

1 公式

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$$\sum_{k=1}^n k^3 = \frac{1}{4}n^2(n+1)^2$$

$$\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r}, r \neq 1$$

$$\sum_{k=1}^n kr^k = r \frac{1-(n+1)r^n + nr^{n+1}}{(1-r)^2}, r \neq 1$$

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots$$

$$\ln x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} (x-1)^n = (x-1) - \frac{(x-1)^2}{2} - \dots$$

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \dots$$

$$\frac{d}{dx} \tan x = \sec^2 x$$

$$\frac{d}{dx} \cot x = -\csc^2 x$$

$$\frac{d}{dx} \sec x = \sec x \tan x$$

$$\frac{d}{dx} \csc x = -\csc x \cot x$$

$$\frac{d}{dx} \arcsin x = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx} \arccos x = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx} \arctan x = \frac{1}{1+x^2}$$

$$\int \frac{1}{x^2+\alpha^2} dx = \frac{\arctan \frac{x}{\alpha}}{\alpha}$$

$$\int \frac{1}{\pm x^2 \mp \alpha^2} dx = \frac{\ln \left(\frac{x \mp \alpha}{x \pm \alpha} \right)}{2\alpha}$$

$$\int \frac{1}{ax^2+b} dx = \frac{1}{\sqrt{ab}} \arctan \frac{\sqrt{ax}}{\sqrt{b}}$$

$$\int \frac{1}{\sqrt{a^2-x^2}} dx = \arcsin \frac{x}{a} = -\arccos \frac{x}{a}$$

$$\int \sec^2 x dx = \tan x$$

$$\int \csc^2 x dx = -\cot x$$

$$\int \sec x \tan x dx = \sec x$$

$$\int \csc x \cot x dx = -\csc x$$

$$\int \tan x dx = -\ln |\cos x| = \ln |\sec x|$$

$$\int \cot x dx = \ln |\sin x|$$

$$\int \sec x dx = \ln |\sec x + \tan x|$$

$$\int \csc x dx = \ln |\csc x - \cot x| = \ln \left| \frac{\tan x - \sin x}{\sin x \tan x} \right|$$

$$\int x^n e^{ax} dx = \frac{1}{a} x^n e^{ax} - \frac{n}{a} \int x^{n-1} e^{ax} dx$$

$$\int_{-\infty}^{\infty} e^{-ax^2+bx+c} dx = \sqrt{\frac{\pi}{a}} e^{\frac{b^2}{4a}+c}$$

2 Probability

2.1 Distribution

Poisson Distribution

本质上是一个 $n \rightarrow \infty$ 的二项分布, $\lambda = np$ 。

性质: $\mathbb{E}(X) = \lambda, \text{Var}(X) = \lambda$

Approximate Bin: n large, p small ($n \geq 50, np \leq 5$)

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。

$\mathbb{E}(X) = n \cdot \frac{m}{N}, \text{Var}(X) = n \cdot \frac{m}{N} \left(1 - \frac{m}{N}\right) \left(1 - \frac{n-1}{N-1}\right)$

Normal Distribution

Approximate Bin: $np(1-p) \geq 10$

$Z \sim N(0, 1), \mathbb{E}(g'(Z)) = \mathbb{E}(Zg(Z))$, assuming that

$\lim_{x \rightarrow \infty} \frac{g(x)}{e^{\frac{x^2}{2}}} = 0$. So $\mathbb{E}(Z^{n+1}) = n\mathbb{E}(Z^{n-1})$.

