14. Likelihood of a Discrete Distribution

Preparation: Equivalent Expressions for the pmf of a Bernoulli Distribution

1/1 point (graded)

Which of the following function $f(x)$, when restricted to the domain $x \in \{0,1\}$, is equal to the pmf f of the probability distribution $\mathrm{Ber}(p)$? Assume that $p \in (0,1)$. (Choose all that apply.) (Recall that if $X \sim \mathrm{Ber}(p)$, then $p = P(X=1)$.)

- [x] $f(x) = \begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$
- [x] $f(x) = p^x(1-p)^{1-x}$
- [x] $f(x) = xp + (1-x)(1-p)$
- [ ] $f(x) = \begin{cases} 1 & \text{if } x=1 \\ 0 & \text{if } x=0 \end{cases}$

Solution:

We will explain in the order of the choices.

- $f(x) = \begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \end{cases}$ is correct. A random variable $X \sim \mathrm{Ber}(p)$, by definition, has sample space $\{0,1\}$ and satisfies $P(X=1) = p$ and $P(X=0) = 1-p$. The given function is just a restatement of that definition.
- $f(x) = p^x(1-p)^{1-x}$ is correct. Note that $f(1) = p$ and $f(0) = 1-p$, so this is the same as the function considered in the first choice.
- $f(x) = xp + (1-x)(1-p)$ is correct. It also satisfies $f(1) = p$ and $f(0) = 1-p$, so f is the same as the function considered in the first choice.
- $f(x) = \begin{cases} 1 & \text{if } x=1 \\ 0 & \text{if } x=0 \end{cases}$ is incorrect. This is actually the probability mass function of $\mathrm{Ber}(1)$, but we have assumed $p \in (0,1)$.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

Review: Statistical Model for a Bernoulli Distribution

3/3 points (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(p^*)$ for some unknown $p^* \in (0,1)$. Let $(E, \{\mathrm{Ber}(p)\}_{p\in\Theta})$ denote the corresponding statistical model. What is the smallest possible set that could be E?

- ( ) $\{0\}$
- ( ) $\{-1,1\}$
- (x) $\{0,1\}$

- $\mathbb{R}$

The parameter space Θ can be written as an interval $[a, b]$. What is the smallest possible interval so that $\{\text{Ber}\,(p)\}_{p\in\Theta}$ represents all possible Bernoulli distributions?

$a =$ 0 Answer: 0.0

$b =$ 1 Answer: 1.0

Solution:

