left(\frac{n}{2}, 2\right)$

### 2.1.7 $t$-Distribution

$T_k = \frac{C}{\sqrt{D/k}}, C \sim \mathcal{N}(0,1), D = \chi_k^2$
$\frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1}$ for $X_i \sim \mathcal{N}(\mu, \sigma^2)$
$\sigma^2 = \frac{v}{v-2}$ for $v > 2$, $\infty$ for $1 < v \leq 2$
$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$

### 2.1.8 $F$-Distribution

$F(m,n) = \frac{U/m}{V/n}, U \sim \chi_m^2, V \sim \chi_n^2$
$f(w) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \cdot \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot w^{\frac{m}{2}-1} \cdot \left(1 + \frac{m}{n}w\right)^{-\frac{m+n}{2}}$

## 2.2 MGF

$M(X) = \mathbb{E}\left[e^{tX}\right]$
$M^{(m)}(X) = \mathbb{E}\left[X^m\right]$

| **Distribution** | **MGF** | **PMF/PDF** |
|---|---|---|
| Bernoulli$(p)$ | $pe^t + 1 - p$ | $p(1) = p$ |
| Binomial$(n,p)$ | $(1 - p + pe^t)^n$ | $\binom{n}{k}p^k(1-p)^{n-k}$ |
| Poisson$(\lambda)$ | $e^{\lambda(e^t-1)}$ | $p(k) = \frac{\lambda^k}{k!}e^{-\lambda}$ |
| Geo$(p)$ | $\frac{pe^t}{1-(1-p)e^t}$ | $(1-p)^{k-1}p$ |
| $\mathcal{N}(\mu, \sigma^2)$ | $e^{t\mu+\frac{1}{2}\sigma^2 t^2}$ | $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ |
| Exp$(\lambda)$ | $\frac{\lambda}{\lambda-t}$ | $\lambda e^{-\lambda x}$ |
| $\Gamma(\alpha, \beta)$ | $(1-\beta t)^{-\alpha}$ | $\frac{1}{\beta^k\Gamma(k)}x^{k-1}e^{-\frac{x}{\beta}}$ |
| $\chi_k^2$ | $(1-2t)^{-\frac{k}{2}}$ | $\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}x^{\frac{k}{2}-1}e^{-\frac{x}{2}}$ |

## 2.3 Central Limit Theorem

### 2.3.1 Inequalities

Markov's inequality: $\mathbb{P}\mathrm{r}\{X \geq t\} \leq \frac{\mathbb{E}(X)}{t}$，要求是 $X \geq 0, t > 0$

Chebyshev's Inequality：$\mathbb{P}\mathrm{r}\{|X - \mathbb{E}(X)| \geq t\} \leq \frac{\mathrm{Var}(X)}{t^2}$

$\mathbb{P}\mathrm{r}\{|X - \mathbb{E}(X)| \geq k\sigma\} \leq \frac{1}{k^2}$

### 2.3.2 Law of Large Numbers

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
Weak: $\lim_{n\to\infty} \mathbb{P}\mathrm{r}\left\{\left|\overline{X}_n - \mu\right| > \epsilon\right\} = 0$
Strong: $\mathbb{P}\mathrm{r}\left\{\lim_{n\to\infty} \overline{X}_n = \mu\right\} = 1$

### 2.3.3 CLT

$S_n = \sum_{i=1}^{n} X_i$
$\lim_{n\to\infty} \mathbb{P}\mathrm{r}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right\} = \Phi(x)$

# 3 Estimation

## 3.1 Maximum Likelihood (MLE)

$L(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta})$
Standard conditions:

1. $L(\theta) > 0$ for all $\theta \in (a, b)$
2. $\frac{\partial L(\theta)}{\partial \theta}$ exists for all $\theta \in (a, b)$
3. $\lim_{\theta\to a^+} L(\theta) = \lim_{\theta\to b^-} L(\theta) = 0$

$\mathrm{Bias}(\theta) = \mathbb{E}_\theta(\theta) - \theta$