基础数理统计 2023 Spring

Discrete dist.	pmf	mean	variance	mgf/moment
Discrete Uniform(n)	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{e^t(1-e^{nt})}{n(1-e^t)} = \frac{1}{n}\sum e^{it}$
Bernoulli(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$(1-p)+pe^t$
Binomial(n,p)	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$((1-p)+pe^t)^n$
Geometric(p)	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Poisson(λ)	$\frac{\lambda^x}{x!}e^{-\lambda}$	λ	λ	$e^{\lambda(e^t-1)}$
Beta-binomial (n,α,β)	$\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\cdot \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$	$\frac{n\alpha}{\alpha+\beta}$	$\frac{n\alpha\beta}{(\alpha+\beta)^2}$	If $X\|P$ is Binomial(n,P), and P is Beta(α,β), then X is Beta-binomial(n,α,β).

Continuous dist.	pdf	mean	variance	mgf/moment
Uniform(a,b)	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Exponential(θ)	$\frac{1}{\theta}e^{-\frac{x}{\theta}}$	θ	θ^2	$\frac{1}{1-\theta t}, t<\frac{1}{\theta}$
Exponential(λ)	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}, t<\lambda$
Normal(μ,σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t+\frac{\sigma^2t^2}{2}}$
Gamma(α,β)	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta t}\right)^\alpha, t<\frac{1}{\beta}$
Beta(α,β)	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$1+\sum_{k=1}^\infty\left(\prod_{i=0}^{k-1}\frac{\alpha+i}{\alpha+\beta+i}\right)\frac{t^k}{k!}$
Cauchy	$1/\left(\pi(1+x^2)\right), x\in\mathbb{R}$	(∞)	n/a, $F_X(x)$	$=\arctan(x)/\pi+1/2$
$\chi_p^2=\sum_{i=1}^p Z_i^2$	$\frac{1}{2^{p/2}\Gamma(p/2)}x^{p/2-1}e^{-x/2}$	p	$2p$	$(1-2t)^{-p/2}, t<1/2$

$(uv)'=uv'+u'v$ $\quad(f(g(x)))'=f'(g(x))g'(x)$ $\quad(\arcsin x)'=\frac{1}{\sqrt{1-x^2}}$ $\quad(\arctan x)'=\frac{1}{1+x^2}$

$(\frac{u}{v})'=\frac{u'v-uv'}{v^2}$ $\quad(a^x)'=(\ln a)a^x$ $\quad(\arccos x)'=-\frac{1}{\sqrt{1-x^2}}$

$\int uv'\,\mathrm{d}x=uv-\int u'v\,\mathrm{d}x$

1.概率

条件概率 $P(A|B)=P(AB)/P(B)$

全概率 $P(B)=\sum_i P(B|A_i)P(A_i)$

贝叶斯 $P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$

作业: 蓝眼 1/4. 对于三个孩子的家庭, 如果至少有一个孩子是蓝眼睛 (A), 至少有两个蓝眼睛 (B) 的概率为 $P(B|A)=P(AB)/P(A)=P(B)/P(A)=10/37$.

作业: $p_{1\ldots5}=0,\frac{1}{4},\frac{1}{2},\frac{3}{4},1$ 代表第 i 枚硬币出现正面的概率. 随机选取一枚硬币, 投掷直到出现正面. 求 $P(C_i\mid B_4)$, B_4 表示第 4 次首次出现正面. 由贝叶斯, $P(C_i|B_4)=\frac{P(B_4|C_i)P(C_i)}{\sum_{i\in[5]}P(B_4|C_i)P(C_i)}=\frac{P(B_4|C_i)}{\sum_{i\in[5]}P(B_4|C_i)}$.

2.随机变量

随机变量 $X:\Omega\mapsto\mathbb{R}$ 对每个样本赋予实值.

CDF $F_X(x)=P(X\le x)$

PDF 1: $f_X(x)\ge 0$, 2: $\int_{-\infty}^{+\infty}f_X(x)\,\mathrm{d}x=1$, 3: $P(a<X<b)=\int_a^b f_X(x)\,\mathrm{d}x$.

$F^{-1}(q)=\inf\{x:F(x)>q\}$, $(\max x: f(x)\le q)$

Poisson 分布: 平均 5 分钟 10 人到店, 问等待第一个顾客两分钟以上的概率. 两分钟实际到店人数 X 服从参数为 $\lambda=4$ 的 Poisson 分布, $P(T>2)=P(X=0)=e^{-4}$. 推广, 等待时间 $F(t)=1-e^{-\lambda t}, f(t)=F'(t)=\lambda e^{-\lambda t}$, 为指数分布, $t\in[0,\infty)$.

