8b. Belief Updating and Bayesian Learning

Duarte Gonçalves

University College London

MRes Microconomics

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

Belief updating via Bayes's rule.

Today: (1) How to get behaviour 'as if' Belief updating = Bayes's rule from choices; (2) Properties of Bayesian updating

Overview

1. Uncertainty
2. Characterising Bayesian Updating
3. Properties of Bayesian Updating
4. More

Overview

1. Uncertainty
2. Characterising Bayesian Updating
3. Properties of Bayesian Updating
4. More

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

- **Update beliefs about state and act accordingly**

  $\forall E \subseteq \Omega$, posterior belief given $A \subseteq \Omega$ is

  $$(\mu \mid A)(E) := \mu(E|A) = \frac{\mu(A \cap E)}{\mu(A)}, \quad \forall A : \mu(A) > 0. \text{ (Bayes' rule)}$$

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

- **Update beliefs about state and act accordingly**

  $\forall E \subseteq \Omega$, posterior belief given $A \subseteq \Omega$ is

  $$(\mu \mid A)(E) := \mu(E|A) = \frac{\mu(A \cap E)}{\mu(A)}, \quad \forall A : \mu(A) > 0. \text{ (Bayes' rule)}$$

  Issue: beliefs were *deduced* from preferences/behaviour. Are deduced beliefs updated according to Bayes rule?

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

- **Update beliefs about state and act accordingly**

  $\forall E \subseteq \Omega$, posterior belief given $A \subseteq \Omega$ is

  $$(\mu \mid A)(E) := \mu(E|A) = \frac{\mu(A \cap E)}{\mu(A)}, \quad \forall A : \mu(A) > 0. \text{ (Bayes' rule)}$$

  Issue: beliefs were *deduced* from preferences/behaviour. Are deduced beliefs updated according to Bayes rule?

  Yes — under some additional restrictions.

Bayesian Updating

Deciding whether or not to carry an umbrella. Look out the window and check the weather now. Informative about weather later.

Stock market tomorrow. Election outcome and polls. Job market today and decision of education.

- **Update beliefs about state and act accordingly**

  $\forall E \subseteq \Omega$, posterior belief given $A \subseteq \Omega$ is

  $$(\mu \mid A)(E) := \mu(E|A) = \frac{\mu(A \cap E)}{\mu(A)}, \quad \forall A : \mu(A) > 0. \text{ (Bayes' rule)}$$

  Issue: beliefs were *deduced* from preferences/behaviour. Are deduced beliefs updated according to Bayes rule?

  Yes — under some additional restrictions.

Continue with Anscombe-Aumann setup (similar conditions for Savage framework)

AA SEU: A Refresher

- Ω: set of states of the world, finite;
- X: set of consequences or outcomes, finite;
- $f : \Omega \to \Delta(X)$: an act;
- $\mathcal{F} := \Delta(X)^{\Omega}$: set of acts;
- $\succsim \subseteq \mathcal{F}^2$: preference relation.

AA SEU: A Refresher

- Ω: set of states of the world, finite;
- X: set of consequences or outcomes, finite;
- $f : \Omega \to \Delta(X)$: an act;
- $\mathcal{F} := \Delta(X)^{\Omega}$: set of acts;
- $\succsim \subseteq \mathcal{F}^2$: preference relation.

Theorem

(1) A pref. rel. $\succsim$ on $\mathcal{F}$ sat. continuity, independence, and separability/monotonicity $\iff$ $\succsim$ admits a SEU representation, i.e., $\exists u : X \to \mathbb{R}$ and $\mu \in \Delta(\Omega) : f \succsim g \iff \mathbb{E}_{\mu}[\mathbb{E}_{f(\omega)}[u]] \geq \mathbb{E}_{\mu}[\mathbb{E}_{g(\omega)}[u]]$.
(2) Moreover, u is unique up to positive affine transformations and, if $\exists f, g \in \mathcal{F} : f \succ g$, μ is unique.

Bayesian Updating

Enrich AA's setup

- Events $A \in \mathcal{E} := 2^{\Omega} \setminus \{\emptyset\}$ are **information**; denote DM learning that state ω lies in A.

Bayesian Updating

Enrich AA's setup

- Events $A \in \mathcal{E} := 2^{\Omega} \setminus \{\emptyset\}$ are **information**; denote DM learning that state ω lies in A.
- Collection of preferences: $\{\succsim_A\}_{A \in \mathcal{E}}$
  Each $\succsim_A$ is preference relation on acts $\mathcal{F}$
  Idea is that $\succsim_A$ describes behaviour upon learning $\omega \in A$.
  $\succsim \equiv \succsim_{\Omega}$: DM's behaviour in absence of any information.

