XXXX 2019
xx.xx – x.xx

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

Bayesian Statistics (Level M) Solutions

"Hand calculators with simple basic functions (log, exp, square root, etc.) may be used in examinations. No calculator which can store or display text or graphics may be used, and any student found using such will be reported to the Clerk of Senate".

NOTE: Candidates should attempt ALL questions.

CONTINUED OVERLEAF/

1. Suppose you are considering investing in a small start-up company that sells umbrellas. Let's consider modelling weekly sales of umbrellas. Define Y as the number of umbrellas sold per week, and assume that $Y \sim \text{Poisson}(\lambda)$. From previous experience, you can assume $\lambda \sim \text{Gamma}(20, 2)$. Inspecting the annual sales report, you find that company sold 898 umbrellas in the last year (52 weeks).

   (a) Calculate the posterior expected number of weekly sales for the coming year.
   [4 MARKS]

Solution

We have the following model:

$$\begin{aligned} Y &\sim \text{Poisson}(\lambda) \\ \lambda &\sim \text{Gamma}(20, 2) \end{aligned}$$

where Y is defined as the number of umbrellas sold per week.
This is a conjugate prior and so $\lambda|Y$ follows a Gamma distribution with the hyperparameters updated as:

$$\alpha_n = \alpha_0 + \sum_{i=1}^{n} y_i = 20 + 898 = 918 \quad \text{and} \quad \beta_n = \beta + n = 2 + 52 = 54$$

So, the parameter posterior is

$$\lambda|Y \sim \text{Gamma}(918, 54)$$

The expected value for a future observation can be derived from the posterior predictive distribution for **weekly** sales:

$$\begin{aligned} p(\tilde{Y}|Y) &= \int p(\tilde{Y}|\lambda)p(\lambda|Y)d\lambda = \frac{54^{918}}{\Gamma(918)} \int_0^\infty \frac{\lambda^{\tilde{Y}} e^{-\lambda}}{\tilde{Y}!} \lambda^{917} e^{-54\lambda} d\lambda \\ &= \frac{54^{918}}{\Gamma(918)\tilde{Y}!} \int_0^\infty \lambda^{917+\tilde{Y}} e^{-55\lambda} d\lambda. = \frac{54^{918}}{\Gamma(918)\tilde{Y}!} \times \frac{\Gamma(918+\tilde{Y})}{55^{918+\tilde{Y}}} \\ &= \frac{(917+\tilde{Y})!}{\tilde{Y}!917!} \left(\frac{54}{55}\right)^{918} \left(\frac{1}{55}\right)^{\tilde{Y}} = \binom{\tilde{Y}+917}{\tilde{Y}} \left(\frac{54}{55}\right)^{918} \left(\frac{1}{55}\right)^{\tilde{Y}} \\ &= \text{NegBin}\left(\tilde{Y} \,\middle|\, 918, \frac{1}{55}\right) \end{aligned}$$

And the mean of this distribution is therefore:

$$\mathbb{E}\left[\tilde{Y}|Y\right] = \mathbb{E}\left[\text{NegBin}\left(\tilde{Y} \,\middle|\, r = 918, p = \frac{1}{55}\right)\right] = \frac{pr}{1-p} = \frac{\frac{918}{55}}{\frac{54}{55}} = \frac{918}{54} = 17.$$

Reasonably high marks are assigned for this question as students are given little hints of how to solve this problem. It requires understanding of how inference and prediction are performed in the Bayesian context without direct instructions of what to do.

CONTINUED OVERLEAF/

(b) Your confidence in investing in this company on the scale from 0 to 5 is defined as the following:

0, if the average weekly sales for the coming year are expected to be within $[5, 10)$,
1, if they are within $[10, 15)$,
2, if they are within $[15, 20)$,
3, if they are within $[20, 25)$,
4, if they are within $[25, 30)$,
5, if they are greater than 30.

Considering the solution obtained in part (a), how confident are you with making an investment on a scale from 0 to 5, a posteriori? **[1 MARK]**

Solution

This question is mostly to hint the students that the answer to the previous question must be a number.

