Temporal coherence

Michael Goldstein
Dept. of Mathematical Sciences
Durham University

Introduction

The subjective Bayesian approach is based on a very simple collection of ideas.

Introduction

The subjective Bayesian approach is based on a very simple collection of ideas.

You are uncertain about many things in the world.

Introduction

The subjective Bayesian approach is based on a very simple collection of ideas.

You are uncertain about many things in the world.

You can quantify your uncertainties as probabilities, for the quantities you are interested in, and conditional probabilities for observations you might make given the things you are interested in.

Introduction

The subjective Bayesian approach is based on a very simple collection of ideas.

You are uncertain about many things in the world.

You can quantify your uncertainties as probabilities, for the quantities you are interested in, and conditional probabilities for observations you might make given the things you are interested in.

When data arrives, Bayes theorem tells you how to move from your prior probabilities to new conditional probabilities for the quantities of interest.

Introduction

The subjective Bayesian approach is based on a very simple collection of ideas.

You are uncertain about many things in the world.

You can quantify your uncertainties as probabilities, for the quantities you are interested in, and conditional probabilities for observations you might make given the things you are interested in.

When data arrives, Bayes theorem tells you how to move from your prior probabilities to new conditional probabilities for the quantities of interest.

If you need to make decisions, then you may also specify a utility function, given which your preferred decision is that which maximises expected utility with respect to your conditional probability distribution.

The Bayes linear approach

The Bayesian approach is hard for complicated problems because

The Bayes linear approach

The Bayesian approach is hard for complicated problems because

(i) thinking about complicated problems is hard!

The Bayes linear approach

The Bayesian approach is hard for complicated problems because

(i) thinking about complicated problems is hard!

(ii) the Bayesian approach for such problems is exhausting!

The Bayes linear approach

The Bayesian approach is hard for complicated problems because

(i) thinking about complicated problems is hard!

(ii) the Bayesian approach for such problems is exhausting!

(i) is unavoidable. How about (ii)?

The Bayes linear approach

The Bayesian approach is hard for complicated problems because

(i) thinking about complicated problems is hard!

(ii) the Bayesian approach for such problems is exhausting!

(i) is unavoidable. How about (ii)?

The Bayes linear approach is concerned with problems in which we want to combine prior judgments of uncertainty with observational data, and we use EXPECTATION rather than probability as the primitive for expressing these judgments (see de Finetti "Theory of Probability", Wiley, 1974).

The Bayes linear approach

The Bayesian approach is hard for complicated problems because
(i) thinking about complicated problems is hard!
(ii) the Bayesian approach for such problems is exhausting!
(i) is unavoidable. How about (ii)?
The Bayes linear approach is concerned with problems in which we want to combine prior judgments of uncertainty with observational data, and we use EXPECTATION rather than probability as the primitive for expressing these judgments (see de Finetti "Theory of Probability", Wiley, 1974).
This distinction is of particular relevance in complex problems with too many sources of information for us to be comfortable in making a meaningful full joint prior probability specification of the type required for a BAYESIAN ANALYSIS.
Therefore, we seek methods of prior specification and analysis which do not require this extreme level of detail.

The Bayes linear approach

The Bayesian approach is hard for complicated problems because
(i) thinking about complicated problems is hard!
(ii) the Bayesian approach for such problems is exhausting!
(i) is unavoidable. How about (ii)?

The Bayes linear approach is concerned with problems in which we want to combine prior judgments of uncertainty with observational data, and we use EXPECTATION rather than probability as the primitive for expressing these judgments (see de Finetti "Theory of Probability", Wiley, 1974).

This distinction is of particular relevance in complex problems with too many sources of information for us to be comfortable in making a meaningful full joint prior probability specification of the type required for a BAYESIAN ANALYSIS.

Therefore, we seek methods of prior specification and analysis which do not require this extreme level of detail.

Thus, the Bayes linear approach is similar in spirit to a full Bayes analysis, but is based on a simpler approach to prior specification and analysis, and so offers a practical methodology for analysing partially specified beliefs for large problems.

Adjusted means and variances

In the Bayes linear approach, we make direct prior specifications for that collection of means, variances and covariances which we are both willing and able to assess, and update these prior assessments by linear fitting.

Adjusted means and variances

In the Bayes linear approach, we make direct prior specifications for that collection of means, variances and covariances which we are both willing and able to assess, and update these prior assessments by linear fitting.

Suppose that we have two collections of random quantities, namely vectors $\boldsymbol{B} = (B_1, ..., B_r)$, $\boldsymbol{D} = (D_0, D_1, ..., D_s)$, where $D_0 = 1$, and we intend to observe $\boldsymbol{D}$ in order to improve our assessments of belief over $\boldsymbol{B}$.

Adjusted means and variances

In the Bayes linear approach, we make direct prior specifications for that collection of means, variances and covariances which we are both willing and able to assess, and update these prior assessments by linear fitting.