Bayesian Updating

Enrich AA's setup

- Events $A \in \mathcal{E} := 2^{\Omega} \setminus \{\emptyset\}$ are **information**; denote DM learning that state ω lies in A.
- Collection of preferences: $\{\succsim_A\}_{A \in \mathcal{E}}$
  Each $\succsim_A$ is preference relation on acts $\mathcal{F}$
  Idea is that $\succsim_A$ describes behaviour upon learning $\omega \in A$.
  $\succsim \equiv \succsim_{\Omega}$: DM's behaviour in absence of any information.
- Start by assuming $\forall A \in \mathcal{E}$ sat. independence, continuity, and monotonicity

Definition

$\{\succsim_A\}_{A\in\mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \Omega$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

Bayesian Updating

Definition

$\{\succsim_A\}_{A\in\mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \Omega$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

(ii) **dynamic consistency** if, for all non-null events (wrt $\succsim$) $A \subseteq \Omega$ and all acts $f, g \in \mathcal{F}$, $fAg \succsim_\Omega g \iff f \succsim_A g$;

Bayesian Updating

Definition

$\{\succsim_A\}_{A\in\mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \Omega$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

(ii) **dynamic consistency** if, for all non-null events (wrt $\succsim$) $A \subseteq \Omega$ and all acts $f, g \in \mathcal{F}$, $fAg \succsim_\Omega g \iff f \succsim_A g$;

(iii) **consequentialism** if, for event $A \subseteq \Omega$, two acts $f, g \in \mathcal{F}$ deliver the same lottery $f(\omega) = g(\omega)$ for every $\omega \in A$, then $f \sim_A g$.

Bayesian Updating

Definition

$\{\succsim_A\}_{A\in\mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \Omega$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

(ii) **dynamic consistency** if, for all non-null events (wrt $\succsim$) $A \subseteq \Omega$ and all acts $f, g \in \mathcal{F}$, $fAg \succsim_\Omega g \iff f \succsim_A g$;

(iii) **consequentialism** if, for event $A \subseteq \Omega$, two acts $f, g \in \mathcal{F}$ deliver the same lottery $f(\omega) = g(\omega)$ for every $\omega \in A$, then $f \sim_A g$.

Constant-act consistency: if two acts whose (distrib. over) consequences is independent from the state, then whatever you learn should not change how you compare them.

Bayesian Updating

Definition

$\{\succsim_A\}_{A\in\mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \Omega$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

(ii) **dynamic consistency** if, for all non-null events (wrt $\succsim$) $A \subseteq \Omega$ and all acts $f, g \in \mathcal{F}$, $fAg \succsim_\Omega g \iff f \succsim_A g$;

(iii) **consequentialism** if, for event $A \subseteq \Omega$, two acts $f, g \in \mathcal{F}$ deliver the same lottery $f(\omega) = g(\omega)$ for every $\omega \in A$, then $f \sim_A g$.

Constant-act consistency: if two acts whose (distrib. over) consequences is independent from the state, then whatever you learn should not change how you compare them.

Constant-act consistency + independence, continuity, and monotonicity $\implies \exists \alpha_A > 0, \beta_A \in \mathbb{R} : u_A = \alpha_A u + \beta_A$.

I.e., constant-act consistency ties in utility functions over consequences.

Bayesian Updating

Definition

$\{\succsim_A\}_{A \in \mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \boldsymbol{\Omega}$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

(ii) **dynamic consistency** if, for all non-null events (wrt $\succsim$) $A \subseteq \boldsymbol{\Omega}$ and all acts $f, g \in \mathcal{F}$, $fAg \succsim_{\boldsymbol{\Omega}} g \iff f \succsim_A g$;

(iii) **consequentialism** if, for event $A \subseteq \boldsymbol{\Omega}$, two acts $f, g \in \mathcal{F}$ deliver the same lottery $f(\boldsymbol{\omega}) = g(\boldsymbol{\omega})$ for every $\boldsymbol{\omega} \in A$, then $f \sim_A g$.

Dynamic consistency: two acts that differ only when A occurs should be ranked in the same way before and after knowing $\boldsymbol{\omega} \in A$.

This is making DM keep relative beliefs across non-null events constant as information arrives.