The answer is obviously 2.

(c) Suppose you want to ignore any previous knowledge for the inference in this situation, and therefore you decide to use Jeffreys' prior for λ. Explain what Jeffreys' prior is, and derive what form it will take for this problem. **[10 MARKS]**

Solution

Jeffreys' prior is an uninformative prior for model parameters that has the key feature that its functional dependance on the likelihood L is invariant under reparametrisation of the model. It is defined as the square root of the determinant of the Fisher information matrix.

For this problem we need to remember that we are working with n observations of Y, and use the likelihood of the following form:

$$\begin{aligned} L(\lambda) &= p(y_1, \ldots, y_n|\lambda) = \prod_{i=1}^{n} p(y_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \\ &= e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i} \prod_{i=1}^{n} \frac{1}{y_i!} \end{aligned}$$

If the student considers only one observation of Y, take away 2 marks.

Jeffreys' prior $p(\lambda) = \sqrt{J(\lambda)}$, where $J(\lambda)$ is the Fisher information.

$$J(\lambda) = -\mathbb{E}\left[\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} \middle| \lambda \right]$$

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^{n} y_i \ln \lambda - \sum_{i=1}^{n} \ln y_i!$$

CONTINUED OVERLEAF/

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \sum_{i=1}^{n} \frac{y_i}{\lambda}$$

$$\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -\sum_{i=1}^{n} \frac{y_i}{\lambda^2}$$

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

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

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

(d) Is this Jeffreys' prior proper? Are we allowed to use this Jeffreys' prior for inference in this problem? If we are not, explain why. If we are, evaluate the expected weekly sales of umbrellas for the coming year using Jeffreys' prior. **[5 MARKS]**

Solution

Jeffreys' prior derived in the previous part is not proper, as

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

$$(\infty - 0) = \infty$$

i.e. the normalisation constant does not exist. Therefore the prior is improper.

To be able to use Jeffreys' prior, we need to demonstrate the the posterior is a proper distribution. The Jeffreys' prior is equivalent to $\text{Gamma}(1/2, 0)$, as

$$\text{Gamma}\,(1/2, 0) \propto \lambda^{1/2-1} e^{-0\lambda}$$

and therefore the posterior is

$$\text{Gamma}\left(1/2 + \sum_{i=1}^{n} y_i, n\right)$$

and as long as $n \geq 1$, it is a proper (normalised) probability distribution. So, the posterior is proper, and we can use this prior.

The posterior for this question is:

$$\lambda | Y \sim \text{Gamma}(898.5, 52)$$

CONTINUED OVERLEAF/

and similarly to the solution in the first part

$$\tilde{Y}|Y \sim \text{NegBin}\left(\tilde{Y}\,\middle|\,898.5, \frac{1}{53}\right)$$

$$\mathbb{E}\left[\tilde{Y}|Y\right] = \frac{\frac{898.5}{53}}{\frac{52}{53}} = \frac{898.5}{52} \approx 17.279$$

2. Consider the hierarchical model

$$\begin{aligned} y_i|\lambda_i &\sim \text{Exp}(\lambda_i), \qquad i = 1, \ldots, n \text{ independently;} \\ \lambda_i|\beta &\sim \text{Gamma}(2, \beta), \qquad i = 1, \ldots, n \text{ independently;} \\ \beta &\sim \text{Gamma}(2, 2), \end{aligned}$$

(a) Derive the joint probability density function of all the random quantities in the model $p(\beta, \boldsymbol{\lambda}, \boldsymbol{y})$. **[4 MARKS]**

Solution

$$\begin{aligned} p(\beta, \boldsymbol{\lambda}, \boldsymbol{y}) &= p(\boldsymbol{y}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\beta)p(\beta) \\ &= \left[\prod_{i=1}^{n} \lambda_i \exp\{-\lambda_i y_i\}\right] \times \left[\prod_{i=1}^{n} \beta^2 \lambda_i \exp\{-\beta\lambda_i\}\right] \times 4\beta \exp\{-2\beta\} \end{aligned}$$

