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

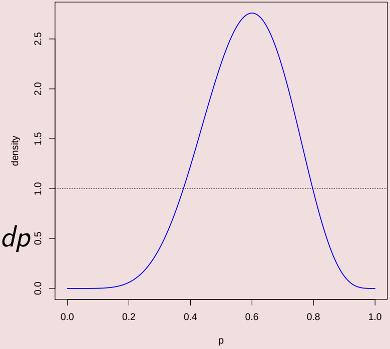
$$p(p|n, y) \propto p^y(1-p)^{n-y}\mathbb{I}_{[0,1]}(p)$$

.

- 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

```
p_grid <- seq(from = 0, to = 1, length.out = 10)
print(p_grid)
```

- ```
  [1] 0.0000000 0.1111111 0.2222222 0.3333333
  [5] 0.4444444 0.5555556 0.6666667 0.7777778
  [9] 0.8888889 1.0000000
  ```
  ```

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 10)
```

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 10)
```

- .

```
prior  <- rep(1, 10)
print(prior)
```

```
[1] 1 1 1 1 1 1 1 1 1 1
```

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 10)
prior  <- rep(1, 10)
```

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 10)
prior  <- rep(1, 10)
```

- .

```
nf = 6
n = 10
likelihood <- dbinom(x = nf, size = n, prob = p_grid)
print(likelihood)
```

```
[1]  0.0000000000 0.0002466915 0.0092547850
[4]  0.0569018950 0.1541821742 0.2409096472
[7]  0.2276075801 0.1133711163 0.0157882546
[10] 0.0000000000
```

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 10)
prior  <- rep(1, 10)
nf = 6
n = 10
likelihood <- dbinom(x = nf, size = n, prob = p_grid)
```

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 10)
prior  <- rep(1, 10)
nf = 6
n = 10
likelihood <- dbinom(x = nf, size = n, prob = p_grid)
```

- .

```
posterior  <- likelihood * prior
print(posterior) # sum(posterior) = 0.8182621
```

```
[4]  0.0569018950  0.1541821742  0.2409096472
[7]  0.2276075801  0.1133711163  0.0157882546
[10]  0.0000000000
```

Sampling from the posterior

- Grid approximation of the posterior:

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

Sampling from the posterior

- Grid approximation of the posterior:

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

- .

```
posterior  <- posterior / sum(posterior)
print(posterior) # sum(posterior) = 1
```

```
[4]  0.0695399334  0.1884263806  0.2944162197
[7]  0.2781597338  0.1385510953  0.0192948614
[10]  0.0000000000
```

Sampling from the posterior

- Grid approximation of the posterior:

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

Sampling from the posterior

- Grid approximation of the posterior:

```
p_grid <- seq(from = 0, to  = 1, length.out = 1000)
prior  <- rep(1, 1000)
nf = 6
n = 10
likelihood <- dbinom(x = nf, 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?

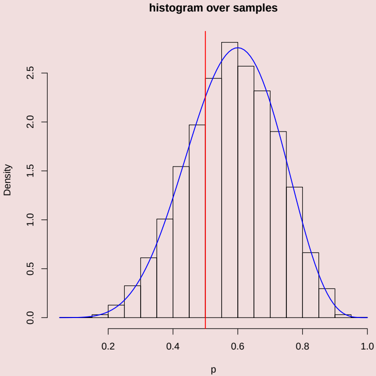


Figure : Histogram of draws from the posterior. Blue line is the posterior distribution

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

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

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

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

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

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

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

- $s = 10$ three runs of the code

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

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

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

Predictions

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

Predictions

- I will denote future/unobserved data with a star.
- n_f^* number tails, n^* number of coin tosses.
- For a given $\hat{p}$, the distribution of n_f^* is the likelihood

$$p(n_f^*|n^*, \hat{p}) = \frac{n^*!}{n^*!(n^* - n_f^*)!} \hat{p}^{n_f^*}(1 - \hat{p})^n$$

Predictions

- I will denote future/unobserved data with a star.
- n_f^* number tails, n^* number of coin tosses.
- For a given $\hat{p}$, the distribution of n_f^* is the likelihood

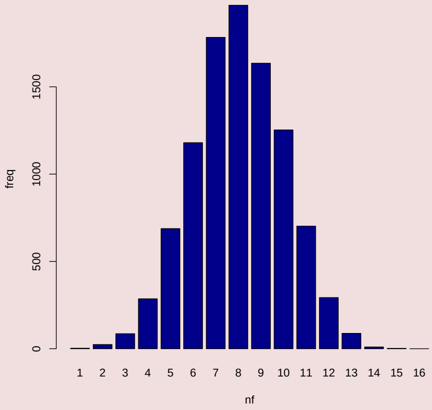
$$p(n_f^*|n^*, \hat{p}) = \frac{n^*!}{n^*!(n^* - n_f^*)!}\hat{p}^{n_f^*}(1 - \hat{p})^n$$

