1 Introduction

Exercises and outline solutions

1. Y has a pack of 4 cards (Ace and Queen of clubs, Ace and Queen of Hearts) from which he deals a random of selection 2 to player X. What is the probability that X receives both Aces conditional on them receiving at least 1 Ace. Suppose now that Y deals X two cards from the pack of 4, after which X says "*I have an Ace*".

   (a) Discuss whether the above information is sufficient to calculate the conditional probability P(X has 2 Aces | X says "I have an Ace").

   (b) If it is not, what other information would be required in order to calculate this conditional probability?

   Solution. For the first part we note that there are 6 possible selections of 2 cards from 4. Five of these selections contain at least one ace. Therefore P(B) = 5/6, where B is the event 'at least one ace'. If A is the event '2 aces' then P(A) = 1/6. It follows that $P(A|B) = 1/5$ from the definition of conditional probability.

   You can't calculate a probability conditional on X saying 'I have an ace' since we have not specified the probability of this as an event. You would need to specify a probabilistic model for what X says when presented with a given hand.

2. An urn is known to contain n differently coloured balls where n can be any integer in the set 1, 2, 3. Your prior information tells you that n is equally likely to be any of these values. A ball is drawn randomly from the urn - it is red. Alice argues that since the probability of the red ball being drawn conditional on there being n balls in the urn is $1/n$, then

$$P(n = 1|red\ ball\ drawn) = \frac{\frac{1}{3} \times \frac{1}{1}}{\frac{1}{3} \times \frac{1}{1} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{3}}$$

   and calculates the posterior probabilities of n being 1, 2, and 3 as 6/11, 3/11 and 2/11 respectively. She then expresses her surprise that the her beliefs regarding N have changed having observed only the colour of a single draw from the urn.

(a) Explain the fallacy in her argument and why the above information alone does not define a posterior probability for n.

Solution. Alice needs to be able to calculate the probability of the event 'n balls in the urn and a red ball drawn' in order to define the conditional probability. This is equal to

$$P(n) \times P(\textit{red ball in the urn}|n) \times \frac{1}{n}$$

She has omitted the middle factor (which has not been defined) in her argument.

(b) Bertie assumes that the n balls placed in the urn are drawn uniformly at random from a large stock of differently coloured balls. Calculate Bertie's posterior probabilities for $n = 1, 2, 3$.

Solution In this case we would have that

$$P(\textit{red ball in the urn}|n) \propto n.$$

This implies that $P(\textit{n and red}) \propto P(n)$ and we see that the posterior probability for n must be the same as the prior (1/3 for each value).

(c) Under what circumstances would Alice's posterior probabilities be correct?

Solution. This would be true if $P(\textit{red ball in the urn}|n)$ was independent of n, e.g. if a red ball was always put in the urn first.

3. Suppose now in the situation of question 2, two balls are drawn from the urn *with replacement* and the event that both are the same colour is observed. Calculate the posterior probabilities for $n = 1, 2, 3$ in this case.

   Solution. Now $P(n|\textit{both balls same}) \propto P(n) \times P(\textit{both balls same}|n) = 1/n, n = 1, 2, 3$. Hence, posterior probabilities of $n = 1, 2, 3$ are 6/11, 3/11 and 2/11 respectively.

4. (Yet more balls and urns) Five balls are drawn uniformly randomly from a very large population of black and white balls where the proportion of black balls is 1/3. You do not know the colours of the balls selected.

(a) Give suitable prior probabilities for the number of black balls in the urn.

Solution The natural choice would be $Bin(5, 1/3)$.

(b) You now select two balls uniformly at random from the urn with replacement. They are both white. Calculate the posterior probabilities for the number of black balls in the urn.

Solution. Since they are drawn with replacement, then the probability that they are both white given the number of black balls n, is $(\frac{5-n}{5})^2$. We can write

$$\pi(n|both\ white) \propto \binom{5}{n}(1/3)^n(2/3)^{5-n}(\frac{5-n}{5})^2$$

Note that this is zero for $n = 5$. Calculate and normalise to get the probabilities.

(c) Suppose that the two balls were selected from the urn without replacement and were both white. Calculate your posterior probabilities for the number of black balls in the urn for this case.

Solution. Same approach as above, but now the probability that both balls are white given n, is $\frac{5-n}{5} \times \frac{5-n-1}{4}$ since sampling is without replacement. Note that this is zero for $n = 5$ and $n = 4$.