(b) Find the full conditional distributions of $\boldsymbol{\lambda}$ and β. These should be proper distributions with known methods to draw samples from them. **[5 MARKS]**

Solution

Dropping irrelevant parts of the joint distribution, we obtain:

$$\begin{aligned} p(\lambda_i|\beta, \boldsymbol{\lambda}_{-i}, \boldsymbol{y}) &\propto \lambda_i \exp\{-\lambda_i y_i\} \times \lambda_i \exp\{-\beta\lambda_i\} \\ &\propto \lambda_i^2 \exp\{-(\beta + y_i)\lambda_i\} \\ &\propto \text{Gamma}(3, \beta + y_i) \end{aligned}$$

CONTINUED OVERLEAF/

$$
\begin{aligned}
p(\beta|\boldsymbol{\lambda}, \boldsymbol{y}) &\propto \beta \exp\{-2\beta\} \times \prod_{i=1}^{n} \beta^2 \exp\{-\lambda_i \beta\} \\
&\propto \beta^{2n+1} \exp\left\{-\left(2 + \sum_{i=1}^{n} \lambda_i\right)\beta\right\} \\
&\propto \text{Gamma}\left(2n+2, 2 + \sum_{i=1}^{n} \lambda_i\right)
\end{aligned}
$$

(c) Explain how the full conditional distributions could be used to implement a Gibbs sampler to draw from $p(\boldsymbol{\lambda}, \beta|\boldsymbol{y})$. **[4 MARKS]**

Solution

Initialise $\lambda = \lambda^{(0)}$ and $\beta = \beta^{(0)}$

for $k = 1, \ldots, K$

  Draw $\lambda_1^{(k)}$ from Gamma$(3, \beta^{(k-1)} + y_1)$,

  Draw $\lambda_2^{(k)}$ from Gamma$(3, \beta^{(k-1)} + y_2)$,

  $\vdots$

  Draw $\lambda_n^{(k)}$ from Gamma$(3, \beta^{(k-1)} + y_n)$,

  Draw $\beta^{(k)}$ from Gamma$\left(2n+2, 2 + \sum_{i=1}^{n} \lambda_i^{(k)}\right)$ *(note $\lambda_i^{(k)}$)*

Discard burn-in and optionally thin-out.

(d) Explain why one usually discards the initial draws from a Gibbs sampler. How can running multiple chains help one to decide how much to discard? **[4 MARKS]**

Solution

The Gibbs sampler builds a Markov chain. Markov chains take time to reach equilibrium, i.e. settle to their limiting distribution. So, one needs to eliminate initial draws which will not yet be coming from the target distribution. These initial samples are called the 'burn-in'.

Multiple chains started from different initial parameter values will all converge to the target distribution along their own trajectories. A time plot of all the chains will indicate when all the chains come together, and samples prior to that can be discarded.

(e) Suppose that a sample $(\boldsymbol{\lambda}^{(t)}, \beta^{(t)})$, $t = 1, \ldots, T$, from the joint posterior distribution of $\boldsymbol{\lambda}$ and β is available. Explain how you can use it to compute an estimate of the posterior predictive distribution of a future observation $\tilde{y}_{n+1}$ for a new experiment $(n+1)$. **[3 MARKS]**

CONTINUED OVERLEAF/

Solution

For each sample from the joint posterior distribution of $\boldsymbol{\lambda}$ and β, take a draw of

$$\lambda_{n+1} \sim \text{Gamma}(2, \beta)$$

using `rgamma` in R, and then we draw

$$\tilde{y}_{n+1} \sim \text{Exp}(\lambda_{n+1}).$$

The set $\left\{\tilde{y}_{n+1}^{(t)} : t = 1, \dots, T\right\}$ are samples from the posterior predictive distribution of $\tilde{y}_{n+1}$.

Note that we never had any samples for λ_{n+1} in the posterior, so the first step is essential too.

