Mathematical Statistics

Anna Janicka

Lecture XV, 3.06.2019

BAYESIAN STATISTICS – CONT.

Plan for Today

1. Bayesian Statistics
   - a priori and a posteriori distributions
   - Bayesian estimation:
     - Maximum a posteriori probability (MAP)
     - Bayes Estimator

Bayesian Model – reminder

- $X_1, \ldots, X_n$ come from distribution P_θ, with density $f_\theta(\mathbf{x})$ – conditional density given a specific value of θ
- P – family of probability distributions P_θ
- General knowledge: distribution Π over the parameter space Θ, given by $\pi(\theta)$ (**prior**)
- Additional knowledge (specific): data values

$$f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, x_2, \ldots, x_n \mid \theta)\pi(\theta)$$

- Based on this distribution, we find the conditional

$$\pi(\theta \mid x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n \mid \theta)\pi(\theta)}{m(x_1, \ldots, x_n)},$$

$$m(x_1, \ldots, x_n) = \int_\Theta f(x_1, \ldots, x_n \mid \theta)\pi(\theta)d\theta$$

Bayesian Model – posterior distribution – reminder

$\pi(\theta \mid x_1,\ldots,x_n)$ is called the **a posteriori/ posterior** distribution, denoted Π_x

The posterior distribution reflects all knowledge: general (initial) and specific (based on the observed data).

Grounds for Bayesian inference and modeling

A priori and a posteriori distributions: examples (reminder)

1. Let $X_1, ..., X_n$ be IID r.v. from a 0-1 distr. with prob. of success θ, let

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

for $\theta \in (0,1)$

Beta(α,β) distr with mean = $\alpha/(\alpha+\beta)$

where $B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

and $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1}\exp(-u)du = (\alpha-1)\Gamma(\alpha-1)$

then the posterior distribution:

$$\text{Beta}\left(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta\right)$$

conjugate prior for Bernoulli distr.

A priori and a posteriori distributions: examples (2)

2. Let $X_1, \ldots, X_n$ be IID r.v. from $N(\theta, \sigma^2)$, and σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the posterior distribution for θ:

$$N\left(\frac{n\frac{1}{\sigma^2}\overline{X} + \frac{1}{\tau^2}m}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2} + \frac{1}{\tau^2}} \right)$$

conjugate prior for a normal distr.

Bayesian Statistics

Based on the Bayes approach, we can

- find estimates
- find an equivalent of confidence intervals
- verify hypotheses

- make predictions

Bayesian Most Probabale (BMP) / Maximum a posteriori Probability (MAP) estimate

Similar to ML estimation: the argument which maximizes the posterior distribution:

$$\pi(\hat{\theta}_{BMP} \mid x_1,\dots,x_n) = \max_\theta \pi(\theta \mid x_1,\dots,x_n)$$

i.e.

$$BMP(\theta) = \hat{\theta}_{BMP} = \operatorname{argmax}_\theta \pi(\theta \mid x_1,\dots,x_n)$$

BMP: examples

1. Let $X_1, ..., X_n$ be IID r.v. from a Bernoulli distr. with prob. of success θ; for $\theta \in (0,1)$

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

Beta(α,β) distr; the mode of this distr = (α-1)/(α+ β-2) for α>1, β>1

We know the posterior distribution:

$$\text{Beta}\left(\sum_{i=1}^{n} x_i + \alpha, n - \sum_{i=1}^{n} x_i + \beta\right)$$

we have max for

$$BMP(\theta) = \frac{\sum_{i=1}^{n} x_i + \alpha - 1}{n + \beta + \alpha - 2}$$

i.e. for 5 successes in 10 trials for a U(0,1) prior (i.e. Beta(1,1) distr.) we have *BMP*(θ)=5/10 = ½; if the prior were Beta(5,5) then *BMP*(θ)=9/18 = ½; if the prior were Beta(1,5) then *BMP*(θ)=5/14

and for 9 successes in 10 trials for a U(0,1) prior we have *BMP*(θ)=9/10; if the prior were Beta(5,5) then *BMP*(θ)=13/18; if the prior were Beta(1,5) then *BMP*(θ)=9/14

BMP: examples (2)

2. Let $X_1, ..., X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the posterior distr. for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$

so

$$BMP(\theta)=\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}$$

i.e. if we have a sample of 5 obs. 1.2; 1.7 ; 1.9 ; 2.1; 3.1 from distr. N(θ, 4) and the prior distr. is $\theta \sim N(1, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1)/(5/4 + 1) = 14/9 \approx 1.56$$

and if the prior distr. were $\theta \sim N(3, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1*3)/(5/4 + 1) = 22/9 \approx 2.44$$

Bayes Estimator

An estimation rule which minimizes the posterior expected value of a loss function

$L(\theta, a)$ – **loss function,** depends on the true value of θ and the decision a.

e.g. if we want to estimate $g(\theta)$:

$L(\theta, a) = (g(\theta) - a)^2$ – quadratic loss function

$L(\theta, a) = |g(\theta) - a|$ – module loss function

Bayes Estimator – cont.

