BAYESIAN MODELING USING SAS MCMC PROCEDURE

03rd Dec2016- PhUSE SDE Mumbai, India

Saidulu Aithagoni

PAREXEL®

YOUR JOURNEY. OUR MISSION.™

Agenda

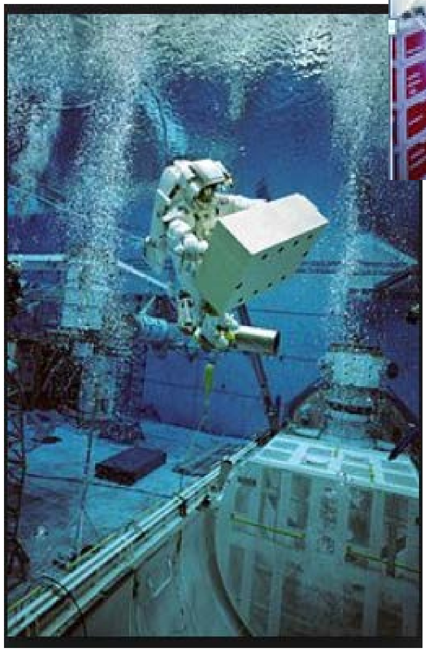
- Introduction
- Simulation & Modeling
- Frequentist approaches
- Bayes Rule (Formula)
- Bayesian Model
- Bayesian Model as a Option & Procedure
- PROC MCMC
- Pros & Cons of Bayesian
- References
- Q & A’s

Introduction

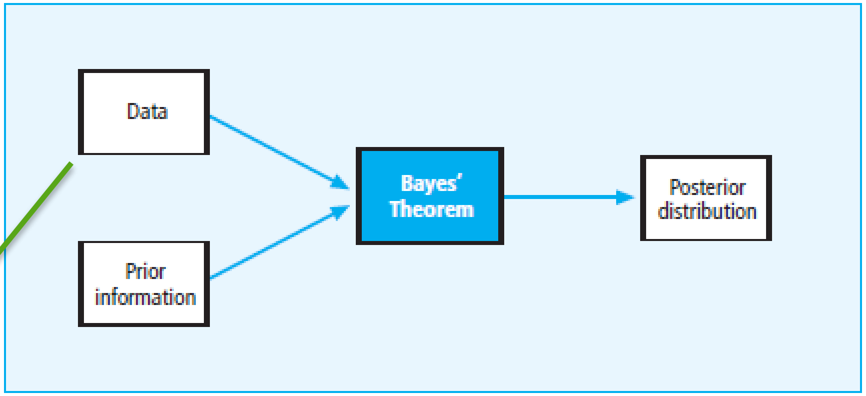
- The term Bayesian refers to Thomas Bayes (1702–1761), who proved a special case of what is now called Bayes theorem.
- However, it was Simon Laplace (1749–1827) who introduced a general version of the theorem and used it to approach problems in mechanics, medical statistics, and reliability.
- In the 1980s, there was a dramatic growth in research and applications of Bayesian methods, mostly attributed to the discovery of **Markov chain Monte Carlo simulation methods**, which removed many of the computational problems, and an increasing interest in nonstandard, complex applications.
- Bayesian methods are widely accepted and used in CT.

Simulation

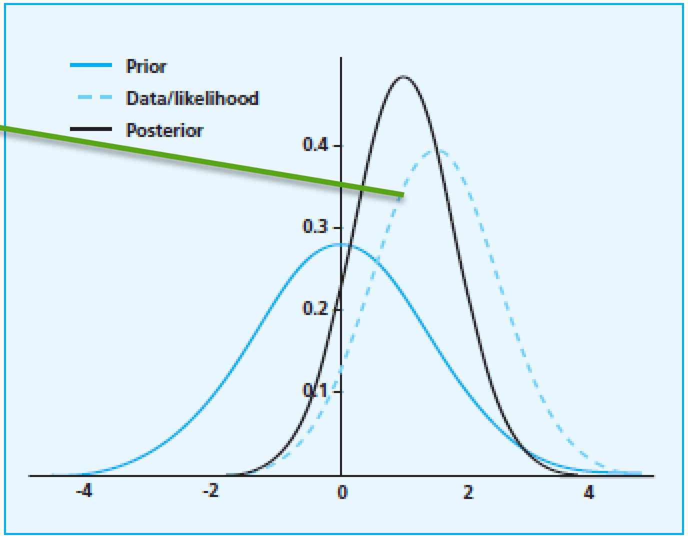
- **Simulation** is the imitation of the operation of a real-world process or system over time. The act of simulating something first requires that a model to be developed; this model represents the key characteristics or behaviors/functions of the selected physical or abstract system or process.



