# **Practice Quiz 7.1 & 7.2**

**1.** Below are a number of proposed explanations for why we typically prefer to work with the log likelihood L(θ, X) rather than the regular likelihood L (θ, X). Mark which one(s) of the proposed explanations that are correct.

* It has closed-form solutions for the maximum more often
* It is often better for optimisation since the log likelihood can be convex
* It is numerically safer 没看到这个选项 quiz里
* It is often better for optimisation since the negative log likelihood can be convex
* It allows us to ignore the normalisation constant of the parametric family during optimisation

**2**.You have a dataset of how long you have had to wait for the bus on your commute. Is a Gaussian distribution a good first choice for modelling this distribution?

* No
* Yes

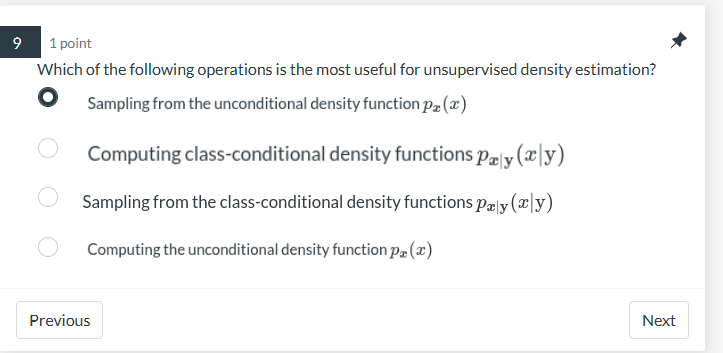
**3**.Which of the following operations is the most useful for estimating the density of a particular class?

