Bayesian Computation for Linear Regression and Analysis of Variance Models

Blaza Toman

National Institute of Standards and Technology

USA

Statistical **Linear Models**
and
Guide to the Expression of Uncertainty in Measurement (GUM)

- Annex H.3: Linear Regression
- Annex H.5: Analysis of Variance (ANOVA)
- Both annexes
  1. specify: **no Type B uncertainty**
  2. use **frequentist** statistical methods

Frequentist and Bayesian Statistics

Frequentists: define probability of an event as a **relative frequency** of its occurrence. Employ the idea of repeated identical experiments.

Bayesians: define probability of an event as a (possibly subjective) **measure of likelihood** of its occurrence. Does not use repeated identical experiments.

Type A and Type B uncertainty

Type A uncertainties are: “evaluated from the statistical distribution of the results of series of measurements and can be characterized by standard deviations”

Type B uncertainties are: “evaluated from assumed probability distributions based on experience or other information”

EURACHEM/CITAC

Quantifying Uncertainty in Analytical Measurement.

Type A and Type B uncertainties

Type A – mostly obtained by frequentist methods

Type B – require the Bayesian interpretation of probability

ANOVA Example

10V Zener voltage standard is calibrated against a stable voltage reference over 2-week period.

v_{jk} is the k^{th} difference measured on day j

Data from the example:

Day	Daily mean $\bar{v}_{j.}$ in Volts	Daily std s_j in μVolts
1	10000.172	60
2	10000.116	77
3	10000.013	111
4	10000.144	101
5	10000.106	67
6	10000.031	93
7	10000.060	80
8	10000.125	73
9	10000.163	88
10	10000.041	86

Motivation for the ANOVA model

It is thought to be likely that there will be *random variation* causing the mean (*measurand*) of the observations *to change* from day-to-day.

Frequentist Model and Analysis
(used in the GUM)

- Assumes that the v_{jk} are realizations of V_{jk} from the model:

$$V_{jk} = \mu + \alpha_j + e_{jk}$$
$$j = 1,\ldots,10$$
$$k = 1,\ldots,5$$

- μ is the measurand,
- α_j are random variables representing the day effects, they have 0 mean and variance σ_α
- e_{jk} are random variables representing within-day variability, they have 0 mean and variance σ

Alternative Frequentist Model Formulation

In hierarchical representation:

$$V_{jk} \mid \theta_j, \sigma^2 \sim N\left(\theta_j, \sigma^2\right)$$

$$\theta_j \mid \mu, \sigma_\alpha^2 \sim N\left(\mu, \sigma_\alpha^2\right)$$

Within-day variability

Between-day variability

$\theta_j \mid \mu, \sigma_\alpha^2$ represents conditioning.

Frequentist Analysis

The *value*, that is the estimate of μ is

$$\bar{v}_{..} = \frac{1}{50}\sum_{j=1}^{10}\sum_{k=1}^{5} v_{jk}$$

The *standard uncertainty* is an estimate of

$$\sqrt{\frac{1}{10}\left(\frac{\sigma^2}{5} + \sigma_\alpha^2\right)}$$

Results

value = 10000.097*V*

standard uncertainty = 18 μ*V*, with 9 degrees of freedom

When no day effect is assumed:

value is the same

standard uncertainty = 13 μ*V*, with 49 degrees of freedom

Add Type B uncertainty

Suppose now that it is necessary to add type B uncertainty in this Example.

That is, that a type B uncertainty for the measurements on day j is given as τ_j.

(For the sake of example, we will assume that 40% of the uncertainty as given in the third column of the table is of type B.)

Bayesian Analysis in general

μ is a random variable with prior distribution *f(μ)*.

Bayes Theorem $$f\left(\mu \mid v\right)=\frac{f\left(v \mid \mu\right)f\left(\mu\right)}{\int f\left(v \mid \mu\right)f\left(\mu\right)d\mu}$$

Used to produce posterior distribution $f\left(\mu \mid v\right)$

Bayesian Analysis - continued

The posterior mean of μ can be used for the *value*.

The posterior standard deviation for the *standard uncertainty*.

Bayesian ANOVA Model with Type B

The experiment in the Example can be represented as:

Likelihood

$$V_{jk} \mid \theta_j, \sigma_j^{\,2} \sim N(\theta_j, \sigma_j^{\,2})$$

Type A within-day

$$\theta_j \mid \delta_j, \tau_j^{\,2} \sim N(\delta_j, \tau_j^{\,2})$$

Type B within-day

$$\delta_j \mid \mu, \sigma_\alpha^2 \sim N(\mu, \sigma_\alpha^2)$$

between-days

Non-informative prior

$$\begin{cases} \sigma_\alpha \sim U\left(0, c_\alpha\right) \\ \mu \sim N\left(m, \omega^2\right) \\ \sigma_j^2 \sim Gamma(a,b) \end{cases}$$

Bayesian ANOVA Model continued

- First level: likelihood as in Frequentist ANOVA
- Second level: type B uncertainty
- Third level: between-days-variability as in Frequentist ANOVA
- Fourth level: prior knowledge about μ, σ_α and σ_i . Usually let c_α, ω, a and b be large to represent no prior knowledge.

