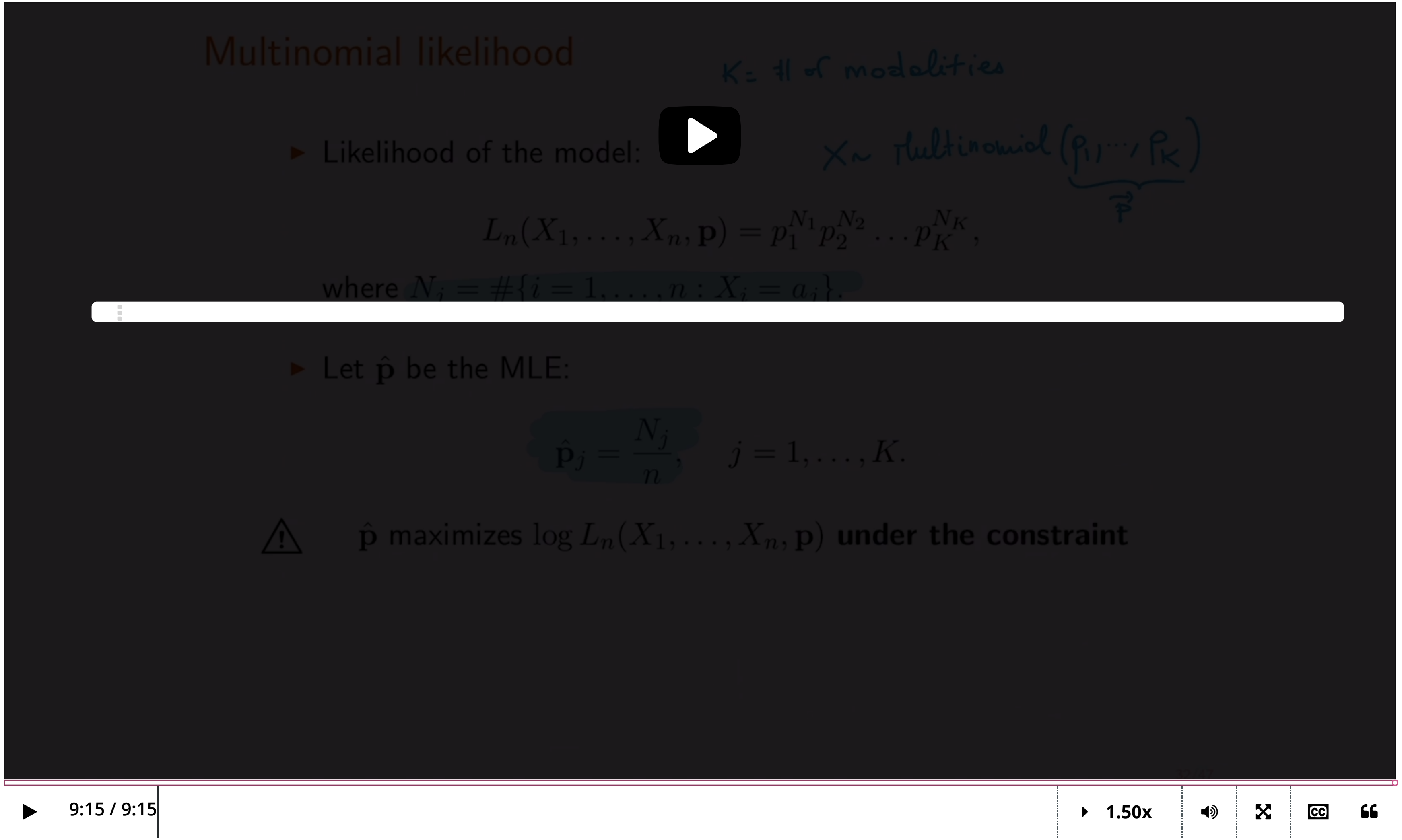
5. Maximum Likelihood Estimator for the Categorical Distribution

Video note: The following video derives the maximum likelihood estimator $\widehat{\mathbf{p}}$ of a categorical statistical model. Note that we have seen this previously in Lecture 10 and in Recitation 6.

