Assignment 1

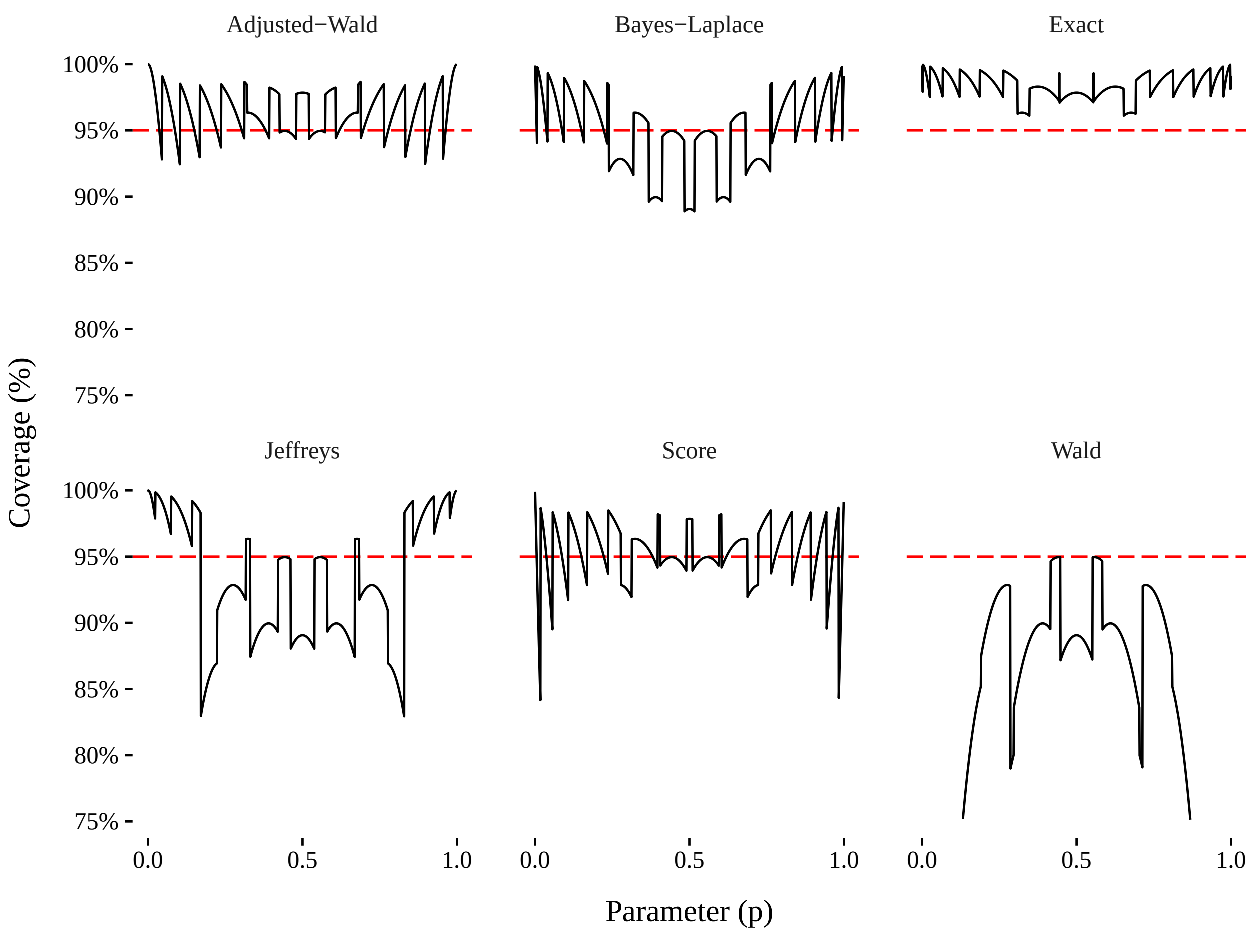
2017-04-05

Question 1: Inference for the binomial parameter:

