Advanced Bayesian Methods: Theory and Applications in R

Posterior Summaries

Nikolaus Umlauf
https://nikum.org/abm.html

Posterior Summaries

- While the ultimate outcome of Bayesian inference is the posterior, this is often compressed into posterior summaries, in particular
  - posterior point estimates and
  - posterior measures of uncertainty.
- Typical point estimates:
  - posterior mean (estimated by averages of samples),
  - posterior median (estimated by empirical median),
  - posterior mode (difficult to determine from samples).
- Typical measures of uncertainty:
  - posterior variance / standard deviation (estimated by empirical analogues),
  - posterior quantiles.

Credible Intervals

- A pointwise Bayesian credible interval $[\theta_{s,\text{low}}, \theta_{s,\text{upp}}]$ for a scalar parameter θ_s is characterized by the posterior coverage probability

$$P(\theta_{s,\text{low}} \leq \theta_s \leq \theta_{s,\text{upp}} | \mathbf{y}) \geq 1 - \alpha$$

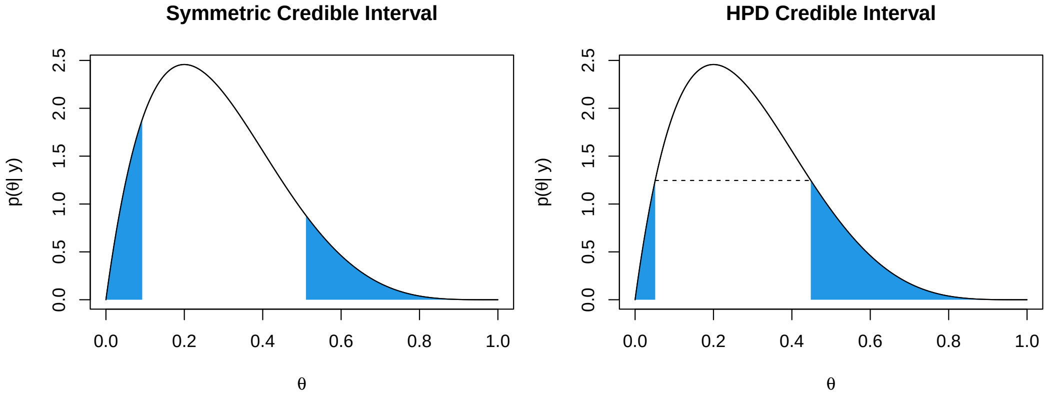
  where $1 - \alpha$ denotes the desired coverage level.

- A simultaneous band for multiple parameters $\{\theta_s, s \in \mathcal{S}\}$ should have

$$P(\theta_{s,\text{low}} \leq \theta_s \leq \theta_{s,\text{upp}}, s \in \mathcal{S} | \mathbf{y}) \geq 1 - \alpha$$

Credible Intervals

- Symmetric and highest posterior density credible intervals:



Credible Intervals

In R:

```
R> library("HDInterval")
```

Simulate beta distributed data (as in the density plot before).

```
R> x <- rbeta(1000, shape1 = 2, shape2 = 5)
```

HDI intervals for samples and the beta distribution.

```
R> i <- hdi(x, credMass = 0.8)
R> j <- hdi(qbeta, credMass = 0.8, shape1 = 2, shape2 = 5)

R> print(rbind(i, j))
       lower     upper
i 0.06738695 0.4619858
j 0.05126051 0.4483045
```

Bayesian Tests

- To test the hypotheses

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \notin \Theta_0$$

  we can compute the posterior probabilities

$$p_0 = P(\boldsymbol{\theta} \in \Theta_0 | \mathbf{y}) \quad \text{and} \quad p_1 = P(\boldsymbol{\theta} \notin \Theta_0 | \mathbf{y}).$$

- The decision can then be based on the ratio

$$\frac{p_1}{p_0}$$

  that measures the evidence in favor of H_1 as compared to H_0.

Bayesian Tests

- H_1 and H_0 are therefore treated symmetrically in the Bayesian context.
- Unfortunately, point hypotheses can not meaningfully be tested in the Bayesian paradigm since then

$$P(\boldsymbol{\theta} = \boldsymbol{\theta}_0 | \mathbf{y}) = 0.$$

- Instead of formally testing hypothesis, the decision between H_0 and H_1 is often made based on model choice procedures in the Bayesian framework.
- As an alternative, one often considers a Bayesian credible interval for $\boldsymbol{\theta}$ and evaluates whether $\boldsymbol{\theta}_0$ is contained in the credible interval or not.

