

Homework 2

Gaurav Kandlikar & Marcel Vaz

October 14, 2016

1a: Algebraically derive the posterior $p(\lambda|y_1)$ given a single observation y_1 . Specify your answer (a) in terms of a named distribution with parameters (ie $\text{Gamma}(a, b)$), and specify a and b , and give the actual density formula.

Note: The work for this question is shown in the Appendix to keep the body clean

Our posterior $p(\lambda|y_1) \sim \text{Gamma}(\alpha + y, \beta + 1)$.

The density of this distribution is:

$$\frac{(\beta + 1)^{(\alpha+y)} * \lambda^{(\alpha+1)} * e^{(-\lambda(\beta+1))}}{\Gamma(\alpha + y)}$$

1b: What is the support (place where density/function is non-negative) of: (i) prior, (ii) posterior, (iii) sampling density, (iv) likelihood (v) prior predictive?

The support for all of these is the space $0, \infty$.

1c: Algebraically derive the posterior given a sample $y_i, i = 1, \dots, n$ of size n . Again specify it both in terms of a named distribution with parameters, and give the actual density formula.

Note: The work for this question is shown in the Appendix to keep the body clean

Given $\mathbf{Y} = [y_1, y_2, \dots, y_n]$ Our posterior $p(\lambda|\mathbf{Y}) \sim \text{Gamma}(\alpha + \sum_{i=1}^n y_i, \beta + n)$.

The density of this distribution is:

$$\frac{(\beta + n)^{(\alpha + \sum_{i=1}^n y_i)} * \lambda^{(\alpha + \sum_{i=1}^n y_i - 1)} * e^{(-\lambda(\beta+n))}}{\Gamma(\alpha + \sum_{i=1}^n y_i)}$$

1d: In the prior $\text{gamma}(a, b)$, which parameter acts like a prior sample size? (Hint: look at the posterior from problem (1c), how does n enter into the posterior density?) You will need this answer in the second part of problem 1(g)iv or you will do that problem incorrectly.

In the prior $\text{gamma}(a, b)$, the parameter β approximates the prior sample size s_0 . This is because the gamma prior for λ when we have $\mathbf{Y} = [y_1, y_2, \dots, y_n]$, $b = b_0 + n$.

1f: Name your store, and the date and time.

I will visit the Santa Monica REI on Friday, October 14 2016 and watch for people coming in between 1:15-1:20 PM.

1g: We are now going to specify the parameters a and b of the gamma prior density.

We will do this in two different ways, giving two different priors. We designate one set of prior parameters as a1 and b1; the other set of prior parameters are a2 and b2.

1. Before you visit the store, make a guess as to the mean number of customers entering the store in one minute. Call this m0. This is the mean of your prior distribution for lambda

I estimate a prior mean m_0 of 7 people entering the store per minute.

2. Make a guess s0 of the prior sd associated with your estimate m0. This s0 is the standard deviation of the prior distribution for λ . Note: most people underestimate s0.

I estimate a standard deviation s_0 of 3 people: sometimes nobody may come in (after all I am going on a weekday afternoon); other times, a big group may come in (this store is right by the Santa Monica promenade, so maybe lots of tourist groups?)

3. Separately from the previous question 1(g)ii, estimate how many data points n0 your prior guess is worth. That is, n0 is the number (strictly greater than zero) of data points (counts of 5 minutes) you would just as soon have as have your prior guess of m0.

For the purpose of this exercise, I will assume that **my prior is worth just 1 data point**. I don't frequent stores and am not tracking customer dynamics when I do, so I am going with what I think is a conservative approach.

4. Solve for a1 and b1 based on m0 and s0.

$$E[\lambda] = \frac{\alpha}{\beta}; Var[\lambda] = \frac{\alpha}{\beta^2}$$

$$7 = \frac{\alpha}{\beta}; 9 = \frac{\alpha}{\beta^2}$$

$$\alpha_1 = 49/9; \beta_2 = 7/9$$

5. Separately solve for a2 and b2 using m0 and n0 only. You usually will not get the same answer each time. This is ok and is NOT wrong. (Note: if you do get the same answer, then please specify a second choice of a2 , b2 to use with the remainder of this problem!)

