Bayesískar aðferðir

Málstofa Viðskipta- og hagfræðideildar H.Í.

Helgi Tómasson

10-03-2004

Thomas Bayes

From Wikipedia, the free encyclopedia.

Thomas Bayes (c. 1702-April 7, 1761) was a British mathematician and Presbyterian minister, known for having formulated a special case of Bayes' theorem. Bayes was elected Fellow of the Royal Society in 1742.

Born in London, England, Bayes died in Tunbridge Wells, Kent. He is interred in Bunhill Fields Cemetery in London, where many Nonconformists are buried.

Table of contents

Works by Thomas Bayes

Bayes is known to have published two works in his lifetime: *Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures* (1731), and *An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of the Analyst* (published anonymously in 1736), in which he defended the logical foundation of Isaac Newton's calculus against the criticism of George Berkeley, author of *The Analyst*. It is speculated that Bayes was elected to the Royal Society on the strength of the *Introduction to the Doctrine of Fluxions*, as he is not known to have published any other mathematical works during his lifetime.

Bayes' solution to a problem of "inverse probability" was presented in the *Essay Towards Solving a Problem in the Doctrine of Chances* (1763), published posthumously by his friend Richard Price in the *Philosophical Transactions of the Royal Society of London.* This essay contains a statement of a special case of Bayes' theorem.

In the first decades of the eighteenth century, many problems concerning the probability of certain events, given specified conditions, were solved. For example, given a specified number of white and black balls in an urn, what is the probability of drawing a black ball? These are sometimes called "forward probability" problems. Attention soon turned to the converse of such a problem: given that one or more balls has been drawn, what can be said about the number of white and black balls in the urn? The *Essay* of Bayes contains his solution to a similar problem, posed by Abraham de Moivre, author of *The Doctrine of*

Yfirlit

- Markmið: Áróður fyrir hugmyndafræði og tækni
- Bakgrunnur: Ályktunarfræði í tölfræði
- Bayes mat í einföldu líkani
- Tæknileg vandamál
- Einfalt tölvudæmi, BUGS og CODA forritin
- Lokaorð

Bakgrunnur

- Regla Bayes:

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

- Líkindfræði er stærðfræði, með skilgreiningum, setningum sönnunum, sýnidæmum. Atburður, líkur, hending(random-variable), dreifing, Ýmsar dreifingar þykja tengjast raunverulegum fyrirbærum og verið heiðraðar með sérstöku nafni, normal, poisson, Bernoulli ,gamma, beta, o.s.frv.

- Hagnýtingar á líkindafræði: Ályktunarfræði (Inference) og ákvörðunarfræði (decision theory).

- Tölfræði = Ályktunarfræði (Inference) + Líkindafræði

- Ýmsir metríu-kúltúrar = Econometrics, biometrics, pschycometrics, sociometrics, chemometrics, technometrics, criminometrics o.s.frv.

- Ályktanir út frá mælingum krefjast þess að líkur séu túlkaðar, þ.e. hvað þýðir $P(A) = 1/2$?

- Túlkun 1, tíðniskóli (frequentistic): Mælikvarði á tíðni atburða.

- Túlkun 2, Bayes-skóli: Mælikvarði á vissu.

- Athugið ályktun út frá mælingum er alltaf í tengslum við líkan:

  $\boldsymbol{x}|\boldsymbol{\theta} \sim \pi(\boldsymbol{x}|\boldsymbol{\theta})$

  þar sem

  $\boldsymbol{x} = (x_1, x_2, \dots, x_n)'$ er vektor af mælingum

  $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ er líkindadreifing, sem stýrt er af stikanum $\boldsymbol{\theta}$

  Til dæmis getum verið að tala um slembiúrtak (random sample) úr normaldreifingu þar sem

  $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ er $N(\boldsymbol{\mu}, \Sigma)$

  $\boldsymbol{\theta} = (\mu, \sigma^2)$ og $\boldsymbol{\mu} = (\mu, \dots, \mu)$ og

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & 1 \end{bmatrix}$$

- Frequentistar líta á $\boldsymbol{\theta}$ sem fasta stærð, í praxís óþekkta, sem ekki hafi neina líkindadreifningu, því endurtekning á tilrauninni myndi ekki breyta $\boldsymbol{\theta}$.

- Bayesíanar líta óþekktan stika, $\boldsymbol{\theta}$ sem hendingu og óvissuna um þann stika skuli setja fram í formi líkindadreifingar.

- Frequentisti setur fram líkan

  $\boldsymbol{X}|\boldsymbol{\theta}$ safnar gögnum $\boldsymbol{x}$ og fær mat ML,MM,OLS ..