二维正态分布: $f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right)$

边际分布: $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}y$

独立定义 $P(X\in A,Y\in B)=P(X\in A)P(Y\in B)$, $f_{X,Y}(x,y)=f_X(x)f_Y(y)$, 独立 iff $\forall x,y, f(x,y)=r(x)g(y), f_X(x)=r(x)/\int_{-\infty}^{\infty}r(x)\,\mathrm{d}x$

条件分布 $f_{X|Y}(x|y)=\frac{P(X=x,Y=y)}{P(Y=y)}=\frac{f_{X,Y}(x,y)}{f_Y(y)}$

多项分布: $P(\ldots,X_k=n_k)=\frac{n!}{n_1!\ldots n_k!}p_1^{n_1}\ldots p_k^{n_k}$

多元正态分布: $f(x_1,\ldots,x_n)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

随机变量的函数: 对于随机变量 X 考虑函数形式 $r(x)$, 计算 $Y=r(X)$ 的分布: 对每个 y 求集合 $A_y=\{x:r(x)\le y\}$; 求 CDF: $F_{r(X)}(y)=P(r(X)\le y)=P(X\in A_y)=\int_{A_y}f_X(x)\,\mathrm{d}x$; 求 PDF: $f_Y(y)=F_Y'(y)$. 当 r 单调时, r 的反函数为 $s=r^{-1}$, 有 $f_Y(y)=f_X(s(y))\left|\frac{\mathrm{d}s(y)}{\mathrm{d}y}\right|$.

例: $f_X(x)=e^{-x}(x>0)$, $F_X(x)=\int_0^x f_X(s)\,\mathrm{d}s=1-e^{-x}$. 令 $Y=r(X)=\log X, A_y=\{x:x\le e^y\}$, $F_Y(y)=P(X\le e^y)=F_X(e^y)=1-e^{-e^y}$, $f_Y(y)=F_Y'(y)=e^ye^{-e^y}$.

作业: 令 $X,Y\sim\text{Uniform}(0,1)$ 且独立, 求 $X-Y$ 的 PDF. 令 $Z=X-Y$, $F_Z(z)=\int_{x+y\le z}f(x,y)\,\mathrm{d}x\,\mathrm{d}y=$

$$\begin{cases}0 & \text{if } z\le -1\\ \frac{(1+a)^2}{2} & \text{if } -1<z<0\\ 1-\frac{(1-a)^2}{2} & \text{if } 0\le z<1\\ 1 & \text{if } z\ge 1\end{cases}\qquad \begin{cases}f_Z(z)= & \\ 1+z, & -1<z<0\\ 1-z, & 0\le z<1\\ 0, & \text{otherwise}\end{cases}$$

3.期望

$\mu_X=E(X)=\int x\,\mathrm{d}F(x)<\infty$, 称期望存在.

懒惰统计学家法则 $E(r(X))=\int r(x)dF_X(x)$.

k 阶矩 $E(X^k)$, 中心矩 $E((X-\mu)^k)$.

$\sigma^2=V(X)=E((X-\mu)^2)=E(X^2)-E(X)^2$

样本 $\overline{X}=1/n\sum X_i, S_n^2=1/(n-1)\sum(X_i-\overline{X})^2$

定理: $E(\overline{X}_n)=\mu, V(\overline{X}_n)=\sigma^2/n, E(S_n^2)=\sigma^2$.

$Cov(X,Y)=E((X-\mu_X)(Y-\mu_Y))=E(XY)-E(X)E(Y)$, 不相关: $Cov(X_i,X_j)=0$, 此时两者方差可加, 但不代表独立（例: $X=U(-1,1), Y=|X|$）

矩母函数 $M_X(t)=E(e^{tX})$. 如果 $Y=aX+b$, 则 $M_Y(t)=e^{bt}M_X(at)$. 意义: $E(X^n)=M_X^{(n)}(0)$.

例: $X\sim\text{Exp}(\lambda), E(X)=M_X'(0)=\frac{\lambda}{(\lambda-t)^2}|_{t=0}=\frac{1}{\lambda}$, $E(X^2)=M_X^{(2)}(0)=\frac{2\lambda}{(\lambda-t)^3}|_{t=0}=\frac{2}{\lambda^2}$.

