ISYE 6420: Homework 3

aelhabr3

1. Estimating the Precision Parameter of a Rayleigh Distribution.

Instructions

If two random variables X and Y are independent of each other and normally distributed with variances equal to σ^2, then the variable $R = \sqrt{X^2 + Y^2}$ follows the Rayleigh distribution. Parameterized with precision parameter $\xi = \frac{1}{\sigma^2}$, the Rayleigh random variable R has a density

$$f(r) = \xi r \exp\{\frac{-\xi r^2}{2}\}, \quad r \geq 0, \xi > 0.$$

An example of such random variable would be the distance of darts from the target center in a dart-throwing game where the deviations in the two dimensions of the target plane are independent and normally distributed.
(a) Assume that the prior on ξ is exponential with the rate parameter λ. Show that the posterior is gamma $\mathcal{G}a(2, \lambda + \frac{r^2}{2})$.
(b) Assume that $R_1 = 3, R_2 = 4, R_3 = 2$, and $R_4 = 5$ are Rayleigh-distributed random observations representing the distance of a dart from the center. Find the posterior in this case for the same prior from (a), and give a Bayesian estimate of ξ.
(c) For $\lambda = 1$, Numerically find 95% Credible Set for ξ.

Hint: In (b) show that if $R_1, R_2, \ldots, r_n$ are observed, and the prior on ξ is exponential $\mathcal{E}(\lambda)$, the the posterior is gamma $\mathcal{G}a(n + 1, \lambda + \frac{1}{2}\sum_{i=1}^{n} r_i^2)$.

Response

a

Since we know that the population is distributed as $f(r) = \xi r e^{-(\xi r^2)/2}$, we can say that the likelihood is

$$\begin{aligned}
\sum_{i=1}^{n} f(r_i) &= \xi r_1 e^{-\frac{1}{2}\xi r_1} + \ldots + \xi r_n e^{-\frac{1}{2}\xi r_n} \\
&\propto \xi e^{-\frac{1}{2}\xi r_1} + \ldots + \xi e^{-\frac{1}{2}\xi r_n} \\
&\propto \xi e^{-\frac{1}{2}\xi \sum_{i=1}^{n} r^2}.
\end{aligned}$$

Given that the prior $\pi(\lambda) = \lambda e^{-\lambda}$ (because it follows the distribution $\mathcal{E}xp(\lambda)$), we deduce that the posterior is

$$\begin{aligned} \text{posterior} &= \text{likelihood} \times \text{prior} \\ &= \xi^n \lambda e^{-\xi(\lambda + \frac{1}{2}\sum_{i=1}^n r_i^2)} \\ &\propto \xi^n e^{-\xi(\lambda + \frac{1}{2}\sum_{i=1}^n r_i^2)} \end{aligned}$$

The above expression is the kernel for the distribution $\mathcal{G}a(n+1, \lambda + \frac{1}{2}\sum_{i=1}^n r_i^2)$. (This is what is indicated by the hint).

Then, for the case of $n = 1$, we see that $\mathcal{G}a(2, \lambda + \frac{1}{2}r^2)$.

b

Using the general form of the posterior shown above with $n = 4$ and $\sum_{i=1}^n r_i^2 = 3^2 + 4^2 + 2^2 + 5^2) = 54$, **we find that the posterior is**

$$\mathcal{G}a((4) + 1, \lambda + \tfrac{1}{2}(54)) = \mathcal{G}a(5, \lambda + 27).$$

The Bayesian estimate of ξ is simply the expected value of the parameter under the posterior distribution.

We note that if $X \sim \mathcal{G}a(\alpha, \beta)$, then

$$\mathrm{E}(X) = \tfrac{\alpha}{\beta}.$$

Thus, given that $\xi \sim \mathcal{G}a(5, \lambda + 27)$, we find that the Bayesian estimate of ξ is

$$\xi_{Bayes} = \mathrm{E}(\xi) = \tfrac{5}{\lambda+27}.$$

c

The equitailed credible set C is based on the posterior distribution. This means that the set $C = [L, U]$ contains the true parameter $1 - \alpha$ probability. Here, the parameter is ξ and $\alpha = 0.05$.