Suppose that we have two collections of random quantities, namely vectors $\boldsymbol{B} = (B_1, ..., B_r)$, $\boldsymbol{D} = (D_0, D_1, ..., D_s)$, where $D_0 = 1$, and we intend to observe $\boldsymbol{D}$ in order to improve our assessments of belief over $\boldsymbol{B}$.

The *adjusted* or *Bayes linear* expectation for B_i given $\boldsymbol{D}$ is the linear combination $\boldsymbol{a}_i^T \boldsymbol{D}$ minimising $\mathrm{E}((B_i - \boldsymbol{a}_i^T \boldsymbol{D})^2)$ over choices of $\boldsymbol{a}_i$.

$$\mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) = \mathrm{E}(\boldsymbol{B}) + \mathrm{Cov}(\boldsymbol{B}, \boldsymbol{D})(\mathrm{Var}(\boldsymbol{D}))^{-1}(\boldsymbol{D} - \mathrm{E}(\boldsymbol{D}))$$

Adjusted means and variances

In the Bayes linear approach, we make direct prior specifications for that collection of means, variances and covariances which we are both willing and able to assess, and update these prior assessments by linear fitting.

Suppose that we have two collections of random quantities, namely vectors $\boldsymbol{B} = (B_1, ..., B_r)$, $\boldsymbol{D} = (D_0, D_1, ..., D_s)$, where $D_0 = 1$, and we intend to observe $\boldsymbol{D}$ in order to improve our assessments of belief over $\boldsymbol{B}$.

The *adjusted* or *Bayes linear* expectation for B_i given $\boldsymbol{D}$ is the linear combination $\boldsymbol{a}_i^T \boldsymbol{D}$ minimising $\mathrm{E}((B_i - \boldsymbol{a}_i^T \boldsymbol{D})^2)$ over choices of $\boldsymbol{a}_i$.

$$\mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) = \mathrm{E}(\boldsymbol{B}) + \mathrm{Cov}(\boldsymbol{B}, \boldsymbol{D})(\mathrm{Var}(\boldsymbol{D}))^{-1}(\boldsymbol{D} - \mathrm{E}(\boldsymbol{D}))$$

The *adjusted variance matrix* for $\boldsymbol{B}$ given $\boldsymbol{D}$, is

$\mathrm{Var}_{\boldsymbol{D}}(\boldsymbol{B}) = \mathrm{Var}(\boldsymbol{B} - \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B})) =$
$\mathrm{Var}(\boldsymbol{B}) - \mathrm{Cov}(\boldsymbol{B}, \boldsymbol{D})(\mathrm{Var}(\boldsymbol{D}))^{-1}\mathrm{Cov}(\boldsymbol{D}, \boldsymbol{B})$

Adjusted means and variances

In the Bayes linear approach, we make direct prior specifications for that collection of means, variances and covariances which we are both willing and able to assess, and update these prior assessments by linear fitting.

Suppose that we have two collections of random quantities, namely vectors $\boldsymbol{B} = (B_1, ..., B_r)$, $\boldsymbol{D} = (D_0, D_1, ..., D_s)$, where $D_0 = 1$, and we intend to observe $\boldsymbol{D}$ in order to improve our assessments of belief over $\boldsymbol{B}$.

The *adjusted* or *Bayes linear* expectation for B_i given $\boldsymbol{D}$ is the linear combination $\boldsymbol{a}_i^T \boldsymbol{D}$ minimising $\mathrm{E}((B_i - \boldsymbol{a}_i^T \boldsymbol{D})^2)$ over choices of $\boldsymbol{a}_i$.

$$\mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) = \mathrm{E}(\boldsymbol{B}) + \mathrm{Cov}(\boldsymbol{B}, \boldsymbol{D})(\mathrm{Var}(\boldsymbol{D}))^{-1}(\boldsymbol{D} - \mathrm{E}(\boldsymbol{D}))$$

The *adjusted variance matrix* for $\boldsymbol{B}$ given $\boldsymbol{D}$, is

$\mathrm{Var}_{\boldsymbol{D}}(\boldsymbol{B}) = \mathrm{Var}(\boldsymbol{B} - \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B})) =$
$\mathrm{Var}(\boldsymbol{B}) - \mathrm{Cov}(\boldsymbol{B}, \boldsymbol{D})(\mathrm{Var}(\boldsymbol{D}))^{-1}\mathrm{Cov}(\boldsymbol{D}, \boldsymbol{B})$

More than you could want to know in

Bayes linear Statistics: Theory and Methods, 2007, (Wiley)

Michael Goldstein and David Wooff

Interpretations of belief adjustment

Interpretations of belief adjustment

[1] Within the usual Bayesian view, adjusted expectation offers a simple, tractable approximation to conditional expectation, which is useful in complex problems, while adjusted variance is a strict upper bound to expected posterior variance, over all prior specifications consistent with the moment structure.

Interpretations of belief adjustment

[1] Within the usual Bayesian view, adjusted expectation offers a simple, tractable approximation to conditional expectation, which is useful in complex problems, while adjusted variance is a strict upper bound to expected posterior variance, over all prior specifications consistent with the moment structure. The approximations are exact in certain important special cases, and in particular if the joint probability distribution of B, D is multivariate normal. Therefore, there are strong formal relationships between Bayes linear calculations and the analysis of Gaussian structures.