5. A fair coin is tossed n times where n can take the values 1, 2, ..., 5 with equal probability. Suppose that 2 heads result from the n tosses.

   Determine the posterior distribution (i.e. work out the probability function) of n and identify the value of n that is *a posteriori* most likely.

   Solution. Clearly the posterior probability is non-zero for values of n greater than 1. For $n \geq 2$ we obtain

$$\pi(n|2\ heads) \propto \frac{1}{5} \times \binom{n}{2}(1/2)^5.$$

   Thus for n = 2, 3, 4, 5, the posterior probabilities are in the ratio 1/4, 3/8, 6/16, 10/32, respectively. Actually from this we see that *both* 3 and 4 are *a posteriori* equally likely.

Suppose now the coin is to be tossed repeatedly until m tails are obtained where the value m is first selected from a *Geometric*(1/3) distribution. Suppose that 2 heads are obtained in the sequence. What is the posterior distribution of m given this information? (It is sufficient to write an expression involving infinite sums!)

Solution. Given m, the number of heads, X, obtained in achieving m tails has a negative binomial distribution with probability function

$$P(x|m) = \binom{x+m-1}{x} \frac{1}{2^{x+m}}.$$

It follows that for $m \geq 1$ the posterior probability of m satisfies

$$\pi(m|X=2) \propto \pi(m)P(2|m) = \frac{2^{m-1}}{3^m} \binom{m+1}{2} \frac{1}{2^{2+m}} \propto \frac{m(m+1)}{3^m}.$$

To obtain these probabilities numerically, you need to sum this quantity over all values of m, to get the normalising constant. This can be done analytically. You need to spot that $\sum_{m=1}^{\infty} \frac{m(m+1)}{3^m}$ is the 2nd derivative (w.r.t. x) of $\sum_{m=1}^{\infty} \frac{x^{m+1}}{3^m}$ evaluated at $x = 1$. Now the latter series is geometric and its sum is defined in a neighbourhood of $x = 1$.

However, you can quickly get a feel for the relative posterior probabilities by evaluating the right hand side of the above. (According to my calculations) the posterior probabilities are for m = 1, 2, 3, 4, 5 are in the ratio 2/3, 2/3, 4/9, 20/81, 10/81, ...

2 Bayesian Inference

Exercises

1. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from the Poisson distribution with mean μ. Show that the conjugate prior is Gamma. Using $\mu \sim Gamma(\alpha, \beta)$ as the prior, determine the posterior distribution of μ.

   Solution You need to show that if you start with a Gamma prior, you end up with a Gamma posterior. Suppose that the prior is $Gamma(\alpha, \beta)$. Then given observations $x_1, x_2, ..., x_n$ the likelihood is

   $$L(\mu) = e^{-n\mu}\mu^{\sum x_i} / \prod x_i!.$$

   Now it follows that the posterior density satisfies

   $$\pi(\mu|\mathbf{x}) \propto e^{-(n+\beta)\mu}\mu^{\alpha+\sum x_i - 1}$$

   for $\mu > 0$. This marks it out as a $Gamma(\alpha + \sum x_i, \beta + n)$ density.

   Suppose you specify $\mu \sim Gamma(1, 0.5)$, $n = 5$ and observe $\sum x_i = 15.0$. Calculate an equal-tailed 95% credible region for μ in this case.

   Solution. Now the posterior for μ is $\Gamma(16, 5.5)$. It follows that $11\mu \sim \chi^2_{32}$. We obtain the limits for our credible region by reading off the 97.5%- and 2.5%- critical values of the χ^2_{32} - 18.29 and 49.48, respectively - and then divide by 11 to get the interval (1.66, 4.50).

2. Let $\mathbf{X}$ be a random sample from the Exponential distribution $Exp(\lambda)$ with mean $1/\lambda$, i.e. $\Gamma(1, \lambda)$. Show that the conjugate prior is Gamma. In particular, if X is a single observation, show that the prior $\Gamma(\alpha, \beta)$ leads to a posterior density for λ being $\Gamma(\alpha + 1, \beta + X)$.

   Solution. The likelihood is $L(\lambda) = \lambda e^{-\lambda X}$. From this we see that the posterior is

   $$\pi(\lambda|X) \propto \lambda^{\alpha} e^{-\lambda(\beta+X)}$$

   and is therefore $Gamma(\alpha + 1, X)$.