The maximum likelihood estimator will form the basis of goodness of fit testing for discrete distributions.

MLE for the Categorical Distribution



Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Concept Check: Examples of the Categorical Distribution

2/2 points (graded)

Consider the distribution $\mathrm{Ber}\,(0.25)$. Consider the categorical statistical model $(\{a_1, \dots, a_K\}, \{\mathbf{P}_\mathbf{p}\})$ for this Bernoulli distribution.

If we let $a_1 = 1$ and $a_2 = 0$, then this corresponds to a categorical distribution $\mathbf{P}_{\mathbf{p}}$ with parameter vector $\mathbf{p}$ given by...

- 0.25
- 0.75
- $[0.25 \;\; 0.75]^T$ (selected)
- $[0.75 \;\; 0.25]^T$

Let $a_i = i$ for $i = 1, \ldots, K$. The uniform distribution on $E = \{1, 2, \ldots, K\}$ can be expressed as a categorical distribution $\mathbf{P}_{\mathbf{p}}$ for some choice of parameter $\mathbf{p}$.

What is $\sum_{i=1}^{K} p_i^2$?

1/K

Answer: 1/K

$\frac{1}{K}$

STANDARD NOTATION

Solution:

Let $X \sim \mathrm{Ber}\,(0.25)$. Observe that

$$p_1 = P\,(X = a_1) = P\,(X = 1) = 0.25$$

and

$$p_2 = P(X = a_2) = P(X = 0) = 0.75.$$

Hence, $\mathbf{p} = [0.25 \;\; 0.75]^T$.

Remark: Observe that $\mathrm{Ber}(p)$ has a one-dimensional parameter and $\mathbf{P_p}$ for this example involves a parameter that is two-dimensional, but such that the second parameter depends on the first one ($p_1 = 1 - p_2$). In general, the categorical distribution for $\mathbf{p} \in \Delta_K$ involves a K-dimensional parameter, but the last parameter p_K, for example, is redundant because $p_K = 1 - \sum_{i=1}^{K-1} p_i$. Even though $\mathbf{p}$ is K-dimensional, the categorical distribution has only $K - 1$ degrees of freedom. This will make our analysis more challenging: the extra constraint on the parameter $\sum_{i=1}^{K} p_i = 1$ implies that the Fisher information for the model as specified **does not exist**. Hence, we cannot apply Wald's test directly.

For the second question, by definition, the uniform distribution weighs all elements in $\{1, \ldots, K\}$ equally. Let $\mathbf{P}$ denote the parameter vector of the uniform distribution on $\{1, 2, \ldots, K\}$. Then

$$p_i = P(X = i) = \frac{1}{K}.$$

Thus,

$$\sum_{i=1}^{K} p_i^2 = \sum_{i=1}^{K} \frac{1}{K^2} = \frac{1}{K}.$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Likelihood for a Categorical Distribution

3/3 points (graded)

Suppose that $K = 3$, and let $E = \{1, 2, 3\}$. Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbf{P}_{\mathbf{p}}$ for some unknown $\mathbf{p} \in \Delta_3$. Let $f_{\mathbf{p}}$ denote the pmf of $\mathbf{P}_{\mathbf{p}}$ and recall that the likelihood is defined to be

$$L_n\left(X_1, \ldots, X_n, \mathbf{p}\right) = \prod_{i=1}^{n} f_{\mathbf{p}}\left(X_i\right).$$

Here we let the sample size be $n = 12$, and you observe the sample $\mathbf{x} = x_1, \ldots, x_{12}$ given by

$$\mathbf{x} = 1, 3, 1, 2, 2, 2, 1, 1, 3, 1, 1, 2, .$$

The likelihood for this data set can be expressed as $L_{12}\left(\mathbf{x}, \mathbf{p}\right) = p_1^A p_2^B p_3^C$.

