Predictive distributions and mixture models

Quiz, 5 questions

5/5 points (100%)

Congratulations! You passed!

Next Item

1 / 1 points

1. Consider the Poisson process model we fit in the quiz for Lesson 10 which estimates calling rates of a retailer's customers. The data are attached below.

callers.csv

Re-fit the model and use your posterior samples to simulate predictions of the number of calls by a new 29 year old customer from Group 2 whose account is active for 30 days. What is the probability that this new customer calls at least three times during this period? Round your answer to two decimal places.

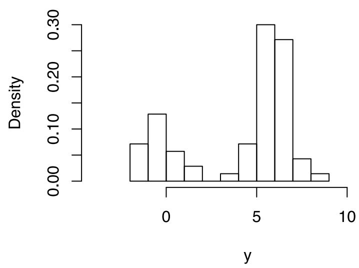
0.23

Correct Response

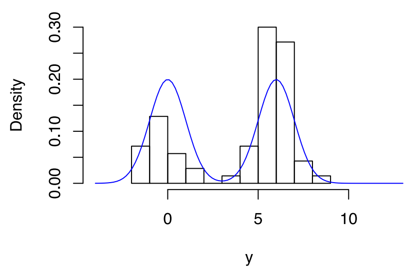
---

1 / 1 points

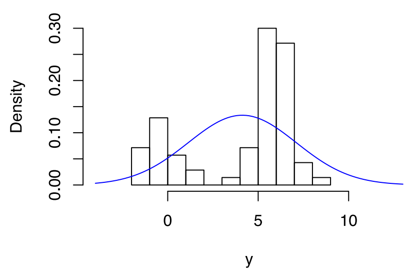
2. Suppose we fit a single component normal distribution to the data whose histogram is shown below.



If we use a noninformative prior for μ and σ^2 and plot the fit distribution evaluated at the posterior means (in blue), what would the fit look like? Is this model appropriate for these data?



The single normal fit accommodates the bi-modality in the dat, but fails to capture the imbalance in the two components. It is not appropriate.



A single normal distribution does not allow bi-modality. Consequently, the fit places a lot of probability in a region with no data. It is not appropriate.

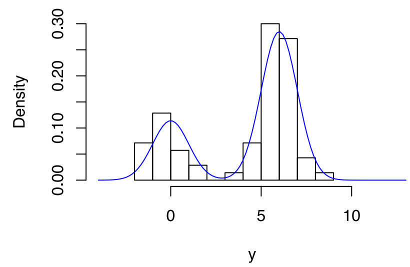
Correct

Without covariates, we would need a more flexible model to capture the bi-modality of these data. A two component mixture of normals would be appropriate.

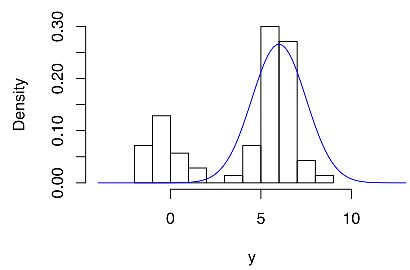
Predictive distributions and mixture models

Quiz, 5 questions

5/5 points (100%)



The single normal fit nicely captures the features of the data. It is appropriate.



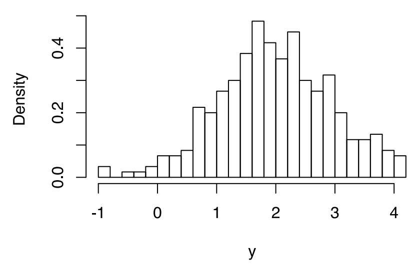
The single normal fit ignores the smaller component, fitting the cluster of points with most data. Consequently, the model places almost no probability in the region of the smaller component.