Interpretations of belief adjustment

[1] Within the usual Bayesian view, adjusted expectation offers a simple, tractable approximation to conditional expectation, which is useful in complex problems, while adjusted variance is a strict upper bound to expected posterior variance, over all prior specifications consistent with the moment structure. The approximations are exact in certain important special cases, and in particular if the joint probability distribution of $\boldsymbol{B}, \boldsymbol{D}$ is multivariate normal. Therefore, there are strong formal relationships between Bayes linear calculations and the analysis of Gaussian structures.

[2] Adjusted expectation is numerically equivalent to conditional expectation in the particular case where $\boldsymbol{D}$ comprises the indicator functions for the elements of a partition, i.e. where each D_i takes value one or zero and precisely one element D_i will equal one, eg, if B is the indicator for an event, then

$$\mathrm{E}_{\boldsymbol{D}}(B) = \sum_i \mathrm{P}(B|D_i) D_i$$

Interpretations of belief adjustment

[1] Within the usual Bayesian view, adjusted expectation offers a simple, tractable approximation to conditional expectation, which is useful in complex problems, while adjusted variance is a strict upper bound to expected posterior variance, over all prior specifications consistent with the moment structure. The approximations are exact in certain important special cases, and in particular if the joint probability distribution of $\boldsymbol{B}, \boldsymbol{D}$ is multivariate normal. Therefore, there are strong formal relationships between Bayes linear calculations and the analysis of Gaussian structures.
[2] Adjusted expectation is numerically equivalent to conditional expectation in the particular case where $\boldsymbol{D}$ comprises the indicator functions for the elements of a partition, i.e. where each D_i takes value one or zero and precisely one element D_i will equal one, eg, if B is the indicator for an event, then

$$\mathrm{E}_{\boldsymbol{D}}(B) = \sum_i \mathrm{P}(B|D_i) D_i$$

Within the Bayes linear view, Bayes analysis is a special case of no greater or lesser interest than any other special case.

The meaning of Bayesian analysis

What do we learn by doing a Bayes linear analysis?

[Special case, what do we learn by doing a Bayesian analysis?]

The meaning of Bayesian analysis

What do we learn by doing a Bayes linear analysis?

[Special case, what do we learn by doing a Bayesian analysis?]

If A and B are both events, what does $\mathrm{P}(B|A)$ mean?

The meaning of Bayesian analysis

What do we learn by doing a Bayes linear analysis?

[Special case, what do we learn by doing a Bayesian analysis?]

If A and B are both events, what does $\mathrm{P}(B|A)$ mean?

$\mathrm{P}(B)$ is your betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise).

The meaning of Bayesian analysis

What do we learn by doing a Bayes linear analysis?

[Special case, what do we learn by doing a Bayesian analysis?]

If A and B are both events, what does $\mathrm{P}(B|A)$ mean?

$\mathrm{P}(B)$ is your betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise).

$\mathrm{P}(B|A)$ is your "'called off"' betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise, if A occurs. If A doesn't occur your price is refunded).

The meaning of Bayesian analysis

What do we learn by doing a Bayes linear analysis?

[Special case, what do we learn by doing a Bayesian analysis?]

If A and B are both events, what does $\mathrm{P}(B|A)$ mean?

$\mathrm{P}(B)$ is your betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise).

$\mathrm{P}(B|A)$ is your "‘called off"’ betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise, if A occurs. If A doesn't occur your price is refunded).

This is NOT the same as the posterior probability that you will have for B if you find out that A occurs.

[Indeed, there is no obvious relationship between the called off bet and posterior judgment at all.]

The meaning of Bayesian analysis

What do we learn by doing a Bayes linear analysis?

[Special case, what do we learn by doing a Bayesian analysis?]

If A and B are both events, what does $\mathrm{P}(B|A)$ mean?

$\mathrm{P}(B)$ is your betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise).

$\mathrm{P}(B|A)$ is your "'called off"' betting rate on B (e.g. your fair price for a ticket that pays 1 if B occurs, and pays 0 otherwise, if A occurs. If A doesn't occur your price is refunded).

This is NOT the same as the posterior probability that you will have for B if you find out that A occurs.

[Indeed, there is no obvious relationship between the called off bet and posterior judgment at all.]

What is the relationship between our current beliefs and our future beliefs?

Temporal rationality

Today, Doctor Jekyll makes certain collections of probabilistic judgments. Tomorrow, as Mister Hyde, he will again make some such collection of judgments. However, while these preferences may be rational at each individual time point, there need be no linkage whatsoever between the two collections of judgments.

Temporal rationality

Today, Doctor Jekyll makes certain collections of probabilistic judgments. Tomorrow, as Mister Hyde, he will again make some such collection of judgments. However, while these preferences may be rational at each individual time point, there need be no linkage whatsoever between the two collections of judgments.