Exponential Distribution

CDF: $F(X) = 1 - e^{-\lambda x}, x \geq 0$

$\Pr(X > x) = e^{-\lambda x}, x \geq 0$

$\mathbb{E}(X^n) = \frac{n!}{\lambda^n} \mathbb{E}(X^{n-1}) = \frac{n!}{\lambda^n}$

$\mathbb{E}(X) = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$

Gamma Distribution

考试都用 $\Gamma(\alpha, \beta)$ 的形式

$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u} du, x > 0$

$\Gamma(x) = (x-1)\Gamma(x-1)$

$\Gamma(n) = (n-1)!$

α : 发生次数

$\mathbb{E}(X) = \alpha\beta, \text{Var}(X) = \alpha\beta^2$

$\mathbb{E}(X^n) = (n + \alpha - 1)\beta \cdot \mathbb{E}(X^{n-1}) = \alpha^n \beta^n$

Chi-Squared Distribution

$X \sim \chi^2(k)$

$\mathbb{E}(X) = k$

$\mathcal{N}(0, 1)^2 \sim \chi_1^2$

$\chi_n^2 \sim \Gamma\left(\frac{n}{2}, 2\right)$

$\frac{1}{\sigma^2} \sum (X_i - \mu)^2 \sim \chi_n^2$

$\frac{1}{\sigma^2} \sum (X_i - \bar{X})^2 \sim \chi_{n-1}^2$

t-Distribution

$$T_k = \frac{C}{\sqrt{D/k}}, C \sim \mathcal{N}(0, 1), D = \chi_k^2$$

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1} \text{ for } X_i \sim \mathcal{N}(\mu, \sigma^2)$$

$$\sigma^2 = \frac{v}{v-2} \text{ for } v > 2, \infty \text{ for } 1 < v \leq 2$$

$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

F-Distribution

$$F(m, n) = \frac{U/m}{V/n}, U \sim \chi_m^2, V \sim \chi_n^2$$

$$f(w) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot w^{\frac{m}{2}-1} \cdot \left(1 + \frac{m}{n}w\right)^{-\frac{m+n}{2}}$$

$$\mu = \frac{n}{n-2} \text{ for } n > 2$$

$$\sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ for } n > 4$$

2.2 MGF

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

$$M^{(m)}(X) = \mathbb{E}[X^m]$$

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(λ)	$e^{\lambda(e^t-1)}$	$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$
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(λ)	$\frac{\lambda}{\lambda-t}$	$\lambda e^{-\lambda x}$
$\Gamma(\alpha, \beta)$	$(1 - \beta t)^{-\alpha}$	$\frac{1}{\beta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\beta}}$
χ_k^2	$(1 - 2t)^{-\frac{k}{2}}$	$\frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$

2.3 Central Limit Theorem

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

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

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

Weak LLN: $\lim_{n \rightarrow \infty} \Pr\{|\bar{X}_n - \mu| > \epsilon\} = 0$

Strong LLN: $\Pr\{\lim_{n \rightarrow \infty} \bar{X}_n = \mu\} = 1$

CLT: $\lim_{n \rightarrow \infty} \Pr\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right\} = \Phi(x)$

3 Estimation

3.1 Method of Moments (MME)

$$\begin{aligned}\mathbb{E}(X_1^j) &= \mu_j \\ \mu_j &= g_j(\boldsymbol{\theta}) \\ \theta_k &= h_k(\boldsymbol{\mu})\end{aligned}$$

3.2 Maximum Likelihood (MLE)

$$L(\mathbf{x} | \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i | \boldsymbol{\theta})$$

Standard conditions:

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

3.3 Estimate an Estimator

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

$$\text{unbiased: } \text{Bias}(\hat{\theta}) = 0$$

$$\text{Standard Error: } \text{SE}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$$

Rule of thumb: 如果 sample 足够大, $\theta \in [\hat{\theta} - \text{SE}(\hat{\theta}), \hat{\theta} + \text{SE}(\hat{\theta})]$ 是 70%, $\theta \in [\hat{\theta} - 2 \cdot \text{SE}(\hat{\theta}), \hat{\theta} + 2 \cdot \text{SE}(\hat{\theta})]$ 是 95%

$$\text{Mean Squared Error: } \text{MSE}(\hat{\theta}) = \mathbb{E} \left[(\hat{\theta} - \theta)^2 \right]$$

$$\text{MSE}(\hat{\theta}) = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) = \text{Bias}(\hat{\theta})^2 + \text{SE}(\hat{\theta})^2$$

