Theory of Statistical Inference - Lecture III.5
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

- conditioning via an ancillary statistic is a necessary part of the frequentist approach to statistical inference and is motivated by the following example

Example III.5.1 *Cox's two measuring instruments*

- a physics experimenter is going to use a voltmeter as part of an experiment they are running

- they make a to the laboratory stores which has two voltmeters

- voltmeter i returns a measurement according to a $N(\mu, \sigma_i^2)$ distribution where μ is the true voltage where $\sigma_1^2 << \sigma_2^2$

- the manager of the stores randomly (prob. $1/2$) allocates the voltmeter and the experimenter knows whic on ethey received, namely voltmeter 1

- after obtaining n measurements the experimenter records $\bar{x} \pm 1.96\sigma_1/\sqrt{n}$ as a measure of the accuracy of the estimate $\bar{x}$ of the true voltage

- but it is claimed that this interval is not correct because they could have received voltmeter 2 and so should have used the mixture distribution $0.5N(\mu, \sigma_1^2) + 0.5N(\mu, \sigma_2^2)$ to determine the correct confidence interval which results in the interval $\bar{x} \pm 1.96\sqrt{(\sigma_1^2 + \sigma_2^2)/2n}$ which suggests much less accuracy in the results because $(\sigma_1^2 + \sigma_2^2)/2 >> \sigma_1^2$

- if we assess accuracy via repeated sampling then strictly speaking this is correct

- this seems absurd ■

Definition III.5.1 A function $A : \mathcal{X} \to \mathcal{A}$ is an *ancillary statistic* for the model $\{f_\theta : \theta \in \Theta\}$ if $P_{\theta A} = P_A$ for every $\theta \in \Theta$. ■

- so the observed value $A(x)$ of an ancillary statistic can tell you nothing about the true value of θ

- it is also claimed that all inferences should be conditional on the value of an ancillary, namely, based on the conditional model $\{f_\theta(\cdot \,|\, A(x)) : \theta \in \Theta\}$ as this leads to more appropriate assessments of accuracy under repeated sampling

Example III.5.1 *Cox's two measuring instruments (continued)*

- the choice of the measuring instrument has a distribution that does not depend on μ and so is ancillary

- conditioning on this choice resolves the problem as the correct CI is then $\bar{x} \pm 1.96\sigma_1/\sqrt{n}$ ■

- if there are more than two anclliaries, which should we condition on?

- suppose A_1 and A_2 are ancillaries and there exists function h such that $A_1(x) = h(A_2(x))$ and h is not 1-1

- then it seems that conditioning on A_2 is more appropriate because because A_2 is ancillary for the model $\{f_\theta(\cdot \mid A_1(x)) : \theta \in \Theta\}$

Definition III.5.2 A function $A : \mathcal{X} \to \mathcal{A}$ is a *maximal ancillary statistic* for the model $\{f_\theta : \theta \in \Theta\}$ if it is ancillary and, if $A(x) = h(A_1(x))$ for any other ancillary A_1 for some function h, then h is 1-1. ■

- there is a problem as a maximum ancillary for a model is not unique, so which one to use

Example III.5.2

- consider the model

(x_1, x_2)	$(1, 1)$	$(1, 2)$	$(2, 1)$	$(2, 2)$
$f_1(x_1, x_2)$	$1/6$	$1/6$	$2/6$	$2/6$
$f_2(x_1, x_2)$	$1/12$	$3/12$	$5/12$	$3/12$

- $U(x_1, x_2) = x_1$ and $V(x_1, x_2) = x_2$ are both maximal ancillary and the conditional models, when $(x_1, x_2) = (1, 1)$ is observed, are as follows

(x_1, x_2)	$(1, 1)$	$(1, 2)$	$(2, 1)$	$(2, 2)$
$f_1(x_1, x_2 \mid U = 1)$	$1/2$	$1/2$	0	0
$f_2(x_1, x_2 \mid U = 1)$	$1/4$	$3/4$	0	0

(x_1, x_2)	$(1, 1)$	$(1, 2)$	$(2, 1)$	$(2, 2)$
$f_{M,1}(x_1, x_2 \mid V = 1)$	$1/3$	0	$2/3$	0
$f_{M,2}(x_1, x_2 \mid V = 1)$	$1/6$	0	$5/6$	0

- so which maximal ancillary to use for conditioning? ■

- there is a resolution of this issue that will be discussed in the next section

- pure likelihood inference (and Bayesian inference) have no need for conditioning via an ancillary as the conditional likelihood is proportional to

$$f_\theta(x \mid T(x)) = \frac{f_\theta(x) J_T(x)}{f_T(T(x))} \propto L(\theta \mid x)$$

- on the other hand an ancillary statistic T provides a natural approach to model checking for if the observed value $T(x)$ is a surprising value form its distribution, namely, out in the tails of f_T, then this is an indcation something is amiss

Example III.5.3

- suppose $x = (x_1, \ldots, x_n)^t$ is a sample from the $\{N(\mu, \sigma_0^2) : \mu \in \mathbb{R}\}$ model (the variance is known)

- then we have $x \leftrightarrow (\bar{x}, x - \bar{x}1_n)$ where $1_n = (1, \ldots, 1) \in \mathbb{R}^n$ so $T(x) = x - \bar{x}1_n$ is the vector of residuals

- note that $x = \mu 1_n + \sigma_0 z$ where $z = (z_1, \ldots, z_n)^t \sim N_n(0, I)$

- therefore

$$T(x) = x - \bar{x}1_n = \mu 1_n + \sigma_0 z - (\mu + \sigma_0 \bar{z})1_n = \sigma_0(z - \bar{z}1_n)$$

so the distribution of $T(x)$ is free of μ and so is ancillary

- functions of $T(x)$ can thus be used for model checking such as $s^2 = (x - \bar{x}1_n)^t(x - \bar{x}1_n) \sim \sigma_0^2$chi-squared$(n-1)$

- in this case, $(x-\mu 1_n)^t(x-\mu 1_n) = n(\bar{x}-\mu)^2 + T^t(x)T(x)$

$$\begin{aligned} f_{\theta\mu}(x \mid T(x)) &= \frac{f_\mu(x)J_T(x)}{f_T(T(x))} \\ &= (2\pi\sigma_0^2)^{-n/2}\exp(-(x-\mu 1_n)^t(x-\mu 1_n)/2\sigma_0^2)\frac{J_T(x)}{f_T(T(x))} \\ &= (2\pi\sigma_0^2)^{-1/2}\exp(-n(\bar{x}-\mu)^2/2\sigma_0^2)\times \\ &\quad \left\{(2\pi\sigma_0^2)^{-(n-1)/2}\exp(-T^t(x)T(x)/2\sigma_0^2)\frac{J_T(x)}{f_T(T(x))}\right\} \end{aligned}$$

where

$$J_T(x) = \left|\det\left(\frac{\partial(x_i-\bar{x})}{\partial x_j}\right)\right|^{-1/2} = \begin{vmatrix} 1-1/n & -1/n & \cdots & \\ -1/n & \ddots & & \\ \vdots & & \cdots & 1-1/n \end{vmatrix}^{-1/2}$$

and so $\bar{x} \sim N(\mu, \sigma_0^2/n)$ statistically independent of $T(x)$