XXXX 2020
xx.xx – x.xx

EXAMINATION FOR THE DEGREES OF M.Sci., M.Sc. and M.Res.

Bayesian Statistics (Level M) Solutions

This paper consists of 12 pages and contains 4 questions.
Candidates should attempt all questions.

Question 1	20 marks
Question 2	20 marks
Question 3	20 marks
Question 4	20 marks
Total	80 marks

The formula sheet and statistical tables are provided in the end of this paper.

"An electronic calculator may be used provided that it is allowed under the School of Mathematics and Statistics Calculator Policy. A copy of this policy has been distributed to the class prior to the exam and is also available via the invigilator."

CONTINUED OVERLEAF/

1. The number of car breakdowns per year of service can be modelled with a Poisson distribution:

$$y_i \sim \text{Poi}(\lambda), \quad i = 1, \ldots, n, \text{ independently,}$$

where y_i is the number of breakdowns in year i, and rate $\lambda > 0$.

(a) Derive Jeffreys' prior for the parameter λ of this model. **[4 MARKS]**

Solution

This was done multiple times in lectures and tutorials. The students were also advised to practice these derivations for most common distributions. This derivation was also covered in past exam papers.

$$p(y|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$

$$\ln p(y|\lambda) = \sum_{i=1}^{n} (y_i \ln \lambda - \lambda - \ln(y_i!))$$

$$\frac{d \ln p(y|\lambda)}{d\lambda} = \sum_{i=1}^{n} \frac{y_i}{\lambda} - 1$$

$$\frac{d^2 \ln p(y|\lambda)}{d\lambda^2} = -\sum_{i=1}^{n} \frac{y_i}{\lambda^2}$$

$$\mathbb{E}\left[\frac{d^2 \ln p(y|\lambda)}{d\lambda^2} \middle| \lambda \right] = -\sum_{i=1}^{n} \frac{1}{\lambda} = -\frac{n}{\lambda}$$

$$J(\lambda) = \frac{n}{\lambda}$$

$$p(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

(b) Is this Jeffreys' prior proper? Explain. **[2 MARKS]**

Solution

Definition of proper and improper priors was covered in the lectures.

The prior is proper when the integral over the domain of the variable is finite.

$$\int_0^\infty \frac{1}{\sqrt{\lambda}} d\lambda = \left[2\sqrt{\lambda} \right]_{\lambda=0}^{\infty} = \infty.$$

So, this prior is improper.

CONTINUED OVERLEAF/

(c) Can this Jeffreys' prior be used for inference and why? **[2 MARKS]**

Solution

This was iterated many times in lecture material.

An improper prior can be used for inference if it leads to a proper posterior. Assuming that $n > 0$, $y_i \geq 0$, and $\sum_{i=1}^{n} y_i > 0$, we observe

$$p(\lambda|y) \propto \frac{1}{\sqrt{\lambda}} \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$$
$$\propto \lambda^{-1/2+\sum y_i} e^{-n\lambda} \propto \text{Ga}(1/2 + \sum y_i, n),$$

which is a proper distribution, and therefore this prior can be used for inference when the data are not trivial.

(d) Assuming Jeffreys' prior and the following observations of the number of breakdowns in the first three years $y = (2, 1, 3)$ derive the posterior distribution of the breakdown rate parameter λ. **[2 MARKS]**

Solution

Using the result obtained in the previous part, we get

$$\lambda|y \sim \text{Ga}(6.5, 3)$$

(e) Derive the probability mass function for the posterior predictive distribution of the number of breakdowns in the fourth year of service given the information available so far. **[5 MARKS]**

Solution

An example of posterior predictive distribution derivation was done in the lectures for the normal and binomial models. Exponential model was covered in past exam papers.

$$\begin{aligned}
p(\tilde{y}|y) &= \frac{3^{6.5}}{\Gamma(6.5)} \int_0^\infty \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!} \lambda^{5.5} e^{-3\lambda} d\lambda \\
&= \frac{3^{6.5}}{\Gamma(6.5)\tilde{y}!} \int_0^\infty \lambda^{\tilde{y}+5.5} e^{-4\lambda} d\lambda \\
&= \frac{3^{6.5}}{\Gamma(6.5)\Gamma(\tilde{y}+1)} \times \frac{\Gamma(\tilde{y}+6.5)}{4^{\tilde{y}+6.5}} \\
&= \frac{\Gamma(\tilde{y}+6.5)}{\Gamma(6.5)\Gamma(\tilde{y}+1)} \left(\frac{3}{4}\right)^{6.5} \left(\frac{1}{4}\right)^{\tilde{y}}
\end{aligned}$$

CONTINUED OVERLEAF/

This is a Negative Binomial distribution, however our standard formula sheet defines the Negative Binomial distribution differently, and students are may not be able to identify it. As long as the p.d.f. is reported correctly, the answer should be considered correct.