We can also define the **accuracy of an estimate** for a given loss function :

$$acc(\Pi, \hat{g}(x)) = E\left(L(\theta, \hat{g}(x)) \mid X = x\right) = \int_{\Theta} L(\theta, \hat{g}(x)) \pi(\theta \mid x) d\theta$$

(the average loss of the estimator for a given a priori distribution and data, i.e. for a specific posterior distribution)

Bayes Estimator – cont. (2)

The **Bayes Estimator** for a given loss function $L(\theta, a)$ is $\hat{g}_B$ such that

$$\forall x \quad acc(\Pi, \hat{g}_B(x)) = \min_a acc(\Pi, a)$$

For a quadratic loss function $(\theta - \text{a})^2$:

$$\hat{\theta}_B = E(\theta \mid X = x) = E(\Pi_x)$$

For a module loss function $|\theta - \text{a}|^2$:

$$\hat{\theta}_B = Med(\Pi_x)$$

more generally: $E(g(\theta)|x)$

Bayes Estimator: Example (1)

1. Let $X_1, ..., X_n$ be IID r.v. from a Bernoulli distr. with prob. of success θ; for $\theta \in (0,1)$

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

Beta(α,β) distr with mean = $\alpha/(\alpha+\beta)$

We know the posterior distribution:

$$\text{Beta}\left(\sum_{i=1}^{n} x_i + \alpha, n - \sum_{i=1}^{n} x_i + \beta\right)$$

so the Bayes Estimator is

$$\hat{\theta}_B = \frac{\sum_{i=1}^{n} x_i + \alpha}{n + \beta + \alpha}$$

i.e. for 5 successes in 10 trials for a U(0,1) prior (i.e. Beta(1,1) distr.) we have $\hat{\theta}_B$ =6/12 = ½; if the prior were Beta(5,5) then $\hat{\theta}_B$ =10/20 = ½; if the prior were Beta(1,5) then $\hat{\theta}_B$ =6/16

and for 9 successes in 10 trials for a U(0,1) prior we have $\hat{\theta}_B$ = 10/12; if the prior were Beta(5,5) then $\hat{\theta}_B$ = 14/20; if the prior were Beta(1,5) then $\hat{\theta}_B$ = 9/16

BMP: examples

1. Let $X_1, \ldots, X_n$ be IID r.v. from a Bernoulli distr. with prob. of success θ; for $\theta \in (0,1)$

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

Beta(α,β) distr; the mode of this distr = (α-1)/(α+ β-2) for α>1, β>1

We know the poster distribution:

$$\text{Beta}\left(\sum_{i=1}^{n} x_i + \alpha, n - \sum_{i=1}^{n} x_i + \beta\right)$$

we have max for

$$BMP(\theta) = \frac{\sum_{i=1}^{n} x_i + \alpha - 1}{n + \beta + \alpha - 2}$$

i.e. for 5 successes in 10 trials for a U(0,1) prior (i.e. Beta(1,1) distr.) we have *BMP*(θ)=5/10 = ½; if the prior were Beta(5,5) then *BMP*(θ)=9/18 = ½; if the prior were Beta(1,5) then *BMP*(θ)=5/14

and for 9 successes in 10 trials for a U(0,1) prior we have *BMP*(θ)=9/10; if the prior were Beta(5,5) then *BMP*(θ)=13/18; if the prior were Beta(1,5) then *BMP*(θ)=9/14

Bayes Estimator: examples (2)

2. Let $X_1, ..., X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the posterior distr. for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$

so

$$\hat{\theta}_B = \frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}$$

i.e. if we have a sample of 5 obs 1.2; 1.7 ; 1.9 ; 2.1; 3.1 from distr. N(θ, 4) and the a priori distr is $\theta \sim N(1, 1)$, then

$$\hat{\theta}_B = (5/4 * 2 + 1)/(5/4 + 1) = 14/9 \approx 1.56$$

and if the a priori distr were $\theta \sim N(3, 1)$, then

$$\hat{\theta}_B = (5/4 * 2 + 1*3)/(5/4 + 1) = 22/9 \approx 2.44$$

BMP: examples (2)

2. Let $X_1, \ldots, X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the a posteriori distr for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$

so

$$BMP(\theta) = \frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}$$

i.e. if we have a sample of 5 obs 1.2; 1.7 ; 1.9 ; 2.1; 3.1 from distr. N(θ, 4) and the a priori distr is $\theta \sim N(1, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1)/(5/4 + 1) = 14/9 \approx 1.56$$

and if the a priori distr were $\theta \sim N(3, 1)$, then

$$BMP(\theta) = (5/4 * 2 + 1*3)/(5/4 + 1) = 22/9 \approx 2.44$$

Example problem

Let 0.38, 0.65, 0.72, 1.00 be independent realizations of a random variable from a uniform distribution over the interval $(0,\theta)$, where $\theta>0$ is an unknown parameter.