3. In a photophysics experiment we are observing photons arriving to a photon counter from a very weak light source. The waiting time, y, in nanoseconds (ns) until the arrival of the next photon is exponentially distributed with unknown rate λ. Assume a Gamma(α, β) prior on λ with $\alpha \geq 1$. Note that the mode of the Gamma(α, β) distribution is $(\alpha - 1)/\beta$, its mean is α/β, and its variance is α/β^2.

   (a) In our first experiment, we observe that there were no photons registered within 10ns, which means the observed arrival time is greater than 10ns. Derive the posterior distribution for λ assuming a general Gamma(α, β) prior. The posterior distribution should be a function of α and β. Write down the maximum *a posteriori* (MAP) estimate of the rate λ. **[6 MARKS]**

Solution

The model for this problem is

$$\begin{aligned} y|\lambda &\sim \text{Exp}(\lambda) \\ \lambda &\sim \text{Gamma}(\alpha, \beta) \end{aligned}$$

With corresponding densities:

$$\begin{aligned} p(y|\lambda) &= \lambda e^{-\lambda y} \\ p(\lambda) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \end{aligned}$$

CONTINUED OVERLEAF/

We know that $y > 10$, and therefore the corresponding posterior distribution should be:

$$\begin{aligned}
p(\lambda|y > 10) &\propto \int_{10}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \lambda e^{-\lambda y} dy \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^\alpha e^{-\beta\lambda} \int_{10}^{\infty} e^{-\lambda y} dy \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^\alpha e^{-\beta\lambda} \left[-\frac{1}{\lambda} e^{-\lambda y} \right]_{y=10}^{\infty} \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^\alpha e^{-\beta\lambda} \left[0 + \frac{1}{\lambda} e^{-10\lambda} \right]
\end{aligned}$$

$$\begin{aligned}
p(\lambda|y > 10) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} e^{-10\lambda} \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-(\beta+10)\lambda} \\
&\propto \text{Gamma}(\alpha, \beta + 10)
\end{aligned}$$

The MAP estimate of λ is the mode of this posterior distribution, and can be calculated as

$$\text{mode}(\lambda|y > 10) = \frac{\alpha - 1}{\beta + 10}$$

(b) You decide to run this experiment once again, using the information you learned in the previous experiment (described in the previous part of this question) as the prior for the new experiment. This time you observe the photon arriving at exactly 10ns. Derive the posterior of λ using this new observation. Write down the maximum *a posteriori* (MAP) estimate of λ using the information from both of your experiments. **[4 MARKS]**

Solution

The model used in this case is

$$\begin{aligned}
y|\lambda &\sim \text{Exp}(\lambda) \\
\lambda &\sim \text{Gamma}(\alpha, \beta + 10)
\end{aligned}$$

In this case we have a direct observation $y_2 = 10$, and using conjugate update, we arrive to the posterior for λ:

$$\lambda|y, y_2 \sim \text{Gamma}(\alpha + 1, \beta + 20)$$

CONTINUED OVERLEAF/

The MAP estimate of λ becomes

$$\text{mode}(\lambda|y > 10, y_2 = 10) = \frac{\alpha}{\beta + 20}$$

This question targets understanding of the concept of iterative learning from data in the Bayesian framework, where we use the posteriors from previous experiments as priors for the new experiments.

(c) Having obtained the posterior for the photon arrival rate using two experiments described in previous parts, and a general Gamma(α, β) prior, derive the probability distribution for the photon waiting time that you expect to observe in a new, third, experiment. **[6 MARKS]**

Solution

The posterior predictive distribution for $\tilde{y}$ given the results of the previous question is