   An important consequence of the conjugacy property is that if observations arrive sequentially then updating the posterior distribution is simple. Suppose that the prior distribution is $Gamma(\alpha, \beta)$ and that

x_1 is observed. Obtain the posterior distribution. Now suppose that x_2 is observed. Find the new posterior distribution by updating the existing posterior. Finally, show that this posterior distribution is the same as that obtained from the original prior if we observe a random sample of size 2 consisting of (x_1, x_2).

Solution. For a random sample (x_1, x_2) and prior $Gamma(\alpha, \beta)$ we get a posterior that is $Gamma(\alpha + 2, \beta + x_1 + x_2)$ and the result follows.

3. The lifetime of a component, T, (measured in days) follows an $Exp(\lambda)$ distribution where *a priori* $\lambda \sim \Gamma(1, 2)$. You select a random sample of 5 components for which $\sum t_i = 3.0/days$. Find the posterior distribution of λ.

   Solution. From question 2, we know that the posterior distribution of λ is $\Gamma(6, 5)$.

   A component of this kind forms part of certain system which is required to function continuously for a period of 6 hrs. What is the probability that the component fails before the end of 6 hrs? (You will have to work out the posterior predictive distribution of the lifetime of a component.)

   Solution. Let Z denote the lifetime of the component. You have to work out the predictive density, $f(z)$ by integrating the $Exp(\lambda)$ density weighted by this posterior density. Thus:

$$f(z) = \frac{5^6}{5!} \int_0^\infty \lambda^5 e^{-5\lambda} \lambda e^{-\lambda z} d\lambda = \frac{5^6 6}{(5+z)^7}$$

   You want $P(Z > 1/4)$. Now $P(Z > z) = \frac{5^6}{(5+z)^6}$ giving a probability of around 0.75.

4. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a random sample from the Pareto distribution with p.d.f. $f(x) = \theta(1+x)^{-(\theta+1)}$, $0 < x < \infty$. Show that the Gamma distribution is the conjugate prior for this distribution by proving that if the prior distribution of θ is $Gamma(\alpha, \beta)$, then the posterior is $G(\alpha + n, \beta + t(\mathbf{x}))$, where $t(\mathbf{x}) = \sum_i \log(1 + x_i)$.

   Solution. As per usual we get from *prior* $\times$ *likelihood*

$$\pi(\theta|\mathbf{x}) \propto \theta^{\alpha-1} e^{-\beta\theta} \theta^n \prod (1 + x_i)^{-(\theta-1)}$$

   The result follows after some algebra (writing $(1 + x_i)$ as $e^{log(1+x_i)}$).

5. *[1997 Statistical Inference Exam, Q4]* In a raid on a coffee shop, Bayesian trading inspectors take a random sample of n packets of coffee, each of nominal weight 125 g. They model these data as independent values $X_1, \ldots, X_n$ from a Normal $\mathrm{N}(\mu, \sigma^2)$ distribution. They take σ^2 to be known, while for μ they assume a prior distribution of $\mathrm{N}(\mu_0, \sigma_0^2)$, where μ_0 and σ_0^2 are specified values.

   (a) Show that the inspectors' posterior distribution is also Normal, and find its mean and variance.

   Solution. 'Posterior $\propto$ Prior $\times$ Likelihood', so the posterior p.d.f. is

$$
\begin{aligned}
&\propto \exp(-\frac{1}{2}(\frac{\mu-\mu_0}{\sigma_0})^2)\exp(-\frac{1}{2}(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}})^2) \\
&= \exp(-\frac{1}{2}(\mu^2(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2})-\mu(\frac{2\mu_0}{\sigma_0^2}+\frac{2n\bar{x}}{\sigma^2})+\text{ constant})) \\
&\propto \exp(-\frac{1}{2}(\mu-(\mu_0 w+\bar{x}(1-w)))^2/\sigma_1^2) \\
&\propto \text{ p.d.f. of } N(\mu_1, \sigma_1^2)
\end{aligned}
$$

   so the posterior distribution is $N(\mu_1, \sigma_1^2)$, where $\sigma_1^2 = (1/\sigma_0^2 + n/\sigma^2)^{-1}$, and $\mu_1 = (\mu_0 w + \bar{x}(1-w))$, where $w = (1/\sigma_0^2)/(1/\sigma_0^2 + n/\sigma^2) = \sigma_1^2/\sigma_0^2$.