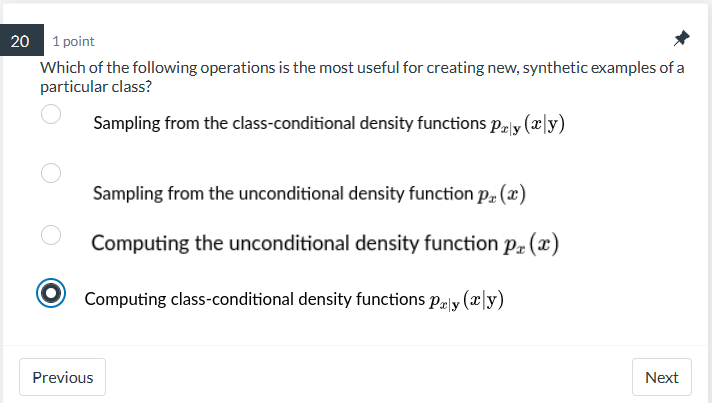
❌

✅

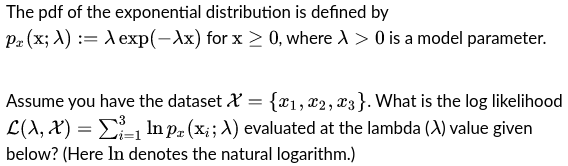
❌

❌



****

**4.**

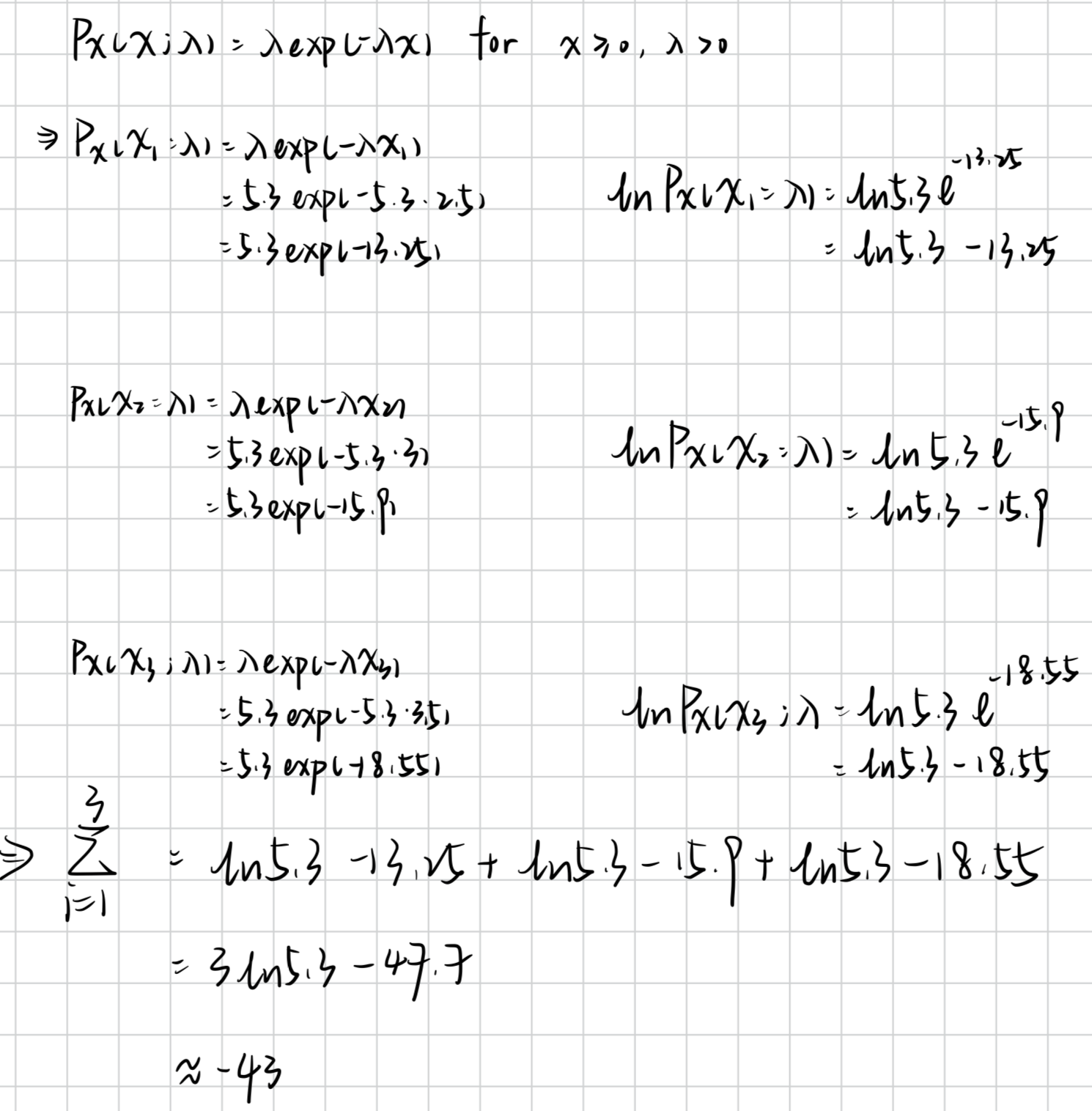
  
Please use the following values in your calculations:

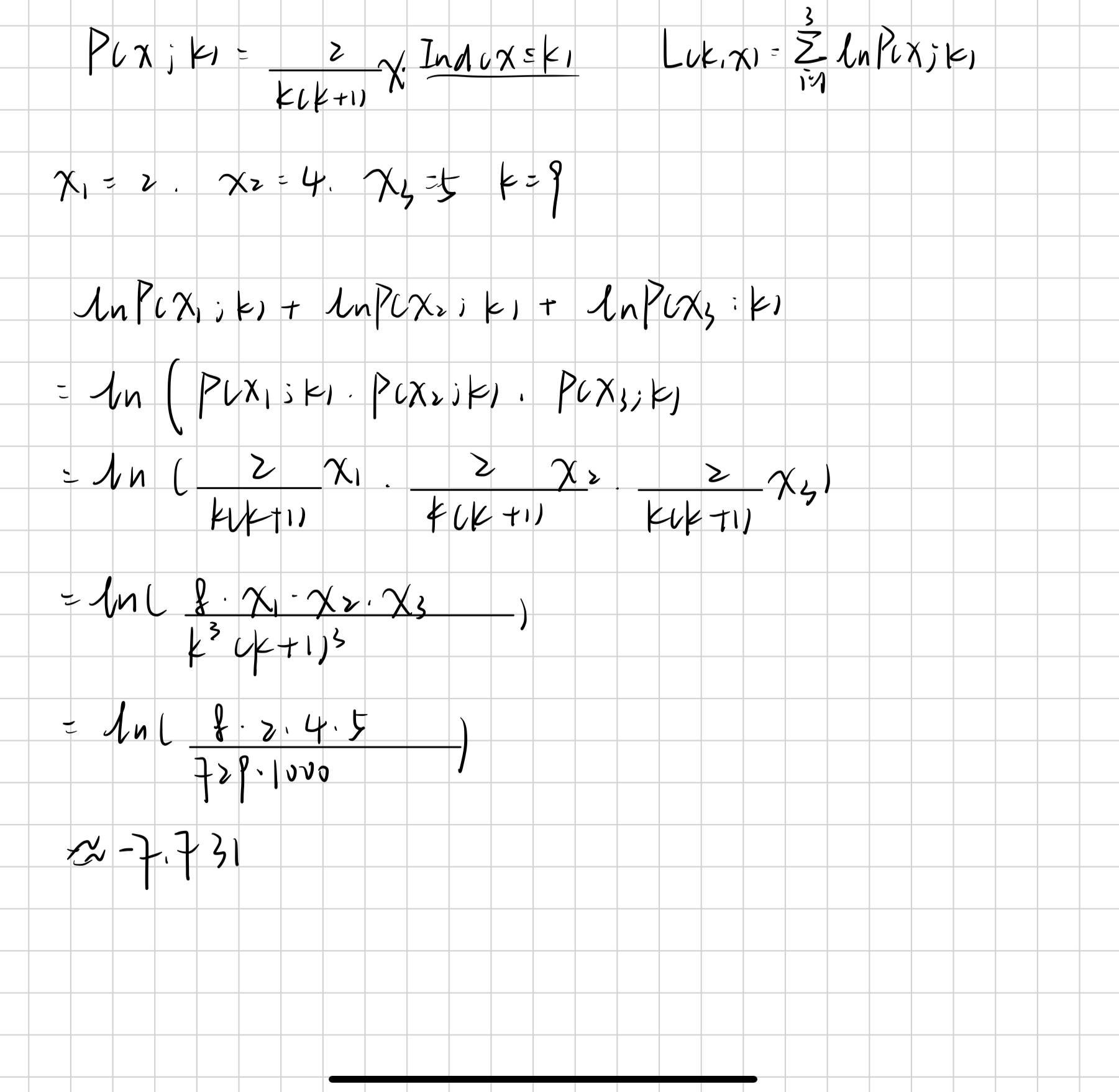
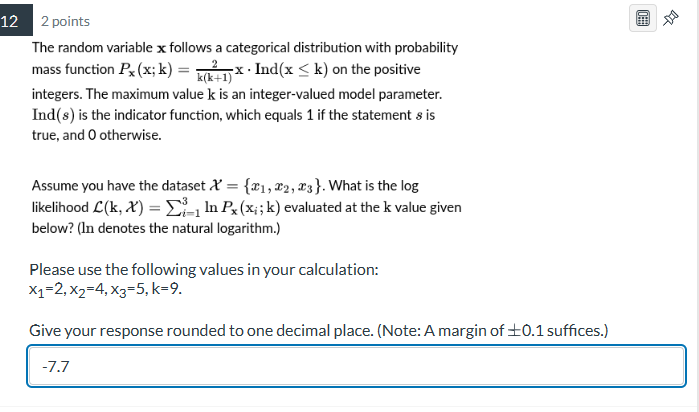
x1=2.8, x2=3.8, x3=3.1, lambda=3.4

