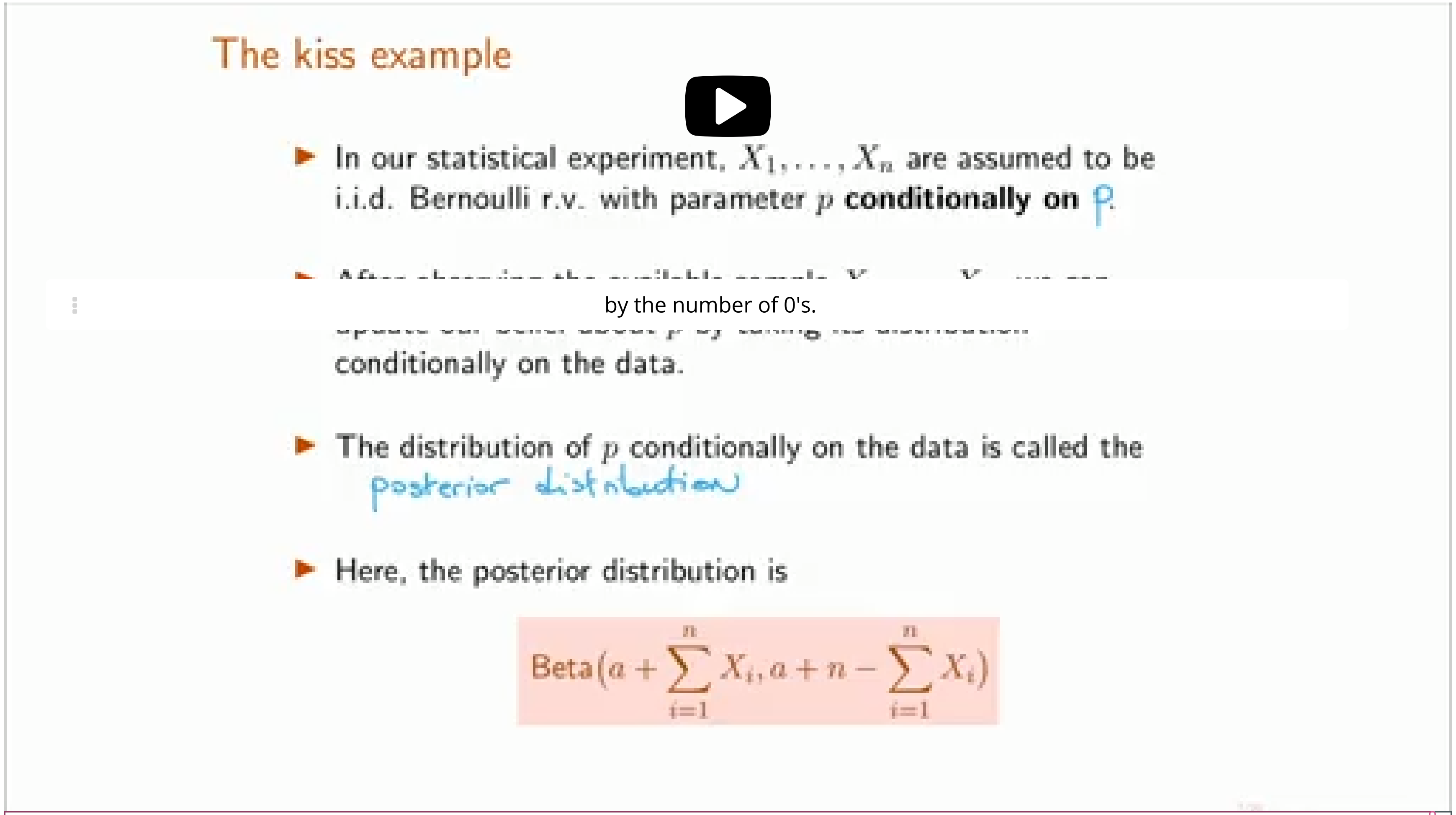
4. Basic Example of the Bayesian Approach

Video note: In the following video, for the last slide: the second argument of the beta distribution on the bottom should be $b + n - \sum X_i$, not $a + n - \sum X_i$.

Basic Example of the Bayesian Approach



9:31 / 9:38 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Frequentist vs. Bayesian II

0/1 point (graded)

Which of the following scenarios require a Bayesian approach, rather than a frequestist approach, in order to utilize all information provided? (Choose all that apply.)

