Data Science Hw 1

工科海洋四 B07505015 梁瑞翔

Problem 1 Random Number Transformation

(1) $F_W(x)$=P[W≤x]=P[max(X,Y)≤x]=P[X≤x]P[Y≤x]= $F_X(x)F_Y(x)$

$f_W(x)$=d$F_W(x)$/dx=d($F_X(x)F_Y(x)$)/dx= $f_x(x)F_Y(x)+f_y(x)F_X(x)$

=2*e^(-x)*(1-e^(-x))

(2) P[min(X,Y)<x]=P[{X<x} or {Y<x}]=P[X<x]+P[Y<x]-P[{X<x} and {Y<x}]

=$F_X(x)+F_Y(x)-F_X(x)F_Y(x)$

$f_W(x)$= $f_x(x)$+ $f_y(x)$-2*e^(-x)*(1-e^(-x))=2e^(-2x)

Problem 2 Statistical Distances

(1) f(x, y) ≥ 0 (non-negativity)

f(x, y) = 0 if and only if x = y (identity of indiscernibles)

f(x, y) = d(y, x) (symmetry)

f(x, z) ≤ d(x, y) + d(y, z) (subadditivity / triangle inequality)

(2) d^2 (x,y)=||x−y||^2=⟨x−y,x−y⟩=||x||^2−||y||^2−⟨2y,x−y⟩

The Bregman distance is defined as: $D_F(p,q) = F(p) - F(q) - \langle \nabla F(q), p - q \rangle_{f(x)}$

Let F(x) = ||x||^2

d^2(x,y)=F(x)−F(y)-⟨∇f(y),x−y⟩)

(3) H(x,y)=H(x|y)+H(y)=H(y|x)+H(x)

y=f(x), so f(y|x)=0

H(x,y)=H(y)+H(x|y)=H(x)

H(y)<=H(x)

Problem 3 Point Estimation

(1) E(X)=$\int_0^{\theta} x\frac{1}{\theta}dx$=(1/2)*θ=X̄

θ̂=2X̄

(2) f(x|θ)=1/θ, for 0<x<θ

g(θ)=1

$\hat{\theta}_{MAP}$(x)=argmax f(x|θ)g(θ)=0

(3)f(x|θ)=1/θ, for 0<x<θ

h(θ)=1, for0<θ<1

u(x, θ)=h(θ)f(x|θ)=1/θ, for 0<x<θ<1

g(x)= $\int_x^1 u(x,\theta)d\theta$=$\int_x^1 \frac{1}{\theta}d\theta$=-ln(x), for 0<x<1

$k(\theta|x)=u(x, \theta)/g(x)=-1/(\theta\ln(x))$, for $0<x<\theta<1$

$\hat{\theta}=E[\theta|x]= \int_x^1 \theta k(\theta|x) d\theta=-1/\ln(x)*\int_x^1 d\theta=(x-1)/\ln(x)$

$\hat{\theta}=(X-1)/\ln(X)$

Problem 4 Goodness of Estimation

(1) set $c_1+c_2=1$, $Var(c_1\theta_1+c_2\theta_2)=c_1^2+2c_2^2+1/2*c_1c_2=c_1^2+2(1-c_1)^2+1/2*c_1(1-c_1)=5/2*c_1^2-7/2*c_1+2$,to minimize the expression, $c_1=7/10$, $c_2=3/10$,
unbiased estimator: $7/10*\theta_1+3/10*\theta_2$
variance: 0.775
(2) set $c_3+c_4=1$, $Var(c_3\theta_3+c_4\theta_4)=c_3^2+2c_4^2+3/2*c_3c_4=c_3^2+2(1-c_3)^2+3/2*c_3(1-c_3)=3/2*c_3^2-5/2*c_3+2$,to minimize the expression, $c_3= 5/6$, $c_4=1/6$,
unbiased estimator: $5/6*\theta_3+1/6*\theta_4$
variance: 0.9583
(3) θ_1,θ_2, whose covariance is lower.

Problem 5 Interval Estimation

(1) $P[Q<=x]=P[X(1)-\theta<=x]=P[\min[X_i]-\theta<=x]=F_{\min(X)-\theta}(x)$

$F_{\min(X)-\theta}(x)=\int_\theta^{x+\theta} e^{-(x-\theta)}\, dx$=-e^(-x)=-e^(-x)+1

Since it is independent of θ, Q is a pivotal quantity.
(2) $P(a\le Q\le b)=1-\alpha$
$\alpha=P(a\le X(1)-\theta\le b)=P(a-X(1)\le -\theta\le b-X(1))=P(b-X(1)\le\theta\le a-X(1))$
$=P(X^2_{1-\alpha/2}-X(1)\le\theta\le X^2_{\alpha/2}-X(1))$
100(1−α)% confidence interval: $[X^2_{1-\alpha/2}-X(1), X^2_{\alpha/2}-X(1)]$