Bias 是要准, SE 是要快, MSE 是成年人我都要

如何走向人生巅峰: $\hat{\theta}'_n = \frac{\theta}{\mathbb{E}(\hat{\theta}_n)} \hat{\theta}_n$

Consistent: $\forall \varepsilon > 0 \Rightarrow \lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) = 0$

咋证: $\lim_{n \rightarrow \infty} \text{Bias}(\hat{\theta}_n) = 0 \wedge \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$

MME 只要是 h 连续一定是 consistent 嘟

3.4 Fisher Information

Log-likelihood function: $\ell = \log L$

Score function: $V(\mathbf{X} | \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{X} | \boldsymbol{\theta})$

Fisher Information: $I_{\mathbf{X}}(\boldsymbol{\theta}) = \mathbb{E} [V(\mathbf{X} | \boldsymbol{\theta})^2]$

Condition (*) (离散同理, 换成 \sum):

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} = \frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\infty}^{\infty} f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} = 0$$

满足 (*) 这个可以推: $\mathbb{E}(V) = 0 \wedge \text{Var}(V) = I$

Fisher Info Alternative Formula

$$I_{\mathbf{X}}(\boldsymbol{\theta}) = n I_{X_1}(\boldsymbol{\theta}) = -n \cdot \mathbb{E} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log L(\mathbf{X}_1 | \boldsymbol{\theta}) \right]$$

证明是首先需要注意到

$$\frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x} | \boldsymbol{\theta}) = \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{x} | \boldsymbol{\theta}) \right] f(\mathbf{x} | \boldsymbol{\theta})$$

然后把 $0 = \frac{\partial}{\partial \boldsymbol{\theta}} \int f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} = \int \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{x} | \boldsymbol{\theta}) \right] f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x}$ 两边再求个偏导

3.5 Cramér-Rao Lower Bound (CRLB)

$$\text{条件: } \frac{\partial}{\partial \boldsymbol{\theta}} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(\mathbf{x}) f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x}_1 \cdots d\mathbf{x}_n \right] =$$

$$\left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(\mathbf{x}) \frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x}_1 \cdots d\mathbf{x}_n \right], \text{ For all } h$$

with $\mathbb{E}(|h(\mathbf{X})|) < \infty$

$$\text{Var}(\hat{\theta}) \geq \frac{\left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbb{E}(T) \right]^2}{\mathbb{E} \left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{X} | \boldsymbol{\theta}) \right)^2 \right]}$$

如果 $\hat{\theta}$ 是 unbiased 的, 那我们有 $\text{Var}(\hat{\theta}) \geq \frac{1}{n I_{X_1}(\boldsymbol{\theta})}$

efficient 1. unbiased, 2. $\text{Var}(\hat{\theta}) = \frac{1}{n I_{X_1}(\boldsymbol{\theta})}$

3.6 Asymptotic Normality

Regularity Conditions ($\hat{\theta}$ 是 MLE)

1. $\frac{\partial^3}{\partial \boldsymbol{\theta}^3} f(\mathbf{x} | \boldsymbol{\theta})$ exists and continuous
2. $\exists (a, b) \subseteq S, \theta_0 \in (a, b)$
3. (support) $x \in \mathbb{R} : f(x | \boldsymbol{\theta}) > 0$ is the same $\forall \boldsymbol{\theta}$

不满足: Bernoulli(p): 找不到区间; $U(0, b)$: 不 support
 $\sqrt{n I_{X_1}(\theta_0)}(\hat{\theta} - \theta_0) \sim \mathcal{N}(0, 1)$