$$\begin{aligned}
p(\tilde{y}|y, y_2) &= \int_0^\infty p(\tilde{y}|\lambda)p(\lambda|y, y_2)d\lambda \\
&= \int_0^\infty \lambda e^{-\lambda\tilde{y}} \frac{(\beta+20)^{(\alpha+1)}}{\Gamma(\alpha+1)} \lambda^\alpha e^{-(\beta+20)\lambda} d\lambda \\
&= \frac{(\beta+20)^{(\alpha+1)}}{\Gamma(\alpha+1)} \int_0^\infty \lambda^{(\alpha+1)} e^{-(\beta+20+\tilde{y})\lambda} d\lambda \\
&= \frac{(\beta+20)^{(\alpha+1)}}{\Gamma(\alpha+1)} \times \frac{\Gamma(\alpha+2)}{(\beta+20+\tilde{y})^{\alpha+2}} = \frac{(\beta+20)^{(\alpha+1)} \cdot (\alpha+1)}{(\beta+20+\tilde{y})^{\alpha+2}}
\end{aligned}$$

which is a Lomax$(\tilde{y}|\alpha+1, \beta+20)$ distribution, but the students are not expected to know the name of this distribution.

note: The posterior predictive distribution for the exponential model wasn't demonstrated in lectures, thus high marks in this part.

(d) To select the values of prior hyperparameters α and β, you decide to use the Empirical Bayes approach. You want to express expert knowledge that the waiting time should on average be 5ns. You want your prior to express uncertainty about this expert knowledge, and therefore decide use a prior for λ with variance of 1. Describe what the Empirical Bayes approach is, and use it to select an appropriate informative prior. **[4 MARKS]**

Solution

The Empirical Bayes approach is used to define informative priors by matching the moments of existing knowledge to the moments of the chosen prior distribution.

CONTINUED OVERLEAF/

In this question, we need to solve a system of equations:

$$\begin{cases} \dfrac{\alpha}{\beta} = \dfrac{1}{5} \\ \dfrac{\alpha}{\beta^2} = 1 \end{cases}$$

$$\begin{cases} \beta = 5\alpha \\ \dfrac{\alpha}{25\alpha^2} = 1 \end{cases}$$

$$\begin{cases} \beta = 5\alpha \\ \dfrac{1}{25\alpha} = 1 \end{cases}$$

$$\begin{cases} \alpha = \dfrac{1}{25} \\ \beta = \dfrac{1}{5} \end{cases}$$

So, our informative prior should be Gamma$\left(\frac{1}{25}, \frac{1}{5}\right)$.
If the student assumes $\alpha/\beta = 5$, and does the rest correctly, mark this part with 2 marks.

---

4. Consider a loss function for estimating a parameter θ with the value ('action') a with the following form:

$$L(\theta, a) = \begin{cases} \beta(\theta - a), & \theta - a \geq 0; \\ \gamma(a - \theta), & \theta - a < 0, \end{cases}$$

where β and γ are two positive constants.

(a) Suppose that the posterior distribution of θ is Un$[0, 1]$. Show that the Bayes expected loss is

$$\rho(\pi, a) = \begin{cases} \dfrac{\beta(1 - 2a)}{2}, & a \leq 0 \\ \dfrac{\gamma + \beta}{2}a^2 - \beta a + \dfrac{\beta}{2}, & 0 < a \leq 1 \\ \dfrac{\gamma(2a - 1)}{2}, & a > 1 \end{cases}$$

[6 MARKS]

Solution

CONTINUED OVERLEAF/

$$\rho(\pi, a) = \mathbb{E}\left[L(\theta, a)|y\right] = \int_{-\infty}^{\infty} L(\theta, a)p(\theta|y)d\theta$$
$$= \int_0^1 \left[\beta(\theta - a)I(\theta \geq a) + \gamma(a - \theta)I(\theta < a)\right] d\theta$$

When $a \leq 0$,

$$\rho(\pi, a) = \int_0^1 \beta(\theta - a)d\theta = \frac{\beta}{2}\left[(\theta - a)^2\right]_{\theta=0}^1 =$$
$$= \frac{\beta}{2}\left((1-a)^2 - a^2\right) = \frac{\beta(1-2a)}{2}$$

When $a > 1$,

$$\rho(\pi, a) = \int_0^1 \gamma(a - \theta)d\theta = \frac{\gamma}{2}\left[-(a - \theta)^2\right]_{\theta=0}^1 =$$
$$= \frac{\gamma}{2}\left(a^2 - (a-1)^2\right) = \frac{\gamma(2a-1)}{2}$$