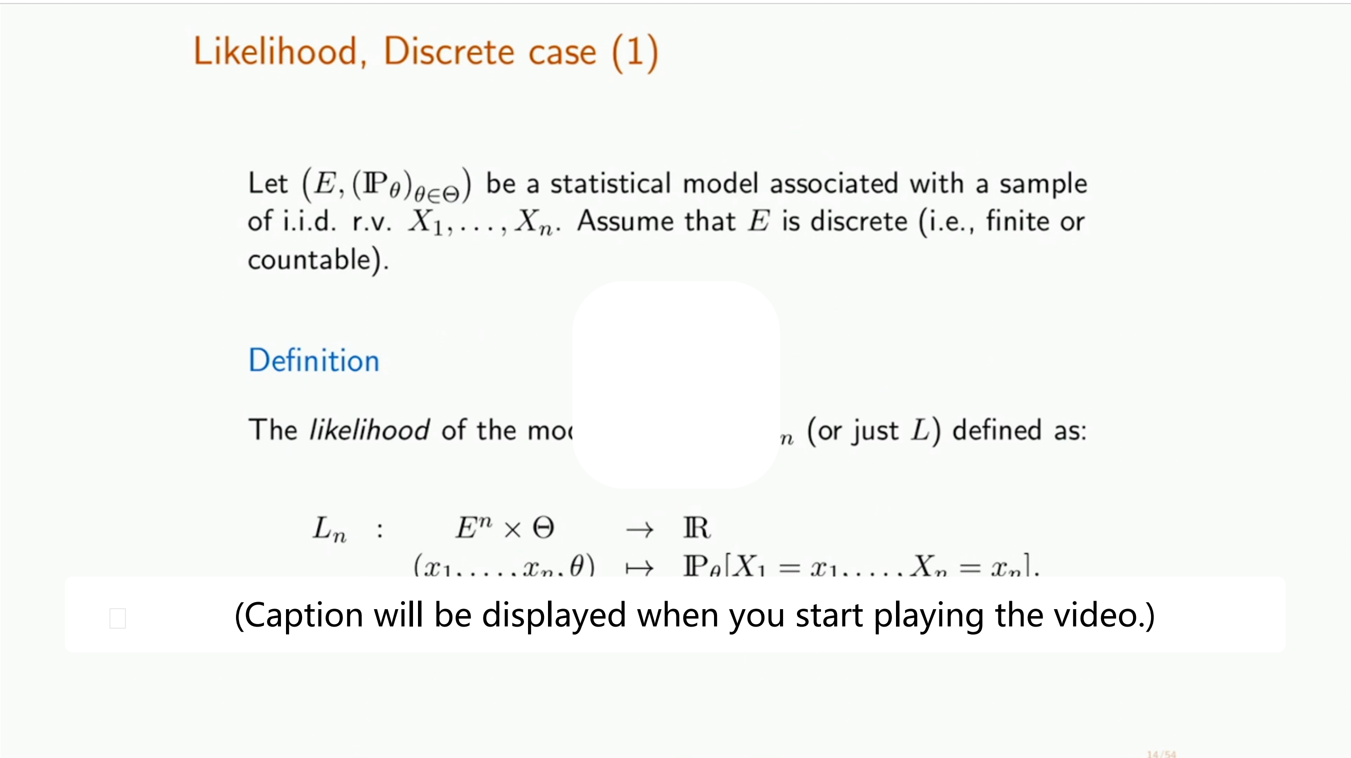
Since a Bernoulli random variable is either 0 or 1, the smallest possible sample space is $\{0, 1\}$.

If $\Theta = [0, 1]$, then $\{\text{Ber}\,(p)\}_{p\in[0,1]}$ is the set of all possible Bernoulli distributions, as desired.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

Likelihood of a Discrete Distribution



for different values of theta.

So this is where you're having this dual vision of what

you've been used to doing in statistics and probability.

When you do the probability, you typically fix this guy,

and compute the value that X1 is equal to x1

for different values of little x1.

Right?

But here, this is given-- this is your data.

And what you're going to move is this.

And it says, oh, for each value you give me--

A simple way to think about this problem is--

I give you two candidate values for theta.

Theta is equal to 0 or theta is equal to 1.

Those are the two possible values you have.

You could try which one is the most likely to have generated

视频
下载视频文件

字幕
下载 SubRip (.srt) file
下载 Text (.txt) file

Concept Check: Interpreting the Likelihood

0/1 point (graded)

Let $(E, \{P_\theta\}_{\theta\in\Theta})$ denote a discrete statistical model. Let p_θ denote the pmf of P_θ. Let $X_1, \dots, X_n \overset{iid}{\sim} P_{\theta^*}$ where the parameter θ^* is unknown. Then the **likelihood** is the function

$$L_n : E^n \times \Theta \to \mathbb{R}$$

$$(x_1, \dots, x_n, \theta) \mapsto \prod_{i=1}^{n} p_\theta(x_i)\,.$$

For our purposes, we think of $x_1, \dots, x_n$ as observations of the random variables $X_1, \dots, X_n$.

Which of the following are true about the likelihood L_n? (Choose all that apply.)

- [x] It is the joint pmf of n iid samples from the distribution P_θ.
- [x] It is a function of the sample $X_1 = x_1, \ldots, X_n = x_n$.
- [ ] It is a function of the parameter θ, where θ ranges over all possible values of in the parameter space Θ.
- [ ] It is the joint pmf of n iid samples from the true distribution P_{θ^*}.

Solution:

We examine the choices in order.

- "It is the joint pmf of n iid samples from the distribution P_θ." is correct. If $Y_1, \ldots, Y_n \overset{iid}{\sim} P_\theta$, then by independence, the joint pmf of these variables is given by a product:

$$P(Y_1 = x_1, \ldots, Y_n = x_n) = \prod_{i=1}^{n} p_\theta(x_i).$$

  Remark 1: We use Y_i to denote these variables to differentiate from the samples X_i that come from the true distribution P_{θ^*}.