4.不等式

Markov $P(X\ge a)\le\frac{E(X)}{a}, X\ge 0, a>0$

Chebyshev $P(|X-E(X)|\ge a)\le\frac{V(X)}{a^2}$

Mill $Z\sim N(0,1), P(|Z|\ge t)\le\frac{2}{\sqrt{2\pi}}\frac{e^{-t^2/2}}{t}$

Hoeffding :

$P(|X-\mu|\ge t)\le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n(b_i-a_i)^2}\right)$

$E\left(e^{\alpha X}\right)\le\exp\left(\frac{\alpha^2(b-a)^2}{8}\right), E(X)=0, \alpha\in\mathbb{R}$

例: $X\sim\text{Bernoulli}(n,p)$, $P(|\overline{X}_n-p|>\epsilon)\le 2\exp(-2n\epsilon^2)$

Cauchy-Schwartz $E(XY)\le\sqrt{E(X^2)E(Y^2)}$

Jensen $E(g(X))\ge g(E(X))$, g 上凸.

5.随机变量的收敛

概率 $X_n\xrightarrow{P}X: P(|X_n-X|\ge\varepsilon)\to 0$, as $n\to\infty$.

分布 $X_n\rightsquigarrow X: F_{X_n}(x)\to F_X(x), \forall x\in$ 连续点.

均方 $X_n\xrightarrow{qm}X: E(X_n-X)^2\to 0$.

均方 $\to$ 概率 $\xrightarrow[\;]{\leftarrow\text{单点}}$ 分布.

依概率但不均方收敛 $X_n=\sqrt{n}I_{(0,1/n)}(U(0,1)); 0$

依分布但不依概率收敛 $X\sim N(0,1), X_n=-X$

1. $X_n+Y_n\to X+Y:(P,qm)$ 2. $X_nY_n\to XY:(P)$ 3. $g(X_n)\to g(X):(P,\rightsquigarrow)$ 4. (Slutzky) $X_n\rightsquigarrow X, Y_n\rightsquigarrow c: X_n+Y_n\rightsquigarrow X+c, X_nY_n\rightsquigarrow cX$.

注: 通常 $X_n\rightsquigarrow X, Y_n\rightsquigarrow Y\nRightarrow X_n+Y_n\rightsquigarrow X+Y$.

$\{X_i\}$ i.i.d., μ,σ^2 存在, $\overline{X}_N=\frac{1}{n}\sum_{i=1}^n X_i$

弱 LLN: $\overline{X}_n\xrightarrow{P}\mu: \lim_{n\to\infty}P(|\overline{X}_n-\mu|\le\varepsilon)=1$

强 LLN: $\overline{X}_n\xrightarrow{a.s.}\mu: P(\lim_{n\to\infty}|\overline{X}_n-\mu|\le\varepsilon)=1$

CLT: $Z_n\equiv\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}\rightsquigarrow N(0,1)$, i.e., $\lim_{n\to\infty}P(Z_n\le z)=\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^z e^{-t^2/2}\,\mathrm{d}t$

变形: $\overline{X}_n\approx N(\mu,\frac{\sigma^2}{n}), \sqrt{n}(\overline{X}_n-\mu)\approx N(0,\sigma^2)$.

样本 $S_n^2=(n-1)^{-1}\sum(X_i-\overline{X}_n)^2$, CLT σ 换为 S_n.

Delta 方法: 求极限分布

$Y_n\sim N(\mu,\frac{\sigma^2}{n})\Rightarrow g(Y_n)\sim N(g(\mu),(g'(\mu))^2\frac{\sigma^2}{n})$

6.模型、统计推断与学习

参数模型: $\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$, 如正态 $\theta=(\mu,\sigma)$

非参数模型: 无法用有限参数表示, 回归/聚类/决策树等

统计量: 完全基于样本所得的量, 是样本的函数.

点估计: $\hat{\theta}=g(X_1,\ldots,X_n)$. 偏差 $\text{bias}(\hat{\theta}_n)=E_\theta(\hat{\theta}_n)-\theta$.

无偏: bias $=0$; 相合: $\hat{\theta}_n\xrightarrow{P}\theta$.

标准误差 $\text{se}(\hat{\theta}_n)=\sqrt{V(\hat{\theta}_n)}$;

MSE $=E_\theta(\hat{\theta}_n-\theta)^2=\text{bias}^2+\text{se}^2$.

如果 bias $\xrightarrow{P}0$ 且 se $\xrightarrow{P}0$, 则 $\hat{\theta}_n\xrightarrow{P}\theta$.