Bayesian Updating

Definition

$\{\succsim_A\}_{A \in \mathcal{E}}$ satisfy

(i) **constant-act consistency** iff preferences over constant acts are consistent: for all lotteries $p, q \in \Delta(X)$ and events $A, B \subseteq \Omega$, $\tilde{p} \succsim_A \tilde{q} \iff \tilde{p} \succsim_B \tilde{q}$;

(ii) **dynamic consistency** if, for all non-null events (wrt $\succsim$) $A \subseteq \Omega$ and all acts $f, g \in \mathcal{F}$, $fAg \succsim_\Omega g \iff f \succsim_A g$;

(iii) **consequentialism** if, for event $A \subseteq \Omega$, two acts $f, g \in \mathcal{F}$ deliver the same lottery $f(\omega) = g(\omega)$ for every $\omega \in A$, then $f \sim_A g$.

Consequentialism is making sure that events $B : B \cap A = \emptyset$ are null events wrt $\succsim_A$.

I.e., the decision-maker believes the information received.

Theorem

Let $\{\succsim_A\}_{A\in\mathcal{E}}$ be collection of preference relations on $\mathcal{F}$ and assume $\exists f, g \in \mathcal{F} : f \succ_\Omega g$.

$\{\succsim_A\}_{A\in\mathcal{E}}$ sat. constant-act consistency, dynamic consistency, and consequentialism, and $\succsim_A$ satisfies continuity, independence, and monotonicity $A \in \mathcal{E}$ *if and only if* $\exists u : X \to \mathbb{R}$ and collection of probability measures $\{\mu_A\}_{A\in\mathcal{E}}$, $\mu_A \in \Delta(\Omega)\ \forall A \in \mathcal{E}\ (A \neq \emptyset)$, s.t.

(i) $\forall f, g \in \mathcal{F}, f \succsim_A g \iff \mathbb{E}_{\mu_A}[\mathbb{E}_{f(\omega)}[u]] \geq \mathbb{E}_{\mu_A}[\mathbb{E}_{g(\omega)}[u]]$; and

(ii) for all non-null events wrt $\succsim_\Omega$, $A \in \mathcal{E}$, $\mu_A(B) = \frac{\mu_\Omega(A\cap B)}{\mu_\Omega(A)}\ \forall B \in \mathcal{E}$.

Moreover, u is unique up to positive affine transformations and μ_Ω is unique.

Bayesian Updating

Proof Sketch

By AA's SEU, $\forall$ non-null event $A \in \mathcal{E}$, $\exists u_A : X \to \mathbb{R}$ and a prior $\mu_A \in \Delta(\Omega)$ s.t. $f \succsim_A g \iff \mathbb{E}_{\mu_A}[\mathbb{E}_{f(\omega)}[u_A]] \geq \mathbb{E}_{\mu_A}[\mathbb{E}_{g(\omega)}[u_A]]$.

1. Show constant-act consistency implies $\forall$ non-null event $A \in \mathcal{E}$, $\exists \alpha_A > 0, \beta_A \in \mathbb{R}$: $u_A = \alpha_A u_\Omega + \beta_A$.
2. Show $\exists x, y \in X : \tilde{\delta}_x \succ_A \tilde{\delta}_y$ for all non-null events $A \in \mathcal{E}$.
3. Use previous step to prove that $\forall$ non-null events $A \in \mathcal{E}$, $\mu_\Omega(A) > 0$.
4. Show $\forall f, g, h \in \mathcal{F}$ and any non-null event $A \in \mathcal{E}$, $f \succsim_A g \iff fAh \succsim_\Omega gAh$.
5. Show that for any acts $f, g, h \in \mathcal{F}$ and any non-null event $A \in \mathcal{E}$, $f \succsim_A g \iff fAh \succsim_\Omega gAh$ implies that $\mu_A(\omega) = \mu_\Omega(\omega)/\mu_\Omega(A)$.
6. Argue for uniqueness claims based on the above and on proof of AA SEU Theorem.
7. Verify 'if' part, i.e., that representation implies the assumptions on $\{\succsim_A\}_{A \in \mathcal{E}}$.

Bayesian Updating

Theorem

Let $\{\succsim_A\}_{A \in \mathcal{E}}$ be collection of preference relations on $\mathcal{F}$ and assume $\exists f, g \in \mathcal{F} : f \succ_\Omega g$.

$\{\succsim_A\}_{A \in \mathcal{E}}$ sat. constant-act consistency, dynamic consistency, and consequentialism, and $\succsim_A$ satisfies continuity, independence, and monotonicity $A \in \mathcal{E}$ *if and only if* $\exists u : X \to \mathbb{R}$ and collection of probability measures $\{\mu_A\}_{A \in \mathcal{E}}$, $\mu_A \in \Delta(\Omega)\ \forall A \in \mathcal{E}\ (A \neq \emptyset)$, s.t.