Formally, an equitailed credible set is calculated as follows.

$$\begin{gathered} \int_{-\infty}^{L} \pi(p|x)dp \leq \tfrac{\alpha}{2}, \int_{U}^{+\infty} \pi(p|x)dp \leq \tfrac{\alpha}{2} \\ \text{s.t. } \Pr(p \in [L, U] \,|X) \geq 1 - \alpha \end{gathered}$$

We can calculate the 95% equitailed credibe set in `R` (given $\lambda = 1$) as follows.

```
compute_equi_credible_set <- function(alpha, beta, level = 0.95) {
  q_buffer <- (1 - level) / 2
  q_l <- (1 - level) - q_buffer
  q_u <- level + q_buffer
  res <-
    c(
      l = qbeta(q_l, shape1 = alpha, shape2 = beta),
      u = qbeta(q_u, shape1 = alpha, shape2 = beta)
    )
}
alpha_1 <- 5
lambda_1 <- 1
beta_1 <- lambda_1 + 27
credible_set_1 <- compute_equi_credible_set(alpha_1, beta_1)
credible_set_1
```

```
##          l          u
## 0.05275056 0.28994842
```

We find that the 95% credible set is [0.0528, 0.2899)].

2. Estimating Chemotherapy Response Rates

Instructions

An oncologist believes that 90% of cancer patients will respond to a new chemotherapy treatment and that it is unlikely that this proportion will be below 80%. Elicit a beta prior on proportion that models the oncologist's beliefs.

Hint: For elicitation of the prior use $\mu = 0.9, \mu - 2\sigma = 0.8$ and expressions for μ and σ. for beta.

During the trial, in 30 patients treated, 22 responded.
(a) What are the likelihood and posterior distributions? What is the Bayes estimator of the proportion?
(b) Using Octave, R, or Python, fine 95% Credible Set for p.
(c) Using Octave, R, or Python, test the hypothesis $H_0 : p \geq 4/5$ against the alternative $H_1 : p < 4/5$.
(d) Using WinBUGS, find the Bayes estimator, Credible Set and test, and compare results with (a-c).

Response

First, we elicit a Beta prior according to the instructions.

The expectation of a Beta distribution $\mathcal{B}e(\alpha, \beta)$ for some random variable X is $\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$. Given $\mu = 0.9, \mu - 2\sigma = 0.8$, we can use algebra (or "trial and error") to find that $\alpha = 0.72, \beta = 0.25\alpha = 0.18$.

We can verify that values with `R`.

```
alpha_0 <- 0.72
beta_0 <- 0.18
compute_beta_mu <- function(alpha, beta) {
  alpha / (alpha + beta)
}
mu_0 <- compute_beta_mu(alpha_0, beta_0)
mu_0
```

```
## [1] 0.8
```

a

Given the problem statement, **we deduce the likelihood follows a binomial $\mathcal{B}in(n,p)$ distribution.**

$$X \sim \mathcal{B}in(n,p)$$
$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

And, given $n = 30$ patients, $x = 22$ of whom responded, the likelihood is

$$f(x|p) = \binom{30}{22} p^{22} (1-p)^8.$$

The posterior distribution $\pi(p|x)$ is proportional to the product of the likelihood $f(x|p)$ and the prior distribution $\pi(p)$.

$$\begin{aligned}\pi(p|x) &\propto f(x|p)\pi(p) \\ &\propto p^x(1-p)^{n-x}\end{aligned}$$

Next, we note that the general form the Beta distribution is

$$X \sim \mathcal{B}e(\alpha,\beta)$$
$$f(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

where $0 \leq x \leq 1, \alpha, \beta > 0$, $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$.

We can re-structure the posterior distribution to look like the general Beta distribution.

$$\begin{aligned}\pi(p|x) &\propto p^x(1-p)^{n-x} \\ &= p^{(x+1)-1}(1-p)^{(n-x+1)-1}.\end{aligned}$$

Then, defining $\alpha' = x+1, \beta' = n-x+1$, **we write the posterior distribution as**

$$\begin{aligned}\pi(p|x) &\sim \mathcal{B}e(\alpha', \beta') \\ &\sim \mathcal{B}e(x+1, n-x+1).\end{aligned}$$

Finally, given $n = 30, x = 22$, **we find that the posterior distribution**

$$\begin{aligned}\mathcal{B}e(x+1, n-x+1) &= \mathcal{B}e(22+1, 30-22+1) \\ &= \mathcal{B}e(23, 9).\end{aligned}$$

The Bayes estimator of the proportion p_{Bayes} is simply the expected value of the parameter under the posterior distribution.