渐进正态性: $\frac{\hat{\theta}_n-\theta}{\text{se}}\rightsquigarrow N(0,1)$.

置信集: 置信区间 $C_n=(a_n,b_n)$, $P_\theta(\theta\in C_n)=1-\alpha$.

注: 置信区间是随机区间, 意思是给定若干组样本, 每组得到的置信区间覆盖真实参数的概率为 $1-\alpha$.

基于正态的置信区间: $\hat{\theta}_n\pm z_{\alpha/2}\text{se}(\hat{\theta}_n)$.

例: $\{X_n\}\sim\text{Bernoulli}(p)$, 由 Chebyshev $P(|\overline{X}_n-p|\ge\epsilon)\le 2\exp(-2n\epsilon^2)$, 令 $\alpha=2\exp(-2n\epsilon)\Rightarrow\epsilon^2=\log(2/a)/(2n)$, 因此 $C_n=(\overline{X}_n-\sqrt{\log(2/a)/(2n)},\overline{X}_n+\sqrt{\log(2/a)/(2n)})$; 由渐进正态, $C_n=(\overline{X}_n-z_{\alpha/2}\sqrt{\overline{X}_n(1-\overline{X}_n)/n},\overline{X}_n+z_{\alpha/2}\sqrt{\overline{X}_n(1-\overline{X}_n)/n})$.

$z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$. $z_{0.05/2} = 1.96, z_{0.1/2} = 1.65, z_{0.025/2} = 2.24, z_{0.01/2} = 2.58, z_{0.005/2} = 2.80$.
假设检验: $H_0: \theta \in \Theta_0, H_1: \theta\in\Theta_1$. 一般把可以否定, 且可根据其构建分布机制的命题作为原假设.
显著性水平: 小概率事件发生的概率 α;
临界值: C, 使得 P_θ(拒绝 H_0, 如 $|\overline{X}-\mu|>C) = \alpha$.
拒绝域: $\mathcal{W}$, 如 $\{(X_1,\dots,X_n): |\overline{X}-\mu|>C\}$.

7.CDF 和统计泛函的估计
经验分布函数 $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i\le x)$ 无偏,
$V(\hat F_n(x)) = \frac{F(x)(1-F(x))}{n}$, $MSE\to 0$, $\xrightarrow{P}$, $F\in(0,1)$: $\sqrt{n}(\hat F_n(x)-F(x)) \rightsquigarrow N(0, F(x)(1-F(x)))$
统计泛函: $T(F)$ 是分布函数 F 的函数.
嵌入式估计量: $\hat\theta_n = T(\hat F_n)$.
线性泛函: $T(F)=\int_{-\infty}^{+\infty} r(x)\,\mathrm{d}F(x)$,
满足 $T(aF+bG) = aT(F)+bT(G)$, 嵌入式估计量: $T(\hat F_n) = \int_{-\infty}^{+\infty} r(x)\,\mathrm{d}\hat F_n(x) = \frac1n\sum_{i=1}^n r(X_i)$,
近似 $1-\alpha$ 置信区间为 $T(\hat F_n)\pm z_{\alpha/2}\hat{\mathrm{se}}$.
例: $\hat\mu = \overline{X}_n, \hat\sigma^2 = \frac1n\sum_{i=1}^n X_i^2 - \overline{X}_n^2$.

8.Bootstrap 方法
从 $\hat F_n$ 中生成 $X_1^*,\dots,X_n^*$ 计算统计量, 重复 B 次.
方差估计: $V_{\mathrm{boot}} = \frac1B\sum_{b=1}^B (T_{n,b}^* - \overline{T}^*)^2$
正态区间法 $T_n \pm z_{\alpha/2}\sqrt{V_{\mathrm{boot}}}$, 除非接近正态否则不准确.
枢轴量法: 定义 $R_n = \hat\theta-\theta$, $\hat\theta^*_{n,1\dots B}$ 为副本,
θ^*_β 为 β 分位数, $C_n = (2\hat\theta_n - \theta^*_{1-\alpha/2}, 2\hat\theta_n - \theta^*_{\alpha/2})$
分位数置信区间: $(\theta^*_{\alpha/2}, \theta^*_{1-\alpha/2})$