(i) $\forall f, g \in \mathcal{F}, f \succsim_A g \iff \mathbb{E}_{\mu_A}[\mathbb{E}_{f(\omega)}[u]] \geq \mathbb{E}_{\mu_A}[\mathbb{E}_{g(\omega)}[u]]$; and

(ii) for all non-null events wrt $\succsim_\Omega$, $A \in \mathcal{E}$, $\mu_A(B) = \frac{\mu_\Omega(A \cap B)}{\mu_\Omega(A)}\ \forall B \in \mathcal{E}$.

Moreover, u is unique up to positive affine transformations and μ_Ω is unique.

Theorem gives conditions on choices that agents must satisfy if they are *behaving like* Bayesian subjective expected utility maximisers.

Note **we cannot observe people's beliefs**, only infer them from behaviour.
Violating assumptions here does not imply beliefs are not updated according to Bayes' rule (not falsifiable).

Overview

1. Uncertainty

2. Characterising Bayesian Updating

3. Properties of Bayesian Updating
   - Setup
   - Consistency of Bayesian Learning
   - Conjugate Priors

4. More

Bayesian Learning

Why do we like Bayesian updating? Because, ultimately, it's Bayes's rule!

Bayesian Learning

Why do we like Bayesian updating? Because, ultimately, it's Bayes's rule!

- Suppose you have a model in which people are Bayesian and they learn from data.
  E.g.:

  traders observing a signal about fundamentals,

  principals who can observe output but not effort,

  consumers who get some information on product quality,

  farmers who observe which fertilisers their neighbours use.

Would people learn the truth if they could get many signals?

Bayesian Learning

Why do we like Bayesian updating? Because, ultimately, it's Bayes's rule!

- Suppose you have a model in which people are Bayesian and they learn from data. E.g.:

  traders observing a signal about fundamentals,

  principals who can observe output but not effort,

  consumers who get some information on product quality,

  farmers who observe which fertilisers their neighbours use.

Would people learn the truth if they could get many signals?

- Suppose you are doing text-analysis (which makes extensive and intensive use of latent Dirichlet classification and Bayesian updating in topic modelling).

  How do you perform inference without a Bayesian law of large numbers (LLN) and something similar to a Bayesian central limit theorem (CLT)?

In about every field of economics, we implicitly or explicitly have to deal with Bayesian learning and its properties and implications.

Bayesian Learning

Why do we like Bayesian updating? Because, ultimately, it's Bayes's rule!

- Suppose you have a model in which people are Bayesian and they learn from data. E.g.:

  traders observing a signal about fundamentals,

  principals who can observe output but not effort,

  consumers who get some information on product quality,

  farmers who observe which fertilisers their neighbours use.

Would people learn the truth if they could get many signals?

- Suppose you are doing text-analysis (which makes extensive and intensive use of latent Dirichlet classification and Bayesian updating in topic modelling).

  How do you perform inference without a Bayesian law of large numbers (LLN) and something similar to a Bayesian central limit theorem (CLT)?

In about every field of economics, we implicitly or explicitly have to deal with Bayesian learning and its properties and implications.

But... **What are its properties and implications?**

Setup

Prior Beliefs: DM entertains hypotheses about the world summarised by (i) parameter $\theta \in \Theta$ and (ii) prior belief, i.e., probability measure $\mu \in \Delta(\Theta)$.

Setup

Prior Beliefs: DM entertains hypotheses about the world summarised by (i) parameter $\theta \in \Theta$ and (ii) prior belief, i.e., probability measure $\mu \in \Delta(\Theta)$.

Data: DM observes sequence of random variables $\{X_n\}_n$. First n observations: $X^n = (X_1, ..., X_n)$.

Likelihood: given a parameter θ, (DM belives) data X^n distributed according to probability P^n_θ (the likelihood).
Often, X_i are iid observations (and P^n_θ is product measure), but generally they need not be. *Assume iid data for convenience.*

Each prior gives rise to joint distrib. of (θ, X^n) (think about what this statement means).

Setup

Prior Beliefs: DM entertains hypotheses about the world summarised by (i) parameter $\theta \in \Theta$ and (ii) prior belief, i.e., probability measure $\mu \in \Delta(\Theta)$.

Data: DM observes sequence of random variables $\{X_n\}_n$. First n observations: $X^n = (X_1, ..., X_n)$.

