Chapter 3

Inference from posterior distribution

Assumption in order:

(1) Likelihood,

(2) Parameters,

(3) Prior.

The assumption results in a Posterior.

Likelihood

- The probability of the data given the parameters. (Assumption)

Likelihood

- The probability of the data given the parameters. (Assumption)
- Which parameters? The parameters you need to define a probability distribution.

Likelihood

- The probability of the data given the parameters. (Assumption)
- Which parameters? The parameters you need to define a probability distribution.
- The likelihood is not a distribution for the parameters.

Example: Placenta praevia

- For the data:

$$X = \begin{cases} 1 & \text{if the child is girl,} \\ 0 & \text{if the child is boy.} \end{cases}$$

What parameter do we need for defining the distribution of the data ? ($\mathbb{P}(X|?)$)

Example: Placenta praevia

- For the data:

$$X = \begin{cases} 1 & \text{if the child is girl,} \\ 0 & \text{if the child is boy.} \end{cases}$$

  What parameter do we need for defining the distribution of the data ? ($\mathbb{P}(X|?)$)

- The parameter is the probability of a child being a girl, p.

Example: Placenta praevia

- If you observe n_F girls of n births, the distribution given the data, is binomial:

$$\mathbb{P}(n_F | n, p) = \frac{n!}{n_F!(n - n_F)!} p^{n_F} (1 - p)^{n - n_F}$$

Priors

- Prior incorporates prior information, before the data.

Priors

- Prior incorporates prior information, before the data.
- Often no information is available:
    - Vague prior - small impact on the posterior.
    - regularizing prior - forces the posterior in certain direction (for more complex model).

Example: Placenta praevia

- If you observe n_F girls of n births, the distribution given the data, is binomial:

$$\mathbb{P}(n_F | n, p) = \frac{n!}{n_F!(n - n_F)!} p^{n_F} (1-p)^{n-n_F}$$

Posterior

From the prior, parameters, and likelihood, one gets posterior:

- By Bayes theorem

$$p(p|n_F, n) = \frac{p(n_F|n, p)p(p)}{p(n_F)}$$

Posterior

From the prior, parameters, and likelihood, one gets posterior:

- By Bayes theorem

$$p(p|n_F, n) = \frac{p(n_F|n, p)p(p)}{p(n_F)}$$

- Or in words:

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Average Likelihood}}$$

Inference from posterior distribution

- The posterior distribution

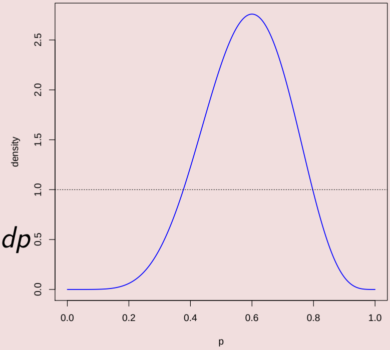
$$p(p|n,y) \propto p^y(1-p)^{n-y}\mathbb{I}_{[0,1]}(p)$$

.

- What is the probability that p is less then 0.5:

$$\mathbb{P}(p < 0.5|n,y) = \int_0^{0.5} p(p|n,y)dp$$

(here we would need the actual density)



Sampling from the posterior

- Grid approximation of the posterior:

```
y = 6
n = 10
p_grid <- seq(from = 0, to  = 1, length.out = 1000)
prior  <- rep(1, 1000)
likelihood <- dbinom(x = y, size = n, prob = p_grid)
posterior  <- likelihood * prior
posterior  <- posterior / sum(posterior)
```

Sampling from the posterior

- Grid approximation of the posterior:

```
y = 6
n = 10
p_grid <- seq(from = 0, to = 1, length.out = 1000)
prior <- rep(1, 1000)
likelihood <- dbinom(x = y, size = n, prob = p_grid)
posterior <- likelihood * prior
posterior <- posterior / sum(posterior)
```

- Sampling:

```
samples <- sample(p_grid,
                  prob = posterior,
                  size = 10000,
                  replace = T)
```

Posterior distribution

What is the probabilility that p is less then 0.5?

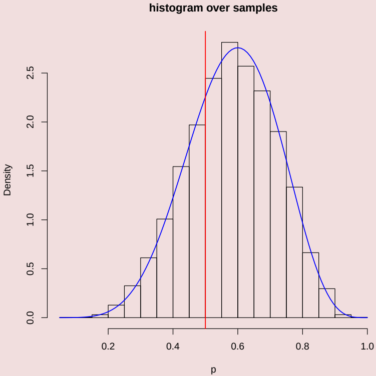


Figure : Histogram of draws from the posterior. Blue line is the posterior distribution

Calculating $\mathbb{P}(p < 0.5)$

- $\mathbb{P}(p|n = 10, y = 6) = \int_0^{0.5} p(p|n = 10, y = 6)dp = 0.274$

Calculating $\mathbb{P}(p < 0.5)$

- $\mathbb{P}(p|n = 10, y = 6) = \int_0^{0.5} p(p|n = 10, y = 6)dp = 0.274$
- Approximation by sampling $(p_1, p_2, \ldots, p_s)$:

$$P_{est} = \frac{1}{10^4} \sum_{i=1}^{10^4} \mathbb{I}_{[0,0.5)}(p_i) = 0.281$$

In code:

```
samples <- sample(p_grid,
                  prob = posterior,
                  size = 10000,
                  replace = T)
P_est <- mean(samples < 0.5)
```

Approximation

```
samples <- sample(p_grid,
prob = posterior,
size = s,
replace = T)
P_est <- mean(samples < 0.5)
```

- What is the effect of the code *s* (the number of samples) on the approximation?

