Climate Change Detection and Attribution: Bayesian view

Slava Kharin
Canadian Centre for Climate Modelling and Analysis,
Environment Canada,
Victoria, British Columbia, Canada

Banff Summer School, May 2008

A Bayesian Climate Change Detection and Attribution Assessment

TERRY C. K. LEE

Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada

FRANCIS W. ZWIERS

Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada, University of Victoria, Victoria, British Columbia, Canada

GABRIELE C. HEGERL

Department of Earth and Ocean Sciences, Duke University, Durham, North Carolina

XUEBIN ZHANG

Climate Monitoring and Data Interpretation Division, Meteorological Service of Canada, Downsview, Ontario, Canada

MIN TSAO

Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada

(Manuscript received 8 July 2004, in final form 2 December 2004)

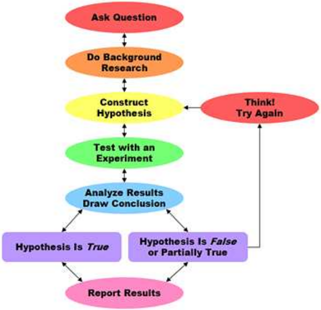
Thomas Bayes (1702-1761)

- British mathematician (University of Edinburgh – logic & theology);
- Formulated a special case of Bayes' theorem
- Bayesian inference has been in use over 200 yrs
- Bayesian view of the probability as "a subjective degree of belief" is more recent.
- The word "Bayesian" is in use since 1950s to distinguish from the "Frequentist" view of probability.

The Scientific Method

In its most basic form, the thought process in the scientific method can consists of the following tasks:

- identify a problem
- form a hypothesis
- design and perform experiments
- collect and analyze data
- formulate conclusions about the hypothesis
- revise the hypothesis and go to step 3



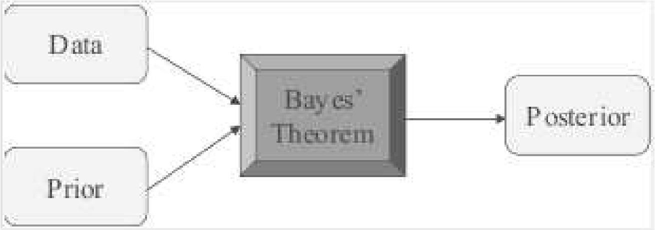
The Bayesian Method

- formulate a problem
- create a statistical model to link data to parameters: $f(x|\theta)$
- formulate prior information about parameters: $f(\theta)$
- collect data (sampling)
- combine the two source of information about θ using Bayes' theorem: $f(\theta|x)$
- use the resulting posterior distribution to derive inferences about parameters
- make decision based on some cost function of θ

The Bayesian Method

- formulate a problem
- create a statistical model to link data to parameters: $f(x|\theta)$
- formulate prior information about parameters: $f(\theta)$
- collect data (sampling)
- combine the two source of information about θ using Bayes' theorem: $f(\theta|x)$
- use the resulting posterior distribution to derive inferences about parameters
- make decision based on some cost function of θ

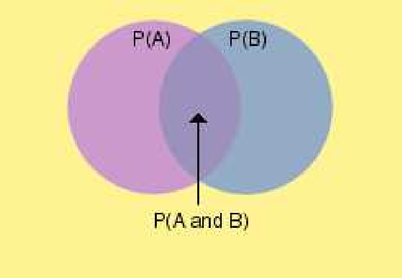
Synthesis of information by Bayes' theorem



Conditional Probability

Conditional probability – probability of an event, based on another event happening.

- $P(B|A)$ - *conditional* probability
- $P(A \cap B)$ - *joint* probability:
- $P(A)$, $P(B)$ *marginal* probabilities



Definition:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, P(A) > 0.$$

Analogously:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0.$$

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

Bayes' Theorem

Let H be a hypothesis to be tested, and
$P(H)$ is the prior probability of H before new evidence E became available. Then

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

$P(H)$ is the prior probability.
$P(E|H)$ is called likelihood.
$P(H|E)$ is the posterior probability.

General notation:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalizing factor}}$$

or

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Bayes' Theorem (continuous distributions)

If a hypothesis is described by a continuous θ, the prior and posterior probabilities can be represented by density functions:

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

$\pi(\theta)$ is the prior density distribution,
$\pi(\theta|x)$ is the posterior density of θ given sample x,
$f(x|\theta)$ is the likelihood function,
$m(x)$ is the marginal (or unconditional) density of x.
Or,

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

Prior information

An example of possible importance of prior information (L.J.Savage 1961):

1. A lady claims to be able to tell whether the tea or milk was poured into cup first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score.
3. A drunken friend says he can predict the outcome of a flip of a fair coin.