(f) You subsequently discover that in the fourth year of service there were fewer than 3 breakdowns. Derive the updated posterior for λ. Make sure that this posterior is properly normalised. You may use the following equations if you need:

$$\Gamma(z+1) = z\Gamma(z), \quad \Gamma(1/2) = \sqrt{\pi}.$$

[5 MARKS]

Solution

Tutorials covered similar examples with exponential distribution and with a binomial case too.

$$\begin{aligned} p(\lambda|y=(2,1,3), z<3) &\propto \lambda^{5.5} e^{-3\lambda} \left[\frac{\lambda^0 e^{-\lambda}}{1} + \frac{\lambda^1 e^{-\lambda}}{1} + \frac{\lambda^2 e^{-\lambda}}{2} \right] \\ &= \lambda^{5.5} e^{-4\lambda} + \lambda^{6.5} e^{-4\lambda} + \frac{1}{2} \lambda^{7.5} e^{-4\lambda} \end{aligned}$$

To normalise this, we solve

$$\begin{aligned} C \int_0^\infty \lambda^{5.5} e^{-4\lambda} d\lambda + C \int_0^\infty \lambda^{6.5} e^{-4\lambda} d\lambda + \frac{C}{2} \int_0^\infty \lambda^{7.5} e^{-4\lambda} d\lambda &= 1 \\ C \frac{\Gamma(6.5)}{4^{6.5}} + C \frac{\Gamma(7.5)}{4^{7.5}} + C \frac{\Gamma(8.5)}{2 \cdot 4^{8.5}} &= 1 \\ C &= \frac{1}{\frac{\Gamma(6.5)}{4^{6.5}} + \frac{\Gamma(7.5)}{4^{7.5}} + \frac{\Gamma(8.5)}{2 \cdot 4^{8.5}}} \end{aligned}$$

And finally, we represent the resulting posterior as a mixture of three gamma distributions:

$$\begin{aligned} p(\lambda|y=(2,1,3), z<3) = C \Bigg(& \frac{\Gamma(6.5)}{4^{6.5}} \mathrm{Ga}(6.5, 4) + \frac{\Gamma(7.5)}{4^{7.5}} \mathrm{Ga}(7.5, 4) \\ & + \frac{\Gamma(8.5)}{2 \cdot 4^{8.5}} \mathrm{Ga}(8.5, 4) \Bigg) \\ & = 0.2410546 \mathrm{Ga}(6.5, 4) + 0.3917137 \mathrm{Ga}(7.5, 4) \\ & + 0.3672316 \mathrm{Ga}(8.5, 4) \end{aligned}$$

It is acceptable to report the weights as fractions and avoid evaluating Gamma functions.

CONTINUED OVERLEAF/

2. Consider the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1/\tau) \quad i = 1, \dots, n, \quad \text{independently},$$

where β_0 is the intercept parameter, β_1 is the slope parameter, and τ is the precision of the normally distributed observation noise. We can rewrite the above model as

$$y_i|\beta_0, \beta_1, \tau, x_i \sim \mathcal{N}\left(\beta_0 + \beta_1 x_i, \frac{1}{\tau}\right), \quad i = 1, \dots, n \quad \text{independently}.$$

Having observed some n data (x_i, y_i), we want to infer the posterior of model parameters β_0, β_1, and τ in a fully Bayesian way, using the following priors:

$$\left.\begin{aligned} \beta_0 &\sim \mathcal{N}(\mu_0, 1/\tau_0) \\ \beta_1 &\sim \mathcal{N}(\mu_1, 1/\tau_1) \\ \tau &\sim \text{Ga}(\alpha, \beta) \end{aligned}\right\} \text{independently}$$

with fixed real values for hyperparameters $\mu_0, \mu_1, \tau_0, \tau_1, \alpha, \beta$.

