Random Variables & Moments

1. Bernoulli $P = p^x(1-p)^{1-x}$

$\mu = p$ var $= p(1-p)$

$M_X(t) = (1-p) + pe^t$

2. Binomial $P(X=k) = \binom{n}{k}p^k(1-p)^k$

$\mu = np$ var $= np(1-p)$

$M_X(t) = [(1-p)+pe^t]^n$

3. Poisson $P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$

$\mu = \lambda$ var $= \lambda$

$M_X(t) = e^{\lambda(e^t-1)}$

4. Uniform $f(x) = \frac{1}{b-a}$

$\mu = \frac{b-a}{2}$ $V = \frac{(b-a)^2}{12}$

$M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}, t \neq 0$

$= 1, t = 0$

5. Exponential $f(x) = \lambda e^{-\lambda x}$

$\mu = \frac{1}{\lambda}$ $V = \frac{1}{\lambda^2}$

$M_X(t) = \frac{\lambda}{\lambda - t}$

6. Gamma $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, x \geq 0$

$\mu = \frac{\alpha}{\lambda}$ $V = \frac{\alpha}{\lambda^2}$

$M_X(t) = (1-\frac{t}{\lambda})^{-\alpha} = (\frac{\lambda}{\lambda - t})^\alpha$

$\Gamma(\alpha) = (\alpha-1)!$ $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$

7. Normal $f(x) = \frac{1}{\delta\sqrt{2\pi}}e^{-(x-\mu)^2/(2\delta^2)}$

$\mu = \mu$ $V = \delta^2$

$M_X(t) = e^{\mu t}e^{\frac{\delta t^2}{2}}$

Method of moment

1. $M_X(t) = E(e^{tX}) = \sum e^{tx}p(x)$

$= \int e^{tx}f(x)dx$

2. $M_X^{(r)}(0) = E(X^r)$ e.g. $M_X''(0) = E(X^2)$

3. ① k-th moment $\mu_k = M_X^{(k)}(0) = E(X^k)$.

② $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$ as estimate.

③ $\theta = f(\mu_1, \mu_2) \leftarrow \hat{\mu}_1, \hat{\mu}_2$

θ 可以是 δ^2、μ 等

Method of MLE

① $Lik(\theta) = f(x_1 \cdots x_n|\theta)$ joind pdf

if X_i iid, $Lik(\theta) = \prod_{i=1}^n f(x_i|\theta)$

② $l(\theta) = \log[Lik(\theta)] = \sum_{i=1}^n \log(f)$

③ $l'(\theta) = 0$ $(\frac{\partial}{\partial\mu}, \frac{\partial}{\partial\delta}) \Rightarrow \max_\theta$

$Var[l'(\theta)] = nI(\theta_0)$

asymtotic var of mle

$Var[\hat{\theta} - \theta_0] = \frac{1}{nI(\theta_0)}$

Fisher info ☆

$I(\theta) = E[\frac{\partial}{\partial\theta}\log f(x|\theta)]^2$

$= -E[\frac{\partial^2}{\partial^2\theta}\log f(x|\theta)]$

$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)$ tends to be $N(0,1)$

sufficient stats

by factorization theorem

$f(x_1 \cdots x_n|\theta) = g[T(x_1 \cdots x_n), \theta] \cdot \overset{\text{可以是1}}{h(x_1 \cdots x_n)}$

$T(x_1 \cdots x_n|\theta)$ is sufficient.

$l(\text{normal}) = \log(\frac{1}{(2\pi)^{\frac{n}{2}}}) - n\log\delta - \frac{1}{2}\frac{\sum(x_i-\mu)^2}{\delta^2}$

Cramer-Rao lower bound (estimator var).

$\frac{1}{nI(\theta)}$ (for normal: $\frac{\delta^2}{n}$).