An experiment of 10 trials is performed and in each case 10 out of 10 guesses turned to be correct.

Classical statistics: $\theta = \mathrm{Prob(correct)}$, H_0: $\theta = 0.5$

Prob(10 out of 10)=2^{-10}

Prior information

An example of possible importance of prior information (L.J.Savage 1961):

1. A lady claims to be able to tell whether the tea or milk was poured into cup first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score.
3. A drunken friend says he can predict the outcome of a flip of a fair coin.

An experiment of 10 trials is performed and in each case 10 out of 10 guesses turned to be correct.
Classical statistics: $\theta = \mathrm{Prob}(\mathrm{correct})$, H_0: $\theta = 0.5$
Prob(10 out of 10)=2^{-10}

Example: normal distribution, known variance

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

Assume $X \sim \mathcal{N}(\theta, \sigma^2)$ where θ is unknown but σ^2 is known.

- Let $\pi(\theta)$ be $\mathcal{N}(\mu, \tau^2)$ (so-called conjugate family).
- The likelihood function $f(x|\theta)$ is given by

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\theta)^2}{2\sigma^2})$$

- It can be shown that the marginal distribution $m(x)$ is $\mathcal{N}(\mu, \sigma^2 + \tau^2)$, and
- the posterior distribution $\pi(\theta|x)$ is $\mathcal{N}(\mu(x), \rho^{-1})$ where

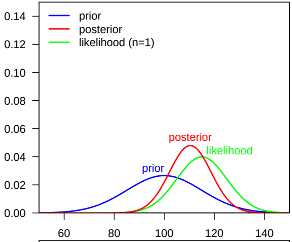
$$\mu(x) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x, \quad \rho = \tau^{-2} + \sigma^{-2}$$

Example: IQ intelligence test

Assume that IQ test result is $\mathcal{N}(\theta, \sigma = 10)$ where θ is the true IQ level of a person (or group of people).

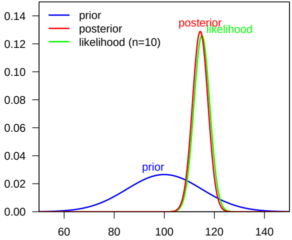
Case 1 (prior st.dev. $\tau = 15$, $n = 1$):

- prior is $\mathcal{N}(\mu = 100, \tau = 15)$
- 1 observation, $x = 115$



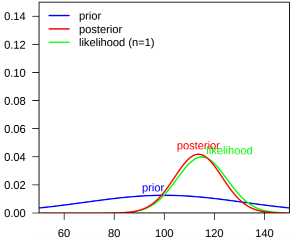
Case 2 (prior st.dev. $\tau = 15$, $n = 10$):

- prior is $\mathcal{N}(\mu = 100, \tau = 15)$
- 10 observations, $\overline{x} = 115$



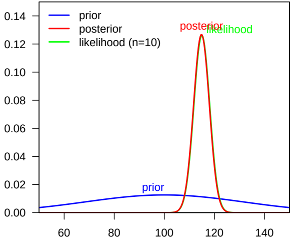
Example: IQ intelligence test – weaker prior beliefs

Assume that IQ test result is $\mathcal{N}(\theta, \sigma = 10)$ where θ is the true IQ level of a person.



Case 3:

- prior is $\mathcal{N}(\mu = 100, \tau = 30)$
- 1 observation, $x = 115$



Case 4:

- prior is $\mathcal{N}(\mu = 100, \tau = 30)$
- 10 observations, $\overline{x} = 115$

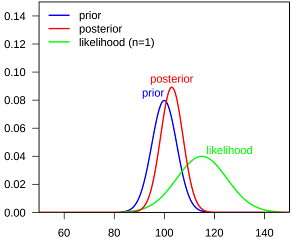
Example: IQ intelligence test – stronger prior beliefs

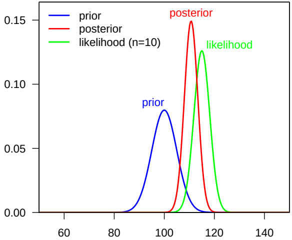
Case 5:

- prior is $\mathcal{N}(\mu = 100, \tau = 5)$
- 1 observation, $x = 115$

Case 6:

- prior is $\mathcal{N}(\mu = 100, \tau = 5)$
- 10 observations, $\overline{x} = 115$





Bayesian Hypothesis Testing

Classical hypothesis testing:

- $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$ are specified.
- A test is evaluated in terms of Type I and Type II errors. (these errors are probabilities to observed a sample for which the test procedure will result in the wrong hypothesis being accepted)

Bayesian hypothesis testing is more straightforward:

- Calculate the posterior probabilities $\alpha_0 = Prob(\Theta_0|x)$ and $\alpha_1 = Prob(\Theta_1|x)$.
- decide between H_0 and H_1 based on α_0 and α_1.