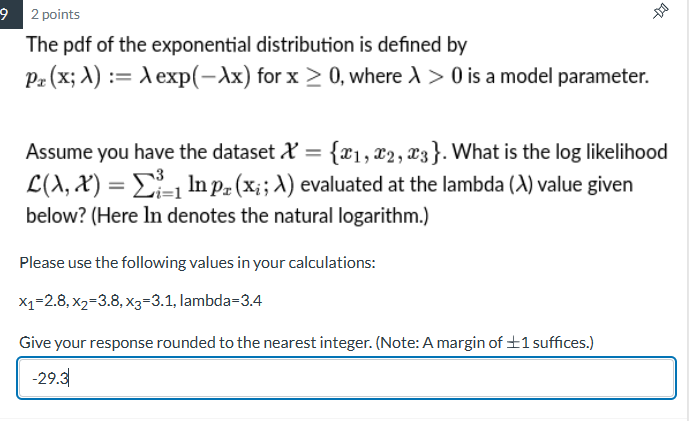
Give your response rounded to the nearest integer. (Note: A margin of

±1 suffices.)

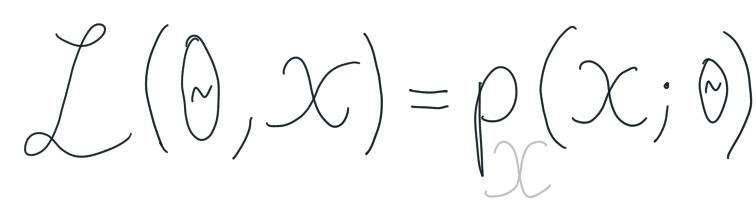
✅Answer：-29







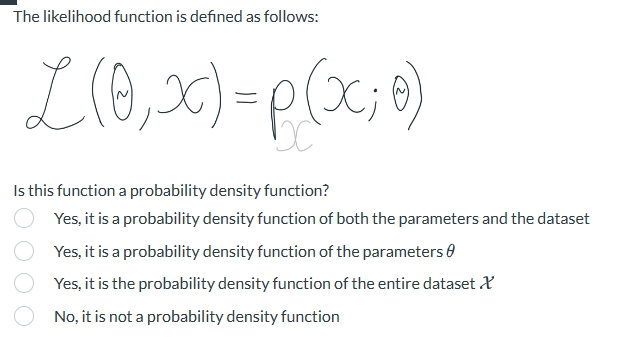
**5.**In the lecture notes, we defined the likelihood function as follows:



Given this definition, is the following statement true or false?

“The likelihood is a joint probability density function of the data."

* True
* False



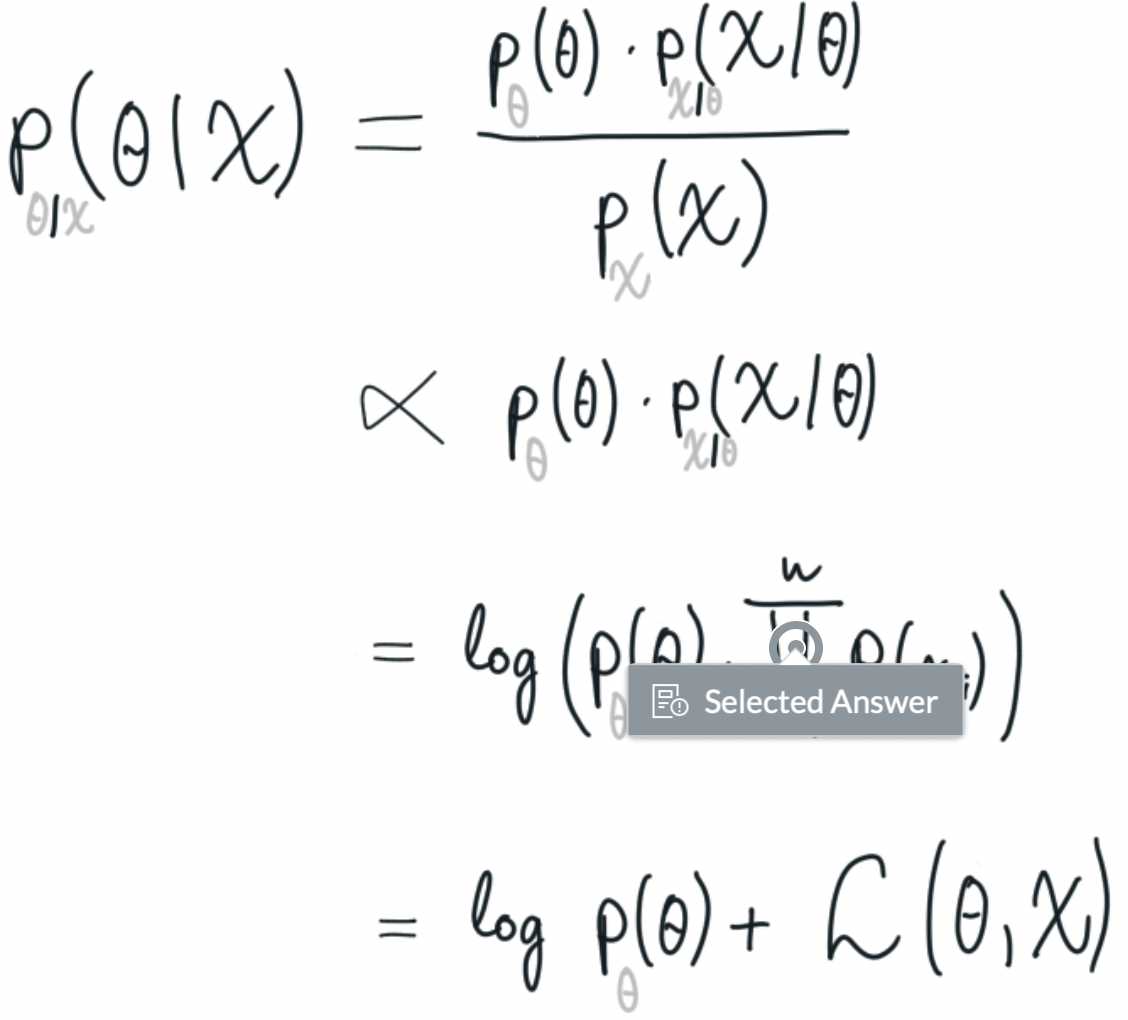
**6.**When is a prior called a conjugate prior in maximum a posteriori estimation?