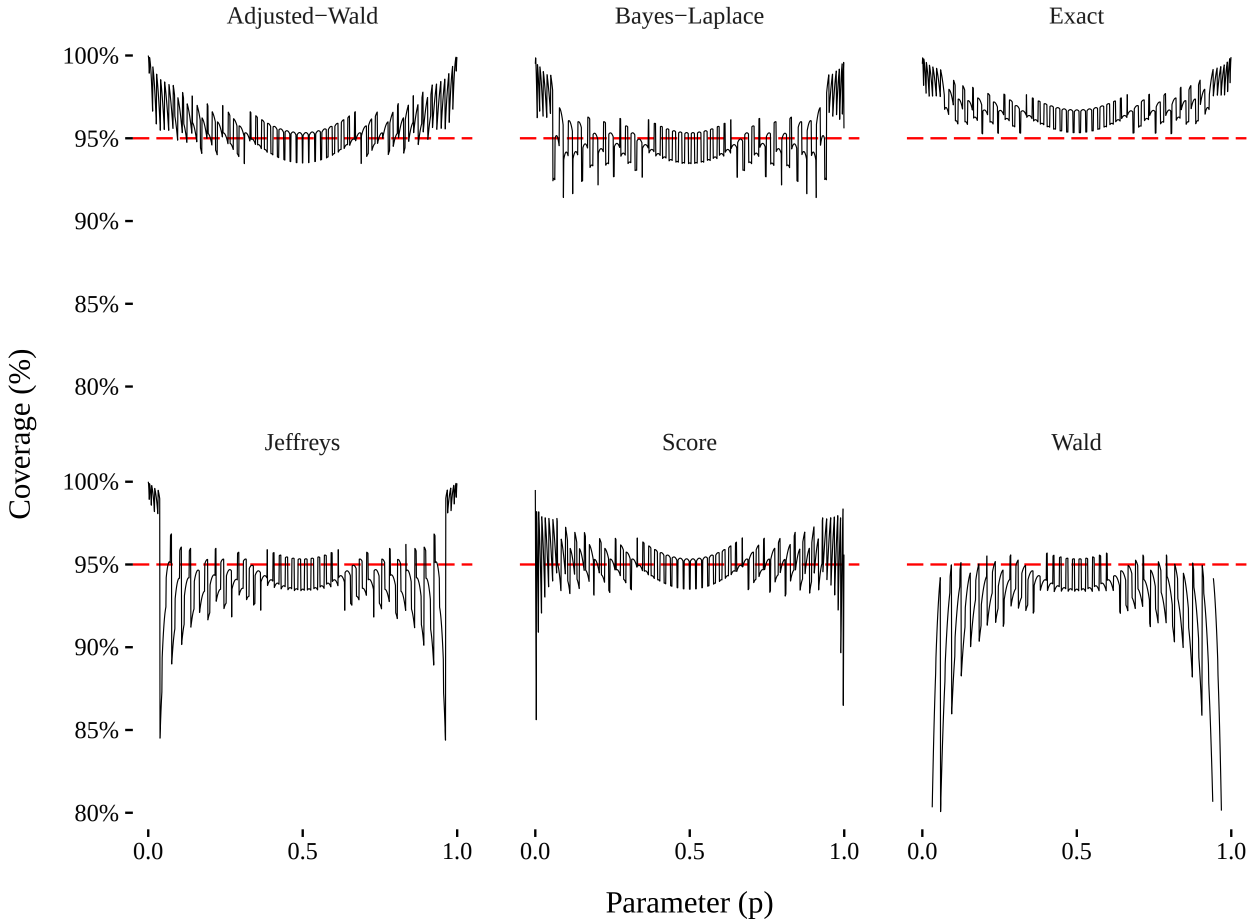
(a) Develop an R function to calculate HPD intervals for data (x, n), given a *beta(a,b)* prior.

Running the above code gives the following output

(b) Reproduce Agresti & Coull's (1998) Figure 4 ($n = 10$), and replicate for the Score and Bayes-Laplace & Jeffreys HPD intervals



(c) Repeat (b) for $n = 50$.



(d) Compare the minimum coverage of the six graphs at (c)

covinterval	mean(dens)	median(dens)	min(dens)
Adjusted-Wald	0.9644291	0.9633510	0.9245493
Bayes-Laplace	0.9498736	0.9518780	0.8888250
Exact	0.9837573	0.9830483	0.9611270
Jeffreys	0.9321782	0.9274306	0.8294769
Score	0.9542317	0.9546219	0.8424326
Wald	0.7698579	0.8770133	0.0099550

covinterval	mean(dens)	median(dens)	min(dens)
Adjusted-Wald	0.9601049	0.9591156	0.9339081
Bayes-Laplace	0.9498598	0.9472447	0.9135598
Exact	0.9733253	0.9730984	0.9536656
Jeffreys	0.9390427	0.9365611	0.8441421
Score	0.9525064	0.9539318	0.8603842
Wald	0.8756749	0.9275039	0.0295690

(e) The adjusted Wald interval appears to perform well with respect to frequentist coverage, if close to nominal combined with reasonable minimum coverage is aimed for. From a Bayesian point of view, performance of individual intervals is just as, if not more, important. Given $x = 0$, compare the adjusted Wald interval with the exact & Score intervals (all two-sided), and with the Bayes-Laplace & Jeffreys

HPD intervals, for a range of values of n and α, comment on its limitations, and give an appropriate graphical illustration.