When $0 < a \leq 1$,

$$\rho(\pi, a) = \int_0^a \gamma(a - \theta)d\theta + \int_a^1 \beta(\theta - a)d\theta =$$
$$= \frac{\gamma}{2}\left[-(a - \theta)^2\right]_{\theta=0}^a + \frac{\beta}{2}\left[(\theta - a)^2\right]_{\theta=a}^1 =$$
$$= \frac{\gamma}{2}\left(-(a-a)^2 + (a-0)^2\right) + \frac{\beta}{2}\left((1-a)^2 - (a-a)^2\right) =$$
$$= \frac{\gamma a^2}{2} + \frac{\beta(1-a)^2}{2} = \frac{\gamma}{2}a^2 + \frac{\beta}{2} - \beta a + \frac{\beta}{2}a^2 = \frac{\gamma + \beta}{2}a^2 - \beta a + \frac{\beta}{2}$$

(b) Show that the Bayes action is $a^\pi = \frac{\beta}{\beta+\gamma}$ and sketch the Bayes expected loss as a function of a. **[6 MARKS]**

Solution

$$\frac{\partial \rho}{\partial a} = \begin{cases} -\beta, & a \leq 0 \\ \beta a + \gamma a - \beta, & 0 < a \leq 1 \\ \gamma, & a > 1 \end{cases}$$

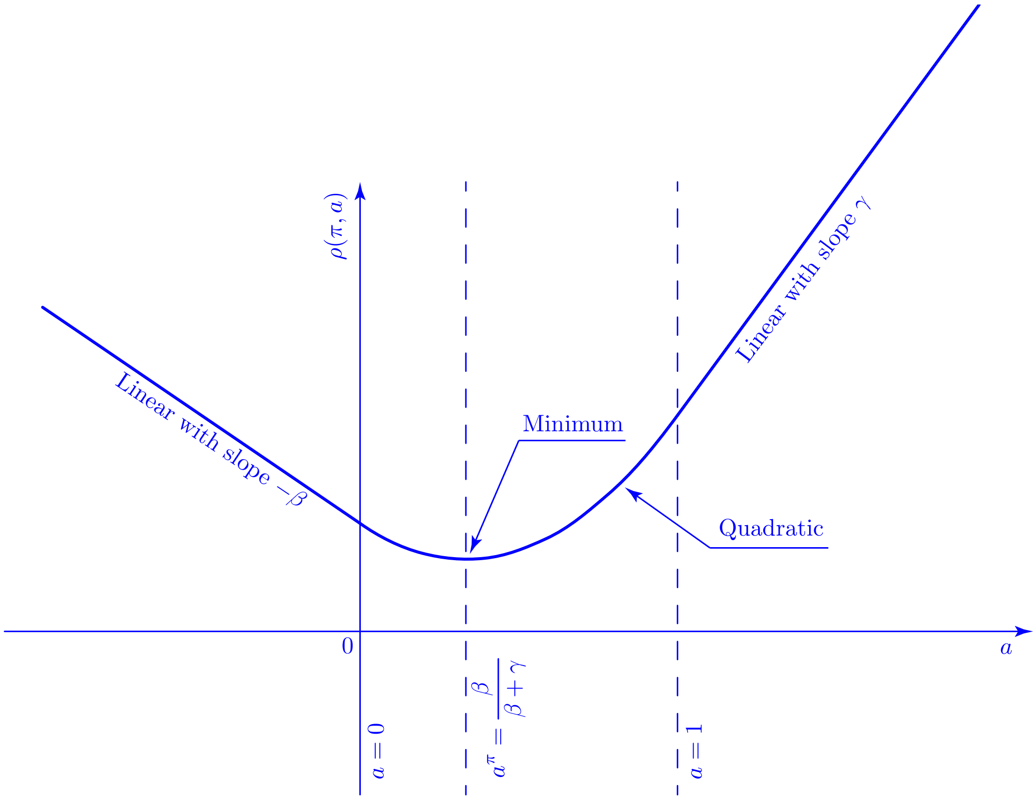
As both β and γ are positive, this means they can't be 0. ρ is continuous as is its first derivative. The only option for the minimum is when $0 < a \leq 1$:

$$\beta a^\pi + \gamma a^\pi - \beta = 0$$
$$(\beta + \gamma)a^\pi = \beta$$
$$a^\pi = \frac{\beta}{\beta + \gamma}$$