- "It is a function of the sample $X_1 = x_1, \ldots, X_n = x_n$." is correct. To construct the likelihood, we observe samples $X_1 = x_1, \ldots, X_n = x_n$ and then compute $L_n(x_1, \ldots, x_n, \theta)$.

- "It is a function of the parameter θ, where θ ranges over all possible values of in the parameter space Θ" is correct. As θ varies over Θ, the likelihood $L_n(x_1, \ldots, x_n, \theta)$ takes on different values. This is evident from the dependence on θ in the definition of the likelihood.

  Remark 2: Later on we will maximize L_n (as a function of θ) to define the **maximum likelihood estimator** . Hence, it is a crucial property that the likelihood is a function of the parameter.

- "It is the joint pmf of n iid samples from the distribution P_{θ^*}." is incorrect. The likelihood takes as input all possible θ, not just the true parameter θ^*. Note how the likelihood is defined for general θ, not just the true parameter θ^*.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

Likelihood of a Bernoulli Statistical Model

1/1 point (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(p^*)$ for some unknown $p^* \in (0,1)$. Let $(E, \{\mathrm{Ber}(p)\}_{p\in\Theta})$ denote the corresponding statistical model constructed in the previous question.

What is the likelihood L_n of this statistical model? (Choose all that apply.)

Hint: Use the pmf's in the second and third choices from the first problem on this page: "Preparation Equivalent Expressions for the pmf of a Bernoulli Distribution".

- [x] $L_n(x_1, \ldots, x_n, p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$.
- [ ] $L_n(x_1, \ldots, x_n, p) = p^{n-\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n x_i}$.
- [ ] $L_n(x_1, \ldots, x_n, p) = p^{\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n x_i}$.

☑ $L_n\left(x_1,\ldots,x_n,p\right)=\prod_{i=1}^n\left(x_ip+\left(1-x_i\right)\left(1-p\right)\right)$

Solution:

We examine the choices in order.

- As shown in the previous problem, we can write the pmf of a Bernoulli as $x\mapsto p^x\left(1-p\right)^{1-x}$. Hence,

$$L_n\left(x_1,\ldots,x_n,p\right)=\prod_{i=1}^n p^{x_i}\left(1-p\right)^{1-x_i}=p^{\sum_{i=1}^n x_i}\left(1-p\right)^{n-\sum_{i=1}^n x_i}.$$

$$\begin{aligned}L\left(x_1,\ldots,x_n,p\right)&=\prod_{i=1}^n p^{x_i}\left(1-p\right)^{1-x_i}\\&=p^{\sum_{i=1}^n x_i}\left(1-p\right)^{n-\sum_{i=1}^n x_i}.\end{aligned}$$

  Hence the first answer choice is correct.

- The second and third choices $L_n\left(x_1,\ldots,x_n,p\right)=p^{n-\sum_{i=1}^n x_i}\left(1-p\right)^{\sum_{i=1}^n x_i}$ and $L_n\left(x_1,\ldots,x_n,p\right)=p^{\sum_{i=1}^n x_i}\left(1-p\right)^{\sum_{i=1}^n x_i}$ are incorrect. Note that they are slight algebraic modifications of the first choice, so these formulas cannot be correct.

- If we use the expression $f\left(x\right)=xp+\left(1-x\right)\left(1-p\right)$ for the pmf of $\mathrm{Ber}\left(p\right)$, then

$$L\left(x_1,\ldots,x_n,p\right)=\prod_{i=1}^n\left(x_ip+\left(1-x_i\right)\left(1-p\right)\right)$$

  is, by definition, the likelihood. Hence, the last answer choice is also correct.

  Remark: Although the last answer choice is formally correct, the formula is much more difficult to work with. It is often more convenient to use $p^{\sum_{i=1}^n x_i}\left(1-p\right)^{n-\sum_{i=1}^n x_i}$ for the likelihood of a Bernoulli statistical model.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

讨论

显示讨论

主题: Unit 3 Methods of Estimation:Lecture 8: Distance measures between distributions / 14. Likelihood of a Discrete Distribution