Likelihood: given a parameter θ, (DM belives) data X^n distributed according to probability P^n_θ (the likelihood).
Often, X_i are iid observations (and P^n_θ is product measure), but generally they need not be. *Assume iid data for convenience.*

Each prior gives rise to joint distrib. of (θ, X^n) (think about what this statement means).

Posterior Beliefs: Upon observing the data, DM updates beliefs about θ using Bayes's rule, forming a posterior belief $\mu_n \equiv \mu \mid X^n$ – the conditional distrib. θ given X^n.

Throughout, consider μ_n is well-defined, in which case 'posterior belief $\propto$ likelihood $\times$ prior belief'.

Consistency of Bayesian Learning

Crucial question: **(when) does DM eventually learn θ?**

Consistency of Bayesian Learning

Crucial question: **(when) does DM eventually learn θ?**

Definition

The posterior distribution μ_n is **consistent** at θ_0 if for any neighbourhood U of θ_0, $\mu_n(U) \to 1$ almost surely under the law determined by θ_0, i.e., the distribution of X^n determined by θ_0.

Consistency of Bayesian Learning

Crucial question: **(when) does DM eventually learn θ?**

Definition

The posterior distribution μ_n is **consistent** at θ_0 if for any neighbourhood U of θ_0, $\mu_n(U) \to \mathbf{1}$ almost surely under the law determined by θ_0, i.e., the distribution of X^n determined by θ_0.

Prior μ is **misspecified** if true model θ_0 is not in the support of μ.

Consistency of Bayesian Learning

Crucial question: **(when) does DM eventually learn θ?**

Definition

The posterior distribution μ_n is **consistent** at θ_0 if for any neighbourhood U of θ_0, $\mu_n(U) \to 1$ almost surely under the law determined by θ_0, i.e., the distribution of X^n determined by θ_0.

Prior μ is **misspecified** if true model θ_0 is not in the support of μ.

Assume: $P_\theta \neq P_{\theta'}$, $\forall \theta \neq \theta'$. (i.e., Identifiability.)

Immediately: consistency requires parameter to be identifiable from data, $P_{\theta_0} \neq P_\theta$ for any $\theta \neq \theta_0$;

(but would be weird to assume this only for true (unknown) parameter θ_0)

Consistency of Bayesian Learning

Theorem (Doob (1948))

$\forall$ prior $\mu \in \Delta(\Theta)$, posterior μ_n is consistent at every θ except possibly on a set of μ-measure zero.

Consistency of Bayesian Learning

Theorem (Doob (1948))

$\forall$ prior $\mu \in \Delta(\Theta)$, posterior μ_n is consistent at every θ except possibly on a set of μ-measure zero.

The Good: DM learns! A sanity check! A triumph for Bayesianism!

Consistency of Bayesian Learning

Theorem (Doob (1948))

$\forall$ prior $\mu \in \Delta(\Theta)$, posterior μ_n is consistent at every θ except possibly on a set of μ-measure zero.

The Good: DM learns! A sanity check! A triumph for Bayesianism!

The Bad (and Ugly): Learning can fail *even if* θ_0 is in supp(μ), and the result says nothing about which θ_0 can be learned...

Consistency of Bayesian Learning

Is it always the case that with a lot of iid data, Bayesian learning always leads to learning?
(insofar as the prior belief is not misspecified and identification is possible)

Consistency of Bayesian Learning

Is it always the case that with a lot of iid data, Bayesian learning always leads to learning?
(insofar as the prior belief is not misspecified and identification is possible)

Unfortunately no...

Consistency of Bayesian Learning

Is it always the case that with a lot of iid data, Bayesian learning always leads to learning?
(insofar as the prior belief is not misspecified and identification is possible)

Unfortunately no...

Example: Learning the wrong thing

θ: pmf on $\mathbb{N}$. True model θ_0 is geometric distrib. with parameter $1/4$.

Consistency of Bayesian Learning

Is it always the case that with a lot of iid data, Bayesian learning always leads to learning?
(insofar as the prior belief is not misspecified and identification is possible)

Unfortunately no...

Example: Learning the wrong thing

θ: pmf on $\mathbb{N}$. True model θ_0 is geometric distrib. with parameter $1/4$.

Freedman (1963): $\exists$ prior μ assigning positive mass to every neighbourhood of θ_0 but with posterior beliefs concentrating on geometric distribution with parameter $3/4$.

DM 'learns', but becomes fully convinced of something that is not true!

Example: Learning the wrong thing (bis)

Consider parameters and priors with support in countably infinite set (as $\mathbb{N}$).