Approximation

```
samples <- sample(p_grid,
prob = posterior,
size = s,
replace = T)
P_est <- mean(samples < 0.5)
```

- What is the effect of the code s (the number of samples) on the approximation?
- $s = 2$ three runs of the code

$$P_{est}^1 = 0.5, P_{est}^2 = 0, P_{est}^3 = 0.$$

Approximation

```
samples <- sample(p_grid ,
prob = posterior ,
size = s ,
replace = T)
P_est <- mean(samples < 0.5)
```

- What is the effect of the code s (the number of samples) on the approximation?
- $s = 2$ three runs of the code

$$P_{est}^1 = 0.5, P_{est}^2 = 0, P_{est}^3 = 0.$$

- $s = 10$ three runs of the code

$$P_{est}^1 = 0, P_{est}^2 = 0.3, P_{est}^3 = 0.2.$$

Approximation

```
samples <- sample(p_grid,
prob = posterior,
size = s,
replace = T)
P_est <- mean(samples < 0.5)
```

- What is the effect of the code s (the number of samples) on the approximation?
- $s = 2$ three runs of the code

$$P^1_{est} = 0.5, P^2_{est} = 0, P^3_{est} = 0.$$

- $s = 10$ three runs of the code

$$P^1_{est} = 0, P^2_{est} = 0.3, P^3_{est} = 0.2.$$

- $s = 10^5$ three runs of the code

$$P^1_{est} = 0.272, P^2_{est} = 0.274, P^3_{est} = 0.274.$$

Predictions

- I will denote future/unobserved data with a star.

Predictions

- I will denote future/unobserved data with a star.
- y^* number tails, n^* number of coin tosses.

Predictions

- I will denote future/unobserved data with a star.
- y^* number tails, n^* number of coin tosses.
- The predictive distribution prior to observing the data (y, n)

$$p(y^*|n^*) = \int p(y^*|n^*, p)p(p)dp.$$

Predictions

- I will denote future/unobserved data with a star.
- y^* number tails, n^* number of coin tosses.
- The predictive distribution prior to observing the data (y, n)

$$p(y^*|n^*) = \int p(y^*|n^*, p)p(p)dp.$$

- The predictive distribution after observing the data (y, n)

$$p(y^*|n^*, n, y) = \int p(y^*|n^*, p)p(p|n, y)dp.$$

Sampling from the posterior

- Prior predictive distribution

```
samples <- runif(n = 10000, min = 0, max =1)
ystar   <- rbinom(n = 10000,
                  size = 16,
                  prob = samples)
```

Sampling from the posterior

- Prior predictive distribution

```
samples <- runif(n = 10000, min = 0, max =1)
ystar   <- rbinom(n = 10000,
                  size = 16,
                  prob = samples)
```

- Posterior predictive distribution

```
y = 6
n = 10
p_grid <- seq(from = 0, to  = 1, length.out = 1000)
prior  <- rep(1, 1000)
likelihood <- dbinom(x = y, size = n, prob = p_grid)
posterior  <- likelihood * prior
posterior  <- posterior / sum(posterior)
samples <- sample(p_grid,
                  prob = posterior,
                  size = 10000,
                  replace = T)
ystar   <- rbinom(n = 10000,
                  size = 16,
                  prob = samples)
```

Bayesian intervals

- a P% Credibility interval (CI_P) is not unique.

Bayesian intervals

- a P% Credibility interval (CI_P) is not unique.
- CI_P for a random variable X is any interval $[a, b]$ such that

$$\mathbb{P}(X \in [a, b]) = P.$$

Bayesian intervals

- a P% Credibility interval (CI_P) is not unique.
- CI_P for a random variable X is any interval $[a, b]$ such that

$$\mathbb{P}(X \in [a, b]) = P.$$

- An approximation of the percentile interval (PI) is obtained by taking quantiles from samples.

Bayesian intervals

- a P% Credibility interval (CI_P) is not unique.
- CI_P for a random variable X is any interval $[a, b]$ such that

$$\mathbb{P}(X \in [a, b]) = P.$$

- An approximation of the percentile interval (PI) is obtained by taking quantiles from samples.
- Highest posterior interval (HPDI) is the smallest possible interval.

Bayesian intervals

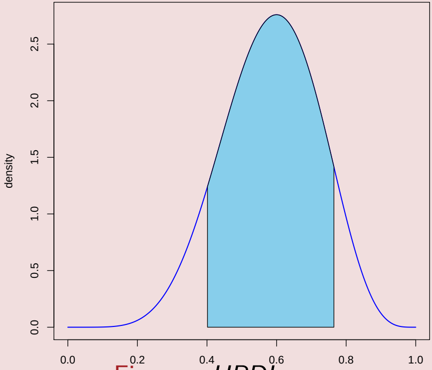


Figure : $HPDI_{80}$

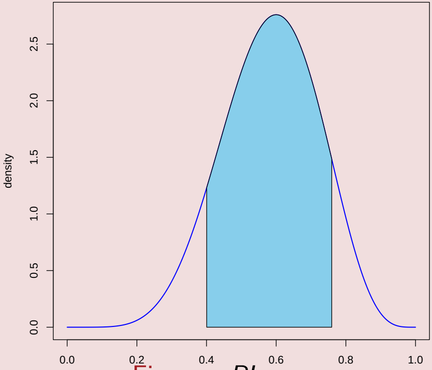


Figure : PI_{80}