We can set $\beta = n_0$, as the value of n modifies the rate parameter in our posterior derived above. Therefore, $\beta = 1$, and we can use this value to calculate α :

$$\begin{aligned} E[\lambda] &= \frac{\alpha}{\beta} \\ 7 &= \frac{\alpha}{1} \\ \alpha &= 7; \beta = 1 \end{aligned}$$

The variance of this would also be 7, as $\frac{\alpha}{\beta^2} = \frac{7}{1^2} = 7$.

1h: Suppose we need to have a single prior, rather than two priors. Suggest 2 distinct methods to settle on a single prior.

First, we might pick between the two priors by calculating the standard deviation from the second approach and, if the two SDs are distinct, ask ourselves which we believe more. If our estimate of s_0 seems more reasonable than that calculated by using m_0 and n_0 , we might choose to stick with the former.

Second, I would spend more time eliciting information from industry experts before setting my priors. I would ask them what an average number of customers entering per minute might be, and also ask for an estimate on the upper bound for the number of customers coming in. I might then treat this "upper bound" as a 95% interval and calculate α and β .

1i: Go to your store and collect your data as instructed in 1e. Report it here.

Minute 1	Minute 2	Minute 3	Minute 4	Minute 5
3	6	1	0	9

1j: Update both priors algebraically using your 5 data points. Give the two posteriors.

Prior 1a

$$\alpha_{posterior} = \alpha_{prior} + \sum_{i=1}^n y_i = \frac{49}{9} + 3 + 6 + 1 + 0 + 9 = 24.444$$

$$\beta_{posterior} = \beta_{prior} + n = \frac{7}{9} + 5 = 5.778$$

So, the posterior for prior 1a is of the form $Gamma(\alpha = 24.444, \beta = 5.778)$

Prior 1b

$$\alpha_{posterior} = \alpha_{prior} + \sum_{i=1}^n y_i = 7 + 3 + 6 + 1 + 0 + 9 = 26$$

$$\beta_{posterior} = \beta_{prior} + n = 1 + 5 = 6$$

So, the posterior for prior 1b is of the form $Gamma(\alpha = 26, \beta = 6)$

1k: Give the posterior mean and variance for your two posteriors.

Prior 1a

$$\mu_{posterior} = \frac{\alpha_{posterior}}{\beta_{posterior}} = \frac{24.444}{5.778} = 4.231$$

