

2 Probability from Bayesian perspective

Diego

Table of contents

1	Highest density interval (HDI)	2
---	--------------------------------	---

1 Highest density interval (HDI)

```
pacman::p_load(tidyverse, brms, janitor, tidybayes)
```

```
HtWtDataGenerator <- function(n_subj, rndsd = NULL, male_prob = 0.50) {  
  
  # Random height, weight generator for males and females. Uses parameters from  
  # Brainard, J. & Burmaster, D. E. (1992). Bivariate distributions for height and  
  # weight of men and women in the United States. Risk Analysis, 12(2), 267-275.  
  # Kruschke, J. K. (2011). Doing Bayesian data analysis:  
  # A Tutorial with R and BUGS. Academic Press / Elsevier.  
  # Kruschke, J. K. (2014). Doing Bayesian data analysis, 2nd Edition:  
  # A Tutorial with R, JAGS and Stan. Academic Press / Elsevier.  
  
  # require(MASS)  
  
  # Specify parameters of multivariate normal (MVN) distributions.  
  # Men:  
  HtMmu <- 69.18  
  HtMsD <- 2.87  
  lnWtMmu <- 5.14  
  lnWtMsD <- 0.17  
  Mrho <- 0.42  
  Mmean <- c(HtMmu, lnWtMmu)  
  Msigma <- matrix(c(HtMsD^2, Mrho * HtMsD * lnWtMsD,  
                    Mrho * HtMsD * lnWtMsD, lnWtMsD^2), nrow = 2)  
  
  # Women cluster 1:  
  HtFmu1 <- 63.11  
  HtFsd1 <- 2.76  
  lnWtFmu1 <- 5.06  
  lnWtFsd1 <- 0.24  
  Frho1 <- 0.41  
  prop1 <- 0.46  
  Fmean1 <- c(HtFmu1, lnWtFmu1)  
  Fsigma1 <- matrix(c(HtFsd1^2, Frho1 * HtFsd1 * lnWtFsd1,  
                    Frho1 * HtFsd1 * lnWtFsd1, lnWtFsd1^2), nrow = 2)  
  
  # Women cluster 2:  
  HtFmu2 <- 64.36  
  HtFsd2 <- 2.49  
  lnWtFmu2 <- 4.86
```

```

lnWtFsd2 <- 0.14
Frho2    <- 0.44
prop2    <- 1 - prop1
Fmean2   <- c(HtFmu2, lnWtFmu2)
Fsigma2  <- matrix(c(HtFsd2^2, Frho2 * HtFsd2 * lnWtFsd2,
                    Frho2 * HtFsd2 * lnWtFsd2, lnWtFsd2^2), nrow = 2)

# Randomly generate data values from those MVN distributions.
if (!is.null(rndsd)) {set.seed(rndsd)}
data_matrix <- matrix(0, nrow = n_subj, ncol = 3)
colnames(data_matrix) <- c("male", "height", "weight")
maleval <- 1; femaleval <- 0 # arbitrary coding values
for (i in 1:n_subj) {
  # Flip coin to decide sex
  sex = sample(c(maleval, femaleval), size = 1, replace = TRUE,
              prob = c(male_prob, 1 - male_prob))
  if (sex == maleval) {datum <- MASS::mvrnorm(n = 1, mu = Mmean, Sigma = Msigma)}
  if (sex == femaleval) {
    Fclust = sample(c(1, 2), size = 1, replace = TRUE, prob = c(prop1, prop2))
    if (Fclust == 1) {datum <- MASS::mvrnorm(n = 1, mu = Fmean1, Sigma = Fsigma1)}
    if (Fclust == 2) {datum <- MASS::mvrnorm(n = 1, mu = Fmean2, Sigma = Fsigma2)}
  }
  data_matrix[i, ] = c(sex, round(c(datum[1], exp(datum[2])), 1))
}

return(data_matrix)

} # end function

```

i Definition of HDI

The HDI indicates which points of a distribution are most credible, and which cover most of the distribution. It summarizes the distribution by specifying an interval that spans most of the distribution, such that every point inside this interval has a higher credibility than any point outside the interval.

