STATISTICAL INFERENCE AND APPLICATIONS 52325
TAKE B 2018

PAVEL CHIGANSKY

Problem 1 and 2

(similar to Home Assignments)

Problem 3

A top with three sides is spun n times independently and the outcomes $X_1, ..., X_n$ are recorded. Denote by N_1 and N_2 the number of 1's and 2's in the sample:

$$N_j = \sum_{m=1}^{n} \mathbf{1}_{\{X_m=j\}}, \quad j = 1, 2.$$

Assume that probability of getting 2 is known and equals $\frac{1}{2}$, while the probability $\theta \in (0, \frac{1}{2})$ to get 1 is unknown.

(1) Find the minimal sufficient statistic for this model

The likelihood function is

$$L(X;\theta) = \theta^{N_1(X)}(\tfrac{1}{2})^{N_2(X)}(\tfrac{1}{2}-\theta)^{n-N_1(X)-N_2(X)}$$

hence (N_1, N_2) is sufficient. To argue for its minimality, consider the ratio

$$\frac{L(x;\theta)}{L(y;\theta)} = \theta^{N_1(x)-N_1(y)}(\tfrac{1}{2}-\theta)^{-(N_1(x)-N_1(y))-(N_2(x)-N_2(y))}$$

This is a constant function of θ if and only if $N_1(x) = N_1(y)$ and $N_2(x) = N_2(y)$. Hence the sufficient statistic is minimal.

(2) Show that the estimator $\widehat{\theta}(X) = N_1/n$ is unbiased for θ and find its m.s.e. risk function.

The estimator is obviously unbiased:

$$\mathbb{E}_\theta \widehat{\theta}(X) = \frac{n\theta}{n} = \theta$$

and its m.s.e. risk equals its variance:

$$\mathrm{Var}_\theta(\widehat{\theta}(X)) = \frac{1}{n}\theta(1-\theta)$$

which does not attain the C-R bound. Hence this estimator is inefficient.

(3) Calculate the Fisher information. Is the estimator from the previous question efficient?

We have

$$\frac{d}{d\theta}\log\left(\theta^{\mathbf{1}_{\{X_1=1\}}}(\frac{1}{2})^{\mathbf{1}_{\{X_1=2\}}}(\frac{1}{2}-\theta)^{\mathbf{1}_{\{X_1=3\}}}\right)=\mathbf{1}_{\{X_1=1\}}\frac{1}{\theta}-\mathbf{1}_{\{X_1=3\}}\frac{1}{\frac{1}{2}-\theta}$$

and

$$\frac{d^2}{d\theta^2}\log\left(\theta^{\mathbf{1}_{\{X_1=1\}}}(\frac{1}{2})^{\mathbf{1}_{\{X_1=2\}}}(\frac{1}{2}-\theta)^{\mathbf{1}_{\{X_1=3\}}}\right)=-\mathbf{1}_{\{X_1=1\}}\frac{1}{\theta^2}-\mathbf{1}_{\{X_1=3\}}\frac{1}{(\frac{1}{2}-\theta)^2}.$$

Hence the Fisher information is

$$I_n(\theta)=n\left(\mathbb{P}_\theta(X_1=1)\frac{1}{\theta^2}+\mathbb{P}_\theta(X_1=3)\frac{1}{(\frac{1}{2}-\theta)^2}\right)=$$
$$n\left(\frac{1}{\theta}+\frac{1}{(\frac{1}{2}-\theta)}\right)=n\frac{\frac{1}{2}}{\theta(\frac{1}{2}-\theta)}$$

The estimator from the previous question is not efficient, that is, it does not attain the C-R bound.

(4) Consider the unbiased estimators of the form

$$\widetilde{\theta}_\alpha(X)=N_1/n+\alpha(N_2/n-\tfrac{1}{2})$$

where $\alpha\in\mathbb{R}$ is a design parameter (to be chosen by the statistician without knowing the true value of θ). Can α be chosen so that $\widetilde{\theta}^\alpha$ has a better risk than $\widehat{\theta}$? If yes, find such value(s), if not, prove your answer.

The m.s.e. risk of this estimator is given by

$$\mathrm{Var}_\theta(\widetilde{\theta}_\alpha)=\mathrm{Var}_\theta(N_1/n)+\alpha^2\,\mathrm{Var}_\theta(N_2/n-\tfrac{1}{2})+2\alpha\,\mathrm{Cov}_\theta(N_1/n,N_2/n-\tfrac{1}{2})=$$
$$\mathrm{Var}_\theta(\widehat{\theta}_n)+\alpha^2\,\mathrm{Var}_\theta(N_2/n)+2\alpha\,\mathrm{Cov}_\theta(N_1/n,N_2/n-\tfrac{1}{2}).$$

Here $\mathrm{Var}_\theta(N_2/n)=\frac{1}{n}\frac{1}{4}$ and

$$\mathrm{Cov}_\theta(N_1/n,N_2/n-\tfrac{1}{2})=\frac{1}{n^2}\mathbb{E}_\theta N_1N_2-\tfrac{1}{2}\mathbb{E}_\theta N_1/n=$$
$$\frac{1}{n^2}\mathbb{E}_\theta\sum_{m=1}^n\sum_{k=1}^n\mathbf{1}_{\{X_m=1\}}\mathbf{1}_{\{X_k=2\}}-\tfrac{1}{2}\theta=(1-1/n)\tfrac{1}{2}\theta-\tfrac{1}{2}\theta=-\frac{1}{n}\frac{1}{2}\theta.$$

Hence

$$\mathrm{Var}_\theta(\widetilde{\theta}_\alpha)-\mathrm{Var}_\theta(\widehat{\theta}_n)=\frac{1}{n}\frac{1}{4}\left(\alpha^2-4\alpha\theta\right)$$

and $\widetilde{\theta}_\alpha$ will be better than $\widehat{\theta}_n$ if

$$\alpha^2-4\alpha\theta<0,\quad\forall\theta\in(0,1/2).$$

Since the left hand side is positive for all $\alpha<0$, this inequality holds if and only if $\alpha\in(0,4\theta)$. But this inclusion fails for any positive choice of α for all θ's on a sufficiently small vicinity of

the origin. Thus all estimators $\widetilde{\theta}_\alpha$ are not comparable with $\widehat{\theta}_n$ and, in particular, none of them improves $\widehat{\theta}_n$.

Department of Statistics, The Hebrew University, Mount Scopus, Jerusalem 91905, Israel
E-mail address: pchiga@mscc.huji.ac.il