countable additivity: disjoint $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i)=\sum_{i=1}^{\infty}\mathbb{P}(A_i)$

Independence: $P(A\cap B)=P(A)P(B)$

Law of total prob. $P(B)=\sum_{i}^{k} P(B|A_i)P(A_i)$

Bayes: $P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum P(B|A_i)P(A_i)}$

Union bound: $\mathbb{P}(\bigcup_{i=1}^{n}A_i)\leq\sum^{n}\mathbb{P}(A_i)$

Bonferroni: $P(A\cap B)\geq P(A)+P(B)-1$

CDF: 1. $\lim_{x\to-\infty}F(x)=0$, $\lim_{x\to\infty}F(x)=1$

2. non-decreasing

3. CDF $\lim_{x\to x_0^+}F(x)=F(x_0)$

$F(X)=F(Y)\Leftrightarrow X,Y$ identically distributed.

$f(x)=P_X(X=x),\forall x$ $\quad F_X(x)=\int_{-\infty}^{x}f_X(t)\,dt$.

Bernoulli: $p(x)=p^{x}(1-p)^{1-x}$, $x\in\{0,1\}$

Binomial: $p(x)=\binom{n}{x}p^{x}(1-p)^{n-x}$ $x\in\{0,1,\dots,n\}$

$p(\lambda)=\frac{\lambda^{x}\exp(-\lambda)}{x!}$ Poi(λ) Poisson

Gaussian: $p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$

Exponential: $p(x)=\lambda\exp(-\lambda x)$, $x\geq0$, Exp(λ)

Transformation $f_Y(y)=f_X(s(y))\left|\frac{ds(y)}{dy}\right|$

k^{th} moments: $\mu_k=\mathbb{E}[X^k]$

central moments α_k: $\mathbb{E}[(X-\mu)^k]$

Bivariate independence: $f_{XY}(x,y)=f_X(x)f_Y(y)$

$f_{XY}(x,y)=h(x)g(y)\to$ independent

$Cov(X,Y)=E((X-\mu_X)(Y-\mu_Y))=\mathbb{E}XY-\mathbb{E}X\mathbb{E}Y$

$\mathbb{E}(f(X)g(Y))=\mathbb{E}(f(X))\mathbb{E}(g(Y))$

Variance$(\sum_{i=1}^{n}a_iX_i)=\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(X_i,X_j)a_ia_j$

$(\sum_{i=1}^{n}X_i)^2=\sum_{i=1}^{n}\sum_{j=1}^{n}X_jX_i$

$E_X[E_{Y|X}[Y|X]]=E[Y]$ Total expectation

$V(Y)=E_X(V(Y|X))+V_X(E(Y|X))$

MGF: $M_X(t)=\mathbb{E}\exp(tX)$

$M_X^{(k)}(0)=E[X^k]$

$M_Y(t)=\prod_{i=1}^{n}M_{X_i}(t)$, $Y=\sum^{n}X_i$

if MGF exists, around 0, X,Y same dist

Markov: $P(X>t)\leq\frac{\mathbb{E}[X]}{t}$ $\boxed{\text{positive}}$ finite mean

Chebyshev: $P(|X-\mu_X|\geq k\sigma_X)\leq\frac{1}{k^2}$ finite mean, variance

Mill's $P(|Z|>t)\leq\sqrt{\frac{2}{\pi}}\frac{\exp(-t^2/2)}{t}$ $Z\sim N(0,1)$

Hoeff $P(|\frac{1}{n}\sum_{i=1}^{n}X_i|\geq t)\leq2\exp(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2})$

$E[X_i]=0$, $a_i\leq X_i\leq b_i$

WLLN: $P(|\frac{1}{n}\sum^{n}X_i-E(X)|\geq\varepsilon)\to0$ as $n\to\infty$

$\leq\frac{Var(X)}{n\varepsilon^2}$

$\varepsilon>0$, $\lim_{n\to\infty}P(|X_n-X|\geq\varepsilon)\to0$ Converge Prob $\boxed{\text{WLLN}}$

$\lim_{n\to\infty}F_n(t)=F(t)$ Converge dist $\boxed{\text{CLT}}$

$\frac{\sqrt{n}(\hat{\mu}-\mu)}{\sigma}\to N(0,1)$ $\boxed{\text{CLT}}$

$q.m.\Rightarrow p\Rightarrow$ dist

Eg. $X\sim U[0,1]$ $X_n=\max_{1\leq i\leq n}X_i$

$P(|X_n-1|\geq\epsilon)=(1-\varepsilon)^n\to0$.