In order to establish links between our judgments at different time points, we need ways of describing 'temporal rationality' which go beyond being internally rational at each time point.

Temporal rationality

Today, Doctor Jekyll makes certain collections of probabilistic judgments. Tomorrow, as Mister Hyde, he will again make some such collection of judgments. However, while these preferences may be rational at each individual time point, there need be no linkage whatsoever between the two collections of judgments.

In order to establish links between our judgments at different time points, we need ways of describing ‘temporal rationality’ which go beyond being internally rational at each time point.

Our description is operational. It concerns preferences between random penalties, as assessed at different time points, considered as small cash penalties or (better) payoffs in probability currency (i.e. tickets in a lottery with a single prize).

Temporal rationality

Today, Doctor Jekyll makes certain collections of probabilistic judgments. Tomorrow, as Mister Hyde, he will again make some such collection of judgments. However, while these preferences may be rational at each individual time point, there need be no linkage whatsoever between the two collections of judgments.

In order to establish links between our judgments at different time points, we need ways of describing 'temporal rationality' which go beyond being internally rational at each time point.

Our description is operational. It concerns preferences between random penalties, as assessed at different time points, considered as small cash penalties or (better) payoffs in probability currency (i.e. tickets in a lottery with a single prize).

[With payoffs in probability currency, expectation for the penalty equals the probability of the reward. Therefore, changes in preferences between penalties A and B over time correspond to changes in probability, rather than utility.]

Constraints on temporal preference

Current preference for random penalty A over penalty B, even when constrained by conditional statements about preferences given possible future evidential outcomes, cannot require you to hold certain future preferences; for example, you may obtain further, hitherto unsuspected, information or insights into the problem before you come to make your future judgments.

Constraints on temporal preference

Current preference for random penalty A over penalty B, even when constrained by conditional statements about preferences given possible future evidential outcomes, cannot require you to hold certain future preferences; for example, you may obtain further, hitherto unsuspected, information or insights into the problem before you come to make your future judgments.

It is more compelling to suggest that future preferences may determine prior preferences. Suppose that you know that tomorrow you will prefer penalty A to B. Should you prefer A to B now?

Constraints on temporal preference

Current preference for random penalty A over penalty B, even when constrained by conditional statements about preferences given possible future evidential outcomes, cannot require you to hold certain future preferences; for example, you may obtain further, hitherto unsuspected, information or insights into the problem before you come to make your future judgments.

It is more compelling to suggest that future preferences may determine prior preferences. Suppose that you know that tomorrow you will prefer penalty A to B. Should you prefer A to B now?

The reasons why current preferences cannot constrain future preferences, based on unanticipated insights, etc., do not apply when the actual future preference is known. What is left are disagreements of a fundamentally different nature, for example, that Doctor Jekyll may consider that his judgments when he turns into Mister Hyde will be intrinsically inferior.

Constraints on temporal preference

Current preference for random penalty A over penalty B, even when constrained by conditional statements about preferences given possible future evidential outcomes, cannot require you to hold certain future preferences; for example, you may obtain further, hitherto unsuspected, information or insights into the problem before you come to make your future judgments.

It is more compelling to suggest that future preferences may determine prior preferences. Suppose that you know that tomorrow you will prefer penalty A to B. Should you prefer A to B now?

The reasons why current preferences cannot constrain future preferences, based on unanticipated insights, etc., do not apply when the actual future preference is known. What is left are disagreements of a fundamentally different nature, for example, that Doctor Jekyll may consider that his judgments when he turns into Mister Hyde will be intrinsically inferior.

Such views are not inherently contradictory. We need operational temporal criteria to determine whether and to what extent your prior analysis may be of value in determining your future judgments (which are weak enough to be compelling in many situations).

Temporal sure preference

Suppose that you must choose between two (probability currency) random penalties, A and B. Suppose that at some future time the values of A and B will be revealed, and you will pay the penalty that you have chosen.

Temporal sure preference

Suppose that you must choose between two (probability currency) random penalties, A and B. Suppose that at some future time the values of A and B will be revealed, and you will pay the penalty that you have chosen.

For your future preferences to influence your current preferences, you must know what your future preference will be.

Temporal sure preference

Suppose that you must choose between two (probability currency) random penalties, A and B. Suppose that at some future time the values of A and B will be revealed, and you will pay the penalty that you have chosen.

For your future preferences to influence your current preferences, you must know what your future preference will be.

You have a **sure preference** for A over B at (future) time t, if you know now, as a matter of logic, that at time t you will not express a strict preference for penalty B over penalty A.

Temporal sure preference

Suppose that you must choose between two (probability currency) random penalties, A and B. Suppose that at some future time the values of A and B will be revealed, and you will pay the penalty that you have chosen.

For your future preferences to influence your current preferences, you must know what your future preference will be.

You have a **sure preference** for A over B at (future) time t, if you know now, as a matter of logic, that at time t you will not express a strict preference for penalty B over penalty A.

The temporal consistency principle that we impose is that future sure preferences are respected by preferences today. We call this the **temporal sure preference principle**, as follows.