  $\hat{\boldsymbol{\theta}}$

- Síðan ályktar hann um $\boldsymbol{\theta}$ út frá $\hat{\boldsymbol{\theta}}$, tjáir sig um nákvæmni með öryggismörkum (confidence interval) og prófar kenningar um $\boldsymbol{\theta}$ með LR (likelihood-ratio), Wald eða LM (Lagrange-multiplier, score) prófum.

- Vandi að túlka p-gildi fyrir frequentista. Neyman og Pearson deildu við Fischer um þau atriði milli 1930 og 1940. Sjá t.d. Spanos og Berger. H_0 hafnað vegna þess að ef H_0 er rétt þá þykir mér mæld útkoma of ótrúleg.

- Bayesíanar vinna allt öðruvísi. Áður en gögnum er safnað er sett fram skoðun á stikanum $\boldsymbol{\theta}$. Þessi skoðun er sett fram í formi líkindadreifingar,

  $\pi(\boldsymbol{\theta})$ a priori dreifningin

  líkanið er $\boldsymbol{X}|\boldsymbol{\theta} \sim \pi(\boldsymbol{x}|\boldsymbol{\theta})$

  gögnum $\boldsymbol{x}$ safnað

  og reiknuð út a posteriori dreifiningin með reglu Bayes

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})}{\int \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{\theta}}$$

- Síðan má skilgreina bayes mat út frá posteriori dreifingunni, t.d:

$$\hat{\boldsymbol{\theta}}_{BAYES} = E_{\pi(\boldsymbol{\theta}|\boldsymbol{x})}(\boldsymbol{\theta})$$

  eða reikna miðtölu, mode eða eitthvað annað út frá posteriori dreifingunni.

- Athugið að nákvæmnina meta Bayesianar með „highest-probability” region fyrir $\boldsymbol{\theta}$ og ályktað eru um kenningar með því að reikna „posterior odds”. Þ.e. líkur eru notaðar beint í ályktuninni.

- Upplagt er að koma sinni hlutdrægni á framfæri með vali á priori dreifingu.

- Hvað á að gera ef lítið er vitað um $\boldsymbol{\theta}$? Hvernig á setja slíkt fram á formi líkindadreifngar?

- Í praxís vel flatan prior. Hugtakið „non-informative-prior” hefur valdið Bayesiönum heilabrotum. Ýmsar lausnir, lagðar til svo sem Jeffrey's prior, minimal-Shannon-information. Shannon-information er skilgreint:

$$I_\pi = \int \pi(\boldsymbol{\theta}) \log(\pi(\boldsymbol{\theta})) d\boldsymbol{\theta}$$

  Finn það fall π þannig að upplýsingaauki úr tilraun séu sem mestur.

$$I_{\boldsymbol{x}} = \int \pi(\boldsymbol{\theta}|\boldsymbol{x}) \log(\pi(\boldsymbol{\theta}|\boldsymbol{x})) d\boldsymbol{\theta}$$

  hámarka

$$I_{\boldsymbol{x}} - I_\pi$$

- Get valið flatan improper-prior, Bayesianar ekki sammála hér.

  Ath. Ef $\pi(log(\sigma)$ er flatt þá er $\pi(\sigma)$ ekki flatt.

- Í praxís vel eitthvað sem er nálægt því að vera flatt fyrir áhugverð gildi á $\boldsymbol{\theta}$.

- Bayesískar aðferðir geta haft góða frequentistic eiginleika, t.d. lítinn mean-square-error. Maximum-likelihood aðferðir er **asymptotísk optimal** en geta haft leiðinlega eiginleika í litlum úrtökum.

- James-Stein og ridge-regression eru aðferðir, sem auka stöðugleika í mati, hafa verið leiddar út með empirical Bayes aðferðum. Empirical Bayes aðferðir ganga út á að nota Bayes formúlur en stinga inn gildum á parametrum úr prior dreifingunni sem er metin út frá mælingunum.

Hagnýting og tæknileg vandamál

- Áður voru Bayes aðferðir fyrst og fremst heimsspekilegar (og stærðfræðilegar).

- Á síðustu árum hefur þetta breyst og nú er Bayes aðferðafræði orðinn praktískur möguleiki í hagnýtum rannsóknum.

- Tæknilega hefur verið nánast ómögulegt að reikna út integralið í posteriori dreifingunni.