$P(n(1-X_n)\leq t)=1-(1-\frac{t}{n})^n=1-\exp(-t)$

$n(1-X_n)\xrightarrow{d}Exp(1)$ CDF: $1-\exp(-\lambda t)$

<u>Lyapunov CLT</u>. $\mu_i=E[X_i]$, $\sigma_i=Var(X_i)$

$S_n^2=\sum_{i=1}^{n}\sigma_i^2$, if satisfy

$\lim_{n\to\infty}\frac{1}{S_n^3}\sum_{i=1}^{n}E|X_i-\mu|^3=0$ then

$\frac{1}{S_n}\sum^{n}|X_i-\mu_i|\xrightarrow{d}N(0,1)$

<u>Multivariate</u>: $\sqrt{n}(\hat{\mu}-\mu)\xrightarrow{d}N(0,\Sigma)$

Delta: $\frac{\sqrt{n}(g(X_n)-g(\mu))}{\sigma}\xrightarrow{d}N(0,[g'(\mu)]^2)$ $g'(\mu)\neq0$ $\Delta_\mu(g)^T\Sigma\Delta_\mu(g)$ $=\begin{pmatrix}\frac{\partial g(x)}{\partial x_1}\\ \vdots\\ \frac{\partial g(x)}{\partial x_d}\end{pmatrix}_{x=\mu}$

$\int(y-E(Y|X))^2f_{Y|X}(y)dy\,f(x)dx+(E(Y|X))^2-(EY)^2=\int y^2f_Y(y)dy-(EY)^2$

Point estimation: $\underbrace{\hat{\theta}_n}_{rv}=g(X_1,\dots,X_n)$

bias: $b(\hat{\theta})=E_\theta(\hat{\theta})-\theta$

$V(\hat{\theta})=E_\theta(\hat{\theta}-E_\theta(\hat{\theta}))^2$

$MSE=b(\hat{\theta})^2+V(\hat{\theta})$

sample variance: $S=\frac{1}{n-1}\sum^{n}(X_i-\frac{1}{n}\sum X_i)^2$

$E[S]=\sigma^2$

Invariance of MLE

$\hat{y}=\arg\sup_y L^*(y|X_1,\dots,X_n)$

$=\underbrace{\arg\sup_y L(\tau^{-1}(y)|X_1,\dots,X_n)}$

$=\tau(\arg\sup_\theta L(\theta|X_1,\dots,X_n))=\tau(\hat{\theta})$

$\arg\sup_y\sup_{\theta:\tau(\theta)=y}L(\theta|X_1,\dots,X_n)$

define $L^*(y|X_1,\dots,X_n)=\sup_{\theta:\tau(\theta)=y}L(\theta|X_1,\dots,X_n)$

<u>Bayes Estimator</u>

prior $\pi(\theta)$, posterior $\pi(\theta|X_1,\dots,X_n)$

Proof? $\pi(\theta|X_1,\dots,X_n)=\frac{\pi(\theta)L(\theta|X_1,\dots,X_n)}{\int_\theta\pi(\theta)L(\theta|X_1,\dots,X_n)}$

posterior mean for point estimation.

consistent estimator $P_\theta(|\hat{\theta}_n-\theta|>\varepsilon)\to0$

for any ε $\quad n\to\infty$.