The temporal sure preference principle *Suppose that you have a sure preference for* A *over* B *at (future) time* t. *Then you should not have a strict preference for* B *over* A *now.*

Temporal sure preference

Suppose that you must choose between two (probability currency) random penalties, A and B. Suppose that at some future time the values of A and B will be revealed, and you will pay the penalty that you have chosen.

For your future preferences to influence your current preferences, you must know what your future preference will be.

You have a **sure preference** for A over B at (future) time t, if you know now, as a matter of logic, that at time t you will not express a strict preference for penalty B over penalty A.

The temporal consistency principle that we impose is that future sure preferences are respected by preferences today. We call this the **temporal sure preference principle**, as follows.

The temporal sure preference principle *Suppose that you have a sure preference for* A *over* B *at (future) time* t. *Then you should not have a strict preference for* B *over* A *now.*

[This is not a rationality requirement. It is a (weak) operationally testable principle which will often appear reasonable and which has important consequences for statistical reasoning.]

Expectation and temporal preference

We treat expectation as the primitive quantification for our approach. We follow the development of de Finetti, and define the expectation of a random quantity, Z as the value $\bar{z}$ that you would choose for z, if faced with the penalty $L = k(Z - z)^2,$ where k is a constant defining the units of loss, and the penalty is paid in probability currency.

Expectation and temporal preference

We treat expectation as the primitive quantification for our approach. We follow the development of de Finetti, and define the expectation of a random quantity, Z as the value $\bar{z}$ that you would choose for z, if faced with the penalty $L = k(Z - z)^2,$ where k is a constant defining the units of loss, and the penalty is paid in probability currency.

For a particular random quantity Z, you specify a current expectation $\mathrm{E}(Z)$ and you intend to express a revised expectation $\mathrm{E}_t(Z)$ at time t.

Expectation and temporal preference

We treat expectation as the primitive quantification for our approach. We follow the development of de Finetti, and define the expectation of a random quantity, Z as the value $\bar{z}$ that you would choose for z, if faced with the penalty $L = k(Z - z)^2$, where k is a constant defining the units of loss, and the penalty is paid in probability currency.

For a particular random quantity Z, you specify a current expectation $\mathrm{E}(Z)$ and you intend to express a revised expectation $\mathrm{E}_t(Z)$ at time t.

As $\mathrm{E}_t(Z)$ is unknown to you, you may express beliefs about this quantity, and in particular make a current expectation for $(Z - \mathrm{E}_t(Z))^2$.

Expectation and temporal preference

We treat expectation as the primitive quantification for our approach. We follow the development of de Finetti, and define the expectation of a random quantity, Z as the value $\bar{z}$ that you would choose for z, if faced with the penalty $L = k(Z - z)^2$, where k is a constant defining the units of loss, and the penalty is paid in probability currency.

For a particular random quantity Z, you specify a current expectation $\mathrm{E}(Z)$ and you intend to express a revised expectation $\mathrm{E}_t(Z)$ at time t.

As $\mathrm{E}_t(Z)$ is unknown to you, you may express beliefs about this quantity, and in particular make a current expectation for $(Z - \mathrm{E}_t(Z))^2$.

Suppose that F is any random quantity whose value you will surely know by time t. Suppose that you assess a current expectation for $(Z - F)^2$.

Expectation and temporal preference

We treat expectation as the primitive quantification for our approach. We follow the development of de Finetti, and define the expectation of a random quantity, Z as the value $\bar{z}$ that you would choose for z, if faced with the penalty $L = k(Z - z)^2$, where k is a constant defining the units of loss, and the penalty is paid in probability currency.

For a particular random quantity Z, you specify a current expectation $\mathrm{E}(Z)$ and you intend to express a revised expectation $\mathrm{E}_t(Z)$ at time t.

As $\mathrm{E}_t(Z)$ is unknown to you, you may express beliefs about this quantity, and in particular make a current expectation for $(Z - \mathrm{E}_t(Z))^2$.

Suppose that F is any random quantity whose value you will surely know by time t. Suppose that you assess a current expectation for $(Z - F)^2$.

To satisfy temporal sure preference you must now assign

$$\mathrm{E}((Z - \mathrm{E}_t(Z))^2) \leq \mathrm{E}((Z - F)^2)$$

Foundational interpretation

What are the implications of a partial collection of prior belief statements about B, D for the posterior assessment that we may make for the expectation of B having observed D?

Foundational interpretation

What are the implications of a partial collection of prior belief statements about $\boldsymbol{B}, \boldsymbol{D}$ for the posterior assessment that we may make for the expectation of $\boldsymbol{B}$ having observed $\boldsymbol{D}$?

The temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies two relations

Foundational interpretation

What are the implications of a partial collection of prior belief statements about $\boldsymbol{B}, \boldsymbol{D}$ for the posterior assessment that we may make for the expectation of $\boldsymbol{B}$ having observed $\boldsymbol{D}$?
The temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies two relations

$$\boldsymbol{B} = \mathrm{E}_T(\boldsymbol{B}) \oplus \boldsymbol{S}$$

Foundational interpretation

What are the implications of a partial collection of prior belief statements about $\boldsymbol{B}, \boldsymbol{D}$ for the posterior assessment that we may make for the expectation of $\boldsymbol{B}$ having observed $\boldsymbol{D}$?

The temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D},$ satisfies two relations

$$\boldsymbol{B} = \mathrm{E}_T(\boldsymbol{B}) \oplus \boldsymbol{S}$$
$$\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) \oplus \boldsymbol{R},$$

Foundational interpretation

What are the implications of a partial collection of prior belief statements about $\boldsymbol{B}, \boldsymbol{D}$ for the posterior assessment that we may make for the expectation of $\boldsymbol{B}$ having observed $\boldsymbol{D}$?

The temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies two relations

$$\boldsymbol{B} = \mathrm{E}_T(\boldsymbol{B}) \oplus \boldsymbol{S}$$
$$\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) \oplus \boldsymbol{R},$$

where $\boldsymbol{S}, \boldsymbol{R}$ each have, a priori, zero expectation and are uncorrelated with each other and with $\boldsymbol{D}$.

Foundational interpretation

What are the implications of a partial collection of prior belief statements about $\boldsymbol{B}, \boldsymbol{D}$ for the posterior assessment that we may make for the expectation of $\boldsymbol{B}$ having observed $\boldsymbol{D}$?

The temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies two relations

$$\boldsymbol{B} = \mathrm{E}_T(\boldsymbol{B}) \oplus \boldsymbol{S}$$
$$\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) \oplus \boldsymbol{R},$$

where $\boldsymbol{S}, \boldsymbol{R}$ each have, a priori, zero expectation and are uncorrelated with each other and with $\boldsymbol{D}$.

Therefore, adjusted expectation is a prior inference for your actual posterior judgments, which resolves a portion of your current variance for $\boldsymbol{B}$.

Foundational interpretation

What are the implications of a partial collection of prior belief statements about $\boldsymbol{B}, \boldsymbol{D}$ for the posterior assessment that we may make for the expectation of $\boldsymbol{B}$ having observed $\boldsymbol{D}$?

The temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies two relations

$$\boldsymbol{B} = \mathrm{E}_T(\boldsymbol{B}) \oplus \boldsymbol{S}$$
$$\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) \oplus \boldsymbol{R},$$

where $\boldsymbol{S}, \boldsymbol{R}$ each have, a priori, zero expectation and are uncorrelated with each other and with $\boldsymbol{D}$.

Therefore, adjusted expectation is a prior inference for your actual posterior judgments, which resolves a portion of your current variance for $\boldsymbol{B}$.

If $\boldsymbol{D}$ represents a partition, $\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) + \boldsymbol{R} = \mathrm{E}(\boldsymbol{B}|\boldsymbol{D}) + \boldsymbol{R}$ where $\mathrm{E}(\boldsymbol{R}|\boldsymbol{D_i}) = 0, \forall i$.

Statistical Modelling using exchangeability

What does it mean to speak of the ”‘true but unknown”’ probability that a spun coin will land heads?

Statistical Modelling using exchangeability

What does it mean to speak of the ”‘true but unknown”’ probability that a spun coin will land heads?

For a finite population, size N, if n people have some property, e.g. they will vote Labour, then it makes sense to say that the probability that a randomly chosen person will vote Labour is n/N.

Statistical Modelling using exchangeability

What does it mean to speak of the ”‘true but unknown”’ probability that a spun coin will land heads?

For a finite population, size N, if n people have some property, e.g. they will vote Labour, then it makes sense to say that the probability that a randomly chosen person will vote Labour is n/N.

Exchangeability is the modelling construction that creates a form of population for coin tosses. A sequence of coin tosses is exchangeable if every subset of the same size has the same probability distribution.

Statistical Modelling using exchangeability

What does it mean to speak of the "'true but unknown"' probability that a spun coin will land heads?

For a finite population, size N, if n people have some property, e.g. they will vote Labour, then it makes sense to say that the probability that a randomly chosen person will vote Labour is n/N.

Exchangeability is the modelling construction that creates a form of population for coin tosses. A sequence of coin tosses is exchangeable if every subset of the same size has the same probability distribution.

De Finetti's representation theorem shows that if coin tosses are exchangeable, then all of our beliefs about coin tosses are exactly the same as if we believed that our observations were a random series of tosses of a coin with a "'true but unknown"' value for the probability of heads.

Statistical Modelling using exchangeability

What does it mean to speak of the "'true but unknown"' probability that a spun coin will land heads?

For a finite population, size N, if n people have some property, e.g. they will vote Labour, then it makes sense to say that the probability that a randomly chosen person will vote Labour is n/N.

Exchangeability is the modelling construction that creates a form of population for coin tosses. A sequence of coin tosses is exchangeable if every subset of the same size has the same probability distribution.

De Finetti's representation theorem shows that if coin tosses are exchangeable, then all of our beliefs about coin tosses are exactly the same as if we believed that our observations were a random series of tosses of a coin with a "'true but unknown"' value for the probability of heads.