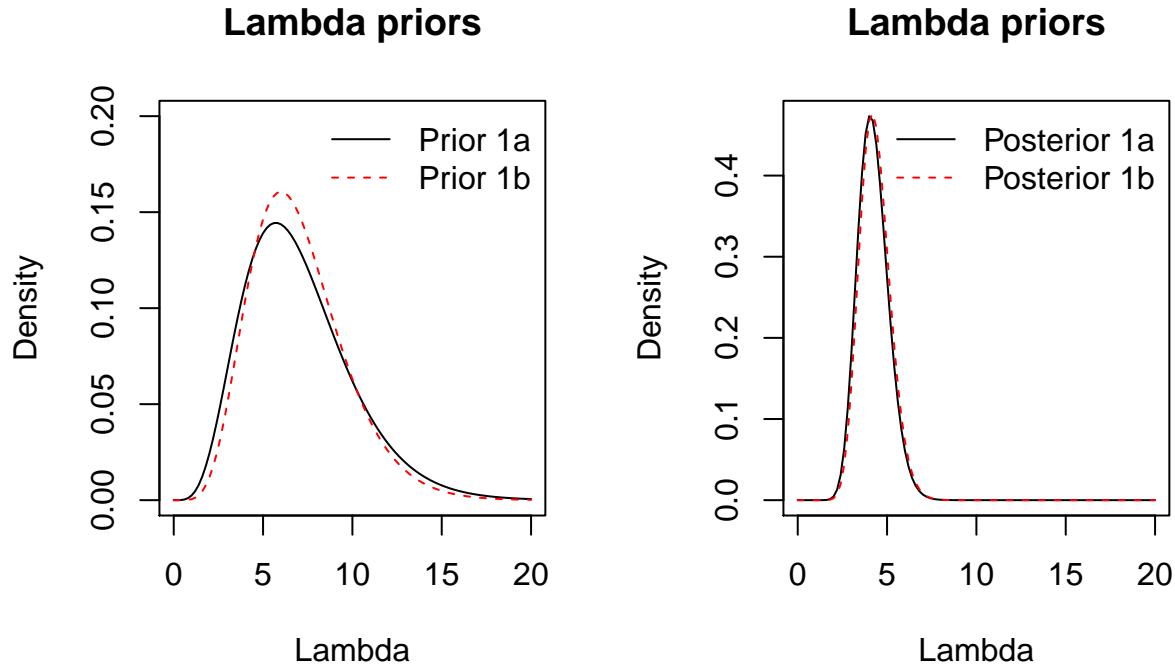
$$Var_{posterior} = \frac{\alpha_{posterior}}{\beta_{posterior}^2} = \frac{24.444}{5.778^2} = 0.732$$

Prior 1b

$$\mu_{posterior} = \frac{\alpha_{posterior}}{\beta_{posterior}} = \frac{26}{6} = 4.33$$

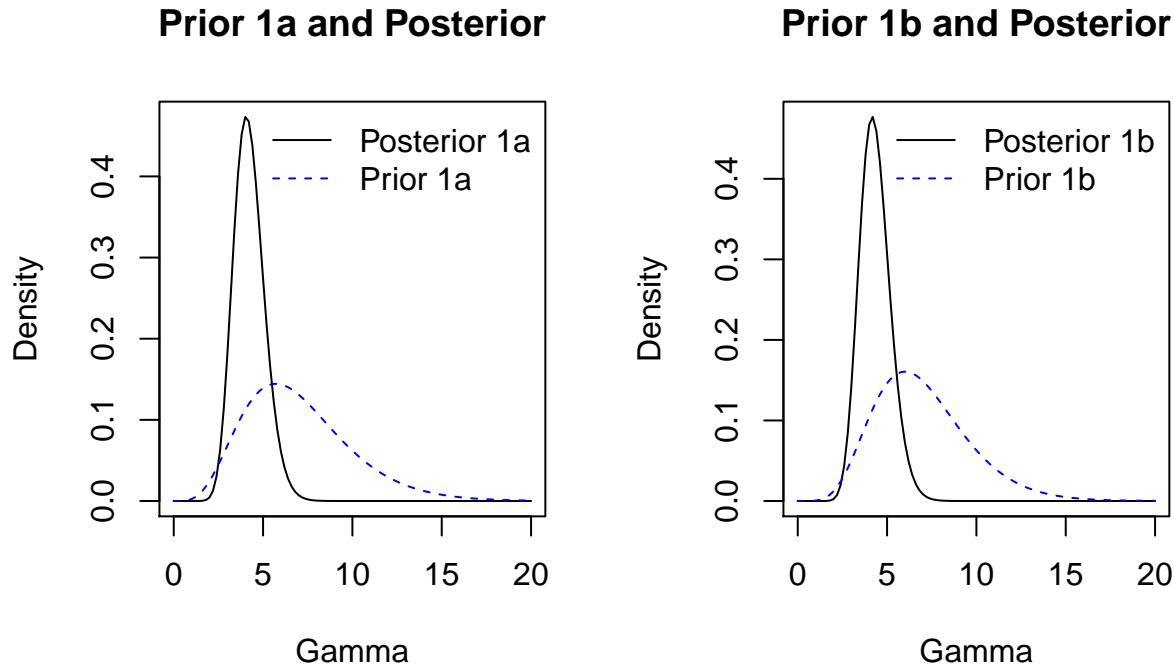
$$Var_{posterior} = \frac{\alpha_{posterior}}{\beta_{posterior}^2} = \frac{26}{6^2} = 0.722$$

11: Plot your two prior densities on one graph. Plot your two posterior densities in another graph. In one sentence for each plot, compare the densities.