Squared loss $L(\alpha,\theta)=(\alpha-\theta)^2$ $\quad\theta$: true

Abs $L(\alpha,\theta)=|\alpha-\theta|$

KL. $L(\alpha,\theta)=KL(f_\theta,f_\alpha)=\mathbb{E}_{X\sim f_\theta}\log(\frac{f_\theta(X)}{f_\alpha(X)})$

<u>Risk: $R(\theta,\hat{\theta}(X))=\mathbb{E}_\theta L(\hat{\theta},\theta)$</u>

$X_1,\dots,X_n\sim Ber(p)$

$\hat{p}_1=\frac{1}{n}\sum X_i$, $\hat{p}_2=\frac{\sum X_i+\alpha}{n+\alpha+\beta}$

$R(p,\hat{p}_1)=0+\frac{p(1-p)}{n}$, $R_2(p,\hat{p}_2)=Var(\frac{\sum X_i+\alpha}{n+\alpha+\beta})+(E[\frac{\sum X_i+\alpha}{n+\alpha+\beta}-p])^2$

$\alpha=\beta=\sqrt{\frac{n}{4}}$, $R_2=\frac{n}{4(n+\sqrt{n})^2}$

Cauchy-Schwarz

$\mathbb{E}[XY]\leq\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$

Jensen

$g(\mathbb{E}[X])\leq\mathbb{E}[g(X)]$

Poisson(λ) $f(k)=\frac{\lambda^ke^{-\lambda}}{k!}$

$R(\theta,\hat{\theta}_n)$ is a constant then $\hat{\theta}_n$ minimax estimator.

KL divergence $D(f\|g)=\int_x f(x)\log(\frac{f(x)}{g(x)})dx$

Maximize $\log L$ equals

$M_n(\theta)=\frac{1}{n}\sum^{n}\log\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)}$

$\to E_{X_i\sim f_{\theta^*}}\log\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)}$

$=-D(f_{\theta^*}\|f_\theta)$

score function:

$S_\theta(x)=\frac{\partial\log f_\theta(x)}{\partial\theta}$

$E_{X\sim f_\theta}[S_\theta(X)]=0$

$0=\int\frac{\partial f_\theta(x)}{\partial\theta}dx$

$1=\int f_\theta(x)dx$ $\quad=\int\frac{\partial\log f_\theta(x)}{\partial\theta}f_\theta(x)dx$

High dimensional mean est

$R(\hat\theta,\theta)=E[\sum_j^d(\hat\theta_j-\theta_j)^2]=\frac{\sigma^2 d}{n}$

Hard: $\hat\theta_i=\bar X_{ni}\mathbb{I}(|\bar X_{ni}|\ge t)$

soft: $\hat\theta_i=\text{sign}(\bar X_{ni})\max\{|\bar X_{ni}|-t,0\}$

$\arg\min_\theta \frac{1}{2n}\sum^n\|X_i-\theta\|_2^2+\frac{t^2}{2}\sum^d\mathbb{I}(\theta_i\neq0)$

$\arg\min_\theta \frac{1}{2n}\sum^n\|X_i-\theta\|_2^2+t\sum^d|\theta_i|$

detour

$\epsilon_1,\dots,\epsilon_d\sim N(0,\sigma^2)$, $\max_i|\epsilon_i|\le\sigma\sqrt{2\log(2d/\delta)}$

$P(|\epsilon|\ge t)\le 2\exp(-t^2/2\sigma^2)$ w.p. $1-\delta$

$P(\max|\epsilon|\ge t)\le 2d\exp(-t^2/2\sigma^2)$ by union bound

$t\gtrsim\sqrt{\frac{\log d/\delta}{n}}$, when $|\theta_i|\le\frac t2$, $(\hat\theta_i-\theta_i)^2=\theta_i^2$ (bias)

When $|\theta_i|>100t$, $(\hat\theta_i-\theta_i)^2=\epsilon_i^2\le\frac{t^2}{4}$

when $\frac t2\le|\theta_i|\le100t$, err $\le\max\{\frac{t^2}{4},\theta_i^2\}\le\cdots$

θ has s non zeros: $\|\hat\theta_i-\theta\|_2^2\lesssim st^2\lesssim s\frac{\log(d/\delta)}{n}$

$\sum^d|\theta_i|\le R_1$, dense, small l1 norm: $\lesssim\sqrt{\frac{R_1\log d/\delta}{n}}$