Bayesian Hypothesis Testing

Classical hypothesis testing:

- $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$ are specified.
- A test is evaluated in terms of Type I and Type II errors. (these errors are probabilities to observed a sample for which the test procedure will result in the wrong hypothesis being accepted)

Bayesian hypothesis testing is more straightforward:

- Calculate the posterior probabilities $\alpha_0 = Prob(\Theta_0|x)$ and $\alpha_1 = Prob(\Theta_1|x)$.
- decide between H_0 and H_1 based on α_0 and α_1.

Odds ratio and Bayes factor

It is convenient to summarize the evidence in terms of *"odds"* ratios:

$$\begin{aligned} \text{prior odds ratio} &= \pi_1/\pi_0 \\ \text{posterior odds ratio} &= \alpha_1/\alpha_0 \end{aligned}$$

For example,

- $\alpha_1/\alpha_0 = 10$ means that H_1 is 10 times more likely than H_0.
- If $\pi_1/\pi_0 = 10$ then available data sample didn't change the odds.
- If $\pi_1/\pi_0 < 10$ then available data sample supports H_1.
- If $\pi_1/\pi_0 > 10$ then available data sample supports H_0.

The *Bayes factor* represents the role of the data in changing the odds:

$$B = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\alpha_1/\alpha_0}{\pi_1/\pi_0}$$

Odds ratio and Bayes factor

It is convenient to summarize the evidence in terms of *"odds"* ratios:

$$\begin{aligned} \text{prior odds ratio} &= \pi_1/\pi_0 \\ \text{posterior odds ratio} &= \alpha_1/\alpha_0 \end{aligned}$$

For example,

- $\alpha_1/\alpha_0 = 10$ means that H_1 is 10 times more likely than H_0.
- If $\pi_1/\pi_0 = 10$ then available data sample didn't change the odds.
- If $\pi_1/\pi_0 < 10$ then available data sample supports H_1.
- If $\pi_1/\pi_0 > 10$ then available data sample supports H_0.

The *Bayes factor* represents the role of the data in changing the odds:

$$B = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\alpha_1/\alpha_0}{\pi_1/\pi_0}$$

Odds ratio and Bayes factor

It is convenient to summarize the evidence in terms of *"odds"* ratios:

$$\begin{aligned} \text{prior odds ratio} &= \pi_1/\pi_0 \\ \text{posterior odds ratio} &= \alpha_1/\alpha_0 \end{aligned}$$

For example,

- $\alpha_1/\alpha_0 = 10$ means that H_1 is 10 times more likely than H_0.
- If $\pi_1/\pi_0 = 10$ then available data sample didn't change the odds.
- If $\pi_1/\pi_0 < 10$ then available data sample supports H_1.
- If $\pi_1/\pi_0 > 10$ then available data sample supports H_0.

The *Bayes factor* represents the role of the data in changing the odds:

$$B = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\alpha_1/\alpha_0}{\pi_1/\pi_0}$$

Bayes Factor as likelihood ratio

- In some simple cases ($\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$), the Bayes factor is simply the ratio of likelihoods:

$$B = \frac{f(x|\theta_1)}{f(x|\theta_0)}$$

  (independent of prior odds).
- More generally, the Bayes factor will depend on the priors. However, B is often relatively insensitive to the choices of priors.
- The advantage of reporting the Bayes factor B is that any user could then determine his own posterior odds by simply multiplying B by his personal prior odds:

$$(\pi_1/\pi_0) = B(\alpha_1/\alpha_0)$$

Bayes Factors classifications

Kass and Raftery (1995):

B	Strength of evidence
1:1 to 3:1	Barely worth mentioning
3:1 to 20:1	Positive
20:1 to 150:1	Strong
>150:1	Very Strong

Bayesian Climate Change Assessment (Lee et al. 2004)

Generalized multiple regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

- $\mathbf{y}$ is the filtered version of observed temperature records,
- the matrix $\mathbf{X}$ contains estimated response patterns (*fingerprints*) to the known forcings ("GS"),
- $\boldsymbol{\beta}$ is the unknown scaling parameters,
- $\boldsymbol{\epsilon}$ accounts for errors in the data and natural climate variability (Gaussian noise).

Classical Detection and Attribution tests

Detection of a postulated climate change signal occurs when its amplitude in observations is significantly different from 0:

- The detection null hypothesis:

$$H_D : \boldsymbol{\beta} = \mathbf{0}$$

Rejection of H_D by a one-sided test leads to detection.