Consistency of Bayesian Learning

Example: Learning the wrong thing (bis)

Consider parameters and priors with support in countably infinite set (as $\mathbb{N}$).

Freedman (1965): Set of pairs of priors and true parameters inducing consistent posteriors is *meager* (countable union of nowhere dense sets).

Freedman (1963) not just a pathological example: the rule rather than exception.

Consistency of Bayesian Learning

Example: Learning the wrong thing (bis)

Consider parameters and priors with support in countably infinite set (as $\mathbb{N}$).

Freedman (1965): Set of pairs of priors and true parameters inducing consistent posteriors is *meager* (countable union of nowhere dense sets).

Freedman (1963) not just a pathological example: the rule rather than exception.

Who cares? YOU!

Care is needed to actually learn the truth (in your models, in your estimation, etc.)

Consistency of Bayesian Learning

Theorem (Diaconis & Freedman (1990))

If (i) X_n are iid and can only take finitely many values, and (ii) $P_\theta \neq P_{\theta'}$ $\forall \theta \neq \theta'$, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.

Consistency of Bayesian Learning

Theorem (Diaconis & Freedman (1990))

If (i) X_n are iid and can only take finitely many values, and (ii) $P_\theta \neq P_{\theta'}$ $\forall \theta \neq \theta'$, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.

Better than that: there are results (see Diaconis & Freedman 1990) providing explicit convergence rates and bounds for how concentrated the posterior is around the empirical mean!

Consistency of Bayesian Learning

Theorem (Diaconis & Freedman (1990))

If (i) X_n are iid and can only take finitely many values, and (ii) $P_\theta \neq P_{\theta'}$ $\forall \theta \neq \theta'$, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.

Better than that: there are results (see Diaconis & Freedman 1990) providing explicit convergence rates and bounds for how concentrated the posterior is around the empirical mean!

And beyond finite cases?

Consistency of Bayesian Learning

Theorem (Diaconis & Freedman (1990))

If (i) X_n are iid and can only take finitely many values, and (ii) $P_\theta \neq P_{\theta'}$ $\forall\theta \neq \theta'$, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.

Better than that: there are results (see Diaconis & Freedman 1990) providing explicit convergence rates and bounds for how concentrated the posterior is around the empirical mean!

And beyond finite cases?

Theorem (Schwartz (1965))

Let Θ be class of densities and X_n be iid with density $\theta_0 \in \Theta$. Let $\mu \in \Delta(\Theta)$: $\forall\varepsilon > 0$, $\mu\left(\{\theta \in \Theta \mid \int \theta_0 \ln(\theta_0/\theta) < \varepsilon\}\right) > 0$. Then, the posterior belief μ_n is consistent at θ_0.

Consistency of Bayesian Learning

Theorem (Diaconis & Freedman (1990))

If (i) X_n are iid and can only take finitely many values, and (ii) $P_\theta \neq P_{\theta'}\ \forall \theta \neq \theta'$, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.

Better than that: there are results (see Diaconis & Freedman 1990) providing explicit convergence rates and bounds for how concentrated the posterior is around the empirical mean!

And beyond finite cases?

Theorem (Schwartz (1965))

Let Θ be class of densities and X_n be iid with density $\theta_0 \in \Theta$. Let $\mu \in \Delta(\Theta) : \forall \varepsilon > 0$, $\mu\left(\{\theta \in \Theta \mid \int \theta_0 \ln(\theta_0/\theta) < \varepsilon\}\right) > 0$. Then, the posterior belief μ_n is consistent at θ_0.

Practical sufficient condition for consistency: if $\{f_\theta, \theta \in \Theta\}$ is family of densities smoothly parametrised by parameter $\theta \in \mathbb{R}^k$ (k finite), and $X_n \overset{iid}{\sim} f_{\theta_0}$, then consistency obtained if and only if θ_0 lies in the support of the prior.

Consistency $\approx$ LLN.

Consistency of Bayesian Learning

Theorem (Diaconis & Freedman (1990))

If (i) X_n are iid and can only take finitely many values, and (ii) $P_\theta \neq P_{\theta'}$ $\forall \theta \neq \theta'$, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.

Better than that: there are results (see Diaconis & Freedman 1990) providing explicit convergence rates and bounds for how concentrated the posterior is around the empirical mean!

And beyond finite cases?

Theorem (Schwartz (1965))

Let Θ be class of densities and X_n be iid with density $\theta_0 \in \Theta$. Let $\mu \in \Delta(\Theta) : \forall \varepsilon > 0$, $\mu\left(\{\theta \in \Theta \mid \int \theta_0 \ln(\theta_0/\theta) < \varepsilon\}\right) > 0$. Then, the posterior belief μ_n is consistent at θ_0.