Regression

$y_i=\langle x_i,\beta^*\rangle+\epsilon_i$, $x_i\in R^d$, $\hat\Sigma=\frac1n\sum x_ix_i^T$

$\hat\beta=\arg\min_\beta\frac12\sum^n(y_i-\langle x_i,\beta\rangle)^2=\hat\Sigma^{-1}[\frac1n\sum x_iy_i]$

when X_i are random $\in R^{n\times d}$, $\hat\beta=(X^TX)^{-1}X^Ty=(X^TX)^{-1}X^T(X\beta^*+\epsilon)$

$\sim N(\beta^*,\sigma^2(X^TX)^{-1})$ $\quad=\beta^*+(X^TX)^{-1}X^T\epsilon$

$X\hat\beta\sim N(X\beta^*,\sigma^2X(X^TX)^{-1}X^T)$

In sample: $E\|\frac{X\hat\beta-X\beta^*}{n}\|_2^2=\frac{\sigma^2}{n}E[tr(X(X^TX)^{-1}X^T)]=\frac{\sigma^2 d}{n}$

out sample: $E[\langle x,\hat\beta\rangle-\langle x,\beta^*\rangle]^2$

l_2: $E\|\hat\beta-\beta^*\|_2^2=\sigma^2E[tr((X^TX)^{-1})]=\frac{\sigma^2}{n}E[tr(\hat\Sigma^{-1})]$

$I_n(\beta)=\frac{n\Sigma}{\sigma^2}=nE[\frac{X^TX}{n\sigma^2}]$ $\quad\le\frac{\sigma^2 d}{cn}$ c: bounded eigen

$\sqrt n(\hat\beta-\beta^*)\xrightarrow{d}N(0,\sigma^2\Sigma^{-1})$

High-dim regression:

$\hat\beta=\arg\min_\beta\frac12\|y-X\beta\|_2^2+\frac{t^2}{2}\sum^d\mathbb{I}(\beta_i\neq0)$ Hard

$\hat\beta=\arg\min_\beta\frac12\|y-X\beta\|_2^2+t\sum^d|\beta_i|$ soft

Non parametric regression $f(x)$. $\hat f\to$ local averaging

1. $x_i=\frac in$ ... 2. $|f(x)-f(y)|\le L|x-y|$, 3. $y_i=f(x_i)+\epsilon_i\sim(0,\sigma^2)$

$R(\hat f,f)=E\int_0^1(\hat f(x)-f(x))^2dx=\int_0^1(f(x)-E\hat f(x))^2dx+$ bias

$h=\frac kn$, k in each bin. $\quad\int_0^1E(\hat f(x)-E\hat f(x))^2dx$ var

$\hat f(x_0)=\frac1k\sum_{i:x_i\in B_{x_0}}y_i$, $|E\hat f(x_0)-f(x_0)|=|\frac1k\sum Ey_i-f(x_0)|$

$=|\frac1k\sum[f(x_i)-f(x_0)]|\le\frac1k\sum|f(x_i)-f(x_0)|\le Lh$

$E[\hat f(x_0)-E\hat f(x_0)]^2=E[\frac1k\sum y_i-f(x_i)]^2=E[\frac1k\sum\epsilon_i]^2=\frac{\sigma^2}{k}$

$b^2+v\le L^2h^2+\frac{\sigma^2}{nh}$, $h=(\frac{\sigma^2}{2nL^2})^{\frac13}$

$R\le(\frac{L\sigma^2}{2n})^{\frac23}$ $\quad\cancel{R\le n^{\frac{-2}{2+d}}}$

Randomized controlled trial

ATE: $E[Y(1)-Y(0)]$, $\frac1n\sum_{i=1}^n[Y_i(1)-Y_i(0)]$

$\hat\tau=\frac{1}{\frac n2}\sum_{T_i=1}Y_i(T_i)-\frac{1}{\frac n2}\sum_{T_i=0}Y_i(T_i)$, $E(\hat\tau)=\tau$