9.参数推断
似然函数: $\mathcal{L}_n(\theta) = \prod_{i=1}^n f(X_i,\theta)$,
对数似然函数 $\ell_n(\theta) = \log L_n(\theta)$
极大似然估计 $\hat\theta_n$: 使 $L_n(\theta)$ 达到最大值的 θ 的值,
可以解方程 $\partial\ell_n(\theta)/\partial\theta = 0$ 得到.
例 1. 正态分布 $\hat\mu = \overline{X}, \hat\sigma = S.$; 2. $\{X_i\}\sim U(0,\theta)$ i.i.d., $f(x;\theta) = 1/\theta$ if $0\le x\le\theta$, $\mathcal{L}_n(\theta) = (1/\theta)^n$ if $\theta \ge \max\{X_i\}$, 因此 $\hat\theta_n = \max\{X_i\}$ 3. $f(x) = \theta x^{\theta-1}$ 求 $\hat\theta$, $\ell_n(\theta) = n\log\theta + (\theta-1)\sum\log X_i$, 解方程 $\partial\ell_n(\theta)/\partial\theta = 0$ 得到 $\hat\theta_n = -\frac{n}{\sum\log X_i}$.
极大似然估计的相合性: $\hat\theta_n \xrightarrow{P} \theta_*$ KL 距离: $D(P,Q) = \int p(x)\log\frac{p(x)}{q(x)}\,\mathrm{d}x$, $D(f,g)\ge 0$, $D(f,g)=0 \Leftrightarrow f=g$.
模型可识别: $\theta\ne\phi \Rightarrow D(\theta,\phi)>0$
极大似然估计的同变性: $\tau = g(\theta)$, $\hat\tau_n = g(\hat\theta_n)$
Fisher $I_n(\theta) = V_\theta(\sum s(X_i;\theta)) = \sum V_\theta(s(X_i;\theta)) = E(s^2(X;\theta)) = nE(s(x_i;\theta))$, $s(X;\theta) = \frac{\partial\log f(X;\theta)}{\partial\theta}$
渐进正态性: 令 se $=\sqrt{V(\hat\theta_n)}$, 适当正则条件下

1. se $\approx \sqrt{1/I_n(\theta)}$, $\frac{\hat\theta_n-\theta}{\mathrm{se}}\rightsquigarrow N(0,1)$,
2. $\hat{\mathrm{se}} = \sqrt{1/I_n(\hat\theta_n)}$, $\frac{\hat\theta_n-\theta}{\hat{\mathrm{se}}}\rightsquigarrow N(0,1)$
3. $C_n = \hat\theta_n \pm z_{\alpha/2}\hat{\mathrm{se}}$, $P_\theta(\theta\in C_n)\to 1-\alpha$

例: $N(\theta,\sigma^2)$, $f(X;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(X-\theta)^2}{2\sigma^2}\right)$,
记分函数 $s(X;\theta) = \frac{\partial\log f(X;\theta)}{\partial\theta} = \frac{X-\theta}{\sigma^2}$, $I_n(\theta) = nE(s^2(X;\theta)) = -nE_\theta(s'(X;\theta)) = -nE_\theta(-\frac{1}{\sigma^2}) = \frac{n}{\sigma^2}$,
$\hat{\mathrm{se}} = \sqrt{\frac{\sigma^2}{n}}$, 故 $\overline{X}$ 近似服从 $N(\theta,\sigma^2/n)$.
极大似然估计 Delta 方法: 如果 $\tau = g(\theta), g'(\theta)\ne 0$, 则 $\frac{\hat\tau_n-\tau}{\hat{\mathrm{se}}(\hat\tau_n)} \rightsquigarrow N(0,1)$, 其中 $\hat\tau_n = g(\hat\theta_n)$, 且 $\hat{\mathrm{se}}(\hat\tau_n) = \sqrt{V(\hat\tau_n)} = |g'(\hat\theta_n)|\hat{\mathrm{se}}(\hat\theta_n)$.