   Show that the mean of this distribution is a weighted average of the prior mean μ_0 and the sample mean $\bar{x}$.

   Here we have already tidied up the answer to indicate that μ_1 is a weighted average of μ_0 and $\bar{x}$, with weights w and $1-w$.

   (b) The data they obtain are (weights in grams):

   105.3, 113.3, 114.5, 121.2, 122.9, 123.7, 124.0, 124.6, 124.9, 124.9, 124.9, 125.1, 125.5, 125.9, 126.8, 127.7, 128.2, 128.3, 128.5, 130.2 ($\sum x_i = 2470.4$, $\sum x_i^2 = 305828.98$).

   The parameter values they assume are $\mu_0 = 126$, $\sigma_0^2 = 1$, $\sigma^2 = 4$.

   The inspectors can impose a fine if their 95% credible interval falls wholly below the claimed value of $\mu = 125$ g.

      i. Show that the inspectors' 95% credible interval for μ for these data does lie wholly below 125 g; they therefore impose a fine on the owners of the coffee shop.

Solution For these data we have $\bar{x} = 123.52$. The inspectors' value of w is $1/(1+20/4) = 1/6$, so their posterior distribution is $\mathrm{N}(\mu_1, \sigma_1^2)$, where $\mu_1 = (1/6)\mu_0 + (5/6)\bar{x} = 123.93$. Their value for σ_1^2 is $1/(1+20/4) = 1/6$, whence their posterior is $\mathrm{N}(123.93, 1/6)$, which gives a 95% credible interval of $\mu_1 \pm 1.96 * \sigma_1 = [123.1, 124.7]$. This is wholly below 125 g., so they are entitled to levy a fine.

ii. Sketch the data (a dotplot or similar), and calculate their sample median and sample variance.
Solution. Any rough plot of the data should indicate that normality is suspect, with the three lowest values giving a pronounced left tail to the data. The median is 124.9, much closer to the desired 125 g. than the mean. Finally, the sample variance is 36.1, much greater than the inspectors' assumed value of 4 (if we omit the 3 lowest values, this reduces to 5.3).

iii. Comment briefly as to whether the inspectors are justified in imposing a fine on the basis of this sample.
Solution. The coffee packets vary much more than the inspectors assumed, perhaps representable as a Normal distribution contaminated with occasional outliers. The substandard mean weight can be accounted for by these outliers (3 in this sample). While the inspectors' modelling is revealed as less than ideal for their job, they are probably correct in fining a shop with such poor quality (?sic - quantity might be the more appropriate word here!) control.

3 Bayesian Inference Continued

Exercises

1. Suppose that $x_1, x_2, ..., x_n$ is a random sample of observations from an $Exp(\lambda)$ distribution where λ is unknown.

   (a) Show that the Jeffreys' prior for λ in this case satisfies $\pi(\lambda) \propto \lambda^{-1}$.

   Solution First you need to calculate the Fisher information

   $$F(\lambda) = -E(\frac{\partial^2}{\partial \lambda^2} \log(L(\lambda; \underline{X}))$$

   Now,

   $$\log(L(\lambda; \underline{X})) = n \log(\lambda) - \lambda \sum X_i$$

   and its second derivative w.r.t. λ is $\frac{-n}{\lambda^2}$. It follows that the $F(\lambda) = \frac{n}{\lambda^2}$ and the Jeffreys prior satisfies

   $$\pi(\lambda) \propto F(\lambda)^{\frac{1}{2}} \propto \frac{1}{\lambda}.$$

   (b) For $n = 5$ and $\sum x_i = 10$ calculate a 95% equal-tailed credible interval for λ using the Jeffreys prior.

   Solution: A posteriori we have $\lambda \sim \Gamma(5, 10)$. Therefore $20\lambda \sim \chi^2_{10}$. We can read off the 97.5% and 2.5% values of 20λ from tables to be 3.247 and 20.48 respectively giving (0.16, 1.02) as the credible interval.

2. Show that if $x_1, x_2, ..., x_n$ is a random sample from a Poisson(λ) distribution, then the Jeffrey's prior for λ is given by $\pi(\lambda) \propto \lambda^{-1/2}$. Comment on this in the light of the connection between the Exp(λ) and the Poisson(λ) distribution.