$\tau=E_XE[Y(1)-Y(0)|X]$ $\quad=\frac{1}{\frac n2}\sum^nE[T_iEY_i(1)]-\frac{1}{\frac n2}\sum^nE((1-T_i)EY_i(0))$

$\mu_1(x)=E[Y^{obs}|X,T=1]$ $\quad\mu_0(x)=$

$\hat\tau=E_X[\mu_1(X)-\mu_0(X)]$

$\|\epsilon\|=\epsilon<\frac t2$ Propensity: $\pi(x)=P(T=1|X=x)$

$\hat\tau=\frac1n\sum[\frac{Y_iT_i}{\pi(X_i)}-\frac{Y_i(1-T_i)}{1-\pi(X_i)}]$

$P(\theta\in C_n(X_1,\dots,X_n))\ge1-\alpha$,

$C(X_1,\dots,X_n)=\{\theta_0:\{X_1,\dots,X_n\}\notin R\}$

$P(FDP\ge\beta)\le\frac{FDR}{\beta}$, by Markov.

FWER. sidak: $p\le1-(1-\alpha)^{1/P}=\alpha t$ indepe.

Bonferroni: $1-(1-\alpha)^{1/P\cdot P}=\alpha=FWER$

Union bound. $P\le\frac\alpha P$, $FWER\le\sum^P\frac\alpha P=\alpha$.

$FDP=\begin{cases}\frac VR, & R>0\\ 0\end{cases}$, $FDR=E\,FDP$,

$FWER=P(V\ge1)$, BH procedure: ① $P_{(1)}\le\dots P_{(T)}$

② $t_i=\frac{i\alpha}{T}$ ③ $i_{max}=\arg\max_i\{i:P_{(i)}\le t_i\}$ ④ reject.

$FDR\le\frac{E[V]}{i^*}=\frac{T\cdot\frac{i^*\alpha}{T}}{i^*}=\alpha$, FWER ≥ FDR when g^T

$T(X_1,\dots,X_m,Y_1,\dots,Y_n)=|\frac1m\sum^mX_i-\frac1n\sum Y_i|$ Permutation

p-value $=\frac{1}{N!}\sum^{N!}\mathbb I(T_j>T_{obs})\approx\frac1B\sum_{j=1}^B\mathbb I(T_j>T_{obs})$

Kolmogrov-Smirnov-Test: $T=\sup_x|\hat F_n(x)-F_0(x)|$

$P(T\ge\epsilon)\le2\exp(-2n\epsilon^2)$, $T\ge\sqrt{\log(2/\alpha)/(2n)}$

$\beta(F_1)=P(\sup|\hat F_n(x)-F_0(x)|\ge\sqrt{\ })$ power

$\le P(\sup|F_1(x)-F_0(x)|+\sup|\hat F_n(x)-F_1(x)|\ge\sqrt{\ })$

$=P(\sup|\hat F_n(x)-F_1(x)|\le\sup|F_1(x)-F_0(x)|-\sqrt{\ })$

χ^2 test: $T(X_1,\dots,X_n)=\sum_{i=1}^k\frac{(Z_i-np_{0i})^2-np_{0i}}{np_{0i}}\to\chi^2_{k-1}-k$

Two sample testing: $X_1,\dots,X_{n_1}\sim P$, $Y_1,\dots,Y_{n_2}\sim Q$ H_0. $P=Q$

$\hat C_i=\frac{Z_i+Z_i'}{n_1+n_2}$, $T_n=\sum^k[\frac{(Z_i-n_1\hat C_i)^2}{n_1\hat C_i}+\frac{(Z_i'-n_2\hat C_i)^2}{n_2\hat C_i}]\sim\chi^2_{k-1}$