* When it leads to a posterior distribution from the same parametric family of distributions as the prior
* When it is a complex-valued function
* When it can only be used to with a likelihood function from the same family of distributions
* When it places hard constraints on the parameter values so that some values have zero probability density

**7.** Please fill in the blanks to create a good description of how to perform unconditional density estimation.Unconditional density estimation requires that I have **a dataset {𝑥ᵢ}.**

I then need to decide on a model family and **an objective function** to optimise. If I am using **maximum a posteriori,** I also need to decide on **a prior** as well. I **can try numerical or analytical optimisation methods** to estimate the parameter values. If I can find a closed-form expression, that means that I **don't need iterative update formulas**.

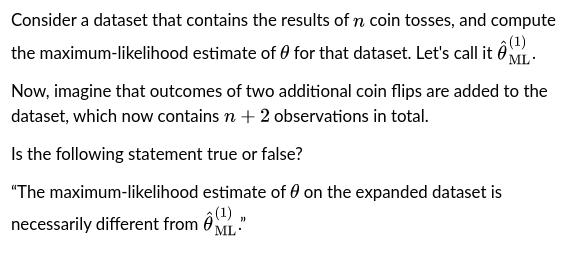
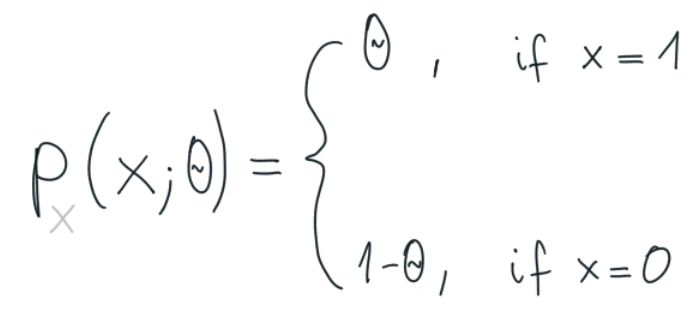
**8.** The following derivation decomposes the posterior, but it contains a mistake. Please mark the first incorrect line by clicking on it.



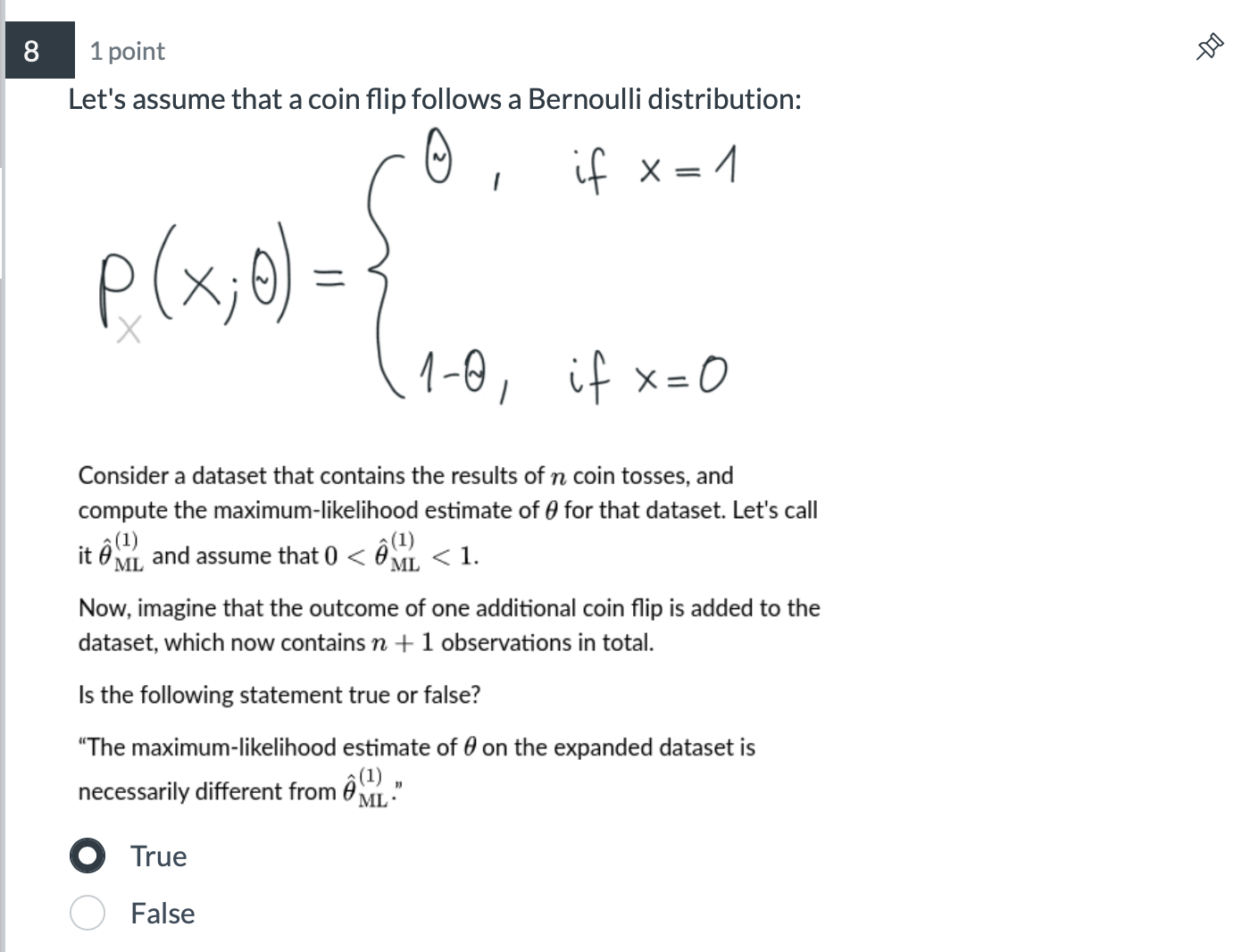
**9.**The prior and posterior distributions in Bayesian probability can be seen as a way to capture…