(a) Define this model for performing sampling from the parameter posterior using `WinBUGS`, i.e., provide the model description that goes within `model{` $\cdots$ `}` in `WinBUGS`. **[4 MARKS]**

Solution

`WinBUGS`work was covered in labs. Students have never been asked to write down a `WinBUGS`model in past exams. Linear regression model has not been covered in the lectures directly, so this question requires reasoning rather than memorisation. All the derivations follow the same scheme that was covered in the lectures.

There is a multitude of possible model descriptions. Two of the most likely ones are:

```
model{
  for (i in 1:n) {
    y[i] ~ dnorm(mu[i],tau)
    mu[i] <- beta0 + beta1*x[i]
  }
  beta0 ~ dnorm(mu0, tau0)
  beta1 ~ dnorm(mu1, tau1)
  tau ~ dgamma(alpha, beta)
}
```

```
model{
  for (i in 1:n) {
    y[i] ~ dnorm(beta0+beta1*x[i],tau)
  }
  beta0 ~ dnorm(mu0, tau0)
  beta1 ~ dnorm(mu1, tau1)
  tau ~ dgamma(alpha, beta)
}
```

(b) Derive necessary full conditional distributions, and describe how you can build a Gibbs sampler to draw samples from the posterior $p(\beta_0, \beta_1, \tau|x, y)$. **[12,4 MARKS]**

Solution

CONTINUED OVERLEAF/

We aim to draw samples from the following posterior:

$$
\begin{aligned}
p(\beta_0, \beta_1, \tau | x, y) &\propto p(y|x, \beta_0, \beta_1, \tau)p(\beta_0)p(\beta_1)p(\tau) \\
&= \tau^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\} \times \\
&\times \exp\left\{-\frac{\tau_0}{2}(\beta_0 - \mu_0)^2\right\} \times \\
&\times \exp\left\{-\frac{\tau_1}{2}(\beta_1 - \mu_1)^2\right\} \times \\
&\times \tau^{\alpha-1}\exp\{-\beta\tau\}.
\end{aligned}
$$

Formulating a Gibbs sampler will require the full conditional distributions of model parameters: *(give up to 4 marks for each full conditional derived, partial marks allowed)*

$$
\begin{aligned}
p(\beta_0|\beta_1, \tau, x, y) &\propto \exp\left\{-\frac{1}{2}\left(\tau_0(\beta_0 - \mu_0)^2 + \tau\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\tau_0\left(\beta_0^2 - 2\mu_0\beta_0\right) + \tau\sum_{i=1}^{n}\left(\beta_0^2 - 2\beta_0(y_i - \beta_1 x_i)\right)\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\beta_0^2(\tau_0 + n\tau) - 2\beta_0\left(\tau_0\mu_0 + \tau\sum_{i=1}^{n}(y_i - \beta_1 x_i)\right)\right)\right\} \\
&= \exp\left\{-\frac{\tau_0 + n\tau}{2}\left(\beta_0^2 - 2\beta_0\frac{\tau_0\mu_0 + \tau\sum_{i=1}^{n}(y_i - \beta_1 x_i)}{\tau_0 + n\tau}\right)\right\} \\
&\propto \exp\left\{-\frac{\tau_0 + n\tau}{2}\left(\beta_0 - \frac{\tau_0\mu_0 + \tau\sum_{i=1}^{n}(y_i - \beta_1 x_i)}{\tau_0 + n\tau}\right)^2\right\} \\
&\propto \mathcal{N}\left(\frac{\tau_0\mu_0 + \tau\sum_{i=1}^{n}(y_i - \beta_1 x_i)}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau}\right)
\end{aligned}
$$

$$
\begin{aligned}
p(\beta_1|\beta_0, \tau, x, y) &\propto \exp\left\{-\frac{1}{2}\left(\tau_1\beta_1^2 - 2\tau_1\mu_1\beta_1 + \tau\sum_{i=1}^{n}\left(\beta_1^2 x_i^2 + 2\beta_0\beta_1 x_i - 2y_i\beta_1 x_i\right)\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\beta_1^2\left(\tau_1 + \tau\sum_{i=1}^{n}x_i^2\right) - 2\beta_1\left(\tau_1\mu_1 + \tau\sum_{i=1}^{n}(y_i - \beta_0)x_i\right)\right)\right\} \\
&\propto \exp\left\{-\frac{\tau_1 + \tau\sum_{i=1}^{n}x_i^2}{2}\left(\beta_1 - \frac{\tau_1\mu_1 + \tau\sum_{i=1}^{n}(y_i - \beta_0)x_i}{\tau_1 + \tau\sum_{i=1}^{n}x_i^2}\right)^2\right\} \\
&\propto \mathcal{N}\left(\frac{\tau_1\mu_1 + \tau\sum_{i=1}^{n}(y_i - \beta_0)x_i}{\tau_1 + \tau\sum_{i=1}^{n}x_i^2}, \frac{1}{\tau_1 + \tau\sum_{i=1}^{n}x_i^2}\right)
\end{aligned}
$$