Here we have a function to generate the HDI:

```

hdi_of_icdf <- function(name, width = .95, tol = 1e-8, ... ) {

  # Arguments:

```

```

# `name` is R's name for the inverse cumulative density function
# of the distribution.
# `width` is the desired mass of the HDI region.
# `tol` is passed to R's optimize function.
# Return value:
# Highest density interval (HDI) limits in a vector.
# Example of use: For determining HDI of a beta(30, 12) distribution, type
# `hdi_of_icdf(qbeta, shape1 = 30, shape2 = 12)`
# Notice that the parameters of the `name` must be explicitly stated;
# e.g., `hdi_of_icdf(qbeta, 30, 12)` does not work.
# Adapted and corrected from Greg Snow's TeachingDemos package.

incredible_mass <- 1.0 - width
interval_width <- function(low_tail_prob, name, width, ...) {
  name(width + low_tail_prob, ...) - name(low_tail_prob, ...)
}
opt_info <- optimize(interval_width, c(0, incredible_mass),
  name = name, width = width,
  tol = tol, ...)
hdi_lower_tail_prob <- opt_info$minimum

return(c(name(hdi_lower_tail_prob, ...),
  name(width + hdi_lower_tail_prob, ...)))
}

```

For a normal distribution, 95% HDI is identical to 95% CI.

```

h <- hdi_of_icdf(name = qnorm,
  mean=0, sd=1)
h
## [1] -1.959964 1.959964

```

Let's visualize 95% HDI in the normal distribution

```

tibble(x = seq(from = -3.5, to = 3.5, by = .05)) %>%
  mutate(d = dnorm(x, mean = 0, sd = 1)) %>%

  ggplot(aes(x = x, y = d)) +
  geom_area(fill = "grey75") +
  geom_area(data = . %>% filter(x >= h[1] & x <= h[2])),

```

```

    fill = "grey50") +
  geom_line(data = tibble(x = c(h[1] + .02, h[2] - .02),
    d = c(.059, .059)),
    arrow = arrow(length = unit(.2, "cm"),
      ends = "both",
      type = "closed"),
    color = "grey92") +
  annotate(geom = "text", x = 0, y = .09,
    label = "95% HDI", color = "grey92") +
  xlim(-3.1, 3.1) +
  ylab("p(x)") +
  theme(panel.grid = element_blank())