The representation theorem allows us to express beliefs about unobservable quantities purely in terms of our beliefs about observable quantities.

Statistical Modelling using exchangeability

What does it mean to speak of the ”‘true but unknown”’ probability that a spun coin will land heads?

For a finite population, size N, if n people have some property, e.g. they will vote Labour, then it makes sense to say that the probability that a randomly chosen person will vote Labour is n/N.

Exchangeability is the modelling construction that creates a form of population for coin tosses. A sequence of coin tosses is exchangeable if every subset of the same size has the same probability distribution.

De Finetti's representation theorem shows that if coin tosses are exchangeable, then all of our beliefs about coin tosses are exactly the same as if we believed that our observations were a random series of tosses of a coin with a ”‘true but unknown”’ value for the probability of heads.

The representation theorem allows us to express beliefs about unobservable quantities purely in terms of our beliefs about observable quantities.

The only (but major) problem with this representation is that, to use the representation theorem, we must specify all of our beliefs over all the outcomes of possible collections of coin tosses of all sample sizes.

Second order exchangeability

In the Bayes linear approach to statistical modelling, models are constructed directly from simple collections of judgments over observable quantities using the second-order exchangeablity representation theorem.

Second order exchangeability

In the Bayes linear approach to statistical modelling, models are constructed directly from simple collections of judgments over observable quantities using the second-order exchangeablity representation theorem.

An infinite sequence of vectors $\boldsymbol{X}_i$ is *second-order exchangeable* or SOE, if the mean, variance and covariance structure is invariant under permutation, namely

$$\mathrm{E}(\boldsymbol{X}_i) = \mu,\ \mathrm{Var}(\boldsymbol{X}_i) = \Sigma,\ \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \Gamma,\ \forall i \neq j$$

Second order exchangeability

In the Bayes linear approach to statistical modelling, models are constructed directly from simple collections of judgments over observable quantities using the second-order exchangeablity representation theorem.

An infinite sequence of vectors $\boldsymbol{X}_i$ is *second-order exchangeable* or SOE, if the mean, variance and covariance structure is invariant under permutation, namely

$$\mathrm{E}(\boldsymbol{X}_i) = \mu,\ \mathrm{Var}(\boldsymbol{X}_i) = \Sigma,\ \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \Gamma,\ \forall i \neq j$$

We may represent each $\boldsymbol{X}_i$ as the sum of an underlying ‘population mean’ $\boldsymbol{M}$, and individual variation $\boldsymbol{R}_i,$, i.e.

$$\boldsymbol{X}_i = \boldsymbol{M} \oplus \boldsymbol{R}_i$$

Second order exchangeability

In the Bayes linear approach to statistical modelling, models are constructed directly from simple collections of judgments over observable quantities using the second-order exchangeablity representation theorem.

An infinite sequence of vectors $\boldsymbol{X}_i$ is *second-order exchangeable* or SOE, if the mean, variance and covariance structure is invariant under permutation, namely

$$\mathrm{E}(\boldsymbol{X}_i) = \mu,\ \mathrm{Var}(\boldsymbol{X}_i) = \Sigma,\ \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \Gamma,\ \forall i \neq j$$

We may represent each $\boldsymbol{X}_i$ as the sum of an underlying ‘population mean’ $\boldsymbol{M}$, and individual variation $\boldsymbol{R}_i$,, i.e.

$$\boldsymbol{X}_i = \boldsymbol{M} \oplus \boldsymbol{R}_i$$

where the vectors $\boldsymbol{M}, \boldsymbol{R}_1, \boldsymbol{R}_2, \ldots$ are mutually uncorrelated, and

$$\mathrm{E}(\boldsymbol{M}) = \mu, \mathrm{Var}(\boldsymbol{M}) = \Gamma, \mathrm{E}(\boldsymbol{R}_i) = 0, \mathrm{Var}(\boldsymbol{R}_i) = \Sigma - \Gamma, \forall i$$

Temporal adjustment of exchangeable vectors

Suppose that $\boldsymbol{X}_i$ is SOE (to you, now) so $\boldsymbol{X}_i = \boldsymbol{M} \oplus \boldsymbol{R}_i$.
Suppose that you will observe, at time T, a sample $(\boldsymbol{X}_{[\boldsymbol{n}]} = \boldsymbol{X}_{\boldsymbol{1}}, ..., \boldsymbol{X}_{\boldsymbol{n}})$ and revise all your judgements about all remaining $\boldsymbol{X}_{\boldsymbol{j}}$

Temporal adjustment of exchangeable vectors

Suppose that $\boldsymbol{X}_i$ is SOE (to you, now) so $\boldsymbol{X}_i = \boldsymbol{M} \oplus \boldsymbol{R}_i$.
Suppose that you will observe, at time T, a sample $(\boldsymbol{X}_{[n]} = \boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ and revise all your judgements about all remaining $\boldsymbol{X}_j$
In particular, you can now evaluate the Bayes linear adjustment $\mathrm{E}_n(\boldsymbol{M})$ for $\boldsymbol{M}$, given $\boldsymbol{X}_{[n]}$. However, by time T, quantity $\boldsymbol{M}$ may not even exist, as at time T your judgements may no longer be SOE.