CONTINUED OVERLEAF/

$$
\begin{aligned}
p(\tau|\beta_0, \beta_1, x, y) &\propto \tau^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right\} \tau^{\alpha-1}\exp\{-\beta\tau\} \\
&= \tau^{\alpha+n/2-1}\exp\left\{-\tau\left(\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{2} + \beta\right)\right\} \\
&\propto \mathrm{Ga}\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{2}\right)
\end{aligned}
$$

Give up to 4 marks for correctly outlining Gibbs sampler

Now we can formulate the algorithm for Gibbs sampling

Initialise $\beta_0 = \beta_0^{(0)}$, $\beta_1 = \beta_1^{(0)}$, and $\tau = \tau^{(0)}$

for $k = 1, \ldots, K$

Draw $\beta_0^{(k)}$ from $\mathcal{N}\left(\frac{\tau_0\mu_0 + \tau\sum_{i=1}^{n}(y_i - \beta_1^{(k-1)}x_i)}{\tau_0 + n\tau^{(k-1)}}, \frac{1}{\tau_0 + n\tau^{(k-1)}}\right)$,

Draw $\beta_1^{(k)}$ from $\mathcal{N}\left(\frac{\tau_1\mu_1 + \tau^{(k-1)}\sum_{i=1}^{n}(y_i - \beta_0^{(k)})x_i}{\tau_1 + \tau^{(k-1)}\sum_{i=1}^{n}x_i^2}, \frac{1}{\tau_1 + \tau^{(k-1)}\sum_{i=1}^{n}x_i^2}\right)$,

Draw $\tau^{(k)}$ from $\mathrm{Ga}\left(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(y_i - \beta_0^{(k)} - \beta_1^{(k)}x_i)^2}{2}\right)$

Discard burn-in and optionally thin-out.

(pay attention to the use of sample indices in the Gibbs sampler when marking).

3. A random sample of n cats is drawn from the population of adult domestic cats in Glasgow and their weights are measured. The average weight of the n sampled cats is $\bar{y} = 6$ pounds. Assume the weights in the population are normally distributed with unknown mean θ, and known standard deviation of 1 pound. Suppose your prior distribution for θ is normal with mean 10 and standard deviation 2.

   (a) Is the suggested prior proper? **[1 MARK]**

Solution

Question 3 covers basic inference procedures, and similar questions were considered in tutorials.

Of course, this normal prior is proper. "Yes" is a sufficient answer.

   (b) Give your posterior distribution for θ. (Your answer should be a function of n). **[6 MARKS]**

Solution

CONTINUED OVERLEAF/

This is a straightforward application of a conjugate update for the normal model with fixed variance.

$$\theta|y \sim \mathcal{N}\left(\mu_n, \tau_n^2\right)$$

$$\mu_n = \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2} = \frac{10/4 + n \cdot 6/1}{1/4 + n/1} = \frac{2.5 + 6n}{0.25 + n}$$

$$\tau_n^2 = \left[\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right]^{-1} = \left[\frac{1}{4} + \frac{n}{1}\right]^{-1} = \frac{4}{1 + 4n}$$

(c) A new cat is sampled at random from the same population. Derive the posterior predictive distribution of this cat's possible weight given the information from your inference result. (Your answer will still be a function of n). **[6 MARKS]**

Solution

$$\tilde{y}|y \sim \mathcal{N}(\mu_n, \sigma^2 + \tau_n^2) = \mathcal{N}\left(\frac{2.5 + 6n}{0.25 + n}, \frac{5 + 4n}{1 + 4n}\right)$$

(d) For $n = 10$, give a 95% posterior interval for θ and a 95% posterior predictive interval for the weight of a newly observed cat. **[3 MARKS]**

Solution

$$n = 10$$

$$\mu_n = \frac{2.5 + 60}{0.25 + 10} = \frac{62.5}{10.25} \approx 6.0976, \qquad \tau_n^2 = \frac{4}{1 + 40} \approx 0.0976$$

$$\tilde{y}|y \sim \mathcal{N}(6.0976, 1.0976)$$

95% posterior interval for θ is therefore

$$6.0976 \pm 1.96\sqrt{0.0976} = 6.0976 \pm 0.6123 = (5.4853, 6.7099) \text{ pounds}$$