例: Poisson(λ), 矩估计: $\hat\lambda = \overline{X}_n$, $f(X;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$, $\mathcal{L}_n = \prod\frac{\lambda^{X_i}e^{-\lambda}}{(X_i)!}$, $\ell_n = \sum(X_i\log\lambda - \lambda - \log(X_i)!)$, $\partial = 0$ 得 $\overline{X}$. $s(X;\lambda) = \frac{\partial\log f(X;\lambda)}{\partial\lambda} = \frac{X-\lambda}{\lambda}$, $I_n(\lambda) = nI(\lambda) = nE(s^2(X;\lambda)) = \frac{n}{\lambda}$, $\hat{\mathrm{se}} = \sqrt{\frac{\lambda}{n}}$, $\overline{X}$ 近似服从 $N(\lambda,\lambda/n)$, $1-\alpha$ 置信区间为 $\hat\lambda\pm z_{\alpha/2}\sqrt{\frac{\hat\lambda}{n}}$.
例: $N(\theta,1)$, $Y_i = I(X_i>0)$, $\psi = P(Y_1=1)$. $\mathcal{L}_n(\theta) = \prod\frac{1}{\sqrt{2\pi}}e^{-\frac{(X_i-\theta)^2}{2}}$, $\ell_n(\theta) = -\frac n2\log 2\pi - \frac12\sum(X_i-\theta)^2$, $\partial\ell_n(\theta)/\partial\theta = \sum(x_i-\theta), = 0 : \hat\theta = \overline{X}$. $I_n(\theta) = nE_\theta(s^2(X_i;\theta)) = n$, $\hat{\mathrm{se}}(\hat\theta) = \frac{1}{\sqrt{I_n(\hat\theta)}} = \frac{1}{\sqrt n}$, $\psi = \Phi(\theta)$,
$\hat\psi = \Phi(\hat\theta) = \Phi(\overline{X})$, $\hat{\mathrm{se}}(\hat\psi) = |\Phi'(\hat\theta)|\hat{\mathrm{se}}(\hat\theta) = \phi(\hat\theta)\hat{\mathrm{se}}(\hat\theta)$.
$\tilde\psi = 1/n\sum Y = E[\overline{Y}] \xrightarrow{P} E[Y] = \psi$, 渐进效率 $V(\tilde\psi)/V(\hat\psi) = \psi(1-\psi)/\phi(\theta) = \Phi(\theta)(1-\Phi(\theta))/\phi(\theta)$.
非正态: 由 LLN $\hat\psi\xrightarrow{P}\Phi(\mu)$, $F_X(0)\ne 1-\Phi(\mu)$ 都不相合.
例: $X_1\sim \mathrm{Binomial}(n_1,p_1), X_2\sim \mathrm{Binomial}(n_2,p_2), \psi = p_1-p_2$, $f(X_i;p_i) = \binom{n_i}{X_i}p_i^{X_i}(1-p_i)^{n_i-X_i}$, $\frac{\partial}{\partial p_i}s(X_i;p_i) = \frac{X_i}{p_i}+\frac{X_i}{1-p_i} = \frac{X_i-np_i}{p_i(1-p_i)}$ $\hat p_i = X_i/n_i, \hat\psi = \hat p_1-\hat p_2 = X_1/n_1 - X_2/n_2$. $I(p_1,p_2) = (\partial^2\log f((X_1,X_2);\psi)/\partial p_i\partial p_j) = \begin{bmatrix}\frac{n_1}{p_1(1-p_1)} & 0\\ 0 & \frac{n_2}{p_2(1-p_2)}\end{bmatrix}$.
多参: $\theta = (\theta_1,\dots,\theta_k)^T$ Fisher 信息矩阵 $I_n(\theta) = (E_\theta(H_{ij}))$, $H_{ij} = \frac{\partial\log f(X;\theta)}{\partial\theta_i\partial\theta_j}$, $J_n(\theta) = I_n^{-1}(\theta)$
多参数 Delta 方法: $\nabla g = (\frac{\partial g}{\partial\theta_1},\dots)^T$ 在 $\hat\theta_n$ 处不为 0, 令 $\hat\tau_n = g(\hat\theta_n)$, 则 $\frac{\hat\tau_n-\tau}{\sqrt{\hat\tau_n^T J_n(\hat\theta_n)\hat\tau_n}}\rightsquigarrow N(0,1)$, 其中 $\hat{\mathrm{se}}(\hat\tau_n) = \sqrt{(\nabla g|^T_{\theta=\hat\theta_n})J_n(\hat\theta_n)(\nabla g|_{\theta=\hat\theta_n})}$.
例: $X_1\dots X_n\sim N(\mu,\sigma^2)$, $\psi = g(\mu,\sigma) = \sigma/\mu$, (μ,σ^2) MLE: $(\overline{X},S^2)$, $I_n(\mu,\sigma^2) = \begin{pmatrix} n/\sigma^2 & 0\\ 0 & 2n/\sigma^2\end{pmatrix}$ $J_n = \begin{pmatrix}\sigma^2/n & 0\\0&\sigma^2/2n\end{pmatrix}$, $\nabla g = (-\frac{\sigma}{\mu^2},\frac1\mu)^T$, $\hat{\mathrm{se}}(\hat\tau) = \frac{1}{\sqrt n}\sqrt{\frac{\hat\sigma^4}{\hat\mu^4}+\frac{\hat\sigma^2}{2\hat\mu^2}}$.

