Lesson 2

Quiz, 8 questions

8/8 points (100%)

Congratulations! You passed!

Next Item

1 / 1
points

1.
Which of the following is one major difference between the frequentist and Bayesian approach to modeling data?

- Frequentists treat the unknown parameters as fixed (constant) while Bayesians treat unknown parameters as random variables.

Correct
The only random variables in frequentist models are the data. The Bayesian paradigm also uses probability to describe one's uncertainty about unknown model parameters.

- Frequentist models are deterministic (don't use probability) while Bayesian models are stochastic (based on probability).

- The frequentist paradigm treats the data as fixed while the Bayesian paradigm considers data to be random.

- Frequentist models require a guess of parameter values to initialize models while Bayesian models require initial distributions for the parameters.

---

1 / 1
points

2.
Suppose we have a statistical model with unknown parameter θ, and we assume a normal prior $\theta \sim \mathrm{N}(\mu_0, \sigma_0^2)$, where μ_0 is the prior mean and σ_0^2 is the prior variance. What does increasing σ_0^2 say about our prior beliefs about θ?

- Increasing the variance of the prior **narrows** the range of what we think θ might be, indicating **less** confidence in our prior mean guess μ_0.

- Increasing the variance of the prior **widens** the range of what we think θ might be, indicating **greater** confidence in our prior mean guess μ_0.

- Increasing the variance of the prior **narrows** the range of what we think θ might be, indicating **greater** confidence in our prior mean guess μ_0.

-

Increasing the variance of the prior **widens** the range of what we think θ might be, indicating **less** confidence in our prior mean guess μ_0.

Correct
This also lowers the "effective sample size" of the prior, so that the data become more influential in determining the posterior for θ.

---

1 / 1 points

3.
In the lesson, we presented Bayes' theorem for the case where parameters are continuous. What is the correct expression for the posterior distribution of θ if it is discrete (takes on only specific values)?

- $p(\theta \mid y) = \frac{p(y|\theta) \cdot p(\theta)}{\int p(y|\theta) \cdot p(\theta) \, d\theta}$
- $p(\theta) = \int p(\theta \mid y) \cdot p(y) \, dy$
- $p(\theta) = \sum_j p(\theta \mid y_j) \cdot p(y_j)$
- $p(\theta_j \mid y) = \frac{p(y|\theta_j) \cdot p(\theta_j)}{\sum_j p(y|\theta_j) \cdot p(\theta_j)}$

Correct
This is the discrete version of Bayes' theorem.

---

1 / 1 points

4.
For Questions 4 and 5, refer to the following scenario.

In the quiz for Lesson 1, we described Xie's model for predicting demand for bread at his bakery. During the lunch hour on a given day, the number of orders (the response variable) follows a Poisson distribution. All days have the same mean (expected number of orders). Xie is a Bayesian, so he selects a conjugate gamma prior for the mean with shape 3 and rate $1/15$. He collects data on Monday through Friday for two weeks.

Which of the following hierarchical models represents this scenario?

- $$y_i \mid \lambda \overset{\text{iid}}{\sim} \text{Pois}(\lambda) \quad \text{for } i = 1, \ldots, 10,$$
  $$\lambda \mid \mu \sim \text{Gamma}(\mu, 1/15)$$
  $$\mu \sim \text{N}(3, 1.0^2)$$
- $$y_i \mid \mu \overset{\text{iid}}{\sim} \text{N}(\mu, 1.0^2) \quad \text{for } i = 1, \ldots, 10,$$
  $$\mu \sim \text{N}(3, 15^2)$$

$$y_i \mid \lambda_i \overset{\text{ind}}{\sim} \text{Pois}(\lambda_i) \quad \text{for } i = 1, \ldots, 10,$$
$$\lambda_i \mid \alpha \sim \text{Gamma}(\alpha, 1/15)$$
$$\alpha \sim \text{Gamma}(3.0, 1.0)$$

$$y_i \mid \lambda \overset{\text{iid}}{\sim} \text{Pois}(\lambda) \quad \text{for } i = 1, \ldots, 10,$$
$$\lambda \sim \text{Gamma}(3, 1/15)$$