Both prior densities are centered around the prior mean of 7 and are fairly wide (prior sd = 3); prior 1b is slightly to the right of prior 1a (higher mean). Both posterior densities are centered around a lower mean of ~4.5 and are much narrower (low posterior SD); again, posterior 1b is slightly to the right of posterior 1a (due to the priors).

1m: Plot each prior density/posterior density pair on the same graph. For each plot, compare the two densities in one sentence.



In both cases, the prior has a higher mean and standard deviation than the posterior mean: this means that I probably overestimated the popularity of REI.

1n: Use WinBUGS (twice) to update your two priors with your data to get your two posteriors. Compare summary statistics between the two posteriors.

```
# Write the model
sink("model_hw2.txt")
cat("model {
  # prior
  lambda ~ dgamma(alpha, beta)

  for (i in 1:N){
    x[i] ~ dpois(lambda)
  }

}", fill = TRUE)
sink()
```

I have used the JAGS model above, along with appropriately set priors, the dataset, and N, to simulate the two posteriors.

```
## module glm loaded
```

We can compare the two posteriors graphically and in a table:

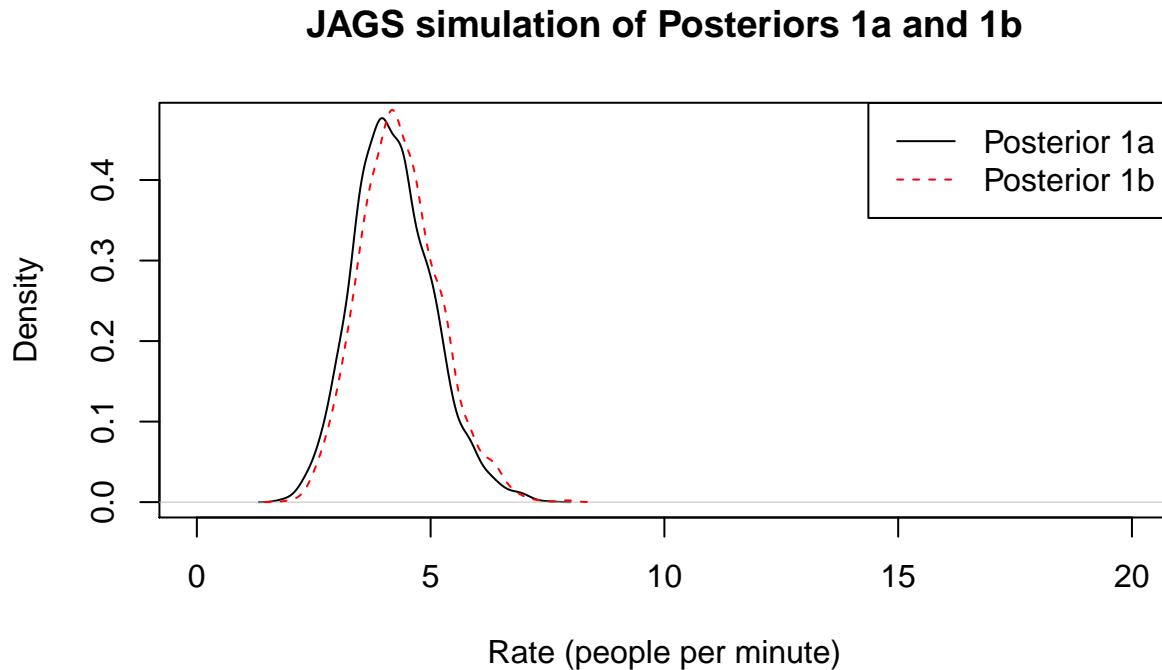


Table 2: Summary of simulations of two posteriors

	mean	sd	2.5%	97.5%
Posterior 1	4.198	0.851	2.670	6.048
Posterior 1b	4.345	0.850	2.809	6.211

We see above that both simulations have very similar values of posterior mean, posterior SD, and CI, and below we will see that the simulated metrics are very close to the algebraic estimates.