---

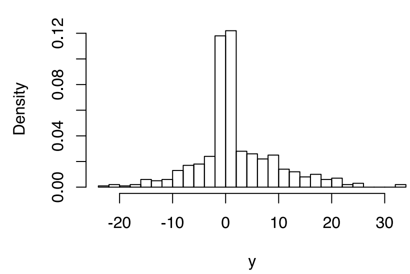
1 / 1 points

3. Which of the following histograms shows data that might require a mixture model to fit?

A)



B)



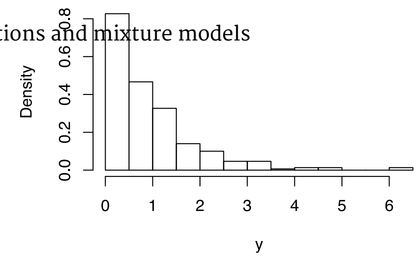
Correct
These data were actually drawn from a two component mixture of normal distributions, one with high variance.

C)

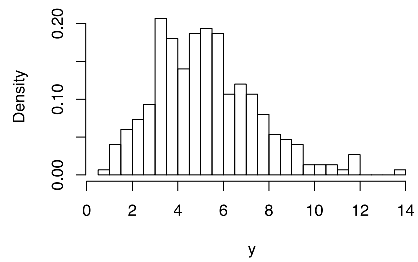
Predictive distributions and mixture models

Quiz, 5 questions

5/5 points (100%)



D)



---

1 / 1 points

4. The Dirichlet distribution with parameters $\alpha_1 = \alpha_2 = \ldots = \alpha_K = 1$ is uniform over its support, the values for which the random vector contains a valid set of probabilities. If θ contains five probabilities corresponding to five categories and has a $\text{Dirichlet}(1,1,1,1,1)$ prior, what is the effective sample size of this prior?

Hint: If θ has a $\text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_K)$ prior, and the counts of multinomial data in each category are $x_1, x_2, \ldots, x_K$, then the posterior of θ is $\text{Dirichlet}(\alpha_1 + x_1, \alpha_2 + x_2, \ldots, \alpha_K + x_K)$. The data sample size is clearly $\sum_{k=1}^5 x_k$.

5

Correct Response

The prior effective sample size of a Dirichlet prior for multinomial data is $\sum_{k=1}^K \alpha_k$. If a less informative prior is desired, we could decrease the α parameters together.

---

1 / 1 points

5. Recall that in the Bayesian formulation of a mixture model, it is often convenient to introduce latent variables z_i which indicate "population" membership of y_i (the "population" may or may not have meaning in the context of the data). One possible hierarchical formulation is given by:

$$\begin{aligned} y_i \mid z_i, \theta &\overset{\text{ind}}{\sim} f_{z_i}(y \mid \theta), \quad i = 1, \ldots, n \\ \Pr(z_i = j \mid w) &= w_j, \qquad j = 1, \ldots, J \\ w &\sim \text{Dirichlet}(w) \\ \theta &\sim p(\theta) \end{aligned}$$

where $f_j(y \mid \theta)$ is a probability density for y for mixture component j and $w = (w_1, w_2, \ldots, w_J)$ is a vector of prior probabilities of membership.

What is the full conditional distribution for z_i?

- $\Pr(z_i = j \mid \cdots) = \frac{f_j(y_i \mid \theta)}{\sum_{\ell=1}^J f_\ell(y_i \mid \theta)}, \quad j = 1, \ldots, J$
- (selected) $\Pr(z_i = j \mid \cdots) = \frac{f_j(y_i \mid \theta)\, w_j}{\sum_{\ell=1}^J f_\ell(y_i \mid \theta)\, w_\ell}, \quad j = 1, \ldots, J$

Correct

Notice the resemblance with this expression and the discrete version of Bayes' theorem. This expression is often used in classification problems, where we need to infer membership of y_i to different candidate populations.

- $\Pr(z_i = j \mid \cdots) = w_j, \quad j = 1, \ldots, J$
- $\Pr(z_i = j \mid \cdots) = \frac{w_j^{I(z_i=j)}(1-w_j)^{1-I(z_i=j)}}{\sum_{\ell=1}^J w_j^{I(z_i=j)}(1-w_j)^{1-I(z_i=j)}}, \quad j = 1, \ldots, J$

Predictive distributions and mixture models

Quiz, 5 questions

5/5 points (100%)