```

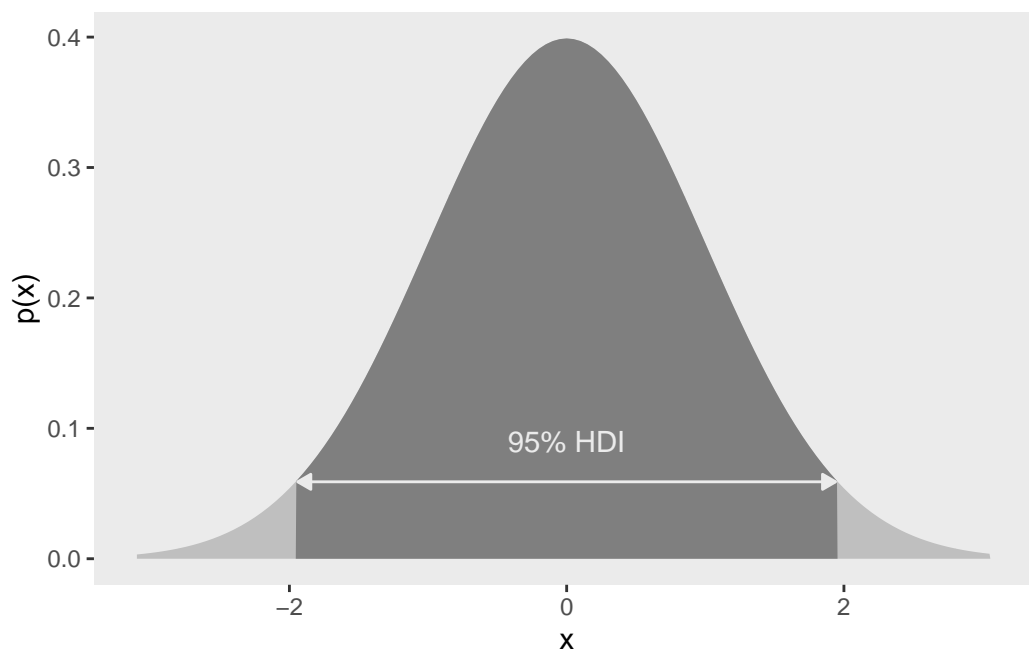


Figure 1: 95% HDI for a normal distribution

However, in a Beta distribution, we will get:

```

h <- hdi_of_icdf(name=qbeta,
  shape1=15, shape2=4)

tibble(x = seq(from = 0, to = 1, by = .01)) %>%
  mutate(d = dbeta(x, shape1 = 15, shape2 = 4)) %>%

```

```

ggplot(aes(x = x, y = d)) +
  geom_area(fill = "grey75") +
  geom_area(data = . %>% filter(x >= h[1] & x <= h[2]),
            fill = "grey50") +
  geom_line(data = tibble(x = c(h[1] + .01, h[2] - .002),
                           d = c(.75, .75)),
            arrow = arrow(length = unit(.2, "cm"),
                           ends = "both",
                           type = "closed"),
            color = "grey92") +
  annotate(geom = "text", x = .8, y = 1.1,
            label = "95% HDI", color = "grey92") +
  xlim(.4, 1) +
  ylab("p(x)") +
  theme(panel.grid = element_blank())

```

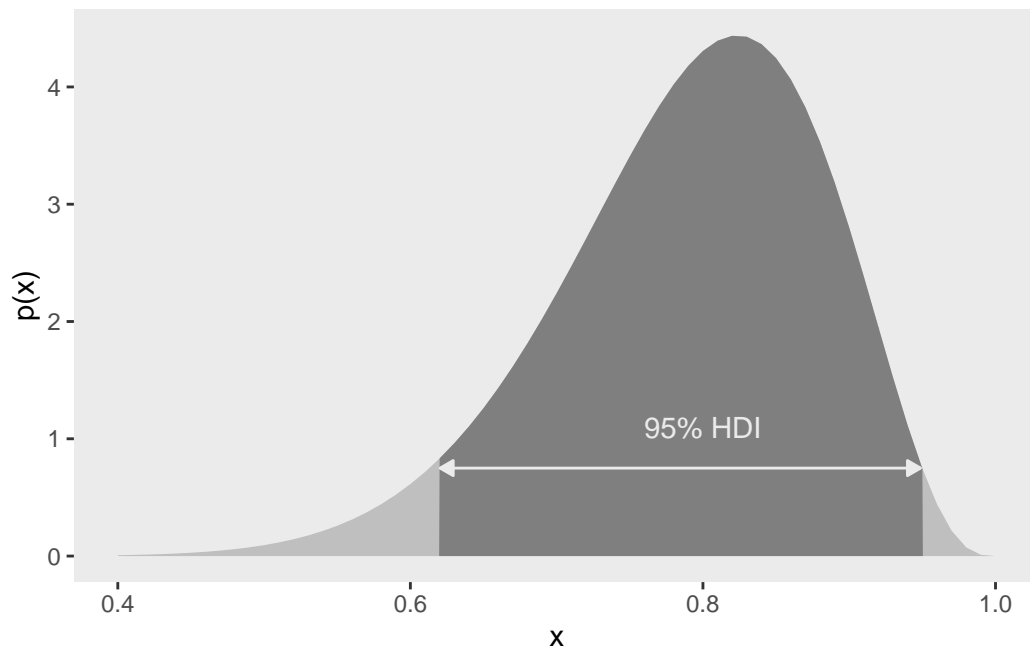


Figure 2: 95% HDI in Beta distribution

If we use “tidybayes” packages

```

set.seed(4)
d <-
  tibble(x = c(rnorm(6e5, mean = 1.50, sd = .5),
               rnorm(4e5, mean = 4.75, sd = .5)))

h <-
  d %>%
    mode_hdi()
h
## # A tibble: 2 x 6
##       x .lower .upper .width .point .interval
##   <dbl> <dbl> <dbl> <dbl> <chr> <chr>
## 1  1.50  0.468  2.54  0.95 mode hdi
## 2  1.50  3.81  5.68  0.95 mode hdi

```