- Það er hægt ef líkanið og priori-dreifingin mynda „conjugate"-par. Þá verða priori dreifingin og posteriori dreifingin bæði úr sömu fjölskyldu. Þ.e. prior-líkan parið er „closed under sampling"

- Dæmi: Vil álykta um meðaltal í normaldreifingu, $N(\mu, \sigma^2)$. Gerum ráð fyrir að σ sé þekkt þá er skref 1 að setja fram prior

  $\pi(\mu) \sim N(a, \tau)$ fæ mælingu x úr

  $N(\mu, \sigma)$ nú má sýna með algebru og reglu Bayes að

  $$\pi(\mu|x) \sim N(x\frac{\tau}{\sigma^2+\tau} + a\frac{\sigma^2}{\sigma^2+\tau}, \frac{1}{1/\sigma^2+1/\tau})$$

  þ.e. a posteriori dreifinging varð ný normaldreifing. Ef σ hefði verið óþekkt hefið mátt nota normal-gamma prior og fá að a posteriori dreifingin er t-dreifing.

- Það hægt að leiða út conjugate prior fyrir exponential family dreifingar. Og þá er hugsanlegt að vinna analýtísk með posteriori dreifinguna. Fyrir aðra a priori dreifingar er þetta yfirleitt erfitt.

- Því verður að grípa til númerískra aðgerða. Það er erfitt að integrera númerískt í mörgum víddum. Til determinstískar reglur, Gaussian quadrature sem ganga í lágum víddum (1,2,3).

- Nútímapraxís byggir á hermunum (simulations). Þ.e. til að meta integralið

  $$\int g(\boldsymbol{\theta}) d\boldsymbol{\theta} \text{ er safnað gildum á } \boldsymbol{\theta})$$

  með random generator og reiknað

  $$\frac{1}{S}\sum_{i=1}^{S} g(\boldsymbol{\theta}_i)$$

- Random sample ekki skynsamlegt. Menn hafa reynt a) importance sampling (svipað og stratified sampling) og Markov-Chain-Monte-Carlo (MCMC) aðferðir. Þekkt prinsíp í MCMC eru Metropolis-Hastings aðferðin og Gibbs sampler.

- Tæknilegur vandi er að búa til random tölur úr a posteriori-dreifingunni. Ef ákveðnum stærðfræðilegum skilyrðum er fullnægt er hægt er að hanna Markov-keðju sem hefur jafnvægisdreifingu sem er sama og posteriori dreifingin.

- Hugmyndin er því sú að setja af stað keðju, eftir ákveðin tíma eru gildi í keðjunni „representatív" fyrir posteriori dreifinguna. Þá er safnað úr keðjunni gögnum þannig að góð hugmynd fæst um dreifinguna.

- Vandi er að vita hvenær er maður kominn nógu langt? Þ.e. hvað þarf burn-in tímabilið að vera langt? Mikilvægt að velja startgildi rétt.

- Þróaðar hafa verið convergence-diagnostic aðferðir.

- Sýni lítið dæmi með notkun á BUGS (Bayesians-Using-Gibbs-Sampler) og CODA (Convergence-Output-Diagnostic-Analysis).

Einfalt dæmi

Einföld aðhvarfsgreining(regression)

x	y
1	1
2	3
3	3
4	3
5	5

Líkanið er

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad \varepsilon \sim N(0, \sigma^2)$$

a priori dreifingar $\alpha \sim N(0, 10000) \qquad \beta \sim N(0, 10000)$

$$\sigma^{-2} \sim gamma(0.00001, 0.00001)$$

Venulegt niðurstaða úr vengulegu tölfræðiforriti er t.d. $\hat{\sigma} = 0.73$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.6000	0.7659	0.78	0.4906
x	0.8000	0.2309	3.46	0.0405

CODA-ouput

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
alpha	0.6719	1.2109	0.017124	0.030980
beta	0.7768	0.3605	0.005099	0.009386
sigma	0.9993	0.6285	0.008889	0.013761
tau	1.8749	1.5220	0.021524	0.028852

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
alpha	-1.672447	0.03057	0.6327	1.225	3.263
beta	.001338	60776	0.7870	0.968	1.468
sigma	0.415499	0.62246	0.8200	1.171	2.649
tau	0.142518	0.72870	1.4874	2.581	5.792