We initially assume that θ is uniformly distributed over the interval [1/2, 2].

Find the posterior distribution.

Find the Bayesian most probable estimator.

Find the Bayes estimator for a quadratic loss function.

Find the Bayes estimator for a modulus loss function.

Example exam problem (1)

3. We continue the topic of unemployment duration in small towns; we assume that in general it follows a $\Gamma(k, \lambda)$ distribution, where k is a fixed, known integer, and $\lambda > 0$ is an unknown parameter. Our sample consists of observations $X_1, X_2, \ldots, X_n$.

- Researcher A constructs the Maximum Likelihood estimator for λ. This estimator has the form $\hat{\lambda}_{ML} =$ ...., and if $k = 2$ and the observations are $X_1 = 1, X_2 = 2, X_3 = 3$, the estimate is equal to $\hat{\lambda}_{ML} =$ ....
- Researcher B constructs the Bayesian Most Probable Estimator, based on a prior distribution with density $f(\lambda) = 2e^{-2\lambda}$ for $\lambda > 0$. This estimator has the form $\hat{\lambda}_{BMP} =$ ...., and if $k = 2$ and the observations are $X_1 = 1, X_2 = 2, X_3 = 3$, the estimate is equal to $\hat{\lambda}_{BMP} =$ ....

Hint. The density of a $\Gamma(k, \lambda)$ distribution is equal to $f(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x}$ for $x > 0$, for integer values of k and $\lambda > 0$.

Example exam problem (2)

8. The time spent on a bus stop waiting for the bus (in minutes) has a uniform distribution over the interval $(0, \theta)$, where $\theta > 0$ is an unknown parameter. Two statisticians use the bus; they have observed the following waiting times (denoted. $X_1, \ldots, X_5$):

$$4, 3, 6, 1, 9.$$

- Let us assume that the first statistician wants to verify the null hypothesis that $\theta = 8$ against the alternative that $\theta = 10$ with a test such that the critical region is $\{X_{5:5} > c\}$ for a constant c. For a test with significance level $\alpha = 0.1$, the constant c should be equal to:

  ....

  and the decision based on the observations is to REJECT /NO GROUNDS TO REJECT H_0 (underline the appropriate).
- Let us assume that the second statistician, based on previous experience, supposes that the distribution of θ is given by a density

$$\pi(\theta) = \frac{4}{6}\left(\frac{6}{\theta}\right)^5 \text{ for } \theta > 6,$$

  and zero otherwise. The Bayesian Most Probable estimate of θ, based on this *a priori* distribution and the observed sample, is equal to

  ....

Highest Posterior Density Credible Interval

A 1-α *HPD* (*Highest Posterior Density*) *credible interval (Bayesian Confidence Interval)* for parameter θ is a set A $\subseteq \Theta$ such that

$$A = \{\theta : \pi(\theta \mid \boldsymbol{x}) > k_\alpha\}$$

and

$$\Pi(A \mid \boldsymbol{x}) \geq 1 - \alpha$$

for k_α highest such that the second condition is fulfilled

The HPD credible interval has the intuitive property of inclusion which the frequentist CI does not have

HPD Credible Interval: example

Let $X_1, \ldots, X_n$ be IID r.v. from $N(\theta, \sigma^2)$, with σ^2 known; $\theta \sim N(m, \tau^2)$ for m, τ known.

Then the posterior distr. for θ: $N\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$

so for $\alpha = 0.05$ we get a HPD CI:

$$\left(\frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}} - 1.96\cdot\frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \frac{n\frac{1}{\sigma^2}\overline{X}+\frac{1}{\tau^2}m}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}} + 1.96\cdot\frac{1}{n\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$$

i.e. if we have a sample of 5 obs. from distr. N(θ, 4) with mean 2 and the a priori distr. is $\theta \sim N(1, 1)$, then due to the fact that $u_{0.975} \approx 1.96$ we have an HPD CI:

$$\approx \left(1.83 - 1.96\cdot \tfrac{1}{6}, 1.83 + 1.96\cdot \tfrac{1}{6}\right) = \left(1.50, 2.15\right)$$