Fill in the values of A, B, and C below.

$A =$ 6 ✔ **Answer:** 6 $B =$ 4 ✔ **Answer:** 4 $C =$ 2 ✔ **Answer:** 2

Solution:

Since $K = 3$ and $E = \{1, 2, 3\}$,

$$f_{\mathbf{p}}\left(i\right) = p_i, \quad i = 1, 2, 3.$$

Next,

$$L_n\left(X_1, \ldots, X_n, \mathbf{p}\right) = \prod_{i=1}^{n} f_{\mathbf{p}}\left(X_i\right) = p_1^{N_1} p_2^{N_2} p_3^{N_3}$$

where

$$N_i = \text{number of times } i \text{ appears in } (X_1, \ldots, X_n), \quad i = 1, 2, 3.$$

In the data set above, 1 appears 6 times, 2 appears 4 times, and 3 appears 2 times. Thus, $A = N_1 = 6$, $B = N_2 = 4$, and $C = N_3 = 2$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Maximum Likelihood Estimator for Categorical Distribution

3/3 points (graded)

As above, under the statistical model $(\{1,2,3\}, \{\mathbf{P}_{\mathbf{p}}\}_{\mathbf{p} \in \Delta_3})$, we have

$$L_{12}(\mathbf{x}, \mathbf{p}) = p_1^A p_2^B p_3^C$$

where

$$\mathbf{x} = 1, 3, 1, 2, 2, 2, 2, 1, 1, 3, 1, 1, 2.$$

In the previous problem, you found the specific values for A, B, and C.

Recall that the MLE is given by

$$\widehat{\mathbf{p}}_n^{MLE} = \text{argmax}_{\mathbf{p} \in \Delta_3} \log L_n(X_1, \ldots, X_n, \mathbf{p}).$$

By the theory of Lagrange multipliers, one can show that the maximum occurs at the point $\mathbf{p}$ such that there exists $\lambda \neq 0$ so that

$$\nabla \log L_n (X_1, \ldots, X_n, \mathbf{p}) = \lambda \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

(The gradient above is taken with respect to the parameter $\mathbf{p}$.)

Using this result and the previous problem, what is the estimate $\widehat{\mathbf{p}}_{12}^{MLE}$ for $\mathbf{p}$ given the data set $\mathbf{x}$?

$(\widehat{\mathbf{p}}_{12}^{MLE})_1 =$ 1/2 ✔ **Answer:** 1/2 $(\widehat{\mathbf{p}}_{12}^{MLE})_2 =$ 1/3 ✔ **Answer:** 1/3 $(\widehat{\mathbf{p}}_{12}^{MLE})_3 =$ 1/6 ✔

Answer: 1/6

Solution:

In the previous problem, we saw that $A = 6$, $B = 4$, and $C = 2$. Thus

$$\log L_n (\mathbf{x}, \mathbf{p}) = 6 \log p_1 + 4 \log p_2 + 2 \log p_3.$$

Hence,

$$\nabla \log L_n (\mathbf{x}, \mathbf{p}) = \begin{bmatrix} \frac{6}{p_1} \\ \frac{4}{p_2} \\ \frac{2}{p_3} \end{bmatrix}.$$

Applying the Lagrange multipliers, we have

$$\begin{bmatrix} \frac{6}{p_1} \\ \frac{4}{p_2} \\ \frac{2}{p_3} \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Therefore,

$$p_1 = \frac{6}{\lambda},\ p_2 = \frac{4}{\lambda},\ p_3 = \frac{2}{\lambda}.$$

By the constraint $p_1 + p_2 + p_3 = 1$, we see that

$$\lambda = 6 + 4 + 2 = 12.$$

Therefore,

$$\widehat{\mathbf{p}}_{12}^{MLE} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}.$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 4 Hypothesis testing:Lecture 15: Goodness of Fit Test for Discrete Distributions / 5. Maximum Likelihood Estimator for the Categorical Distribution

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