CODA diagnostics options menu

1:Display current diagnostic options

2:Window sizes for Geweke's diagnostic

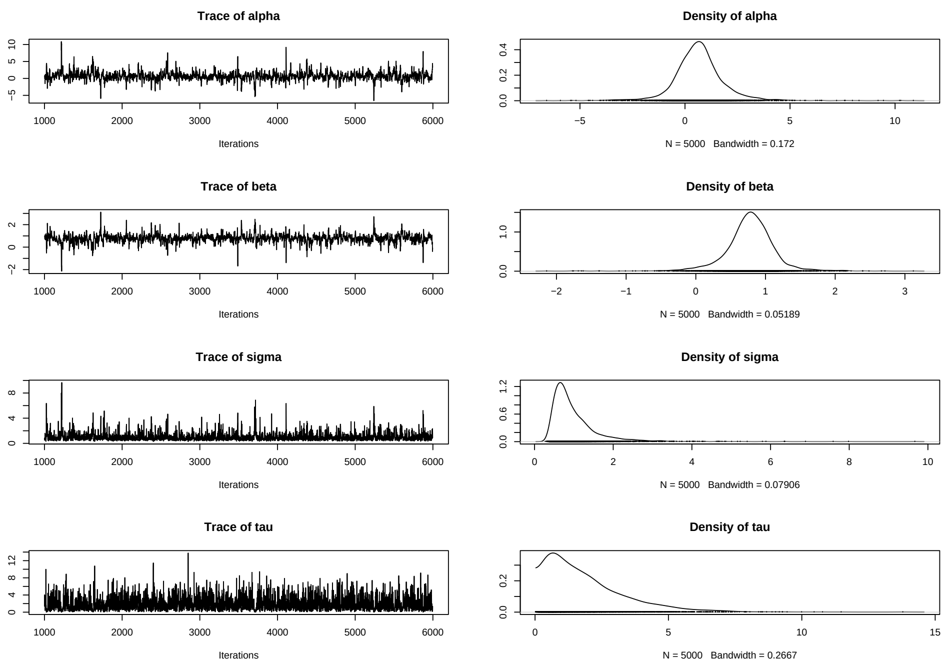
3:Bin size for plotting Geweke's diagnostic

4:Bin size for plotting Gelman & Rubin's diagnostic

5:Parameters for Raftery & Lewis' diagnostic

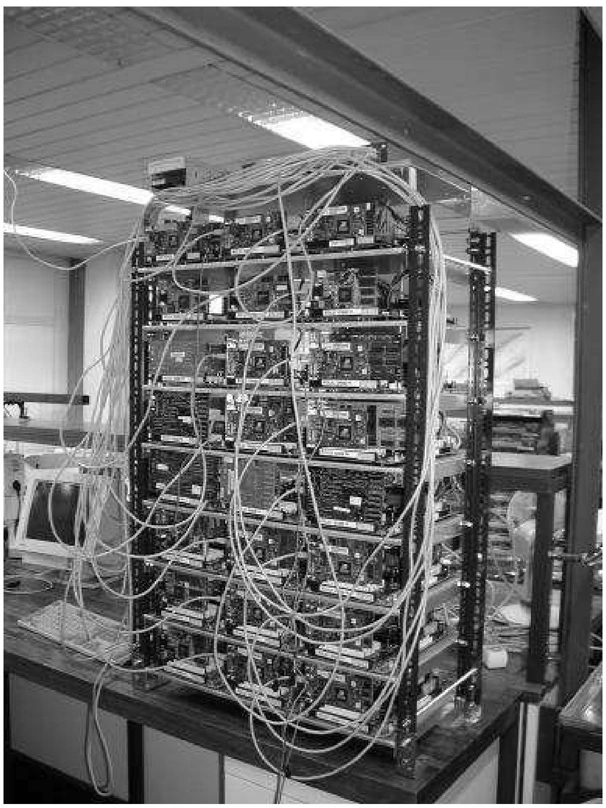
6:Halfwidth precision for Heidelberger & Welch's diagnostic