Attribution is the process of establishing a cause and effect relationship between the observed change and forcings.

- detection is required;
- elimination of other plausible causes;
- *attribution consistency test*:

$$H_A : \boldsymbol{\beta} = \mathbf{1}$$

Attribution is claimed when H_A *cannot* be rejected.

Classical Detection and Attribution tests

Detection of a postulated climate change signal occurs when its amplitude in observations is significantly different from 0:

- The detection null hypothesis:

$$H_D : \beta = \mathbf{0}$$

Rejection of H_D by a one-sided test leads to detection.

Attribution is the process of establishing a cause and effect relationship between the observed change and forcings.

- detection is required;
- elimination of other plausible causes;
- *attribution consistency test*:

$$H_A : \beta = \mathbf{1}$$

Attribution is claimed when H_A *cannot* be rejected.

Bayesian Detection and Attribution tests

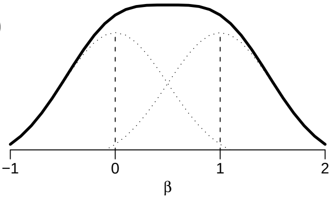
1. Construct prior distribution(s) of β.
   - noniformative priors
   - robust Bayes analysis (test the sensitivity of the results to a wide range of various priors)
   - report results for several alternative priors.
2. Generalized linear regression is performed to obtain an estimate of the signal amplitudes $\widehat{\beta}$ (Hegerl et al., 1997; Allen and Tett 1999; Zwiers and Zhang 2003; Allen et al. 2004)
3. The posterior distribution of β is determined given the estimate of $\widehat{\beta}$.
4. Define detection and attribution criteria and inspect the posterior distribution whether these criteria are met.

The prior distribution for β

- Mixture of normal distributions:

$$\pi_1(\beta) = p\phi(0, \tau_1^2) + (1-p)\phi(1, \tau_2^2)$$

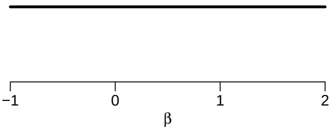
$p = 1/2$, $\tau_1^2 = \tau_2^2 = 1/4$
Equal weights to values at $\beta = 0$ and $\beta = 1$ (noncommittal)
Uncertainty is guessed by using a simple Energy Budget Model.



Prior odds for $\beta \in (0.1 \ldots \infty)$ are $\approx 9:4$.

- Noniformative prior:

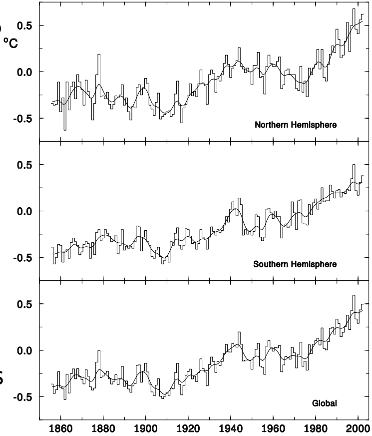
$$\pi_2(\beta) \propto const$$



Prior odds for $\beta \in (0.1 \ldots \infty)$ are $1:1$.

Observed and model data

- Observational data set is the HadCRUT dataset on a 5×5 grid (Jones et al 2001)
- Only one climate forcing is used: combined GHG and the direct effect of sulphate aerosol ("GS" signal.)
- Two models are used: CGCM1 and CGCM2, 6 members.
- Covariance matrix of noise is estimated from 1600 yrs of control simulations



The posterior distribution for β

Reduction of dimensionality of data is an essential step in DA analysis.

- Posterior distribution of β in 1950–99 when 15, 20, 25 EOFs are retained.

(see Lee et al. (2005) for mathematical details.)

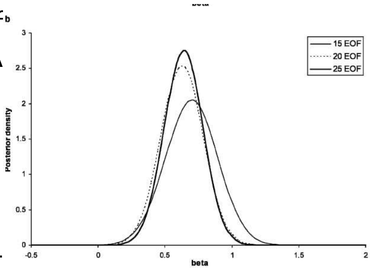


FIG. 2. Posterior distribution derived using prior $\pi_1(\beta)$ for (a) 1900–49 and (b) 1950–99 when 15, 20, and 25 EOFs are retained.