We note that if $X \sim \mathcal{B}e(\alpha, \beta)$, then

$$\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}.$$

Thus, we find that

$$\begin{aligned}p_{Bayes} = \mathrm{E}(X) &= \frac{(23)}{(23)+(9)} \\ &= \frac{23}{32} \\ &= 0.71875.\end{aligned}$$

We can verify the posterior mean μ_1 using `R`.

```
n <- 30
x <- 22
alpha_1 <- x + 1
beta_1 <- n - x + 1
mu_1 <- compute_beta_mu(alpha_1, beta_1)
mu_1
```

```
## [1] 0.71875
```

b

We can calculate the 95% credible set using `R` (using the function created in 1(c)).

```
credible_set_1 <- compute_equi_credible_set(alpha_1, beta_1)
credible_set_1
```

```
##         l         u 
## 0.5538661 0.8577715
```

We find that the 95% credible set is [0.5539, 0.8578].

c

To test the null hypothesis $H_0 : p \geq 4/5$ against the alternative $H_1 : p < 4/5$, let's calculate the posterior probability of the alternative first (using `R`).

```
p_h1 <- 4/5
p_h1 <- pbeta(p_h1, alpha_1, beta_1, lower.tail = TRUE)
p_h1
```

```
## [1] 0.8492375
```

Analytically, the calculation for $\Pr(H_1)$ looks as follows.

$$\begin{aligned}
\Pr(H_1 : p < 4/5) &= \int_0^{4/5} \mathcal{B}e(23, 9) dp \\
&= \frac{\Gamma(23+9)}{\Gamma(23)\Gamma(9)} \int_0^{4/5} p^{23-1}(1-p)^{9-1} dp \\
&= \frac{\Gamma(34)}{\Gamma(23)\Gamma(9)} \int_0^{4/5} p^{22}(1-p)^{10} dp \\
&\approx 0.8492.
\end{aligned}$$

Then we find that the probability of the null hypothesis is

$$\Pr(H_0 : p \geq 4/5) = 1 - \Pr(H_1 : p < 4/5) = 0.1508.$$

Since we find that the alternative hypothesis H_1 has a probability greater than 0.5, we choose to accept the alternative hypothesis over the null hypothesis H_0.

Note that Bayesian hypothesis testing is different than Frequentist hypothesis testing. Vidakovic explains in ***Engineering Biostatistics*** (http://statbook.gatech.edu/statb4.pdf).

> *“In frequentist tests, it was customary to formulate H_0 as $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ instead of $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$, as one might expect. The reason was that we calculated the p-value under the assumption that H_0 is true, and this is why a precise null hypothesis was needed.*
>
> *”Bayesian testing is conceptually straightforward: The hypothesis with a higher posterior probability is favored. There is nothing special about the “null” hypothesis, and for a Bayesian, H_0 and H_1 are interchangeable."*

d

Below is OpenBUGS code to evaluate each of the results found above.

OpenBUGS

File Edit Attributes Tools Info Model Inference D

hw03-q2

```
model {
   x ~ dbin(p, n)
   p ~ dbeta(alpha, beta)
   h_1 <- step(4/5 - p)
}

data
list(alpha = 0.72, beta = 0.18, x = 22, n = 30)
```

Note the following about the implementation in OpenBUGS.

- Initial values were generated via “gen inits”;
- 1,000 burn-in samples were used;
- 10,000 samples were used for estimation.

The output is shown below.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
h_1	0.7913	0.4064	0.004145	0.0	1.0	1.0	1001	10000
p	0.7343	0.07828	8.5E-4	0.5673	0.74	0.8697	1001	10000

The Bayes estimator corresponds to the mean of `p`, which is 0.7352. This is very close to the value that we found analytically—0.71875.

The 95% credible set corresponds to the 2.5% and 97.5% values for `p`, which are 0.5693 and 0.8719 respectively. These values are reasonably close to the values we found with `R` –0.5539 and 0.8578.

The probability of the alternative hypothesis H_1 corresponds to the posterior mean of the `h_1` variable —0.7913. This is relatively close to the value that we found with `R` —0.8492.

Aside

We can use the `R` package `{R2OpenBUGS}` to achieve the same results found directly in OpenBUGS.

```
model <- function() {
  x ~ dbin(p, n)
    p ~ dbeta(alpha, beta)
    h_1 <- step(4/5 - p)
}
data <- list(alpha = 0.72, beta = 0.18, x = 22, n = 30)
inits <- NULL
params <- c('h_1', 'p')
res_sim <-
  R2OpenBUGS::bugs(
    data = data,
    inits = inits,
    model.file = model,
    parameters.to.save = params,
    DIC = FALSE,
    n.chains = 1,
    n.iter = 10000,
    n.burnin = 1000
  )
res_sim$summary
```

	mean	sd	2.5%	25%	50%	75%	97.5%
h_1	0.7912222	0.4064578	0.0000000	1.0000	1.00	1.000	1.0000
p	0.7343966	0.0781395	0.5678975	0.6837	0.74	0.791	0.8697