Jacobian

X, Y jointly variable

$u = g_1(X,Y) = a + bX$

$v = g_2(X,Y) = c + dY$

$x = (u-a)/b$

$y = (v-c)/d$

$J(x,y) = \det\begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{pmatrix}$

Joint density of U and V

$f_{UV}(u,v) = f_{XY}(x,y)|J^{-1}(x,y)|$

where $x = \frac{u-a}{b}$ $y = \frac{v-c}{d}$

1. Permutation test

T can be any statistic, usually t-stats

$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2(\frac{1}{n_1}+\frac{1}{n_2})}}$

2. Mann-Whitney U-test

$T(X_1 \cdots X_n, Y_1 \cdots Y_n) = \sum_{i=1}^n R_i$ (sum of ranks for x)

now do permutation test

3. Wilcoxon signed rank test

define $R_i = \text{rank}(|x_i - y_i|)$

$T(x_1 \cdots x_n, y_1 \cdots y_n) = \sum_{i=1}^n \text{sign}(x_i - y_i) R_i$

sign $= 1$, $x_i - y_i > 0$

$= -1$, $x_i - y_i < 0$ (bernolli with $p = \frac{1}{2}$)

p-value $= P(T \geq T_{obs})$.

4. Histogram.

sample $X_1 \cdots X_n \sim$ i.i.d. F. want to esti p

no of bins: M, divide $[0,1]$ to M bins

① estimator

$\hat{p}_n(x) = \frac{x \text{ in } B_l}{n} \times \frac{1}{\text{length of bin}}$

$= \frac{M}{n}\sum_{i=1}^n 1(X_i \in B_l)$

② $MSE(x) = E\{[\hat{p}_n(x) - p(x)]^2\}$

$= \underbrace{E\{\hat{p}_n(x) - E\hat{p}_n(x)\}}_{var} + \underbrace{[E\hat{p}_n(x) - p(x)]^2}_{bias^2}$

↓ 也可以用普通方法算.

③ rate of convergence $= n^{-\frac{2}{3}}$ (para: n^{-1})

④ $MISE = \int E\{[\hat{p}_n(x) - p(x)]^2\}dx$

Theorem: if f' continuous & $\int (f'(u))^2 du < \infty$

then $MISE = \frac{h^2}{12}\int (f'(u))^2 du + \frac{1}{nh} + o(h^2) + O(\frac{1}{n})$

$h^* = n^{-\frac{1}{3}}(\frac{6}{\int (f'(u))^2 du})^{\frac{1}{3}}$ (optimal bin width)

under h^*, $MISE = \frac{C}{n^{\frac{2}{3}}}$

(4 $C = (\frac{\sqrt{3}}{\ })^{\frac{2}{3}}$ $(\int f'(u))^2 du)^{\frac{1}{3}}$

Kernel density estimate.

① estimator

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$$

② MSE

$$R_x = \frac{1}{4}\sigma_K^4 h^4 (f''(x))^2 + \frac{f(x)\int K^2(x)dx}{nh} + O(\tfrac{1}{n}) + O(h^6)$$

MISE

$$R = \underbrace{\frac{1}{4}\sigma_K^4 h^4 \int (f''(x))^2 dx}_{bias^2} + \underbrace{\frac{\int K^2(x)dx}{nh}}_{var} + O(\tfrac{1}{n}) + O(h^6)$$

$$\sigma_K = \int x^2 K(x)dx$$

实操上 bias $= E\hat{f}_n(x) - f(x)$.

var $= var(\hat{f}_n(x))$

③ optimal bin width

$h^* = C n^{-\frac{1}{5}}$; $\frac{d}{dh}\widehat{MSE} \to h$

Rate $= O(n^{-\frac{4}{5}})$

Nadaraya-Watson kernel regression

① estimator (X_i, Y_i) pairs

$$\hat{r}_n(x) = \sum_{i=1}^{n} l_i(x) Y_i$$

$$l_i(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)}$$

$$\frac{\sum K\left(\frac{x-X_i}{h}\right) Y_i}{\sum K\left(\frac{x-X_i}{h}\right)}$$

② $bias^2 = \frac{h^4}{4}\left(\int x^2 K(x)dx\right)^2 \cdot \int \left(r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right)^2 dx$

var $= \frac{\sigma^2 \int K^2(x)dx}{nh} \int \frac{1}{f(x)} dx + O(nh^{-1}) + O(h^4)$

③ h^* : LOOCV

Bootstrap.