1o: How close are the WinBUGS numerical calculations to the actual algebraically calculated posterior means?

Table 3: Comparing calculted vs. simulated Posterior statistics

	mean	2.5%	97.5%	sd
Calculated Posterior 1	4.231	2.519	5.942	0.856
Simulated Posterior 1	4.198	2.670	6.048	0.851
Calculated Posterior 2	4.333	2.634	6.033	0.850
Simulated Posterior 2	4.345	2.809	6.211	0.850

For both priors, the simulated and calculated posterior Mean and SD were very close (10^{-3} accuracy) to the algebraically calculated Mean and SD. That said, the 95% CI around the mean was quite wide in the simulations, ranging from 50% to 150% of the calculated mean.

Question 2

2a: Give algebraic formulas for the relationships between (i) lambda5 and lambda1, (ii) the prior mean of lambda5 and lambda1, (iii) prior variances, (iv) prior standard deviations, (v) prior a-parameters, and (vi) b-parameters.

1. Relationship between Lambda5 and Lambda1

λ_5 is the estimated rate of people entering the store per 5 minutes, whereas λ_1 is the rate of people entering the store per minute. So, the expectation is that $\lambda_5 = \lambda_1 * 5$.

2. Prior mean of Lambda5 and Lambda1

$$\text{Prior mean of } \lambda_5 = 5 * (\text{Prior mean of } \lambda_1)$$

3. Prior variances of Lambda5 and Lambda 1

$$\text{Prior Var of } \lambda_5 = 5^2 * (\text{Prior Var of } \lambda_1)$$

4. Prior SD of Lambda5 and Lambda1

$$\text{Prior SD of } \lambda_5 = 5 * (\text{Prior SD of } \lambda_1)$$

5. Prior alpha of Lambda5 and Lambda1

$$\alpha_{\lambda_5} = \alpha_{\lambda_1}$$

6. Prior beta of Lambda5 and Lambda1

$$\beta_{\lambda_5} = \frac{1}{5} * \beta_{\lambda_1}$$

2b: Give the two priors for the parameter lambda5 that correspond to your priors for lambda1

Prior 5.1

$$\lambda_5 \sim \text{Gamma} \left(\alpha = \frac{49}{9}, \beta = \frac{1}{5} * \frac{7}{9} \right) = \text{Gamma}(5.444, 0.1556)$$

Prior 5.2

$$\lambda_5 \sim \text{Gamma} \left(\alpha = 7, \beta = \frac{1}{5} * 1 \right) = \text{Gamma}(7, 0.2)$$

2c: Give the two resulting posteriors for lambda5 .

Posterior 1

$$\lambda_5 \sim \text{Gamma} \left(\frac{49}{9} + 19, \frac{1}{5} * \frac{7}{9} + 1 \right) = \text{Gamma}(24.444, 1.1556)$$

Posterior 2

$$\lambda_5 \sim \text{Gamma} \left(7 + 19, \frac{1}{5} * 1 + 1 \right) = \text{Gamma}(26, 1.2)$$

2d: Explain the relationship between the posterior means of lambda5 and lambda1. Repeat for the posterior variance, posterior standard deviation,posterior a-parameters and finally posterior b parameters.