Temporal adjustment of exchangeable vectors

Suppose that $\boldsymbol{X}_i$ is SOE (to you, now) so $\boldsymbol{X}_i = \boldsymbol{M} \oplus \boldsymbol{R}_i$.
Suppose that you will observe, at time T, a sample $(\boldsymbol{X}_{[\boldsymbol{n}]} = \boldsymbol{X}_{\boldsymbol{1}}, ..., \boldsymbol{X}_{\boldsymbol{n}})$ and revise all your judgements about all remaining $\boldsymbol{X}_{\boldsymbol{j}}$
In particular, you can now evaluate the Bayes linear adjustment $\mathrm{E}_n(\boldsymbol{M})$ for $\boldsymbol{M}$, given $\boldsymbol{X}_{[\boldsymbol{n}]}$. However, by time T, quantity $\boldsymbol{M}$ may not even exist, as at time T your judgements may no longer be SOE.
Theorem Suppose, you now judge the adjustments $\mathrm{E}_T(\boldsymbol{X}_{\boldsymbol{j}})$ to be SOE ($j > n$). Then, you can construct a further quantity , $\mathrm{E}_T(\boldsymbol{M})$ so that

Temporal adjustment of exchangeable vectors

Suppose that $\boldsymbol{X}_i$ is SOE (to you, now) so $\boldsymbol{X}_i = \boldsymbol{M} \oplus \boldsymbol{R}_i$.
Suppose that you will observe, at time T, a sample $(\boldsymbol{X}_{[n]} = \boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ and revise all your judgements about all remaining $\boldsymbol{X}_j$
In particular, you can now evaluate the Bayes linear adjustment $\mathrm{E}_n(\boldsymbol{M})$ for $\boldsymbol{M}$, given $\boldsymbol{X}_{[n]}$. However, by time T, quantity $\boldsymbol{M}$ may not even exist, as at time T your judgements may no longer be SOE.
Theorem Suppose, you now judge the adjustments $\mathrm{E}_T(\boldsymbol{X}_j)$ to be SOE ($j > n$). Then, you can construct a further quantity , $\mathrm{E}_T(\boldsymbol{M})$ so that

$$
\begin{aligned}
\boldsymbol{X}_j - \mathrm{E}(\boldsymbol{X}) &= \boldsymbol{M} - \mathrm{E}(\boldsymbol{M}) + \boldsymbol{R}_j \\
&= \quad [\boldsymbol{M} - \mathrm{E}_T(\boldsymbol{M})] \\
&\quad \oplus [\mathrm{E}_T(\boldsymbol{M}) - \mathrm{E}_n(\boldsymbol{M})] \\
&\quad \oplus [\mathrm{E}_n(\boldsymbol{M}) - \mathrm{E}(\boldsymbol{M})] \\
&\quad \oplus [\boldsymbol{R}_j - \mathrm{E}_T(\boldsymbol{R}_j)] \\
&\quad \oplus [\mathrm{E}_T(\boldsymbol{R}_j)]
\end{aligned}
$$

Concluding comments

The relationships between actual belief revisions and formal analysis based on partial prior specifications are entirely derived through stochastic relations.

Concluding comments

The relationships between actual belief revisions and formal analysis based on partial prior specifications are entirely derived through stochastic relations. This is no different than any other relationship between a real quantity and a model for that quantity.

Concluding comments

The relationships between actual belief revisions and formal analysis based on partial prior specifications are entirely derived through stochastic relations.

This is no different than any other relationship between a real quantity and a model for that quantity.

Bayes linear (and full Bayes) analysis is a model for our actual reasoning.

The model is special because there is a clear and well-defined relationship between the model inference and our actual inference.

Concluding comments

The relationships between actual belief revisions and formal analysis based on partial prior specifications are entirely derived through stochastic relations.

This is no different than any other relationship between a real quantity and a model for that quantity.

Bayes linear (and full Bayes) analysis is a model for our actual reasoning.

The model is special because there is a clear and well-defined relationship between the model inference and our actual inference.

The subjectivist approach offers a coherent language and tool set for analysing all of the uncertainties in complicated problems, and therefore provides the best method that I know for analysing uncertainty in complex real world problems.

References

Temporal coherence (with discussion) (1985). In Bayesian Statistics 2, ed. Bernardo, J.M., De Groot, M.H., Lindley, D.V., Smith, A.F.M., 231-248, North-Holland.

Revising Exchangeable Beliefs: Subjectivist foundations for the inductive argument, (1994), in Aspects of uncertainty: a tribute to D.V. Lindley, eds. Freeman, P.R. and Smith, A.F.M., John Wiley, 201-222.

Prior inferences for posterior judgements (1997), in Structures and norms in Science, M.C.D. Chiara et. al. eds., 55-71, Kluwer.