* Statistical (i.e., aleatoric) uncertainty
* Neither of the two
* Systematic (i.e., epistemic) uncertainty

10.Let's assume that a coin flip follows a Bernoulli distribution:

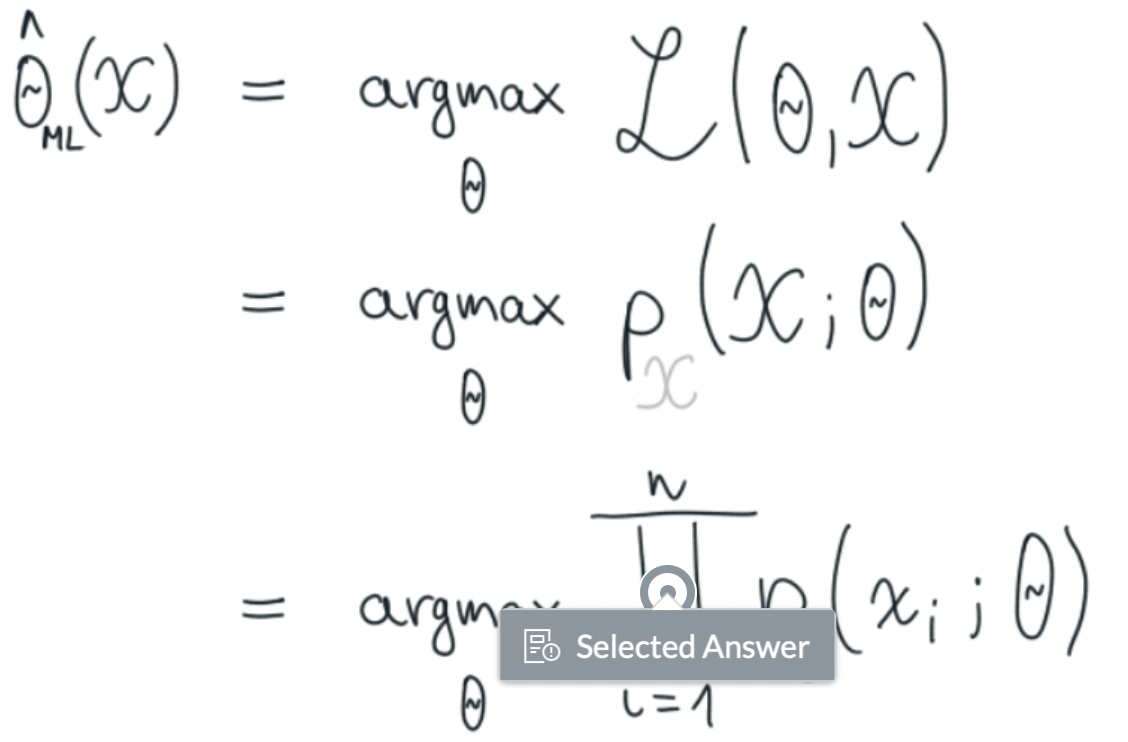
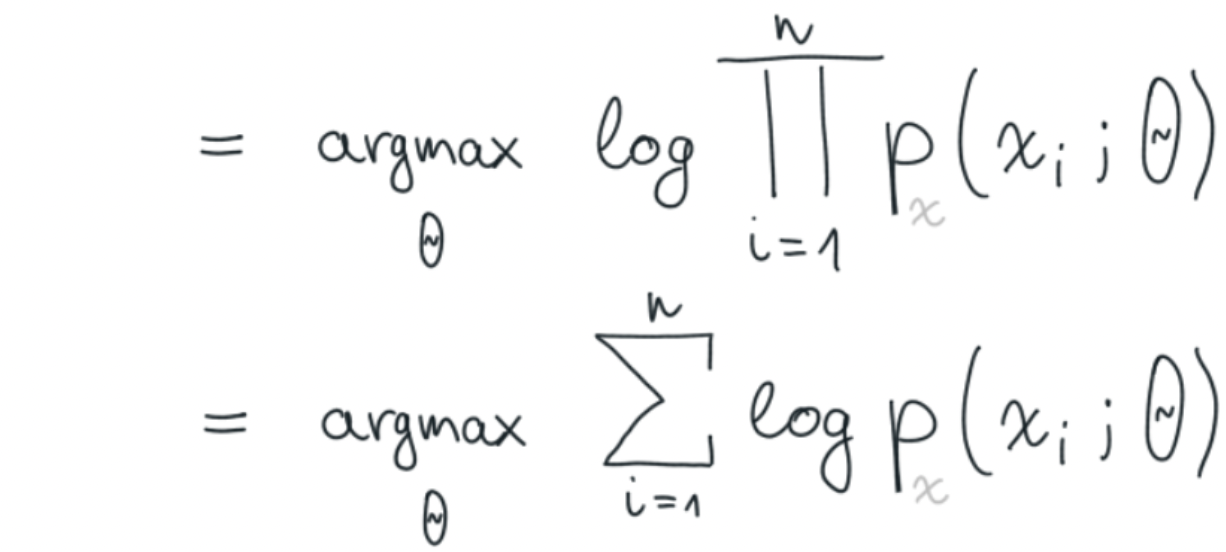