Table 4: Parameters and moments for the four posterior distributions

	alpha	beta	mean	sd	var
Lambda1a	24.444	5.778	4.231	0.732	0.536
Lambda1b	26.000	6.000	4.333	0.722	0.522
Lambda5a	24.444	1.156	21.154	18.306	335.117
Lambda5b	26.000	1.200	21.667	18.056	326.003

Table 4 above shows that α parameters are equal for each λ_1 and its corresponding λ_5 , and that the β values for both λ_5 s are 1/5th of the corresponding λ_1 β s. This makes sense- since β can be considered as an analogue of sample size, reducing the independent sample size from 5 to 1 would lead to this decrease. We see that mean of λ_5 s is 5 times the mean of the corresponding λ_1 .

2e: Do you need to redraw your plots (of priors and posteriors) that you drew in the previous problem? How could you alter them without redrawing to make them conform to the new data structure?

We don't strictly need to redraw the plots of posteriors and priors, since the shapes of the densities should remain the same. However, we will need to change the scale of both axes: the X-mean will now be around 5*old_mean, and the Y-axis will end up being shorter, since the density values at any single value of X will be lower.

2f: Do your conclusions change if you consider your data as a single 5 minute observation or as 5 one minute observations? That is, do your recommendations to the store on staffing levels change?

My recommendations do somewhat change based on which type of data I work with (5 x 1 minute datapoints or 1 x 5 minute datapoint): I'd recommend that REI ought to have enough staff on hand to handle a mean of 240 customers per hour on Friday afternoons, but if I am working only with 1 datapoint, I might suggest that they also have enough staff on hand to handle a surge in visits (high standard deviation in rate).

Appendix

Question 1a

(1)

Getting the posterior for Poisson dataset y and a Gamma prior.

$$p(\lambda|y) = \frac{\textcircled{1} p(y|\lambda) \textcircled{2} p(\lambda)}{\textcircled{3} P(y)}$$

$$\textcircled{1} p(y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!}$$

$$\textcircled{2} p(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda(\beta)}}{\Gamma(\alpha)} \quad \text{for } \text{Gamma}(\lambda; \alpha, \beta)$$

$$\textcircled{3} P(y) = \int_0^\infty p(y|\lambda) p(\lambda) d\lambda$$

$$= \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda(\beta)}}{\Gamma(\alpha)} \cdot \frac{e^{-\lambda} \lambda^y}{y!} d\lambda = \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}}{y! \Gamma(\alpha)} d\lambda$$

We multiply the integral by two convenience terms:

$$= \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}}{y! \Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+y)}{\Gamma(\alpha+y)} \cdot \frac{(1+\beta)^{\alpha+y}}{(1+\beta)^{\alpha+y}} d\lambda$$

and we factor out some terms that are constant WRT λ :

$$\frac{\beta^\alpha \Gamma(\alpha+y)}{y! \Gamma(\alpha) (1+\beta)^{\alpha+y}} \int_0^\infty \frac{(1+\beta)^{\alpha+y} \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}}{\Gamma(\alpha+y)} d\lambda$$

As the term inside the integral is the PDF of Gamma ($\lambda; \alpha+y, \beta+1$), integrating across the range of $0 \rightarrow \infty$ yields 1.

(2)

⑤ Cont'd

Once we set the integral equal to 1, the denominator term ③ reduces to

$$\frac{\beta^\alpha \Gamma(\alpha+y)}{y! \Gamma(\alpha) (1+\beta)^{\alpha+y}}$$

Putting together ①, ②, and ③:

$$\begin{aligned} & \frac{\frac{e^{-\lambda} \lambda^y}{y!} \cdot \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda \beta}}{\Gamma(\alpha)}}{\frac{\beta^\alpha \Gamma(\alpha+y)}{y! \Gamma(\alpha) (1+\beta)^{\alpha+y}}} \\ &= \frac{e^{-\lambda} \lambda^y}{y!} \frac{\cancel{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda \beta}}}{\cancel{\Gamma(\alpha)}} \cdot \frac{\cancel{y! \Gamma(\alpha)} (1+\beta)^{\alpha+y}}{\cancel{\beta^\alpha \Gamma(\alpha+y)}} \\ &= \frac{(1+\beta)^{\alpha+y} \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}}{\Gamma(\alpha+y)} \end{aligned}$$

which is the PDF of Gamma ($\lambda; \alpha+y, \beta+1$)