Analysis

No closed form solution for:

$$f(\mu \mid \underline{v}) = \frac{\prod_j f(\underline{v}_j \mid \theta_j, \sigma_j) f(\theta_j \mid \delta_j, \tau_j) f(\delta_j \mid \mu, \sigma_\alpha) f(\mu) f(\sigma_\alpha) f(\sigma_j)}{\int \ldots \int \prod_j f(\underline{v}_j \mid \theta_j, \sigma_j) f(\theta_j \mid \delta_j, \tau_j) f(\delta_j \mid \mu, \sigma_\alpha) f(\mu) f(\sigma_\alpha) f(\sigma_j) d\mu \, d\sigma_\alpha d\sigma_j}$$

Use Markov Chain Monte Carlo methods.

Gibbs Sampler

Basic idea:

Given: random variables X and Y

probability distributions f(X|Y) and f(Y|X)

Algorithm:

1. For a given staring x_1, generate y_1 from f(Y|X)
2. For y_1, generate x_2 from f(X|Y)
3. Continue….
4. For large enough sample, under certain regularity conditions on f(X|Y) and f(Y|X) we obtain random draws from the joint distribution of X and Y.

Gibbs Sampler

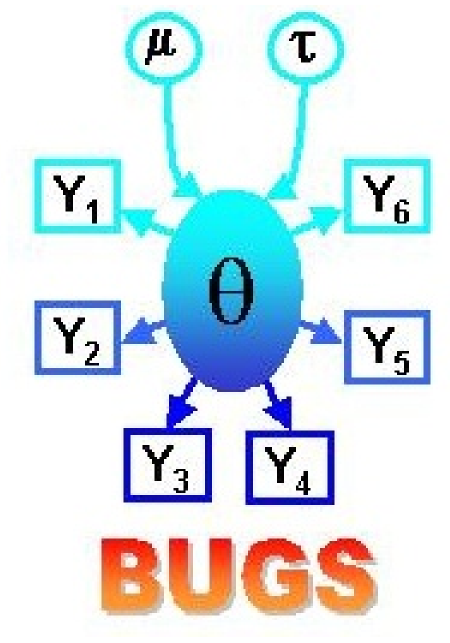
Key References:

S. Geman and D. Geman (1984) IEEE Trans. On Pattern Recognition and Machine Intelligence, 6, 721-741

Casella and George (1992) Am. Statistician, 46, 167-174

Programming

Use free software WinBUGS



www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml

Computation for the EXAMPLE

The data is given in terms of the summary statistics: $\bar{v}_j$ and s_j.

The model needs to be *modified* to:

$$\bar{V}_j \mid \theta_j, \sigma_j^{\,2} \sim N(\theta_j, \frac{\sigma_j^{\,2}}{k})$$

$$\theta_j \mid \delta_j, \tau_j^{\,2} \sim N(\delta_j, \tau_j^{\,2})$$

$$\delta_j \mid \mu, \sigma_\alpha^2 \sim N(\mu, \sigma_\alpha^2)$$

$$(k-1)s_j^2 \sim Gamma\left(\frac{(k-1)}{2}, \frac{1}{2\sigma_j^2}\right)$$

Computation in WinBUGS

Need: 1. Program script

2. Data file

3. Initial input values

WinBUGS Program for the EXAMPLE

```
{ mu~dnorm(0,1.0E-4)
  sigm~ dunif(0,1)
  sigpm<-1/(sigm*sigm)
```

} $f(\mu), f(\sigma_{\alpha})$

```
        for(i in 1:n){
            typ[i]<-1/(tyb[i])
            theta[i]~dnorm(delta[i],typ[i])
            delta[i]~dnorm(mu,sigpm)
            v[i]~dnorm(theta[i],sg[i])}
```

} Modeling of the mean

```
for(i in 1:n){ sig[I]~dunif(0,0.1)
            sg[I]<-1/(sig[I]*sig[I])
            df[i]<-(sampl[i]-1)/2.
            ssq[i]<-(sampl[i]-1)*s[i]
            sampl[i]<-5
             pg[i]<-sg[i]/2
            ssq[i]~dgamma(df[i],pg[i])}
```

} Modeling of the s_i^2

```
For(I in 1:n){ vp[I]~dnorm(theta[I],sg[I])
            dif[I]<-(vp[I]-v[I])
            p[I]<-step(dif[I])}}
```

} Predictive p-values

Data File

```
list(s=c(0.00216,0.0036,0.0074,0.0061,0.002,
  0.0052,0.0038,0.0032,0.0047,0.0044),
  tyb=c(0.0014,0.0024,0.0049,0.0041,0.0018,
  0.0035,0.0026,0.0021,0.0031,0.00295),
  v=c(10000.172,10000.116,10000.013,
  10000.144,10000.106,10000.031,10000.06,
  10000.125,10000.163,10000.041),n=10)
```

Initial input values

list(mu=0, sigm=.1,sig=c(0.002, 0.002,0.002,0.002,0.002,0.002,0.002,0.002, 0.002,0.002))

Results for the EXAMPLE

Posterior mean of μ

value : 10000.104*V*

Posterior standard deviation of μ

standard uncertainty : 23 μ*V*.

(Compare to:

value = 10000.097*V*

standard uncertainty = 18 μ*V*, with 9 degrees of freedom)

Conclusions

- Linear Models can be used with Type B uncertainty
- Requires state-of-knowledge distributions and Bayesian statistics
- Computations can be done using MCMC computation in WinBUGS.