Theory of Statistical Inference - Lecture II.2
STA422 and STA2162

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

2026

- given probability model $(\Omega, \mathcal{A}, P)$ where $P(A)$ measures the belief that the unknown $\omega \in A \in \mathcal{A}$

- suppose it is learned that $\omega \in C \in \mathcal{A}$ where $P(C) > 0$

- how does this change the belief that $\omega \in A$?

> *Principle of Conditional Probability: if $P(A)$ is the initial probability assigned to event A and event C is observed to be true where $P(C) > 0$, then our belief in the truth of A is now given by $P(A \,|\, C) = \frac{P(A \cap C)}{P(C)}$, the conditional probability of A given that C is true.*

- this is a basic axiom of inference but there are some cautions to bear in mind

Example II.5.1 *Three Prisoners Problem*

- suppose there are 3 prisoners I, II and III

- they are told two will be executed and one will live and that this has been decided

- prisoner I asks the jailer to name one of the other prisoners who will be executed and is told II as the jailer doesn't believe this information is of any value to I

- I's intial belief that they will live is, via the Principle of Insufficient Reason, $1/3$ but after receiving this information concludes this probability is now $1/2$ and so feels some relief

- is that correct?

- possible outcomes

$$\omega \in \Omega = \{(I, II, III), (I, III, II), (II, III, I)\}$$

where first two coordinates indicate those to be executed

- $A = \{(II, III, I)\}$

- since I believes the assignment as to who will live is is random (meaning uniform here) so $P(A) = 1/3$

- I assumes $C = \{(I, II, III), (II, III, I)\}$ where $P(C) = 2/3$ and so

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A)}{P(C)} = \frac{1/3}{2/3} = \frac{1}{2}$$

- but consider different scenarios for how the jailer generates the information "II will be executed"

- suppose, the jailer generates a random variable U where $P(U = II) = p, P(U = III) = 1 - p$ and let $\Xi(\omega, U)$ be the report then

$$\begin{aligned} \Xi((I, II, III), U) &= II \text{ for all } U \\ \Xi((I, III, II), U) &= III \text{ for all } U \\ \Xi((II, III, I), II) &= II \\ \Xi((II, III, I), III) &= III \end{aligned}$$

- then by TTP

$$\begin{aligned} P(\text{"}II \text{ is reported"}) &= \frac{1}{3}.1 + \frac{1}{3}.0 + \frac{1}{3}.p = \frac{1}{3}(1 + p) \\ P(\text{"}II \text{ is reported and } I \text{ lives"}) &= \frac{1}{3}p \\ P(\text{"}I \text{ lives"} \,|\, \text{"}II \text{ is reported"}) &= \frac{p}{1 + p} \in \left[0, \frac{1}{2}\right] \end{aligned}$$

and the conditional probability is only $1/2$ when $p = 1$, is $1/3$ when $p = 1/2$ (Principle of Insufficient Reason again?) and is $1/11$ when $p = 1/10$, etc.

- moral

When conditioning on information we need to know exactly how the information was generated, namely, we condition on the value of a known function (the information generator).

Exercise II.5.1 Suppose in the Three Prisoners Problem the Principle of Insufficient Reason is applied to p so that $p \sim U[0,1]$, (p is a nuisance parameter). What is the value of $P(\text{"}I \text{ lives"} \,|\, \text{"}II \text{ is reported"})$ under these circumstances?

Exercise II.5.2 *Monte Hall Problem*

A contestant in a game show is given the opportunity to open one of three doors. Behind one door is a pot of gold and behind each of the other doors is a goat. The contestant picks a door but before opening it, the host (Monte Hall) opens one of the other doors which reveals a goat and the offers the contestant the opportunity to switch. Is there any benefit to switching? Analyze the situation as we did with the Three Prisoners Problem.

- what about when $P(C) = 0$?

- suppose $\mathbf{X} \in \mathbb{R}^k$ has a.c. distribution with density function $f_{\mathbf{X}}$ and $T : \mathbb{R}^k \rightarrow R^l$ is smooth $(k \geq l)$

- then the conditional density function of $\mathbf{X}$ given $T(\mathbf{X}) = \mathbf{y}$ is, when $\mathbf{x} \in C = T^{-1}\{\mathbf{y}\}$,

$$f_{\mathbf{X}\,|\,\mathbf{Y}}(\mathbf{x}\,|\,\mathbf{y}) = \lim_{\delta_1 \downarrow 0, \delta_2 \downarrow 0} \frac{P_{\mathbf{X}}(B_{\delta_1}(\mathbf{x})\,|\,B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_1}(x))}$$

$$= \lim_{\delta_1 \downarrow 0, \delta_2 \downarrow 0} \left\{ \frac{\frac{P_{\mathbf{X}}(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))} \frac{Vol(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_1}(x)Vol(B_{\delta_2}(\mathbf{y}))}}{\frac{P_{\mathbf{Y}}(B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_2}(\mathbf{y}))}} \right\}$$

$$\overset{\text{fact}}{=} \frac{f_{\mathbf{X}}(\mathbf{x})J_T(x)}{f_T(\mathbf{y})} \text{ where}$$

$$J_T(\mathbf{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_l(\mathbf{x})}{\partial x_k} \end{pmatrix} \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_l(\mathbf{x})}{\partial x_k} \end{pmatrix}' \right|^{-1/2}$$

- problems do arise where $P(C) = 0$ and we are not told how C was generated

Example II.5.2 *Borel paradox*

- $\Omega = \mathbb{R}^2$ and $\mathbf{X} \sim N_2(\mathbf{0}, I)$

- $C = \{\mathbf{x} : x_1 = x_2\}$

- what is the conditional distribution of $\mathbf{X}$ given $\mathbf{X} \in C$?

- this is an ill-specified problem as it has not been specified how C was determined

- e.g. $T(\mathbf{X}) = X_2 - X_1 \sim N(0, 2)$ and $C = T^{-1}\{0\}$ gives conditional density

$$f(\mathbf{x} \,|\, C) = \frac{(2\pi)^{-1}\exp(-\mathbf{x}'\mathbf{x}/2)}{(2\pi)^{-1/2}2^{-1/2}}$$

with $\mathbf{x}'\mathbf{x} = 0$

- but if $T(\mathbf{X}) = X_2/X_1$ then a completely different conditional distribution of $\mathbf{X}$ on C is determined (**Exercise II.5.3**) ■

- in the continuous case you need to fully specify a partition to determine the conditional distribution uniquely

- this is equivalent to specifying how the information was generated