10.假设检验与 p 值
检验统计量: $T_n = T(X_1,\dots,X_n)$,
拒绝域 $\mathcal{A} = \{T_n > c\}$, c 为临界值.
第一类错误: 假阳性; 第二类错误: 假阴性.
势函数 $\beta(\theta) = P_\theta(X\in\mathcal{A})$, 容度 $\beta = \sup_{\theta\in\Theta_1}\beta(\theta)$
容度 $\le\alpha$ 称检验水平为 α. 简单 $\theta=\theta_0$ 复合 $\theta\ge\theta_0$
双边 $H_0:\theta=\theta_0$ vs ...单边 $H_0:\theta\ge\theta_0$ vs ...
例: $X_1\dots X_n\sim N(\mu,\sigma^2)$, 检验 $H_0:\mu=\mu_0$ vs $H_1:\mu\ne\mu_0$, 检验统计量 $T_n = \sqrt n\frac{|\overline{X}_n-\mu_0|}{S_n}$,
拒绝域 $\{T_n > z_{\alpha/2}\}$, 势函数 $\beta(\mu) = P_\mu(T_n > z_{\alpha/2})$
如检验 $H_0:\mu\le 0$ vs $H_1:\mu>0$, 拒绝域 $\mathcal{A} = \{\overline{X}>c\}$, 势函数 $\beta(\mu) = P_\mu(\overline{X}>c) = P(\sqrt n\frac{(\overline{X}-\mu)}{\sigma} > \sqrt n\frac{c-\mu}{\sigma}) = 1-\Phi(\sqrt n\frac{c-\mu}{\sigma})$, 检验水平 $\alpha = \sup_{\mu\le\mu_0}\beta(\mu) = 1-\Phi(\sqrt n\frac{c-\mu_0}{\sigma})$, 设定显著性水平 α, $c = \mu_0 + \frac{\sigma}{\sqrt n}z_{\alpha/2}$.
Wald 检验: $H_0:\theta=\theta_0$ vs $H_1:\theta\ne\theta_0$, 若 $\hat\theta$ 渐进正态, $W = \frac{\hat\theta-\theta_0}{\hat{\mathrm{se}}(\hat\theta)}$, 当 $|W|>z_{\alpha/2}$ 时拒绝 H_0. 势函数:
$\beta(\theta_*) \approx 1-\Phi(z_{\alpha/2}+\frac{\theta_0-\theta_*}{\hat{\mathrm{se}}(\hat\theta)}) + \Phi(-z_{\alpha/2}+\frac{\theta_0-\theta_*}{\hat{\mathrm{se}}(\hat\theta)})$
例: 给定样本 $\{X_m\},\{Y_n\}$, 检验均值是否相等.
$H_0:\mu_X-\mu_Y=0$ vs $H_1:\mu_X-\mu_Y\ne 0$,
$V(\hat\theta) = V(\overline{X}-\overline{Y}) = \frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}$, $\hat{\mathrm{se}} = \sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}}$,
拒绝域为 $|\overline{X}-\overline{Y}|>z_{\alpha/2}\hat{\mathrm{se}}$,
例: 给定样本 $\{(X_n,Y_n)\}$ 检验均值是否相等.
$H_0:\mu_X=\mu_Y$ vs $H_1:\mu_X\ne\mu_Y$, 构造 $\theta=\mu_1-\mu_2$, 估计为 $\hat\theta = \overline{X}-\overline{Y}$. 方差: $V(\hat\theta) = \sigma^2/n$, 其中 σ 是 X_i-Y_i 总体标准差, 用样本标准差估计 $\hat{\mathrm{se}} = \sqrt{S^2/n}$, 拒绝域为 $|\overline{X}-\overline{Y}|>z_{\alpha/2}\hat{\mathrm{se}}$.
Wald 置信区间: $C = \left(\hat\theta - z_{\alpha/2}\hat{\mathrm{se}}(\hat\theta), \hat\theta+z_{\alpha/2}\hat{\mathrm{se}}(\hat\theta)\right)$,
p 值: $\inf\{\alpha:\theta_0\in C\}$, 可以拒绝 H_0 的最小检验水平.
例: 计算出了 $W = \left|\frac{\overline{X}-\overline{Y}}{\hat{\mathrm{se}}}\right|$, p 值: $P(|Z|>W)_{(Z\sim N(0,1))}$