① use plug-in estimator for $T(X_1 \cdots X_n)$

② $V_{boot} = \frac{1}{B}\sum_{b=1}^{B}\left(T_{n,b}^* - \frac{1}{B}\sum_{r=1}^{B} T_{n,r}^*\right)^2$

③ CIs

1) Normal interval: $T_n - ET_n \sim N(0, \hat{se}_{boot}^2)$

$CI_\alpha = (T_n \mp z_{\alpha/2}\hat{se}_{boot})$

2). pivotal: $C_n = (2\hat{\theta}_n - \hat{\theta}^*_{((1-\alpha/2)B)},$

$2\hat{\theta}_n - \hat{\theta}^*_{((\alpha/2)B)})$ B is no. bootstraps

3) studentized pivotal:

$(T_n - z^*_{1-\alpha/2}\hat{se}_{boot},\ T_n - z^*_{\alpha/2}\hat{se}_{boot})$

z^*_β is β-th quantile of $z^*_{n,1} \cdots z^*_{n,B}$

$z^*_{n,b} = \frac{T^*_{n,b} - T_n}{\hat{se}^*_b}$

4) percentile $C_n = (T^*_{(B\alpha/2)}, T^*_{(B(1-\alpha/2))})$

Monte Carlo sampling

1. Direct sampling : inverse CDF

$Pr(F^{-1}(u) \le x) = F(x)$.

2. Rejection sampling

$h(\theta) = c\, \underbrace{f(\theta)}_{true}$, proposal density $g(\theta)$

$h(\theta) \le M g(\theta)$.

Acceptance Rate $= \frac{\text{Area under } h(\theta)}{\text{Area under } Mg(\theta)} = \frac{\int_0^1 h(\theta)d\theta}{M}$

① method : generate $\theta \sim g(\theta)$, $u \sim uni[0,1]$

if $u < \frac{f(\theta)}{Mg(\theta)}$, accept θ, else reject

② select M

$\log M = \max_\theta \{\log h(\theta) - \log g(\theta)\}$

③ normalization const c

$c = M \cdot AR$

Importance sampling

① weights: $w(\theta) = \frac{h(\theta)}{g(\theta)}$

② estimator: $\frac{\frac{1}{n}\sum^n w(\theta_i)\, k(\theta_i)}{\frac{1}{n}\sum^n w(\theta_i)}$

③ ESS $= \frac{1}{\sum^n \hat{w}_i}$

$\hat{w}_i = \frac{w(\theta_i)}{\sum^n w(\theta_i)}$

Markov chains

$P(x,y)$, $\pi(y) \to$ long-term stationary distr

①. simple random walk

$p(x,y) = \frac{1}{deg(x)}$, if $y \sim x$, else 0

$\pi(y) = \frac{deg(y)}{2|E|}$

② convergence to a unique $\pi(y)$

conditions: irreducible & aperiodic

沿绝路　不certain何时回原点

Markov chain monte carlo (MCMC)

stochastic process:

$$Q(x,y) = \begin{cases} P(x,y)a(x,y), & y \ne x \\ 1 - \sum_{z \ne x} P(x,z)a(x,z), & y = x \end{cases}$$

1. Metropolis-Hasting algo

① acceptance prob: $a(x,y)$

$a(x,y) = \min\left\{\frac{\pi(y)P(y,x)}{\pi(x)P(x,y)}, 1\right\}$

② implementation

1) choose starting point $\theta^{(0)}$ for $p(\theta^0|y) > 0$

2) for $t = 1, 2, \cdots$ sample θ^* from $g(\theta^*|\theta^{t-1})$

$r = \frac{f(\theta^*)\, g(\theta^{(t-1)}|\theta^*)}{f(\theta^{(t-1)})\, g(\theta^*|\theta^{(t-1)})}$

$\theta^{(t)} = \begin{cases} \theta^* & \text{with } p = \min(r, 1) \\ \theta^{(t-1)} & \text{else} \end{cases}$

2. Initialization: high p start, & close scale f

3. MCMC for inference

①. burn-in: throw away initial portion of trajectory. Solves marginal dist $\mu^{(0)}, \mu^{(1)} \ldots \mu^{(T)}$ take time to converge and $\theta^{(0)} \ldots \theta^{(T)}$ not independent

②. ESS

Lag L autocorrelation for MC & f

$\rho_L = corr(f(\theta^0), f(\theta^L))$

Sample $\hat{\rho}_L$ =

$$\frac{\sum_{t=0}^{T-L}[f(\theta^t) - \overline{f(\theta)}][f(\theta^{t+L}) - \overline{f(\theta)}]}{\sum_{t=0}^{T}(f(\theta^t) - \overline{f(\theta)})^2}$$

$\overline{f(\theta)} = \frac{1}{T+1}\sum_{t=0}^{T} f(\theta^t)$

$ESS = T\left(1 + 2\sum_{L=1}^{\infty} \rho_L\right)^{-1}$