Practical sufficient condition for consistency: if $\{f_\theta, \theta \in \Theta\}$ is family of densities smoothly parametrised by parameter $\theta \in \mathbb{R}^k$ (k finite), and $X_n \overset{iid}{\sim} f_{\theta_0}$, then consistency obtained if and only if θ_0 lies in the support of the prior.

Consistency $\approx$ LLN. Can we get **Bayesian CLT**? Yes! See Ghosal (1997) and references there.

Conjugate Priors

$\theta \in \Theta$; X rv with likelihood P_θ, taking values in $\mathcal{X}$

A set $M \subseteq \Delta(\Theta)$ is a **conjugate prior** if $\forall \mu \in M$, posterior $\mu \mid X = x \in M$, $\forall x \in \mathcal{X}$.

Conjugate Priors

$\theta \in \Theta$; X rv with likelihood P_θ, taking values in $\mathcal{X}$

A set $M \subseteq \Delta(\Theta)$ is a **conjugate prior** if $\forall \mu \in M$, posterior $\mu \mid X = x \in M, \forall x \in \mathcal{X}$.

E.g., Θ finite, $\Delta(\Theta)$ is a conjugate prior.

Not very useful, but important reminder: no such thing as *the* conjugate prior.

Conjugate Priors

Useful Examples

- Likelihood: Bernoulli, $X \sim \text{Bernoulli}(\theta)$.
  Beta distribution family is a conjugate prior: $\theta \sim \text{Beta}(\alpha_0, \alpha_1) \implies \theta | X = x \sim \text{Beta}(\alpha_0 + (1 - x), \alpha_1 + x)$.

Conjugate Priors

Useful Examples

- Likelihood: Bernoulli, $X \sim \text{Bernoulli}(\theta)$.
  Beta distribution family is a conjugate prior: $\theta \sim \text{Beta}(\alpha_0, \alpha_1) \implies \theta | X = x \sim \text{Beta}(\alpha_0 + (1 - x), \alpha_1 + x)$.
- Likelihood: categorical (generalization of Bernoulli), $X \in \{1, \ldots, k\}$, $X \sim \text{Categorical}(\theta)$ where $\theta = (\theta_1, \ldots, \theta_k) \in \Delta^{k-1}$.
  Dirichlet distribution family (generalization of Beta) is a conjugate prior: $\theta \sim \text{Dirichlet}(\alpha) \implies \theta | X = x \sim \text{Dirichlet}(\alpha + e_x)$, where $\alpha = (\alpha_i)_{i=1,\ldots,k}$ and $e_x = (1_{x=1}, \ldots, 1_{x=k})$.

Conjugate Priors

Useful Examples

- Likelihood: Bernoulli, $X \sim \text{Bernoulli}(\theta)$.
  Beta distribution family is a conjugate prior: $\theta \sim \text{Beta}(\alpha_0, \alpha_1) \implies \theta | X = x \sim \text{Beta}(\alpha_0 + (1 - x), \alpha_1 + x)$.
- Likelihood: categorical (generalization of Bernoulli), $X \in \{1, \dots, k\}$, $X \sim \text{Categorical}(\theta)$ where $\theta = (\theta_1, \dots, \theta_k) \in \Delta^{k-1}$.
  Dirichlet distribution family (generalization of Beta) is a conjugate prior: $\theta \sim \text{Dirichlet}(\alpha) \implies \theta | X = x \sim \text{Dirichlet}(\alpha + e_x)$, where $\alpha = (\alpha_i)_{i=1,\dots,k}$ and $e_x = (1_{x=1}, \dots, 1_{x=k})$.
- Likelihood: Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$.
  Normal distribution family is a conjugate prior: $\mu \sim \mathcal{N}(\mu_0, \Sigma_0) \implies \mu | X = x \sim \mathcal{N}(\mu_1, \Sigma_1)$, where $\mu_1 = \Sigma_1(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}x)$ and $\Sigma_1 = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$.