3.7 Confidence Intervals

upper percentage point: $z_{\alpha}, \mathbb{P}(X > z_{\alpha}) = \alpha$

$\forall \theta_0 \in S, \mathbb{P}(L \leq \theta_0 \leq U) = 1 - \alpha$

exact $100(1 - \alpha)\%$ confidence interval for θ

$1 - \alpha$: **confidence level**

需要注意的是 L 和 R 才是尊 Random Var

Pivotal Quantity: The **distribution** of $Q(\mathbf{X}, \boldsymbol{\theta})$

does not depend on any unknown parameter

如果 $X_i \sim N(\mu, 1)$, \bar{X} 是不可以的 (因为 μ 不知道), 但是 $\sqrt{n}(\bar{X} - \mu)$ 是可以的

How to prove: PDF/CDF/MGF 都可以

asymptotically pivotal quantity: $Q_{n \rightarrow \infty} \rightarrow \Psi$

approximate / large sample confidence intervals

$$\left[\hat{\theta} - \frac{z_{\alpha/2}}{\sqrt{n I(\hat{\theta})}}, \hat{\theta} + \frac{z_{\alpha/2}}{\sqrt{n I(\hat{\theta})}} \right] \text{ to approx } 100(1 - \alpha)\%$$

4 Hypothesis Testing

如果 H_0 发生了, 那 \mathbf{X} 发生的概率有多小

p-value $p = \mathbb{P}(T(\mathbf{X}) \geq s | H_0)$

significance level: α , critical value: $t(\alpha)$

$p < \alpha \Leftrightarrow s > t(\alpha) \Leftrightarrow \text{reject } H_0$

算 significance level:

T 越大越 against H_0 : $\max_s (\mathbb{P}(T(\mathbf{X}) \geq s | H_0) \geq \alpha)$

T 越小越 against H_0 : $\min_s (\mathbb{P}(T(\mathbf{X}) \leq s | H_0) \geq \alpha)$

Neyman-Pearson tests

Parameter Space: $\Omega = \Omega_0 \cup \Omega_1$

$H_0 : \theta \in \Omega_0, H_1 : \theta \in \Omega_1$

如果 $|\Omega_0| = 1$ 叫 simple hypothesis, 否则 composite
我其实想要的是 H_1 :

1. rejecting the H_0 in favor of the H_1 .

2. there is not enough evidence to support the H_1 .

Type I error (α): H_0 本不该被 reject, 却 reject 了, FP

Type I error (β): H_0 该被 rej, 却没 rej, FN

rejection region $R \subseteq \mathbb{R}^n$: 样本在这里就选 H_1

size α 所有 H_0 成立的情况下, rej 的最大可能性

$\sup_{\theta \in \Omega_0} \mathbb{P}(H_0 \text{ is rejected} | \boldsymbol{\theta}) = \alpha$

level H_0 成立, rej 的可能性小于等于他 $\sup \leq \alpha$

power H_1 成立, 有多大可能性拒绝 H_0

$\text{Power}(\boldsymbol{\theta}) = \mathbb{P}(H_0 \text{ rejected} | \boldsymbol{\theta} \in \Omega_1) = 1 - \beta(\boldsymbol{\theta})$

Neyman-Pearson Lemma

simple vs simple

$$\text{Likelihood ratio } \Lambda(\mathbf{x}) = \frac{L(\mathbf{x} | \theta_0)}{L(\mathbf{x} | \theta_1)}$$

rejection region $R = \{\mathbf{x} \in \mathbb{R}^n : \Lambda \leq t\}$

Monotone Likelihood Ratio (MLR)

对于 $\theta < \theta'$, 存在 $T(\mathbf{x})$ 使得 $\frac{p_{\theta'}(\mathbf{x})}{p_{\theta}(\mathbf{x})}$ 对 $T(\mathbf{x})$ 是不降的

simple vs composite

uniformly most powerful (UMP)

满足 MLR, UMP test 存在, sup 在 $\theta = \theta_0$ 时候取到

如果 (L, U) 是一个 $100(1 - \alpha)\%$ 的 confidence interval,

那么 reject $H_0 : 0 = \theta_0 \Leftrightarrow \theta_0 \notin (L, U)$ 是 α size 的

反过来也可以说