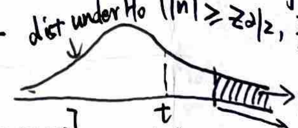
p-value: $\inf\{\alpha:T_n\in R_\alpha\}$

rej H_0, if $T_n>t_\alpha$

$\sup_{\theta\in\Theta_0}P_{X_n\sim\theta}(T(X_n)\ge T_n)$

pvalue for wald test:

dist under H_0 $\quad|T_n|\ge z_{\alpha/2}$, p-value $=2(1-\Phi(T_n))=2\Phi(-T_n)$



reject if $T>t$ $\quad T(X_1,\dots,X_n)$

$\{R_\alpha\subseteq R_{\alpha'}$, if $\alpha\le\alpha'\}$

p-values $=\inf\{\alpha,\{X_1,\dots,X_n\}\in R_\alpha\}$

$=1-\Phi(T(X_1,\dots,X_n))$

$P_{\theta_0}(1-\Phi(T(X_1,\dots,X_n))\le u)=P(T(X_1,X_n)\ge\Phi^{-1}(1-u))=1-(1-u)=u$.

level α test: $\sup_{\theta\in\Theta_0}\beta(\theta)\le\alpha$, $P_\theta(X_1,\dots,X_n\in R)$

LRT: $T(X_1,\dots,X_n)=\frac{\sup_{\theta\in\Theta_0}L(\theta;X_1,\dots,X_n)}{\sup_{\theta\in\Theta}L(\theta;X_1,\dots,X_n)}$

Wilks: $-2\log T\to\chi^2_1$,

$l(\theta)\approx l(\hat\theta)+l'(\hat\theta)(\theta-\hat\theta)+l''(\hat\theta)\frac{(\theta-\hat\theta)^2}{2}$

power: $\int\phi(x)f_1(x)dx$, size $\int\phi(x)f_0(x)dx$

NP test: $T(X)=\frac{L(X;\theta_0)}{L(X;\theta_1)}=\frac{f_0(x)}{f_1(x)}$ (simple vs simple)

NP Lemma: $\int(\phi_{NP}(x)-\phi_A(x))(f_1(x)-\frac{f_0(x)}{t^*})dx\ge0$

NP test most powerful at size α $P_0(T(X)\le t^*)\ge\alpha$

Wald: $\sqrt{nI(\theta_0)}(\hat\theta-\theta_0)\to N(0,1)$, reject if

$|T_n|\ge\Phi^{-1}(1-\frac\alpha2)$

$r_n(T(X_1,\dots,X_n)-\theta_0)\xrightarrow{d}F$

reject $r_n(T(X_1,\dots,X_n)-\theta_0)\ge F_\alpha$

power of wald: $P(\sqrt{nI(\theta_0)}(\hat\theta-\theta_0)\ge\Phi^{-1}(1-\frac\alpha2))$

$=P(\sqrt{nI(\theta_1)}(\hat\theta-\theta_1)+(\theta_1-\theta_0))\ge\dots\Phi^{-1}(1-\frac\alpha2))$

$=P(\sqrt{nI(\theta_1)}(\hat\theta-\theta_1)\ge\sqrt{nI(\theta_1)}(\theta_0-\theta_1)+\Phi^{-1}(1-\frac\alpha2))$

$=1-\Phi(\dots+\Phi^{-1}(1-\frac\alpha2))+\Phi(\dots-\Phi^{-1}(1-\frac\alpha2))$

p-value: $|T_n|\ge z_{\alpha/2}$, $p=2\Phi(-T_n)=2(1-\Phi(T_n))$

$P_0(p\le u)=P_0(\inf\{\alpha:\{X_1,\dots,X_n\}\in R_\alpha\}\le u)$

$=P_0(\{X_1,\dots,X_n\}\in R_l,\forall l>u)$

$=P_0(\bigcap_{l>u}\{X_1,\dots,X_n\}\in R_l)$

$=\lim_{l\to u}P_0(\{X_1,\dots,X_n\}\in R_l)$

$\le P_0(\{X_1,\dots,X_n\}\in R_u)\le u$ level.