7:Combine chains to calculate correlation matrix

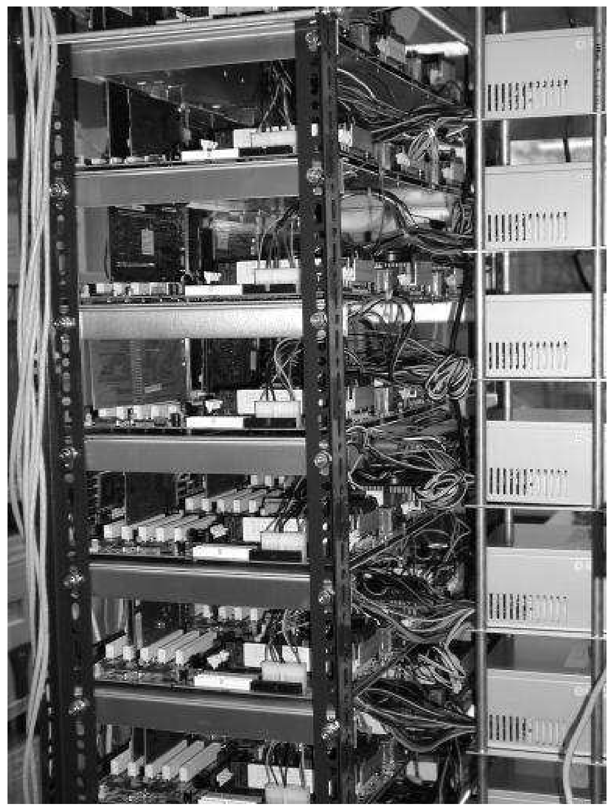


Mynd 1: Útkoma úr CODA

HIP Software and Physics project



The 24 nodes of the Celeron ATX blade server [1].



Here the power supply tower can be seen.

23.10.2003 8 Tomas Lindén

Mynd 2: Linux-cluster í smíðum

1... GENERALISED LINEAR MODELS
1.3. GENERAL PRINCIPLES OF MODELLING

- IT IS ART
- ALL MODELS ARE WRONG
  - SOME OF THEM ARE MORE USEFUL THAN OTHERS
  - WE SEEK FOR MODELS WHICH DESCRIBE REALITY
  - WE FIT AND CHECK MANY DIFFERENT MODELS
- ALWAYS USE SOME DIAGNOSTICS FOR CHECKING THE GOODNESS OF FIT

2... CLASSIC BUGS

- **2.1. INTRODUCTION: WHAT IS BUGS?**
- **2.2. A SIMPLE EXAMPLE**
- **2.3. EXAMPLE 2: BINOMIAL DATA**
- **2.4. EXAMPLE 3: BERNOULLI DATA**
- **2.5 EXAMPLE 4: MODELS FOR 2X2 CONTINGENCY TABLES**
- **2.6 EXAMPLE 5: MODEL FOR 3-WAY TABLES**

2... CLASSIC BUGS
2.1. INTRODUCTION: What is BUGS

- **BUGS**: **B**ayesian inference **U**sing **G**ibbs **S**ampling
- Computing Language for definition of the Model (likelihood, prior)
- Computes the full Conditional Distributions and generates samples from log-concave posterior distributions

2... CLASSIC BUGS
2.1. INTRODUCTION: What is BUGS

- The project started at (about) 1995
- 1998: 1st version of Windows (Winbugs)
- MRC Biostatistics Unit in Cambridge (Spiegelhalter, Gilks, Best, Thomas)
- Now: developed jointly with the Imperial College
- Free from:
  `http://www.mrc-bsu.cam.ac.uk/bugs/`

2... CLASSIC BUGS
2.1. INTRODUCTION: What is BUGS

BEWARE:

GIBBS SAMPLING CAN BE DANGEROUS

2... CLASSIC BUGS
2.1. INTRODUCTION: What is BUGS

Why?

- Wrong Results due to lack of convergence
- Slow convergence
- Bad Starting Values
- Bad Construction of Model
- Over-parameterised Model

Words: B. Carlin

Music: Neil Diamond/The Monkees/Smashmouth (“I’m a Believer”)

Intro: (key/guitar lick)

V1: I thought inference was just a fairy tale,

Confused by stats and probability,

Frequentist approaches (doo-doot doo-doot)

made no sense to me (doo-doot doo-doot)

Summarizing evidence by p?!

Chorus: Then I saw Tom Bayes – Now I’m a believer,

Without a trace – of doubt in my mind,

[I’m a] Bayesian (ooooh) – Oh, I’m a believer –

I couldn’t p now if I tried!

V2: I thought likelihood was just the only thing,

Turn the crank and get the MLE,

What's the use of thinking (doo-doot doo-doot)

Disconnect your brain (doo-doot doo-doot)

Play along and minimize the pain...

Chorus: (repeat)

Solo: (keys/guitar)

V3: `SAS` was out to get me (doo-doot doo-doot)

(partial) – that's the way it seemed (doo-doot doo-doot)

Fixed effects and forced normality...

Chorus: (repeat 2x w/assorted hollering and out!)

Lokaorð

- Bayes hugmyndafræði er frábrugðin tíðintúlkun.
- Oft er hægt að taka tíðnitúlkunarniðurstöður og sjá hvers konar prior sá rannsakandi hefði haft ef hann hefði verið Bayesian.
- Tölvutækni, MCMC aðferðir gera Bayesíska vinnu að praktískum möguleika.
- Með MCMC tækninni er hugsanlega hægt að meta flókin líkön,
- Fræðin um óvissu eru „exact” vísindi.