* True
* False

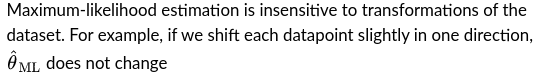


**11.**The attached image contains a derivation of the maximum-likelihood parameter estimate.

Please click on the step that requires the i.i.d. assumption, i.e., that the datapoints are independent and identically distributed. (You may click anywhere on the corresponding line.)

**12.**Mark all true statements about the properties of maximum-likelihood estimation.

❌ 

* Maximum-likelihood estimation can only be applied to one-dimensional problems. In higher dimensions, we need to use the maximum a posteriori approach
* The maximum-likelihood parameter estimate always has an analytic (closed-form) solution
* In the limit of infinite data, the maximum-likelihood estimate converges to the correct parameter value, assuming no model misspecification

**13.**Which of the following operations is the most useful for performing classification with minimum misclassification rate?

* Sampling from the class probabilities P(Y = y)
* Computing conditional class probabilities P(Y=y | X=x)
* Computing class probabilities P(Y = y)
* Sampling from the conditional class probabilities P(Y=y | X=x)

**14.** You are modelling the distribution of the number of purchases of gold pocket watches each day in a jewellery store. Is a Gaussian distribution a good first choice for modelling this distribution?

* No
* Yes

Practice Quiz 7.3

1. Is the following statement true or false?

A mixture of Gaussians can be used to model more complex datasets than a single Gaussian.

* True
* False

**2**. Mark all correct completions of the sentence below:

The Gaussian mixture model…

* ...can only be trained with expectation maximisation.
* ...is usually trained with backpropagation.
* ...can only be used when the input features are at least two-dimensional.
* ...can be used for classification and regression problems as well.

3. Let x > 0 and z := lnx. Which of the following statements are true?

✅

✅If **z** is Gaussian then **x** is lognormal

❌

❌If **x** is Gaussian then **z** is lognormal

**4.**Mark all true statements.

* We do not know a closed-form, analytic solution to the maximum likelihood parameter estimate of a GMM.
* The sum of two Gaussian density functions is a Gaussian mixture model.
* Gaussian mixture models are latent variable models, because the component identities are unobserved.
* The pdf of a GMM cannot be unimodal because there are always at least two Gaussian components in it.

**5.** Is the following statement true or false?

A Gaussian mixture model can have at most as many components as there are datapoints in the dataset.

* True
* False

**6.** Consider a 2D GMM with 3 components, each with different covariance matrices that need not be diagonal. Call the number of modes (local maxima of the density function) m. How many modes can this GMM have?

Hint: Can a 2D Gaussian have a non circular shape?

* m can be both smaller than 3, equal to 3, and larger than 3
* m=3
* m >= 3
* m <= 3

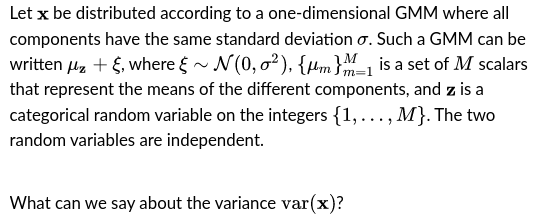
**7.** Is the following statement true or false?

In order to use the EM algorithm, we need to initialise the parameters of the model first. Whether a particular initialisation is good depends on the problem we're trying to solve, so there isn't a universal best choice for the initialisation.

* True
* False

**8.** Mark the true statements about the EM algorithm.

* The log-likelihood of the EM algorithm must converge
* The EM algorithm is guaranteed to find a global optimum, if one exists
* Each EM-algorithm update is guaranteed to never decrease the likelihood

9. Let x be distributed according to a one-dimensional GMM where all 

✅

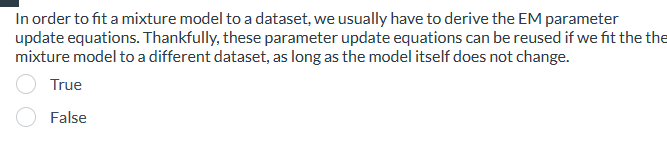
❌

❌

**10.** One of the main advantages of Gaussian mixture models over just a single Gaussian distribution is that they **can adapt to more complex datasets with multiple modes.** At the same time, they have some downsides, for example **the training process can get stuck in local minima.**

**11.** In order to fit a mixture model to a dataset, we usually have to derive the EM parameter update equations. However, these update equations can never be used to fit the same model to a different dataset, because they depend on the observations.

* True
* False

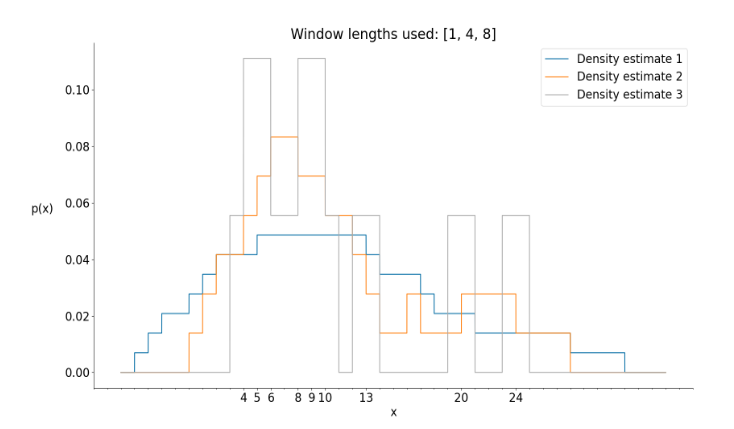


**12.** Gaussian mixture models are especially good for modelling high-dimensional problems, for example audio waveforms, where each measurement is typically about 100-dimensional.