CONTINUED OVERLEAF/

The second derivative at a^π is $\frac{\partial^2 \rho}{\partial a^2} = \beta + \gamma > 0$, so this is a global minimum of the Bayes expected loss.

The sketch should look like the following:



(c) The following loss function:

$$L(\theta, a) = \frac{(\theta - a)^2}{\theta(1-\theta)}$$

if frequently used for deciding on the probability of success in binomial experiments.

i. Prove that the Bayes action for this loss function is:

$$a^\pi = \frac{\mathbb{E}\left[\frac{1}{1-\theta}|y\right]}{\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right]}.$$

[4 MARKS]

Solution

CONTINUED OVERLEAF/

$$\rho(\pi, a) = \int_0^1 \frac{(\theta - a)^2}{\theta(1-\theta)} p(\theta|y) d\theta = \mathbb{E}\left[\frac{(\theta-a)^2}{\theta(1-\theta)}|y\right]$$
$$= \mathbb{E}\left[\frac{\theta}{1-\theta}|y\right] - 2a\mathbb{E}\left[\frac{1}{(1-\theta)}|y\right] + a^2\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right]$$
$$\frac{\partial \rho}{\partial a} = -2\mathbb{E}\left[\frac{1}{1-\theta}|y\right] + 2a\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right]$$

To find the Bayes action we minimise $\rho(\pi, a)$ by setting $\frac{\partial \rho}{\partial a} = 0$ and solving for a:

$$-2\mathbb{E}\left[\frac{1}{1-\theta}|y\right] + 2a\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right] = 0$$
$$a^\pi = \frac{\mathbb{E}\left[\frac{1}{1-\theta}|y\right]}{\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right]}.$$

which is a minimum, since

$$\frac{\partial^2 \rho}{\partial a^2} = 2\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right] > 0, \text{when } 0 \leq \theta \leq 1.$$

---

ii. Considering the model

$$y \sim \text{Bin}(n, \theta)$$
$$\theta \sim \text{Beta}(0.5, 0.5)$$

Evaluate the Bayes action for this loss function. **[4 MARKS]**

Solution

$$a^\pi = \frac{\mathbb{E}\left[\frac{1}{1-\theta}|y\right]}{\mathbb{E}\left[\frac{1}{\theta(1-\theta)}|y\right]},$$

and the posterior for this conjugate setup is going to be

$$\theta|y = \text{Beta}(0.5 + y, 0.5 + n - y).$$

Therefore

$$a^\pi = \frac{\frac{\Gamma(n+1)}{\Gamma(0.5+y)\Gamma(0.5+n-y)} \int_0^1 \theta^{0.5+y-1}(1-\theta)^{0.5+n-y-2} d\theta}{\frac{\Gamma(n+1)}{\Gamma(0.5+y)\Gamma(0.5+n-y)} \int_0^1 \theta^{0.5+y-2}(1-\theta)^{0.5+n-y-2} d\theta}$$
$$= \frac{B(0.5 + y, 0.5 + n - y - 1)}{B(0.5 + y - 1, 0.5 + n - y - 1)}$$
$$= \frac{\frac{\Gamma(0.5+y)\Gamma(0.5+n-y-1)}{\Gamma(n)}}{\frac{\Gamma(0.5+y-1)\Gamma(0.5+n-y-1)}{\Gamma(n-1)}} = \frac{\Gamma(0.5 + y)\Gamma(n-1)}{\Gamma(0.5 + y - 1)\Gamma(n)}$$
$$= \frac{y - 0.5}{n - 1}$$

CONTINUED OVERLEAF/

Total: 80

END OF QUESTION PAPER.