Frequentist approaches



Frequentist statistics



Bayes Rule (Formula)

posterior density of θ given x

sampling density of x given θ

$$p(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{m(x)}$$

prior density for θ

marginal density of x

Bayesian Model

- Bayesian inference is a *method of statistical inference* in which evidence is used to update the uncertainties of *competing probability models*
- It provides uncertainty quantification of parameters by its conditional distribution in the minimal availability of data.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- P(A) is the prior probability of event A.
- P(B|A) is the conditional probability of event B given event A.
- P(B) is the prior or marginal probability of event B.
- P(A|B) is the conditional probability of event A given event B. It is called the posterior probability because it is derived from the specified value of event B.

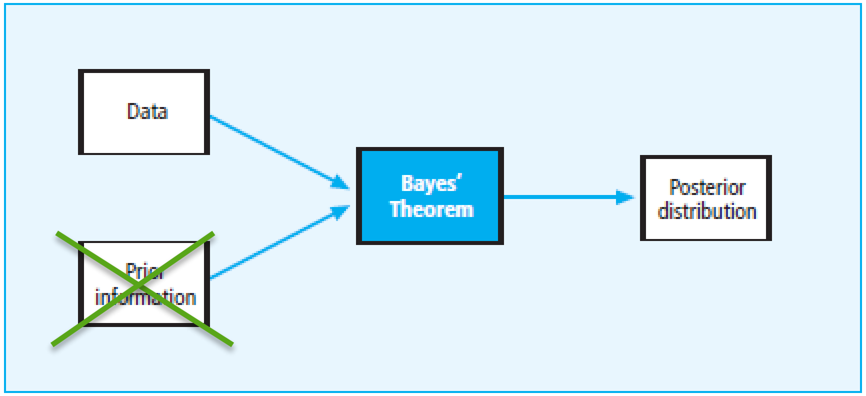
Bayesian Model (Conti. 1)

- The Bayesian approach to statistical inference treats parameters as random variables.
- It includes the incorporation of prior knowledge and its uncertainty in making inferences on unknown quantities (model parameters, missing data, and so on).
- The probability distribution of the parameter, known as the *prior* distribution, is formulated.
- Given the observed data, you choose a statistical model that describes the distribution of the data given the parameters.
- By combining prior distribution and the observed data we are going to calculate the posterior distribution.
- This is carried out by using Bayes' theorem; hence the term Bayesian analysis.

Bayesian- Non-informative & Conjugate priors

- There is no competition with the prior ⟹ the posterior will just be the likelihood.
- This is treated as Non-informative

  or flat prior



- A posterior distribution is not necessarily a normal, binomial, … distribution
- Ex: normal prior for μ + normal data a will give you normal posterior for μ
- If prior & posterior are from same family of distribution then we can call it as conjugate distributions; prior is called conjugate prior.

Bayesian as a Option & Procedure

- Bayesian as an option in general SAS Procedures
  - PROC MIXED
  - PROC GENMOD
  - PROC PHREG
  - PROC LIFEREG

- Bayesian as a SAS Procedure
  - PROC MCMC (Markov chain Monte Carlo)

  This procedure is used to performs the posterior sampling and statistical inference.

PROC MCMC

- Introduced in SAS/STAT 9.2 & enhanced in 9.3 with additional features.
- PROC MCMC enables you to fit a wide range of statistical models
  - linear, generalized linear, nonlinear models
  - nested or non-nested random-effects models
  - hierarchical models.
  - handle various missing data problems, perform variable selection, and make predictions.
- Syntax

```
PROC MCMC options;
   PARMS;
   PRIOR;
   Programming statements;
   MODEL;
   RANDOM;
   PREDDIST;
Run;
```

PROC MCMC (Cont. 1)

```
data xxx;
 input height weight;
  datalines;
   49.0 122.5
   59.5 74.0
     ........
   ;
Run;

proc mcmc data=SasHelp.class seed=1 nbi=5000 nmc=10000 outpost=regOut;
    parms beta0 beta1 sigma2;
    prior beta: ~ normal(0, var=100);
    prior sigma2 ~ igamma(shape=2, scale=2);
    mu = beta0+beta1*height;
    model weight ~ normal(mu, var=sigma2);
run;
```

- **proc mcmc data=xxx seed=1 nbi=5000 nmc=10000 outpost=regOut;**
- The SEED option sets the random number seed to ensure Markov chain reproducibility.