* True
* False

Practice Quiz 7.4 & 7.5

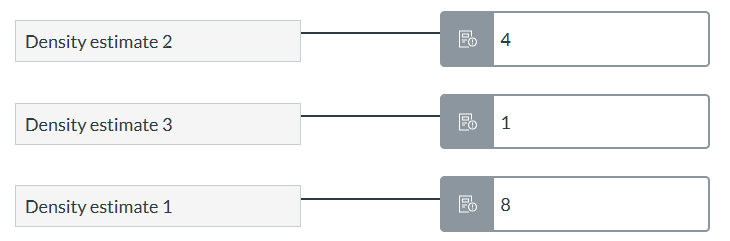
1. False - Kernel functions are always symmetric.
2. False - The Epanechnikov kernel has more favorable theoretical properties than the Gaussian kernel, therefore it is important to always use the Epanechnikov kernel in practice



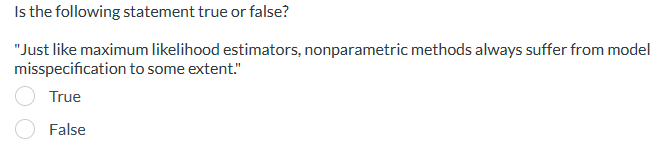
We computed the naive density estimate with 3 different window widths: 1, 4, and 8, using the dataset X = {4, 5, 6, 8, 9, 10, 13, 20, 24}

Please match the density estimates with the window widths used to create them.