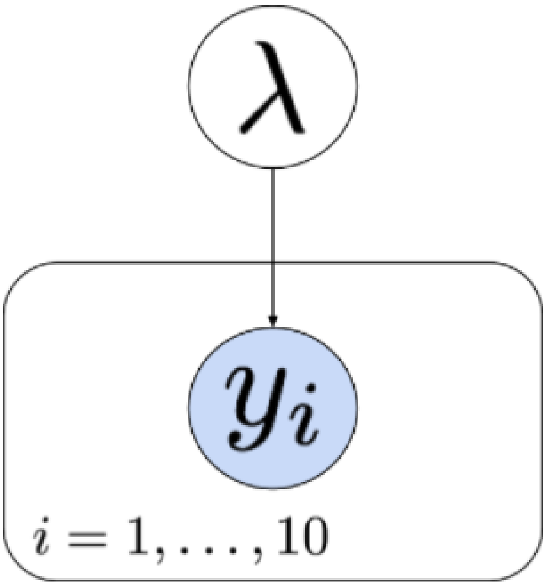
Correct
The likelihood is Poisson with the same mean for all observations, called λ here. The mean λ has a gamma prior.

---

1 / 1
points

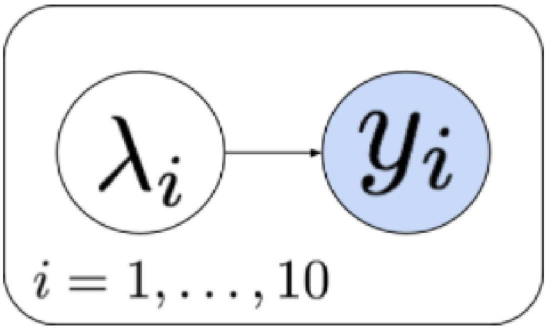
5.
Which of the following graphical depictions represents the model from Xie's scenario?

a)

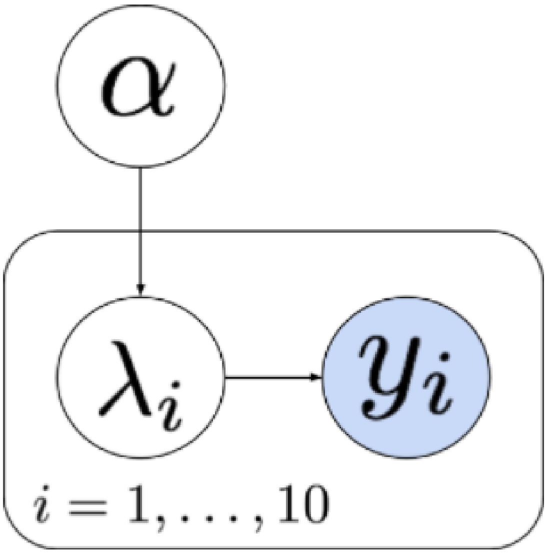


Correct
The observed data variables each depend on the mean demand.

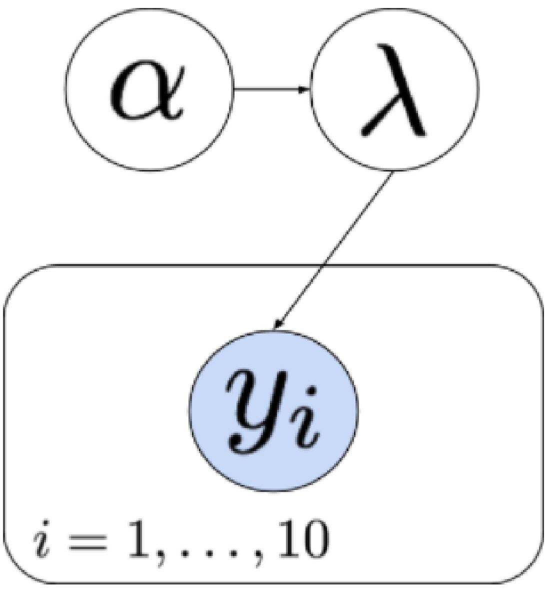
b)


λ_i
y_i
$i = 1, \ldots, 10$

c)