95% prediction interval for $\tilde{y}$ is therefore

$$6.0976 \pm 1.96\sqrt{1.0976} = 6.0976 \pm 2.0534 = (4.0442, 8.1510) \text{ pounds}$$

(e) Repeat your calculations for $n = 100$, i.e. give a 95% posterior interval for θ and a 95% posterior predictive interval for the weight of a newly observed cat. Comment on any difference of these results from the results for $n = 10$. **[3,1 MARKS]**

Solution

$$n = 100$$

$$\mu_n = \frac{2.5 + 600}{0.25 + 100} = \frac{602.5}{100.25} \approx 6.009975, \qquad \tau_n^2 = \frac{4}{1 + 400} \approx 0.009975062$$

CONTINUED OVERLEAF/

$$\tilde{y}|y \sim \mathcal{N}(6.009975, 1.009975062)$$

95% posterior interval for θ is therefore

$$6.009975 \pm 1.96\sqrt{0.009975062} = 6.009975 \pm 0.1957555 = (5.81422, 6.20573)$$

95% prediction interval for $\tilde{y}$ is therefore

$$6.009975 \pm 1.96\sqrt{1.009975062} = 6.009975 \pm 1.969751 = (4.040224, 7.979726)$$

Both, the central posterior interval, and the prediction interval for the second case are tighter than in the first one. This is because we are using a larger data sample, and therefore obtaining more information from the data. This in turn shifts the posterior mean closer to the sample mean, and reduces posterior variance. This in turn results in more confident predictions.

---

4. (a) Imagine you are running a radio manufacturing company. You receive a shipment of transistors from a supplier to be used in the assembly of the final product. Checking the performance of each transistor in the shipment is too expensive, so you establish a sampling plan to either accept or reject the shipment. A random sample of 100 transistors from the whole shipment will be tested. Based upon the number y of defective transistors in that random sample, you will take one of the two decisions: a_1 = *accept the shipment*, or a_2 = *reject the shipment*.

As the test sample is relatively small in comparison to the shipment size, we are safe to assume:

$$y \sim \mathrm{Bi}(n = 100; \theta),$$

where θ is the proportion of the defective transistors in the shipment.

Working with a new supplier, you decide to use a uniform prior on θ between 0 and 1.

You select the following loss function:

$$L(\theta, a) = \begin{cases} 10\ \theta, & a = a_1 \\ 1, & a = a_2 \end{cases}$$

as, if the shipment is rejected (a_2 decided), the loss is a constant cost due to inconvenience, delay, and testing of a replacement shipment; while, if the shipment is accepted (a_1 decided), the loss is proportional to θ, since θ will also reflect the proportion of defective radios produced. The factor 10 indicates the relative costs involved in the two types of errors.

You receive a shipment that has $y = 3$ defective transistors in a random sample of 100.

CONTINUED OVERLEAF/

i. What is your posterior for θ given this latest test result with $y = 3$?
[2 MARKS]

Solution

This is a simple question, but it lays out a necessary step for future decision making.
The posterior for θ can be found through the conjugate update:

$$\begin{aligned} y|\theta &\sim \text{Bi}(n = 100, \theta) \\ \theta &\sim \text{Be}(1, 1) \\ \theta|y &\sim \text{Be}(1 + 3, 1 + 100 - 3) = \text{Be}(4, 98) \end{aligned}$$

ii. Find the Bayes expected loss $p(\pi, a)$ for the decision on accepting this shipment.
[5 MARKS]

Solution

A similar question was in past exam papers. Usually students struggle with decision making for a discrete problem.