PROC MCMC (Cont. 2)

- The NBI= option specifies the number of burn-in iterations, and the NMC= option controls the number of Markov chain iterations after burn-in. The OUTPOST= option saves all posterior samples to the output data set regOut.
- The MODEL statement specifies the likelihood function for the response variable **weight**.
- model weight ~ **normal** (mu, var=sigma2);
- **mu = beta0+beta1*height; =>** programming statement to define the regression mean, which is a function of the regression coefficients β_0 and β_1
- PARMS statement to tell PROC MCMC what names you have created for the model parameters (in this example, beta0, beta1, sigma2)
- **parms beta0 beta1 sigma2;**

PROC MCMC (Cont. 3)

- You can use multiple PARMS statements, where each statement defines a block of parameters that are updated sequentially in each iteration. Specifying multiple PARMS statement is convenient if you are working with large number of parameters.
- next step is to specify the prior distributions for the model parameters in the PRIOR statements

- **prior beta: ~ normal(0, var=100);**
- **prior sigma2 ~ igamma(shape=2, scale=2);**
- **Beta co-efficients: means** beta0 and beta1 in this model follows normal distribution with mean 0 and variance 100 as prior distribution.
- Variance parameter sigma2 follows inverse gamma as prior distribution.
- Mu must be appear before the model statement.

PROC MCMC (Cont. 4)

Model distribution	Parameter	Prior Distribution
Normal with known μ	σ^2	Inverse gamma
Normal with known μ	τ	Gamma
Normal with known scale (σ^2 or $\tau = 1/\sigma^2$)	μ	Normal
Multivariate normal with known Σ	$\boldsymbol{\mu}$	Multivariate normal
Multivariate normal with known $\boldsymbol{\mu}$	Σ	Inverse Wishart
Multinomial	$\mathbf{p}$	Drichelet

PROC MCMC (Cont. 5)

Posterior Summaries and Intervals					
Parameter	N	Mean	Standard Deviation	95% HPD Interval	
beta0	10000	-8.4939	9.5277	-28.1827	9.1692
beta1	10000	1.7539	0.1607	1.4414	2.0734
sigma2	10000	217.5	73.8757	102.4	362.6

- These results are came from the simulation model with posterior distribution normal.
- Interpretation of the results: on an average, a one-unit increase in height causes a 1.7539-unit increase in weight, and that the probability that the slope coefficient (1) lies between 1.4414 and 2.0704 with 95% CI.

Pros

- Bayesian analysis is useful when you have prior information, either expert opinion or historical knowledge, that you want to incorporate into the analysis.
- It is useful if you want to communicate your findings in terms of probability notions that can be more easily understood by non-statisticians.
- The simulations make the computations tractable even for complex hierarchical models.
- Combine prior info with data & provide exact estimate w/o reliance on large sample size
- Can be applied to a wide range of models, e.g. hierarchical models, missing data...
- Now a day its widely used in CT(proof of concept trials) since we have sufficient prior information, it would minimize the cost of CT & time.

Cons

- It does not tell you how to select a prior and there is no one correct way to choose a prior…
  - **Bayesian inferences require skills to translate subjective prior beliefs into a mathematically formulated prior**. If you do not proceed with caution, you can generate misleading results.
- It can produce posterior distributions that are heavily influenced by the priors (informative prior+ small data size).
- It often comes with a high computational cost, especially in models with a large number of parameters.
- Simulation provide slightly different answers
- No guarantee of Markov Chain Mote Carlo (MCMC) convergence

References

- http://www.lexjansen.com/
- http://support.sas.com
- https://support.sas.com/resources/papers/proceedings09/257-2009.pdf

Q & A's