α
λ_i
y_i
$i = 1, \ldots, 10$

d)

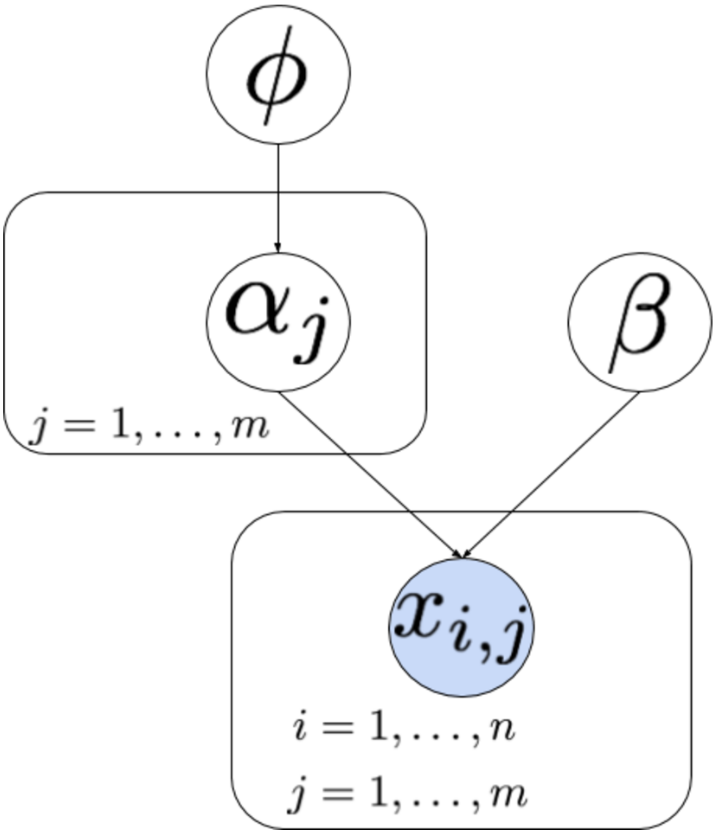

α
λ
y_i
$i = 1, \ldots, 10$

1 / 1
points

6.

Graphical representations of models generally do not identify the distributions of the variables (nodes), but they do reveal the structure of dependence among the variables.

Identify which of the following hierarchical models is depicted in the graphical representation below.



- ( ) $x_{i,j} \mid \alpha_i, \beta_j \overset{\text{ind}}{\sim} \text{Gamma}(\alpha_i, \beta_j), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m$
  $\beta_j \mid \phi \overset{\text{iid}}{\sim} \text{Exp}(\phi), \quad j = 1, \ldots, m$
  $\alpha_i \mid \phi \overset{\text{iid}}{\sim} \text{Exp}(\phi), \quad i = 1, \ldots, n$
  $\phi \sim \text{Exp}(r_0)$

- ( ) $x_{i,j} \mid \alpha, \beta \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m$
  $\beta \sim \text{Exp}(b_0)$
  $\alpha \sim \text{Exp}(a_0)$
  $\phi \sim \text{Exp}(r_0)$

- ( )

$$x_{i,j} \mid \alpha_j, \beta \overset{\text{ind}}{\sim} \text{Gamma}(\alpha_j, \beta), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m$$
$$\beta \sim \text{Exp}(b_0)$$
$$\alpha_j \sim \text{Exp}(a_0), \quad j = 1, \ldots, m$$
$$\phi \sim \text{Exp}(r_0)$$

- $$x_{i,j} \mid \alpha_j, \beta \overset{\text{ind}}{\sim} \text{Gamma}(\alpha_j, \beta), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m$$
$$\beta \sim \text{Exp}(b_0)$$
$$\alpha_j \mid \phi \overset{\text{iid}}{\sim} \text{Exp}(\phi), \quad j = 1, \ldots, m$$
$$\phi \sim \text{Exp}(r_0)$$

Correct

$x_{i,j}$ depends on α_j and β. β doesn't depend on anything. α_j depends on ϕ.

Notice that the $x_{i,j}$ variables are independent (denoted $\overset{\text{ind}}{\sim}$) rather than independent and identically distributed ($\overset{\text{iid}}{\sim}$) because the distribution of $x_{i,j}$ changes with the index j (they have different shape parameters α_j).