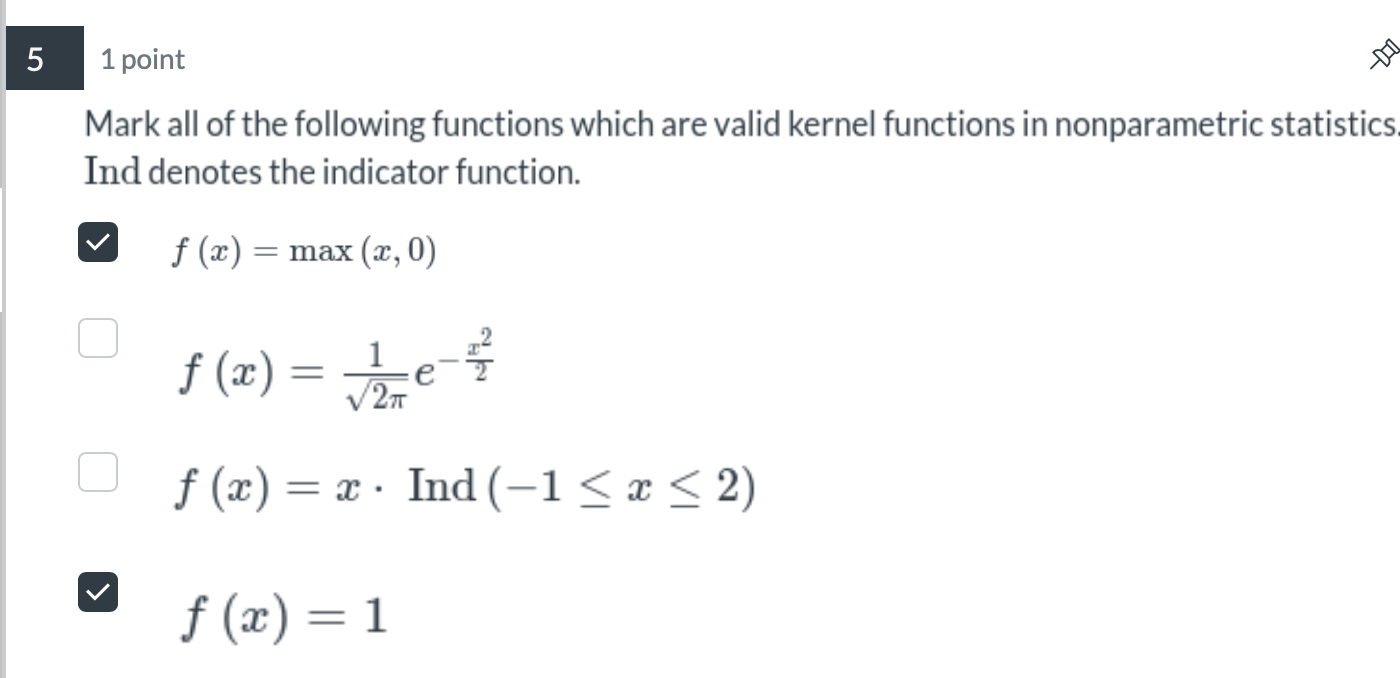
每个颜色的线组成的框越大，数越小



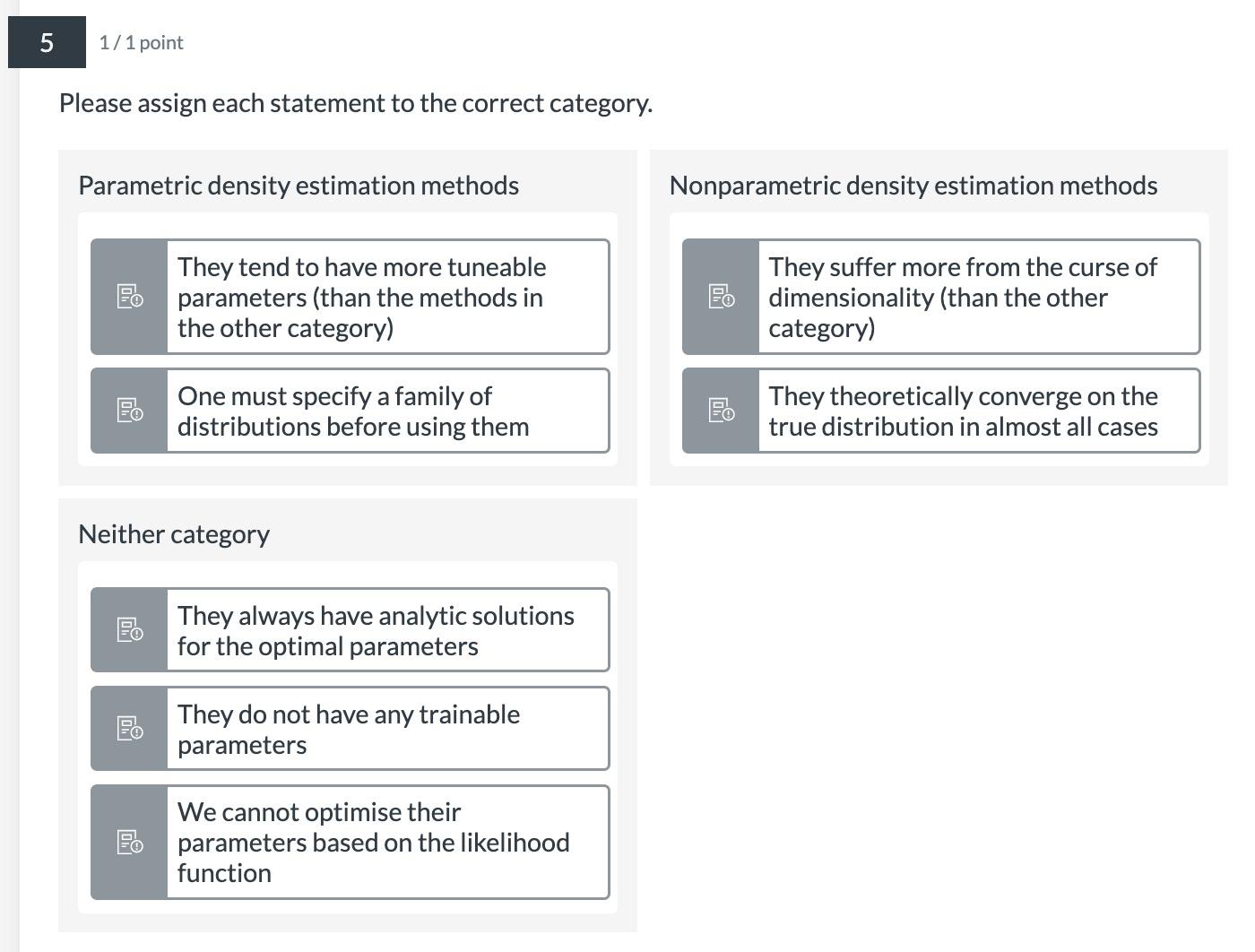
1. false - Unlike nonparametric methods, maximum likelihood estimators never suffer from model misspecification.



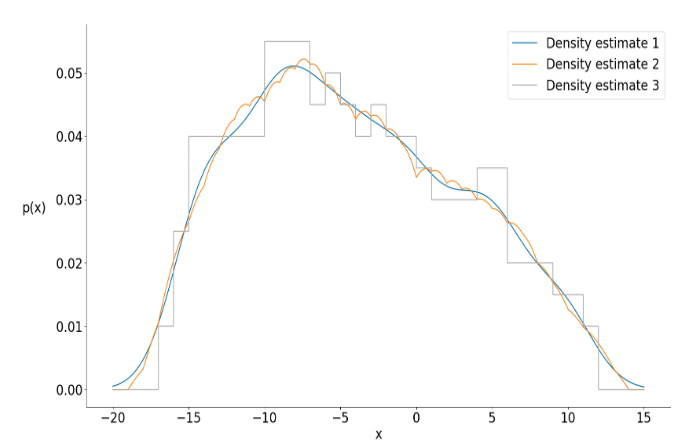
1. Mark all of the following functions which are valid kernel functions. **Ind** denotes the indicator function.

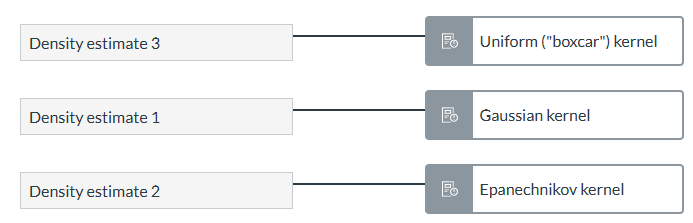
* 

1. Please assign each statement to the correct category.

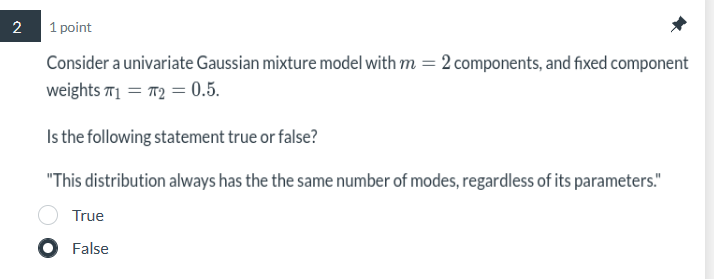


1. The above figure contains kernel density estimates using three different kernels. Please match each of the density estimates with the kernel that was used to generate them.





1. False - Constant functions can be valid kernel functions

-chatgpt