Question 2: Inference for the Cauchy parameter:

(a) Develop an R function to find percentiles of a (general) Cauchy posterior as discussed by Jaynes (1976, Example 6) and Box & Tiao (1973, p.64), to be used for the examples below.

```
source("Question2/2a.R")
CauchyPercentage <- function(x, # a vector of samples
                             p, # a vector of possible parameters
                             y = NULL # a value to test Pr[p < y]
) {
  cauchydens <- function(x,p){

    cauchydist <- function(x,p) {
      H = vector(mode = "numeric",length = length(x))
      for(i in 1:length(x)){
        H[i] = (1+(x[i] - p)^2)^(-1)
      }
      return(prod(H))
    }
    dens = vapply(p,cauchydist,min(p), x = x)
    c = (integrate(dens, -Inf, Inf)$value)^(-1)
    data.frame(vparams = p, postdens = c*dens(p))
  }

  df = cauchydens(x,p)

  yind = ifelse(length(which(df$vparams == y))==1,
                which(df$vparams == y),
                max(which(df$vparams < y)))

  plot(df$vparams, df$postdens, type = 'l')
  abline(v = df$vparams[yind])
  abline(h = df[yind,"postdens"])

  cumdist = c*integrate(dens,-Inf,y)$value

  return((c(yind,df[yind,"postdens"],cumdist)))
}

CauchyHPD <- function(x, # vector of samples
                      p, # vector of possible parameter values
                      alpha = 0.95, # HPD interval value
                      tol = 0.0001) { # level of tolerance for exact HPD interval

  cauchydens <- function(x,p){

    cauchydist <- function(x,p) {
      H = vector(mode = "numeric",length = length(x))
      for(i in 1:length(x)){
        H[i] = (1+(x[i] - p)^2)^(-1)
      }
      return(prod(H))
    }
    dens = function(p) vapply(p,cauchydist,min(p), x = x)
```

```
    c = (integrate(dens, -Inf, Inf)$value)^(-1)
    data.frame(vparams = p, postdens = c*dens(p))
  }

  df = cauchydens(x,p)

  cumdist = cumsum(df$postdens)*diff(df$vparams)[1]
  post_median = which.min(abs(cumdist-0.5))

  HPDlimits <- function(post_dens) { ## find lower and upper values for which
    ## prob dens is closest to target value
    lower = which.min(abs(df$postdens[1:post_median]-post_dens))
    upper = which.min(abs(df$postdens[(post_median+1):length(df$postdens)]-post_dens))+post_median
    limits = c(lower,upper)
  }

  HPDlimitarea <- function(post_dens) {
    limitints = HPDlimits(post_dens)
    limitarea = sum(df$postdens[limitints[1]:limitints[2]])*diff(df$vparams)[1]
  }
  ## find credible interval
  v2 = seq(0,max(df$postdens),by=tol)
  vals = sapply(v2,HPDlimitarea)
  w = which.min(abs(vals-alpha))
  r = c(df$vparams[HPDlimits(v2[w])])
  names(r) = c("lower","upper")
  par(mfrow = c(1,2))
  plot(df$vparams, cumdist, type = 'l')
  abline(h = 0.5)
  abline(v = df$vparams[post_median], col = 'red')
  abline(v = r["upper"], col = 'blue')
  abline(v = r["lower"], col = 'blue')
  plot(df$vparams, df$postdens, type = 'l')
  abline(v = df$vparams[post_median], col = 'red')
  abline(h = df[HPDlimits(v2[w])[1],"postdens"])
  abline(v = r["upper"], col = 'blue')
  abline(v = r["lower"], col = 'blue')
  return(r)
}
```

(b) Consider Jaynes' example of $n = 2$ observations $(3, 5)$: plot the posterior and calculate the 90% central credible interval. Explain why it is quite different from the confidence interval derived by Jaynes (p.202).

```
source("Question2/2b.R")
```

(c) Consider Box & Tiao's example of $n = 5$ observations $(11.4, 7.3, 9.8, 13.7, 10.6)$: plot the posterior and calculate 95% central and HPD credible intervals and check $Pr[\theta < 11.5]$ given by Box & Tiao.

```
source("Question2/2c.R")
```

(d) Consider Berger's (1985, p.141) example of $n = 5$ observations $(4.0, 5.5, 7.5, 4.5, 3.0)$: calculate 95% central and HPD credible intervals, with and without Berger's restriction $(\theta > 0)$.

```
source("Question2/2d.R")
```

(e) Clearly, Berger's restriction $(\theta > 0)$ will sometimes lead to a posterior quite different from the unrestricted posterior. Plot this restricted posterior for the hypothetical negative version of Berger's example: i.e. $(-4.0, -5.5, -7.5, -4.5, -3.0)$, and calculate the 95% HPD interval.

```
source("Question2/2e.R")
```