---

1 / 1
points

7.

Consider the following model for a binary outcome y:

$$y_i \mid \theta_i \overset{\text{ind}}{\sim} \text{Bern}(\theta_i), \quad i = 1, \ldots, 6$$
$$\theta_i \mid \alpha \overset{\text{iid}}{\sim} \text{Beta}(\alpha, b_0), \quad i = 1, \ldots, 6$$
$$\alpha \sim \text{Exp}(r_0)$$

where θ_i is the probability of success on trial i. What is the expression for the joint distribution of all variables, written as $p(y_1, \ldots, y_6, \theta_1, \ldots, \theta_6, \alpha)$ and denoted by $p(\cdots)$? You may ignore the indicator functions specifying the valid ranges of the variables (although the expressions are technically incorrect without them).

Hint:

The PMF for a Bernoulli random variable is $f_y(y \mid \theta) = \theta^y (1-\theta)^{1-y}$ for $y = 0$ or $y = 1$ and $0 < \theta < 1$.

The PDF for a Beta random variable is $f_\theta(\theta \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$ where $\Gamma()$ is the gamma function, $0 < \theta < 1$ and $\alpha, \beta > 0$.

The PDF for an exponential random variable is $f_\alpha(\alpha \mid \lambda) = \lambda \exp(-\lambda\alpha)$ for $\lambda, \alpha > 0$.

- $p(\cdots) = \prod_{i=1}^{6} \left[\theta_i^{y_i} (1-\theta_i)^{1-y_i} \frac{\Gamma(\alpha+b_0)}{\Gamma(\alpha)\Gamma(b_0)} \theta_i^{\alpha-1} (1-\theta_i)^{b_0-1} \right]$
- $p(\cdots) = \prod_{i=1}^{6} \left[\theta_i^{y_i} (1-\theta_i)^{1-y_i} \frac{\Gamma(\alpha+b_0)}{\Gamma(\alpha)\Gamma(b_0)} \theta_i^{\alpha-1} (1-\theta_i)^{b_0-1} r_0 \exp(-r_0\alpha) \right]$
-

$$p(\cdots) = \prod_{i=1}^{6} \left[\theta_i^{y_i}(1-\theta_i)^{1-y_i}\right] \cdot \frac{\Gamma(\alpha+b_0)}{\Gamma(\alpha)\Gamma(b_0)}\, \theta^{\alpha-1}(1-\theta)^{b_0-1} \cdot r_0 \exp(-r_0\alpha)$$

$$p(\cdots) = \prod_{i=1}^{6} \left[\theta_i^{y_i}(1-\theta_i)^{1-y_i} \frac{\Gamma(\alpha+b_0)}{\Gamma(\alpha)\Gamma(b_0)}\, \theta_i^{\alpha-1}(1-\theta_i)^{b_0-1}\right] \cdot r_0 \exp(-r_0\alpha)$$

Correct

This expression is proportional to the posterior distribution $p(\theta_1, \ldots, \theta_6, \alpha \mid y_1, \ldots, y_6)$. Unfortunately, it does not correspond with a common distribution, so evaluating this posterior would be very challenging (at least until we learn MCMC techniques in the next module).

---

1 / 1 points

8. In a Bayesian model, let y denote all the data and θ denote all the parameters. Which of the following statements about the relationship between the joint distribution of all variables $p(y, \theta) = p(\cdots)$ and the posterior distribution $p(\theta \mid y)$ is true?

- They are proportional to each other so that $p(y, \theta) = c \cdot p(\theta \mid y)$ where c is a constant number that doesn't involve θ at all.

Correct

This fact allows us to work with the joint distribution $p(y, \theta)$ which is usually easier to compute. MCMC methods, which we will learn in the next module, only require us to know the posterior up to proportionality.

- They are actually equal to each other so that $p(y, \theta) = p(\theta \mid y)$.
- The joint distribution $p(y, \theta)$ is equal to the posterior distribution times a function $f(\theta)$ which contains the modification (update) of the prior.
- Neither is sufficient alone--they are both necessary to make inferences about θ.