$Exp(\lambda)$, $\lambda e^{-\lambda x}$, $1-e^{-\lambda x}$, $E=\frac1\lambda$, $V=\frac{1}{\lambda^2}$, $\beta=\frac1\lambda$

$Poi(\lambda)$. $f(k)=\frac{\lambda^ke^{-\lambda}}{k!}$

KL-divergence: $D(f\|g)=\int_xf(x)\log(\frac{f(x)}{g(x)})dx$

MLE ⟺ $M_n(\theta)=\frac1n\sum_{i=1}^n\log\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)}\to E_{X_i\sim f_{\theta^*}}\log\frac{f_\theta(X_i)}{f_{\theta^*}(X_i)}=-D(f_{\theta^*}\|f_\theta)$

score: $S_\theta(X)=\frac{\partial\log f_\theta(X)}{\partial\theta}$, $E_{X\sim f_\theta}S_\theta(X)=0$

proof. $\int f_\theta(x)dx=1\to0=\int\frac{\partial\log f_\theta(x)}{\partial\theta}f_\theta(x)dx$

MLE: $\hat\theta\to\frac1n\sum^nS_{\hat\theta}(X)=0$, by LLN $\quad E_{\theta^*}S_\theta(X)=0$, $\hat\theta=\theta^*$

Fisher I: $I(\theta)=Var_\theta(S_\theta(X))=E(S_\theta(X)^2)=-E_\theta[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)]$

$\frac{\partial}{\partial\theta}E_\theta[\frac{\partial\log f_\theta(X)}{\partial\theta}]=0$

$\int_x[\frac{\partial}{\partial\theta}f_\theta(x)\frac{\partial}{\partial\theta}\log f_\theta(x)]dx=E_\theta[\frac{\partial^2}{\partial\theta^2}\log f_\theta(x)]+E[\frac{\partial}{\partial\theta}\log f_\theta(x)]^2=0$

$\sqrt{nI(\hat\theta)}(\hat\theta-\theta^*)\to N(0,1)$

observed $I_n(\hat\theta)=\frac1n\sum^n[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X_i)|_{\theta=\hat\theta}]$

Proof: $0=l'(\hat\theta)=l'(\theta^*)+(\hat\theta-\theta^*)l''(\theta^*)$

$\sqrt n(\hat\theta-\theta^*)=\frac{\frac{1}{\sqrt n}\sum^nS_{\theta^*}(X_i)}{-\frac1n\sum\frac{\partial^2}{\partial\theta^2}\log f_\theta(X_i)|_{\theta=\theta^*}}\to I(\theta^*)$

Cramer-Rao: $X_1\dots X_n\sim p(X,\theta^*)$, $\hat\theta$ unbias

$E_{X_1\dots X_n\sim p(\theta^*)}\hat\theta=\theta^*$, $Var(\hat\theta)\ge\frac{1}{nI(\theta^*)}$

$\sqrt{nI(\theta^*)}(\hat\theta_{MLE}-\theta^*)\xrightarrow{d}N(0,1)$

$E=\nabla_\theta^2P=0$

$\nabla_\theta^2\log p(X;\theta)=\nabla_\theta\frac{\nabla_\theta p(X;\theta)}{p(X;\theta)}$

$=\frac{\nabla_\theta^2p(X;\theta)}{p(X;\theta)}-\frac{\nabla_\theta p(X;\theta)\nabla_\theta p(X;\theta)^T}{p(X;\theta)^2}=\frac{\nabla_\theta^2p(X;\theta)}{p(X;\theta)}-S(\theta)S(\theta)^T$

Estimating CDF

$\hat F_n(x)=\frac{\sum\mathbb I(X_i\le x)}{n}$

$E[\hat F_n(x)]=F(x)$, bias=0, $Var[\hat F_n(x)]=\frac{P(X\le x)(1-P(X\le x))}{n}$

MSE $=0+Var[\hat F_n(x)]$

Glivenko-Cantelli: $\sup_x|\hat F_n(x)-F(x)|\to0$

DKW $P(\sup_x|\hat F_n(x)-F(x)|\ge\epsilon)\le2\exp(-2n\epsilon^2)$

$T(F)=T(\hat F)$ Plug-in $\hat\sigma^2=\int x^2d\hat F_n(x)-(\int xd\hat F_n(x))^2$

$\hat\mu=T(\hat F)=\frac1n\sum X_i$ $\quad=\frac1n\sum X_i^2-(\frac1n\sum X_i)^2$ $\quad\frac1n\sum(X_i-\bar X)^2$