Bayesian Tests

In R:

```
R> library("bamlss")
```

Simulate some data.

```
R> set.seed(123)
R> n <- 1000
R> x1 <- rnorm(n)
R> x2 <- rnorm(n)
R> y <- 1.2 + 0.5 * x1 + rnorm(n, sd = 0.2)
```

Estimate linear model using MCMC.

```
R> b <- bamlss(y ~ x1 + x2)
```

Bayesian Tests

95% Credible intervals.

```
R> confint(b, model = "mu", level = 0.95)
                    2.5%      97.5%
(Intercept)  1.183167435 1.20773499
x1           0.483302348 0.50753724
x2          -0.006737225 0.01728374
```

Inference for Derived Quantities

- Goal: Conduct Bayesian inference for a derived quantity

$$\boldsymbol{\eta} = g(\boldsymbol{\theta}).$$

- Convenient feature of MCMC: If $\boldsymbol{\theta}^{[1]}, \ldots, \boldsymbol{\theta}^{[T]}$ is a sample from the posterior of $\boldsymbol{\theta}$, $g(\boldsymbol{\theta}^{[1]}), \ldots, g(\boldsymbol{\theta}^{[T]})$ will be a sample from the posterior of the transformed parameter.
- No restrictions on the transformation $g(\cdot)$ and no need to deal with asymptotic considerations

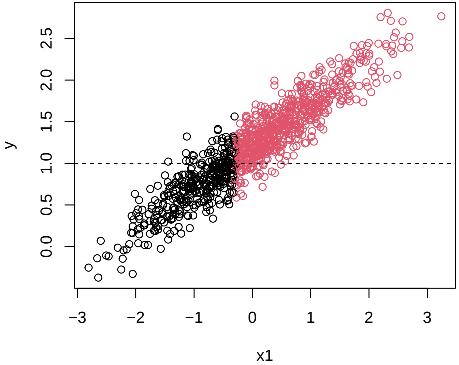
Inference for Derived Quantities

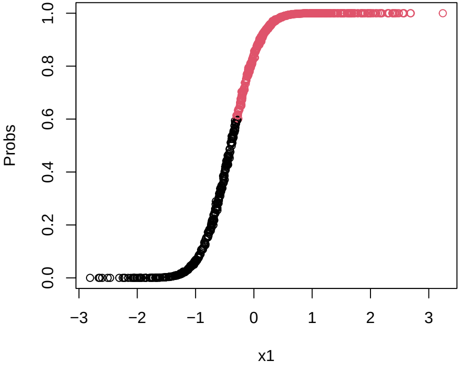
Example in R:

```
R> nd <- data.frame("x1" = x1, "x2" = x2)
R> par <- predict(b, newdata = nd, type = "parameter", FUN = identity)
R> print(names(par))
R> print(dim(par$mu))
R> Probs <- NULL
R> for(i in 1:ncol(par$mu)) {
+    tpar <- list("mu" = par$mu[, i], "sigma" = par$sigma[, i])
+    p <- 1 - family(b)$p(1, tpar) ## Same as 1 - pnorm(1, ...).
+    Probs <- cbind(Probs, p)
+  }
R> Probs <- apply(Probs, 1, mean)
R> col <- rep(1, n)
R> col[Probs > 0.6] <- 2
R> par(mfrow = c(1, 2), mar = c(4, 4, 1, 1))
R> plot(x1, y, col = col)
```

Inference for Derived Quantities

```
R> abline(h = 1, lty = 2)
R> plot(x1, Probs, col = col)
```





Bayesian Information Criteria

- Bayesian information criterion (BIC)

$$\mathrm{BIC}(M_l) = -2\ell(\hat{\theta}_l) + \log(n)\mathrm{df}_l$$

where df_l is the number of parameters in model l.

Bayesian Information Criteria

- Deviance information criterion (DIC)

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + \text{pd}_{\text{DIC}}$$

where

$$\overline{D(\boldsymbol{\theta})} = -2\log(p(\mathbf{y}|\boldsymbol{\theta})) = \frac{1}{T}\sum_{t=1}^{T} D(\boldsymbol{\theta}^{[t]})$$

denotes the model deviance and

$$\text{pd}_{\text{DIC}} = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T} D(\boldsymbol{\theta}^{[t]}) - D\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\theta}^{[t]}\right)$$

provides an estimate for the effective parameter count.

Bayesian Information Criteria

- Widely applicable information criterion (WAIC)

$$\text{WAIC} = 2\left(D_{\text{WAIC}} + \text{p}_{\text{WAIC}}\right) \quad \text{with} \quad D_{\text{WAIC}} = -\sum_{i=1}^{n} \log\left(\frac{1}{T}\sum_{t=1}^{T} p\left(y_i|\boldsymbol{\theta}^{[t]}\right)\right)$$

as the measure of model fit,

$$\text{p}_{\text{WAIC}} = \sum_{i=1}^{n} \widehat{\text{Var}}\left(\log\left(p(y_i|\boldsymbol{\theta})\right)\right)$$

as the measure of model complexity, and the empirical variance

$$\widehat{\text{Var}}(a) = \frac{1}{T-1}\sum_{t=1}^{T}\left(a_t - \bar{a}\right)^2 .$$

Bayesian Information Criteria

Example in R:

```
R> library("bamlss")
R> data("cars")
```

Two models, simple linear and a polynomial model

```
R> m1 <- bamlss(dist ~ speed, data = cars)
R> m2 <- bamlss(dist ~ poly(speed, 3), data = cars)
```

Bayesian Information Criteria

Compute information criteria.

```
R> BIC(m1, m2)
         df      BIC
m1 3.077717 428.3615
m2 5.014228 434.6961
R> DIC(m1, m2)
        DIC       pd
m1 419.3991 3.077717
m2 420.0945 5.014228
R> WAIC(m1, m2)
      WAIC1    WAIC2       p1       p2
m1 419.4495 420.1763 3.128132 3.491537
m2 419.5214 420.6983 4.441135 5.029594
```