- [x] On any given day, the weather is either rainy with probability p, or not rainy, with probability $1-p$. If the weather is rainy, then the commute times of individuals are i.i.d. exponential random variables, with parameter λ_1. If it is not rainy, then the commute times of individuals are, again, i.i.d. exponential random variables, with parameter λ_2. Assume that, you observe commute time of n different individuals to the workplace, and based on this, you want to understand whether the weather is rainy or not. ✔
- [x] On any given day, if the weather is rainy, then the commute times of individuals are i.i.d. exponential random variables, with parameter λ_1. If it is not rainy, then the commute times of individuals are, again, i.i.d. exponential random variables, with parameter λ_2. Assume that, you observe commute time of n different individuals to the workplace, and based on this, you want to understand whether the weather is rainy or not.
- [x] A professor's mood often swings, and he is either happy with probability α, or sad, with probability $1-\alpha$, independent of everything else. If he is happy, he solves any problem he is asked, in a time which has a p.d.f. $f_1(x)$ supported on $(0,\infty)$. If his mood is sad, then he solves a problem, in a time, which has p.d.f. $f_2(x)$, again, supported on $(0,\infty)$. Assuming that his attempts to each problem are i.i.d. trials; his students ask him n questions to determine whether he is happy or sad. ✔
- [x] A professor's mood often swings. If he is happy, he solves any problem he is asked, in a time which has a p.d.f. $f_1(x)$ supported on $(0,\infty)$. If his mood is sad, then he solves a problem, in a time, which has p.d.f. $f_2(x)$, again, supported on $(0,\infty)$. Assuming that his attempts to each problem are i.i.d. trials; his students ask him n questions to determine whether he is happy or sad.

✖

Solution:

The Bayesian views are first and third choices. Recall that, the Bayesian view allows us to reflect our prior belief about the hypotheses that are being considered. In the first choice, the first hypothesis is, weather is rainy; and the second hypothesis is it is not rainy; and we have a prior. In the second choice, however, we only state the distribution of the observation under each hypotheses, while not stating our prior belief about hypotheses; hence, it is **not** Bayesian.

Similarly, in the third choice, there are two hypotheses; namely, "Professor is happy" and "Professor is sad". We put a prior on them, and stated how the observations (i.e., problem solving time) are distributed under each hypothesis, which again represents a Bayesian approach. Finally, the last part is **not** Bayesian due to the absence of a prior.

You have used 2 of 2 attempts

Submit

Answers are displayed within the problem

Mode of the Beta Distribution

5/5 points (graded)

Recall that the **Beta distribution** in x is defined as the distribution with support $[0,1]$ and pdf

$$C(\alpha,\beta)\,x^{\alpha-1}(1-x)^{\beta-1},$$

where α and β are parameters that satisfy $\alpha > 0$, $\beta > 0$. Here, $C(\alpha,\beta)$ is a normalization constant that does not depend on x.

The Beta distribution can take many shapes depending on the chosen parameters α and β. As a result, the highest point (mode) of this distribution can vary wildly. Due to the different overall shapes depending on parameter values, there isn't also a consistent formula for the mode. Compute the correct mode for each of the parameter sets. (A mode of the distribution is the value(s) of x where the pmf attains its highest value in the entire support of the distribution.)

You may use the variables α and β in your answer. If there is no unique mode, enter -1. Note that it is possible for the mode to have a “probability" of infinity, which would be a mode if this happens only once.

Case 1: $\alpha < 1$ and $\beta < 1$.

-1

✔ **Answer:** -1+alpha*0+beta*0

-1

Case 2: $\alpha \leq 1$ and $\beta \geq 1$ (but excluding $\alpha = \beta = 1$).

0

✔ **Answer:** 0+alpha*0+beta*0

0

Case 3: $\alpha \geq 1$ and $\beta \leq 1$ (but excluding $\alpha = \beta = 1$).

1 ✔ **Answer:** 1+alpha*0+beta*0

1

Case 4: $\alpha = \beta = 1$

-1 ✔ **Answer:** -1+alpha*0+beta*0

-1

Case 5: $\alpha > 1$ and $\beta > 1$.

(alpha-1)/(alpha+beta-2) ✔ **Answer:** (alpha-1)/(alpha+beta-2)

$\frac{\alpha-1}{\alpha+\beta-2}$

STANDARD NOTATION

Solution:

We ignore the normalization constant $C(\alpha, \beta)$, focusing on the component that's variable in x, $x^{\alpha-1}(1-x)^{\beta-1}$. First, note that over the interval $[0,1]$, the monomial $x^{\alpha-1}$ is strictly increasing in x for $\alpha > 1$, constant in x (which counts as “weakly increasing" and also as “weakly decreasing") for $\alpha = 1$, and strictly decreasing in x for $\alpha < 1$. Similarly, $(1-x)^{\beta-1}$ is strictly decreasing in x for $\beta > 1$, constant in x for $\beta = 1$, and strictly increasing in x for $\beta > 1$. It is also clear that both $x^{\alpha-1}$ and $(1-x)^{\beta-1}$ are nonnegative for $0 \leq x \leq 1$.

- Case 1: In this case, $x^{\alpha-1}(1-x)^{\beta-1}$ will tend towards infinity as $x \to 0$ or $x \to 1$. This happens because one of x or $1-x$ will tend to 0 and the other will tend to 1, so the corresponding $x^{\alpha-1}$ or $(1-x)^{\beta-1}$ term will go to infintity, while the other will tend to 1, making their product tend to infinity. Hence both 0 and 1 can be considered “modes".

- Case 2: In this case, both $x^{\alpha-1}$ and $(1-x)^{\beta-1}$ are at least weakly decreasing, and due to $\alpha = \beta = 1$ being excluded, we are sure that at least one is strictly decreasing. Hence their product is strictly decreasing, and thus the mode is at 0.