   Solution. This time we have

   $$\log(L(\lambda; \underline{X})) = -n\lambda + \log \lambda \sum X_i - \log(\prod X_i!)$$

   and its second derivative w.r.t. λ is equal to $-\frac{\sum X_i}{\lambda^2}$. Taking the expectation we find that $F(\lambda) = \frac{n}{\lambda}$ (since $E(\sum X_i) = n\lambda$). This gives a Jeffreys prior proportional to $\lambda^{-1/2}$.

This is unsatisfactory from the following point of view. Consider a Poisson process with rate λ. One observer measures the intra-arrival times of n events. The other counts the number of arrivals in n disjoint time windows of length 1, to estimate λ. These are experiments on the same process and λ has the same interpretation in each experiment, yet the Jeffreys formulation leads to different priors for λ for the 2 experiments.

3. An educationalist is interested in the distribution of the number of exam attempts required by individuals to qualify in a certain profession. They believe that it follows a negative binomial distribution with p.m.f.

$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

for $x = r, r+1, r+2, ...$, where r is a positive integer and p a probability between 0 and 1. Suppose that they place a Uniform(0, 1) prior on p, an improper prior on r proportional to $1/r$ and assume that the parameters are a priori independent of each other. They select a random sample of 5 qualified individuals and count their exam attempts. These are in order: 4, 4, 6, 8, 12

(a) Construct the likelihood $L(r,p)$ for these data.

Solution. The likelihood looks like

$$\begin{aligned} L(r,p) &= \prod_{i=1}^{5} (\binom{x_i-1}{r-1} p^r (1-p)^{x_i-r}) \\ &\propto \frac{p^{5r}(1-p)^{5(\bar{x}-r)}}{(r-1)!^5 \prod_i (x_i-r)!}, \end{aligned}$$

ignoring terms not involving r or p.

(b) Show that the marginal posterior probability mass function of r satisfies

$$p(r|\mathbf{x}) \propto \frac{1}{r} \frac{(5r)!(5(\bar{x}-r))!}{((r-1)!)^5 \prod_i (x_i-r)!}$$

for $r = 1, 2, 3, 4$ and is zero for larger values of r.

Solution. Since the range of the distribution is the set $r, r+1, \ldots$ we see that the data cannot arise for values of r greater than 4. This implies that the likelihood is zero for such values of r and, hence, so is the posterior.

To get the the marginal posterior for $r = 1, 2, 3, 4$ we must integrate the joint posterior of r and p with respect to p. Now

$$p(r, p|\mathbf{x}) \propto \frac{1}{r} \frac{p^{5r}(1-p)^{5(\bar{x}-r)}}{(r-1)!^5 \prod_i (x_i - r)!}.$$

On integrating it with respect to p we obtain:

$$\begin{aligned} p(r|\mathbf{x}) &\propto \frac{1}{r} \frac{Beta(5r+1, 5(\bar{x}-r)+1)}{((r-1)!)^5 \prod_i (x_i - r)!} \\ &\propto \frac{1}{r} \frac{(5r)!(5(\bar{x}-r))!}{((r-1)!)^5 \prod_i (x_i - r)!}. \end{aligned}$$

(c) Calculate (using a computer) the r.h.s. of the above expression for $r = 1, 2, 3, 4$ and use the calculated values to make inference on the value of r.

Solution. Not got round to doing this yet!

4. Suppose that a random sample of size 8 from a normally distributed population of mean μ and variance ϕ results in the values

3.1, 3.3, 3.6, 4.2, 4.3, 4.8, 5.4, 5.7.

Assuming that you take independent priors, constant for μ and proportional to ϕ^{-1} for ϕ, calculate:

(a) the posterior probability that μ exceeds 5.0;

(b) the posterior probability that ϕ is less than 1.

Solution. From lecture notes, with this choice of prior, we know that a posteriori

$$\frac{\bar{x} - \mu}{s/\sqrt{8}} \sim t_7,$$

and that

$$7\frac{s^2}{\phi} \sim \chi^2_7,$$

where $\bar{x} = 4.3$ and $s^2 = 0.91$ are the sample mean and variance respectively.