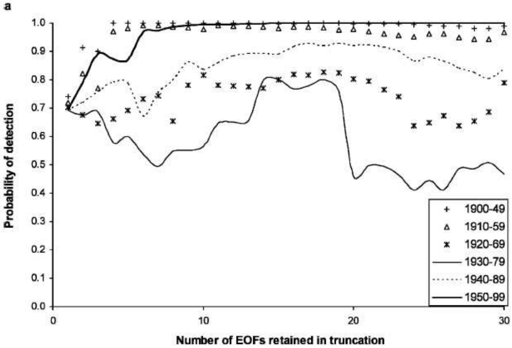
Detection and Attribution

Posterior distribution are used to calculate probabilities that β lies in predefined detection ($\mathcal{D}$) and attribution ($\mathcal{A}$) regions.

- $\mathcal{D} : \beta \in (0.1, \infty)$
  - Prior odds of detection for π_1 are 9:4.
  - "Positive evidence" $(3 \leq B \leq 20) \Rightarrow Prob(\mathcal{D}|\widehat{\beta}) > 0.871$
  - "Strong evidence" $(20 \leq B \leq 150) \Rightarrow Prob(\mathcal{D}|\widehat{\beta}) > 0.978$
  - "Very strong evidence" $(B > 150) \Rightarrow Prob(\mathcal{D}|\widehat{\beta}) > 0.997$
- $\mathcal{A} : \beta \in (0.8, 1.2)$
  - Prior odds of attribution for π_1 are 1:4.
  - "Positive evidence" $(3 \leq B \leq 20) \Rightarrow Prob(\mathcal{A}|\widehat{\beta}) > 0.395$
  - "Strong evidence" $(20 \leq B \leq 150) \Rightarrow Prob(\mathcal{A}|\widehat{\beta}) > 0.813$
  - "Very strong evidence" $(B > 150) \Rightarrow Prob(\mathcal{A}|\widehat{\beta}) > 0.970$

The posterior probability of *detection*

- Posterior probability of detection for overlapping 50-yr periods: 1900-49,. . . , 1950–99.
- Strong (> 0.978) to very strong (> 0.997) evidence of detection for in 1900-49 and 1950–99.
- Similar but weaker evidence in 1910–59 and 1940–89

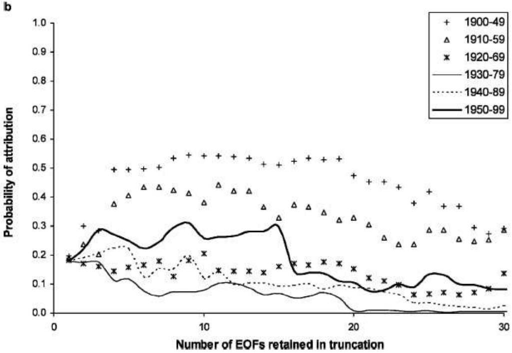


- "Barely worth mentioning" for the middle of the century.

Natural forcing factors (solar and volcanic) are important in explaining the non-monotonic nature of waring in the 20th century (Stott et al, 2000, 2001; Hegerl et al. 2003).

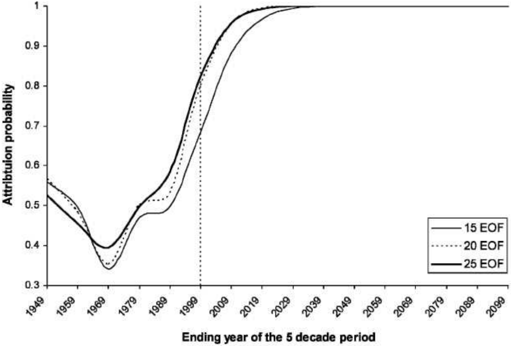
The posterior probability of *attribution*

- Probability of $\mathcal{A} : \beta \in (0.8, 1.2)$ is less than 0.5 ($B \lesssim 3$). There is no strong evidence for attribution.
- Missing forcings and possible model bias don't allow a narrow posterior probability at $\beta = 1$.
- If $\mathcal{A} : \beta \in (0.5, 1.5)$ then $3.2 < B < 12.4$ ('positive evidence')



Estimated future attribution probability

- Future estimates of $\widehat{\sigma}^2_{\widehat{\beta}}$ are obtained from a 21st century run.
- The attribution region is centered at the mode of the posterior distribution.



Bayesian Climate Change Assessment

Advantages of the Bayesian approach:

- ability to incorporate prior knowledge (prior beliefs) into the analysis
- Bayesian assessment of detection and attribution is more explicit than in the frequentist framework.
  - frequentists claim detection when the H_0 is rejected. It is not clear what it means in probabilistic terms.
  - *attribution* is claimed as *failure* to reject H_A.
- In Bayesian setting one must give, and justify, clear definitions for detection and attribution.
- Posterior probabilities can be used in a decision cost/benefit analysis to minimize expected costs and losses.