Question 1c

Getting the posterior for Poisson distributed dataset $Y = [y_1, y_2, \dots, y_n]$ and a Gamma prior:

$$P(\lambda|Y) = \frac{\textcircled{1} P(Y|\lambda) P(\lambda)}{\textcircled{3} P(Y)}$$

$$\textcircled{1} P(Y|\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda} \lambda^{\sum y_i}}{\prod_{i=1}^n y_i!}$$

$$\textcircled{2} P(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} \quad \text{for } P(\lambda; \alpha, \beta)$$

$$\textcircled{3} \int_0^\infty \frac{\beta^\alpha \lambda^{\sum y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha) \prod_{i=1}^n y_i!} d\lambda$$

We multiply the integral by two convenience terms:

$$= \int_0^\infty \frac{\beta^\alpha \lambda^{\sum y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha) \prod_{i=1}^n y_i!} \cdot \frac{\Gamma(\alpha + \sum y_i)}{\Gamma(\alpha + \sum y_i)} \cdot \frac{(\beta+n)^{\alpha + \sum y_i}}{(\beta+n)^{\alpha + \sum y_i}} d\lambda$$

Factor out some terms that are constant w.r.t λ :

$$\frac{\beta^\alpha \Gamma(\alpha + \sum y_i)}{\Gamma(\alpha) \prod_{i=1}^n y_i!} \cdot (\beta+n)^{\alpha + \sum y_i} \int_0^\infty \frac{(\beta+n)^{\alpha + \sum y_i} \lambda^{\sum y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\Gamma(\alpha + \sum y_i)} d\lambda$$

as the term inside integral is the PDF of Gamma($\lambda; \alpha + \sum y_i, \beta+n$), integrating over the whole range of (λ) , i.e $0 \rightarrow \infty$, will yield 1

③, cont

Once we set the integral to 1, the denominator term ③ is:

$$\frac{\beta^\alpha \Gamma(\alpha + \sum_i y_i)}{\Gamma(\alpha) \cdot \prod_{i=1}^n y_i! \cdot (\beta+n)^{\sum_i y_i + \alpha}}$$

Putting together ①, ②, and ③:

$$\frac{\frac{\stackrel{(1)}{e^{-\lambda n}} \lambda^{\sum_i y_i}}{\prod_i y_i!} \cdot \frac{\stackrel{(2)}{\beta^\alpha \lambda^{\alpha+1} e^{-\lambda \beta}}}{\Gamma(\alpha)}}{\frac{\stackrel{(3)}{\beta^\alpha \Gamma(\alpha + \sum_i y_i)}}{\Gamma(\alpha) \cdot \prod_{i=1}^n y_i! \cdot (\beta+n)^{\sum_i y_i + \alpha}}}$$

$$= \frac{\beta^\alpha \lambda^{\sum_i y_i + \alpha + 1} e^{-\lambda(\beta+n)}}{\prod_i y_i! \Gamma(\alpha)}$$

$$= \frac{\beta^\alpha \lambda^{\sum_i y_i + \alpha - 1} e^{-\lambda(\beta+n)}}{\prod_i y_i! \Gamma(\alpha)} \cdot \frac{\Gamma(\alpha) \cdot \prod_{i=1}^n y_i! (\beta+n)^{\sum_i y_i + \alpha}}{\beta^\alpha \Gamma(\alpha + \sum_i y_i)}$$

$$= \frac{(\beta+n)^{\sum_i y_i + \alpha}}{\Gamma(\alpha + \sum_i y_i)} \lambda^{\sum_i y_i + \alpha - 1} e^{-\lambda(\beta+n)}$$

which is the pdf of Gamma $(\lambda; \alpha + \sum_{i=1}^n y_i, \beta+n)$