- In R:

```
nf_star <- rbinom(n = 1000, #number of samples
                  size = 16,
                  prob = p_hat)
```

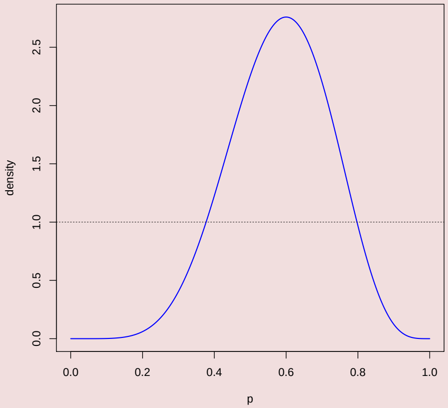
samples n_f^*

Distribution of $p(n_f^*|n = 16, \hat{p} = 0.5)$



which $\hat{p}$

Which $\hat{p}$?



Predictions

- The answer no single $\hat{p}$!

Predictions

- The answer no single $\hat{p}$!
- The predictive distribution prior to observing the data (y, n)

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

Predictions

- The answer no single $\hat{p}$!
- The predictive distribution prior to observing the data (y, n)

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

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

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

Sampling from the posterior

- Prior predictive distribution

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

Sampling from the posterior

- Prior predictive distribution

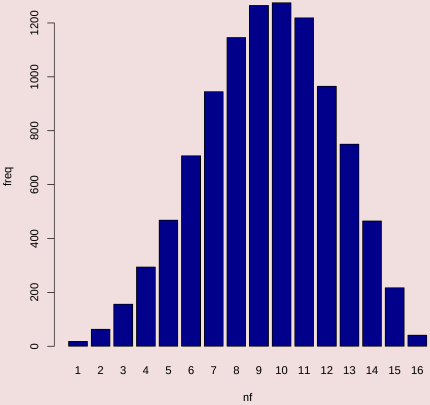
```
samples <- runif(n = 10000, min = 0, max =1)
nf_star   <- rbinom(n = 10000,
                  size = 16,
                  prob = samples)
```

- Posterior predictive distribution

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

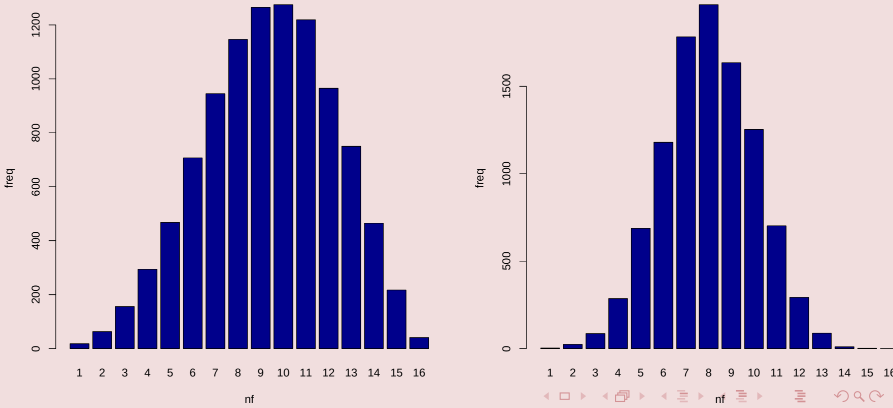
samples n_f^*

Distribution of $p(n_f^* | n^* = 16, n = 10, n_f = 6)$



samples n_f^*

Distribution of $p(n_f^* | n^* = 16, n = 10, n_f = 6)$



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

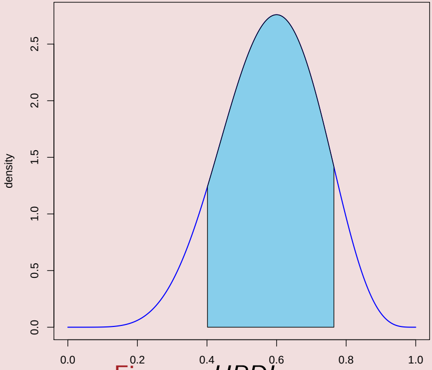


Figure : $HPDI_{80}$

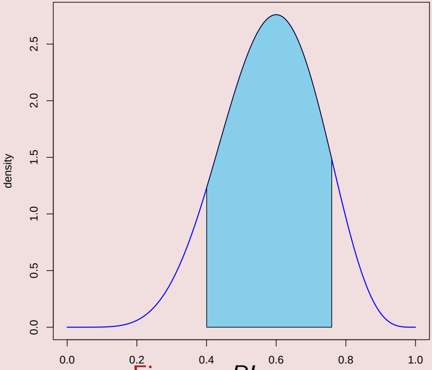


Figure : PI_{80}