A Reconsideration of Mathematical Statistics

江口 真透 数理 推論研究系

Abstract

Standard mathematical Statistics is based on the assumption of **random sample**. Under the assumption **the mathematical statistics** is formulated such as **mean**, **variance**, **exponential model**, **unbiasedness**, **sufficiency**.
However, the assumption of random sample is **away from practical problems**.
Currently the framework of mathematical statistics is not useful for researchers, which should **be reformulated towards the current interests in statistics**.
For this we need **to change the definition** of mean, variance, exponential model.

Sample mean, exponential model, maximum likelihood

Let $X_1,\dots,X_n$ be a random sample from $f(x)$. Then the sample mean $\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$ is unbiased for $\mu=\mathrm{E}_f(X)$.

Assume that the density f belongs to an exponential model $M^{(\mathrm{e})}=\{f_\theta(x):=f_0(x)\exp\{\theta^{\mathrm{T}}x-\kappa(\theta)\}:\theta\in\Theta\}$

Then $\bar{X}$ is MLE for μ, and the minmum variance unbiased estimator for μ. Fisher (1912, 1922), Wilks (1938), Cramer (1945), Cox-Hinkley (1979),

Standard information geometry

m-geodesic $\mathrm{C}^{(\mathrm{m})}_{f,g}=\{f^{(\mathrm{m})}_t(x):=(1-t)g(x)+tf(x):t\in[0,1]\}$ e-geodesic $\mathrm{C}^{(\mathrm{e})}_{f,g}=\{f^{(\mathrm{e})}_t(x)=e^{(1-t)\log g(x)+t\log f(x)-\kappa(t)}:t\in[0,1]\}$

KL divergence $D_0(f,g)=\mathrm{E}_f(\log f-\log g)$ satisfies $\mathrm{C}^{(\mathrm{m})}_{f,g}\perp_g\mathrm{C}^{(\mathrm{e})}_{g,h}\Leftrightarrow D_0(f,h)=D_0(f,g)+D_0(g,h)$ (Amari-Nagaoka, 2001)

Selective sample, IPW estimate

We are not given a random sample $\{X_i\}_{1\le i\le n}$ from $f_\theta(x)$ But we are given a selected sample $\{X^*_i\}_{1\le i\le n}$ from

$$f^*_\theta(x,S=s)=P(S=s\mid x)f_\theta(x)=p(s)f_\theta(x)\,r(x,s)\quad\text{where}\quad p(s)=1/\int r(y,s)f_\theta(x)dx$$

Inverse probability weighted estimator $\hat{\mu}=\dfrac{\sum_{i=1}^n\frac{X_i}{r(X_i,S_i)}}{\sum_{i=1}^n\frac{1}{r(X_i,S_i)}}$ **Consistency** $\hat{\mu}\xrightarrow{\text{a.s.}}\dfrac{\mathrm{E}_f(\frac{X}{r(X,S)})}{\mathrm{E}_f(\frac{1}{r(X,S)})}=\mu$

Cf. Horovitz-Thompson (1952), Heckman (1979), Rubin (1976), Rosenbaum-Rubin (1983), Bang-Robins (2004)

Generalized Information geometry

Generalized m-geodesic $C^{(\phi*)}_{f,g}=\{f^{(\phi*)}_t(x):t\in[0,1]\}$ s.t.

$$\frac{\frac{1}{\phi'(f^{(\phi*)}_t(x))}}{\int\frac{1}{\phi'(f^{(\phi*)}_t)}dP}=(1-t)\frac{\frac{1}{\phi'(f(x))}}{\int\frac{1}{\phi'(f)}dP}+t\frac{\frac{1}{\phi'(g(x))}}{\int\frac{1}{\phi'(g)}dP}$$

Generalized e-geodesic $C^{(\phi)}_{f,g}=\{f^{(\phi)}_t(x):=\phi^{-1}((1-t)\phi(f(x))+t\phi(g(x))-\kappa^{(\phi)}(t)):t\in[0,1]\}$

Generalized KL divergence $D_\phi(f,g)=\dfrac{\int\{\phi(f)-\phi(g)\}/\phi'(f)dP}{\int 1/\phi'(f)dP}$ satisfies $\mathrm{C}^{(\phi*)}_{f,g}\perp_g\mathrm{C}^{(\phi)}_{g,h}\Leftrightarrow D_\phi(f,h)=D_\phi(f,g)+D_\phi(g,h)$

Generalized mean and variance

$$\mathrm{E}^{(\phi)}_f\{a(X)\}=\frac{\int\frac{a}{\phi'(f)}dP}{\int\frac{1}{\phi'(f)}dP}\qquad \mathrm{Cov}^{(\phi)}_f\{a(X),b(X)\}=\frac{\int\frac{-\phi''(f)(a-\mathrm{E}^{(\phi)}a)(b-\mathrm{E}^{(\phi)}b)^{\mathrm{T}}}{\{\phi'(f)\}^2}dP}{\int\frac{1}{\phi'(f)}dP}$$

Maximum ϕ - estimation

Let $\{X_i\}_{1\le i\le n}$ be a selective sample from $f^*_\theta(x,S=s)=P(S=s\mid x)f_\theta(x)=p(s)f_\theta(x)r(x,s)$

ϕ-utility function $L^{(\phi)}(\theta)=-\sum_{i=1}^n\dfrac{\phi(1/n)}{\phi'(r(X_i,S_i))}\log f_\theta(X_i)$ **Maximum ϕ-estimator** $\hat{\theta}^{(\phi)}=\underset{\theta\in\Theta}{\arg\max}\,L^{(\phi)}(\theta)$

If $\phi(t)=\frac{1}{2}t^2$, then $\hat{\mu}^{(\phi)}=\dfrac{\sum_{i=1}^n\frac{t(X_i)}{r(X_i,S_i)}}{\sum_{i=1}^n\frac{1}{r(X_i,S_i)}}$ **(Max ϕ - estimator = Inverse probability weighted estimator)**