Pearson χ^2 统计量: $X = (X_1,\dots,X_k)$ 服从多项分布 $M(n,p=(p_1,\dots,p_k))$, 检验 $H_{0/1}: p? = p_0$.
检验统计量 $\chi^2 = \sum_{i=1}^k\frac{(X_i-np_{0i})^2}{np_{0i}} = \sum_{i=1}^k\frac{(X_j-E_j)^2}{E_j}$. 在 H_0 下 $T\rightsquigarrow\chi^2_{k-1}$, 拒绝域为 $\{T>\chi^2_\alpha(k-1)\}$,
p 值为 $P(\chi^2(k-1)>T)$.
置换检验: 检验分布是否相同, shuffle 看统计量分位数
似然比检验: $H_0:\theta\in\Theta_0$ vs $H_1:\theta\notin\Theta_0$, $\lambda = 2\log\left(\frac{L(\hat\theta)}{L(\hat\theta_0)}\right)$, $\hat\theta_0$ 表示限制 Θ_0. $\lambda\rightsquigarrow\chi^2_{\dim\Theta-\dim\Theta_0}$,
拒绝域 $\{\lambda>\chi^2_\alpha(\dim\Theta-\dim\Theta_0)\}$, p 值 $P(\chi^2(\dots)>\lambda)$.
例: 豌豆实验自由度为 3 − 0 而非 4, 因为概率和为 1.
多重假设检验: Bonferroni: α/m (Union bound). BH: 排序后 $p_i \le \frac{i\alpha}{mC_m}$, C_m 不独立 $\sum_{i=1}^m\frac1i$, 独立 1.

13.线性回归和 Logistic 回归
简单线性回归: $Y = \beta_0+\beta_1X+\epsilon$, $\epsilon: E=0, V=\sigma^2$, $\hat y_i = \hat\beta_0+\hat\beta_1 x$, $\hat\epsilon = y_i-\hat y_i$.
最小二乘: 使 RSS $=\sum\hat\epsilon_i^2$ 最小, E(RSS) $=(n-2)\sigma^2$
$\ell_{xy} = \sum(x_i-\overline{x})(y_i-\overline{y}) = \sum x_iy_i - n\overline{xy}$,
$\ell_{xx} = \sum(x_i-\overline{x})^2 = \sum x_i^2 - n\overline{x}^2$,
$\hat\beta_1 = \ell_{xy}/\ell_{xx}$, $\hat\beta_0 = \overline{y}-\hat\beta_1\overline{x}$, $\hat\sigma^2$ 无偏估计为 $\frac{1}{n-2}$RSS.
若 $\epsilon_i|x_i\sim N(0,\sigma^2)$, 则
$L(\beta_0,\beta_1,\sigma^2) \propto \sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum(y_i-\beta_0-\beta_1x_i)^2\right)$,
$\ell(\beta_0,\beta_1,\sigma^2) \propto -n\log\sigma - \frac{1}{2\sigma^2}\sum(y_i-\beta_0-\beta_1x_i)^2$.
正态性假设下, 最小二乘即极大似然. $\hat\sigma^2 = 1/n\sum\hat\epsilon_i^2$.
$\hat{\mathrm{se}}(\hat\beta_1) = \sqrt{\frac{\hat\sigma^2}{\ell_{xx}}}$, $\hat{\mathrm{se}}(\hat\beta_0) = \hat{\mathrm{se}}(\hat\beta_1)\sqrt{\frac1n\sum x_i^2}$, 适当条件下, 相合, 渐进正态 $(\frac{\hat\beta_i-\beta_i}{\hat{\mathrm{se}}(\hat\beta_i)}) \rightsquigarrow N(0,1)$, 渐进置信区间 $\hat\beta_i\pm z_{\alpha/2}\hat{\mathrm{se}}(\hat\beta_i)$, Wald 检验 $W = \hat\beta_1/\hat{\mathrm{se}}(\hat\beta_1)$.
近似预测区间: $\hat\xi_n^2 = \hat\sigma^2\left(\frac1n+\frac{(x_0-\overline{x})^2}{\ell_{xx}}\right)$, $\hat y_*\pm z_{\alpha/2}\hat\xi_n$.
Logistic 回归: 分类问题损失函数不连续, 接一个 Sigmoid 函数 (Logistic) $g(x) = \frac{1}{1+e^{-x}}$. 没有显式解.