- Case 3: In this case, both $x^{\alpha-1}$ and $(1-x)^{\beta-1}$ are at least weakly increasing, and due to $\alpha=\beta=1$ being excluded, we are sure that at least one is strictly increasing. Hence their product is strictly increasing, and thus the mode is at 1.

- Case 4: In this (very specific) case, the $x^{\alpha-1}(1-x)^{\beta-1}$ is identically 1, so any point in $[0,1]$ can be considered the mode.

- Case 5: In this case, $x^{\alpha-1}$ is increasing while $(1-x)^{\beta-1}$ is decreasing. Furthermore, note that at the endpoints $x=0$ and $x=1$, $x^{\alpha-1}(1-x)^{\beta-1}=0$. (Note that for this case, we used the fact that the exponent for x or $1-x$ is positive, hence this reasoning could not have been done in the previous cases). Thus, the maximum must be at an interior point. Taking the log gives that this is equivalent to finding x that maximizes $(\alpha-1)\log(x)+(\beta-1)\log(1-x)$, then setting the derivative to zero gives $\frac{\alpha-1}{x}=\frac{\beta-1}{1-x}$. Finally, solving gives $x=\frac{\alpha-1}{\alpha+\beta-2}$.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Beta Distribution Probability Example

7/7 points (graded)

Suppose that you have a coin with unknown probability p of landing heads; assume that coin toss outcomes are i.i.d Bernoulli random varaiables. You flip it 5 times and it lands heads thrice. Our parameter of interest is p. Compute the likelihood function for the first five tosses X_1, ..., X_5.

$L(X_1,\ldots,X_5;p)=$

p^3*(1-p)^2 ✔ **Answer:** p^3*(1-p)^2

$p^3\cdot(1-p)^2$

STANDARD NOTATION

The function you computed above is equivalent up to a constant of proportionality to a Beta distribution over $0\le p\le 1$. What are its parameters?

$\alpha=$

4

Answer: 4

$\beta =$

3

Answer: 3

Suppose that you flip it 5 more times and in the next five tosses it lands heads four times. Compute the likelihood function for the first ten tosses $X_1, \ldots, X_{10}$.

$L(X_1, \ldots, X_{10}; p) =$

p^7*(1-p)^3

Answer: p^7*(1-p)^3

$p^7 \cdot (1-p)^3$

STANDARD NOTATION

Again, the function you computed above is equivalent up to a constant of proportionality to a Beta distribution over $0 \leq p \leq 1$. What are its parameters?

$\alpha =$

8

Answer: 8

$\beta =$

4

Answer: 4

Using your result from the previous problem ("Mode of the Beta Distribution"), what is the MLE in the frequentist view?

0.7

Answer: 0.7

Solution:

Represent heads as “1" and tails as “0". We break down $L(X_1, \ldots, X_5; p)$ into the product $L(X_1; p) \ldots L(X_5; p)$. For any $1 \leq i \leq 5$, $L(X_i; p)$ is p if $X_i = 1$ and $1-p$ if $X_i = 0$.

As we have three 1's and two 0's in our first five tosses, our likelihood function is $p^3(1-p)^2$. The beta distribution is of the form $Cx^{\alpha-1}(1-x)^{\beta-1}$, so we can see that this corresponds to Beta $(4,3)$. In general, if an expression is of the form $p^a(1-p)^b$, where the context gives freedom up to a constant of proportionality, then considering the exponents of p and $1-p$ gives $\alpha = a+1$ and $\beta = b+1$, respectively.

In our first ten tosses in total we have seven 1's and three 0's. Thus our likelihood function is $p^7(1-p)^3$, and similarly this gives Beta $(8,4)$.

This falls into the case where $\alpha > 1$ and $\beta > 1$. Note that the MLE, which is defined as the maximum of the likelihood function, is the same as the mode of the distribution Beta $(8,4)$. Using the formula for the relevant case from the previous problem gives MLE $= \frac{8-1}{8+4-2} = \frac{7}{10} = 0.7$.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Bayesian statistics:Lecture 17: Introduction to Bayesian Statistics / 4. Basic Example of the Bayesian Approach

Add a Post

Show all posts ▼ by recent activity ▼

Practical and Simple Descriptions of the Beta Distribution
I made the post below in the probability course. The links below may be helpful for improving your understanding of the beta distribution and how α and β are utilized and u... 5

[Philosophical warning] The universe may actually flip a coin
I'm just reading this book by Sean Carroll which has this intriguing point on the Many World theory that points to a kind of Bayesian world. For a short and nice description: ht... 5
Community TA

Likelihood combinations
Why is there no account for this in the Likelihood functions? I thought there should be some accounting for the ways we can get X from N. I know this can drop out later as a c... 2

Just wondered why the derivative of the Beta function was not differentiated in 1

2

I thought it more natural to do it this way, factorize and look at some of the examples. Sure it popped out a useful expression that I could then toy with easily for the specific c...

Beta Distribution Probability Example_Typo? 4