$$\begin{aligned} p(\pi, a) &= \int_0^1 L(\theta, a) p(\theta|y) d\theta \\ &= \begin{cases} \int_0^1 \frac{10}{B(4, 98)} \theta^4 (1-\theta)^{97} d\theta, & a = a_1 \\ \int_0^1 \frac{1}{B(4, 98)} \theta^3 (1-\theta)^{97} d\theta, & a = a_2 \end{cases} \\ &= \begin{cases} \frac{10B(5, 98)}{B(4, 98)}, & a = a_1 \\ \frac{B(4, 98)}{B(4, 98)}, & a = a_2 \end{cases} \\ &= \begin{cases} 10\frac{\Gamma(5)\Gamma(98)}{\Gamma(103)} \times \frac{\Gamma(102)}{\Gamma(4)\Gamma(98)}, & a = a_1 \\ 1, \quad a = a_2 \end{cases} \\ &= \begin{cases} \frac{40}{102}, & a = a_1 \\ 1, \quad a = a_2 \end{cases} \approx \begin{cases} 0.3922, & a = a_1 \\ 1, \quad a = a_2 \end{cases} \end{aligned}$$

iii. Should you accept or reject this shipment? **[2 MARKS]**

Solution

This test the ability of drawing conclusions from numerical results.

CONTINUED OVERLEAF/

As $p(\pi, a)$ takes its minimum at a_1, that should be the preferred decision as a_1 is the Bayes action for this problem.
Therefore, you should accept the shipment.

iv. What is the minimal number of defective transistors in a random sample of 100 required to reject the shipment according to the above cost function using the same prior? **[6 MARKS]**

Solution

This is a never seen before question in a decision making setup, while being similar to 4(a)ii, it test the ability to generalise the case.
Let's assume there are n defective transistors in a sample of 100. The posterior for θ is going to be

$$\theta|y \sim \text{Be}(1+n, 101-n)$$

Similarly to the previous question, the Bayes expected loss $\rho(\pi, a)$ is

$$\begin{aligned}
\rho(\pi, a) &= \int_0^1 L(\theta, a)p(\theta|y)d\theta \\
&= \begin{cases} \dfrac{10}{B(1+n, 101-n)} \displaystyle\int_0^1 \theta^{1+n}(1-\theta)^{100-n}d\theta, & a = a_1 \\ \dfrac{1}{B(1+n, 101-n)} \displaystyle\int_0^1 \theta^{n}(1-\theta)^{100-n}d\theta, & a = a_2 \end{cases} \\
&= \begin{cases} \dfrac{10B(2+n, 101-n)}{B(1+n, 101-n)}, & a = a_1 \\ \dfrac{B(1+n, 101-n)}{B(1+n, 101-n)}, & a = a_2 \end{cases} \\
&= \begin{cases} \dfrac{10\Gamma(2+n)\Gamma(101-n)\Gamma(102)}{\Gamma(1+n)\Gamma(101-n)\Gamma(103)}, & a = a_1 \\ 1, & a = a_2 \end{cases} \\
&= \begin{cases} \dfrac{10(1+n)}{102}, & a = a_1 \\ 1, & a = a_2 \end{cases}
\end{aligned}$$

In this question we are looking for the smallest integer n such that

$$\begin{aligned}
&\frac{10(1+n)}{102} > 1 \\
&10 + 10n > 102 \\
&10n > 92 \\
&n > 9.2
\end{aligned}$$

CONTINUED OVERLEAF/

Which means that the minimum number of defective transistors in a sample of 100 to reject the shipment is 10.

(b) Explain the concept of a conjugate prior, and demonstrate this in the context of data coming from the Poisson model. **[5 MARKS]**

Solution

This derivation was previously covered in lectures.

A prior is called conjugate if the posterior comes from the same family of distributions as the prior. Conjugate prior allow easy analytical inference through updating distribution hyperparameters.

As an example consider the Poisson model with a Gamma prior:

$$\begin{aligned} y_i|\lambda &\sim \text{Poi}(\lambda), \quad i = 1, \ldots, n, \quad \text{independently} \\ \lambda &\sim \text{Ga}(\alpha, \beta) \end{aligned}$$

The posterior in this case is going to be:

$$\begin{aligned} p(\lambda|y) &\propto \lambda^{\alpha-1} \exp\{-\beta\lambda\} \cdot \lambda^{\sum_{i=1}^n y_i} \exp\{-n\lambda\} \\ &= \lambda^{\alpha+\sum_{i=1}^n y_i - 1} \exp\{-(\beta+n)\lambda\} \\ &\propto \text{Ga}(\alpha + \sum_{i=1}^n y_i, \beta + n) \end{aligned}$$

which is a $\text{Ga}(\alpha_n, \beta_n)$ distribution just like the prior, with updated hyperparameters:

$$\begin{aligned} \alpha_n &= \alpha + \sum_{i=1}^n y_i \\ \beta_n &= \beta + n. \end{aligned}$$

Total: 80 MARKS

END OF QUESTION PAPER.