It follows that a posteriori

$$P(\mu > 5.0) = P(t_7 < \frac{\bar{x} - 5.0}{s/\sqrt{8}}) = P(t_7 < -2.08) = 0.038.$$

Furthermore we have that

$$P(\phi < 1) = P(\chi^2_7 > 7s^2) = 1 - P(\chi^2_7 < 6.36) = 0.5.$$

4 Introduction to simulation techniques

Exercises

1. *Simulating from the Cauchy distribution.* The Cauchy distribution is defined by the density

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, -\infty < x < \infty$$

   (a) By first deriving the c.d.f. of X and its inverse function, describe how samples from a U(0, 1) random number generator could be transformed to give samples from the Cauchy distribution.

   Solution The c.d.f. is

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds\frac{1}{\pi}(tan^{-1}x + \frac{\pi}{2}).$$

   It follows that if $u = F_X(x)$ then $x = tan(\pi(u - 0.5))$. This gives the necessary transformation to a Cauchy random variable X from a U(0, 1) random variable.

   (b) By appealing to the circular symmetry of the standard bivariate normal distribution, show how samples from a Cauchy distribution could be generated from independent N(0, 1) samples.

   Solution. From circular symmetry it follows that if (X, Y) represent a sample from the bivariate normal, then $tan^{-1}\frac{X}{Y} \sim U(-\frac{\pi}{2}, \frac{\pi}{2}) \sim \pi(U - 0.5)$ where $U \sim U(0, 1)$.

   It follows from the first part that $\frac{X}{Y}$ must follow a Cauchy distribution.

   (*Comment.* The Cauchy distribution is the same as the t_1 distribution.)

2. *Simulating from the Beta distribution.*

   (a) Show how U(0, 1) random variates can be transformed by inversion of the c.d.f. to generate samples from a Beta(n, 1) distribution.

   Solution. The p.d.f. of the Beta(n, 1) distribution is

$$f_X(x) = nx^{n-1},$$

and the c.d.f. is therefore $F_X(x) = x^n$. It follows that if $U = F_X(X)$ then $X = U^{\frac{1}{n}}$. Therefore if $U \sim U(0,1)$ then $X = U^{\frac{1}{n}} \sim Beta(n,1)$.

(b) How would you generate samples from a Beta(2, 2) distribution by inversion of the c.d.f.?

Solution For the $Beta(2,2)$ distribution the pdf is $f_X(x) = 6x(1-x) = x - x^2, 0 < x < 1$ and the cdf is $3x^2 - 2x^3$. Now if $F_X(x) = u, 0 < u < 1$ then $3x^2 - 2x^3 - u = 0$. We need to find the root of this equation that lies between 0 and 1. It has 2 other real roots lying outside this region. Since cubics can be solved by radicals (see method of Cardano and Tartaglia) then the root $r \in (0,1)$ could be calculated.

(c) Given that if $X \sim Gamma(n,1)$ and $Y \sim Gamma(m,1)$ are independent where n and m are positive integers then $\frac{X}{X+Y} \sim Beta(n,m)$, describe an algorithm for simulating samples with a Beta(m, n) distribution from independent samples from a U(0, 1) random number generator.

Solution. Note that we can generate a $Gamma(n,1)$ random variable by summing n i.i.d. $Exp(1)$ random variables. If follows that if $U_1, ..., U_{n+m}$ are i.i.d. $U(0,1)$. Then

$$Z = \frac{\sum_1^n \log U_i}{\sum_1^{n+m} \log U_i} \sim Beta(n,m).$$

3. *Simulating from the Beta distribution using rejection sampling.* Design an algorithm to simulate samples from the $Beta(\alpha,\beta)$ distribution where $\alpha, \beta > 1$, using the U(0, 1) as the density $q(x)$,

(a) Derive an expression for the probability that a value generated from the U(0, 1) is accepted for your algorithm.

Solution. The beta density is defined on (0, 1) as

$$\begin{aligned} p(x) &= \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &< \frac{1}{B(\alpha,\beta)} (\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}(\alpha+\beta-2)^{2-\alpha-\beta} = C(\alpha,\beta), \end{aligned}$$

where the r.h.s. is obtained by maximisation.

It follows that the probability that a value is accepted is equal to

$$
\begin{aligned}
p_a &= \int_0^1 q(x) \frac{p(x)}{C(\alpha, \beta)} dx \\
&= \frac{1}{C(\alpha, \beta)}
\end{aligned}
$$

(b) How does this expression behave as α and β become large?

Solution. If we fix the value of $\frac{\alpha}{\beta}$ and let α become large then we can show that the acceptance probability tends to zero. That is the rejection algorithm becomes increasingly inefficient.