Conjugate Priors

Useful Examples

- Likelihood: Bernoulli, $X \sim \text{Bernoulli}(\theta)$.
  Beta distribution family is a conjugate prior: $\theta \sim \text{Beta}(\alpha_0, \alpha_1) \implies \theta | X = x \sim \text{Beta}(\alpha_0 + (1 - x), \alpha_1 + x)$.
- Likelihood: categorical (generalization of Bernoulli), $X \in \{1, \ldots, k\}$, $X \sim \text{Categorical}(\theta)$ where $\theta = (\theta_1, \ldots, \theta_k) \in \Delta^{k-1}$.
  Dirichlet distribution family (generalization of Beta) is a conjugate prior: $\theta \sim \text{Dirichlet}(\alpha) \implies \theta | X = x \sim \text{Dirichlet}(\alpha + e_x)$, where $\alpha = (\alpha_i)_{i=1,\ldots,k}$ and $e_x = (\mathbf{1}_{x=1}, \ldots, \mathbf{1}_{x=k})$.
- Likelihood: Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$.
  Normal distribution family is a conjugate prior: $\mu \sim \mathcal{N}(\mu_0, \Sigma_0) \implies \mu | X = x \sim \mathcal{N}(\mu_1, \Sigma_1)$, where $\mu_1 = \Sigma_1(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}x)$ and $\Sigma_1 = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$.

  Also: reparameterize Normal distrib. with precision matrix $\tau = \Sigma^{-1}$, $\mu \sim \mathcal{N}(\mu_0, \tau_0)$ and $\mu | x \sim \mathcal{N}(\mu_1, \tau_1)$, where $\mu_1 = \tau_1^{-1}(\tau_0\mu_0 + \tau x)$ and $\tau_1 = \tau_0 + \tau$.

Conjugate Priors

Useful Examples

- Likelihood: Bernoulli, $X \sim \text{Bernoulli}(\theta)$.
  Beta distribution family is a conjugate prior: $\theta \sim \text{Beta}(\alpha_0, \alpha_1) \implies \theta | X = x \sim \text{Beta}(\alpha_0 + (1 - x), \alpha_1 + x)$.
- Likelihood: categorical (generalization of Bernoulli), $X \in \{1, ..., k\}$, $X \sim \text{Categorical}(\theta)$ where $\theta = (\theta_1, ..., \theta_k) \in \Delta^{k-1}$.
  Dirichlet distribution family (generalization of Beta) is a conjugate prior: $\theta \sim \text{Dirichlet}(\alpha) \implies \theta | X = x \sim \text{Dirichlet}(\alpha + e_x)$, where $\alpha = (\alpha_i)_{i=1,...,k}$ and $e_x = (\mathbf{1}_{x=1}, ..., \mathbf{1}_{x=k})$.
- Likelihood: Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$.
  Normal distribution family is a conjugate prior: $\mu \sim \mathcal{N}(\mu_0, \Sigma_0) \implies \mu | X = x \sim \mathcal{N}(\mu_1, \Sigma_1)$, where $\mu_1 = \Sigma_1(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}x)$ and $\Sigma_1 = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$.
  Also: reparameterize Normal distrib. with precision matrix $\tau = \Sigma^{-1}$, $\mu \sim \mathcal{N}(\mu_0, \tau_0)$ and $\mu | x \sim \mathcal{N}(\mu_1, \tau_1)$, where $\mu_1 = \tau_1^{-1}(\tau_0\mu_0 + \tau x)$ and $\tau_1 = \tau_0 + \tau$.
- Also exist tractable families for unknown precision $\tau \sim \mathit{Gamma}(\alpha, \beta)$.

Conjugate Priors

Useful Examples

- Likelihood: Bernoulli, $X \sim \text{Bernoulli}(\theta)$.
  Beta distribution family is a conjugate prior: $\theta \sim \text{Beta}(\alpha_0, \alpha_1) \implies \theta | X = x \sim \text{Beta}(\alpha_0 + (1 - x), \alpha_1 + x)$.
- Likelihood: categorical (generalization of Bernoulli), $X \in \{1, \dots, k\}$, $X \sim \text{Categorical}(\theta)$ where $\theta = (\theta_1, \dots, \theta_k) \in \Delta^{k-1}$.
  Dirichlet distribution family (generalization of Beta) is a conjugate prior: $\theta \sim \text{Dirichlet}(\alpha) \implies \theta | X = x \sim \text{Dirichlet}(\alpha + e_x)$, where $\alpha = (\alpha_i)_{i=1,\dots,k}$ and $e_x = (\mathbf{1}_{x=1}, \dots, \mathbf{1}_{x=k})$.
- Likelihood: Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$.
  Normal distribution family is a conjugate prior: $\mu \sim \mathcal{N}(\mu_0, \Sigma_0) \implies \mu | X = x \sim \mathcal{N}(\mu_1, \Sigma_1)$, where $\mu_1 = \Sigma_1(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}x)$ and $\Sigma_1 = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$.
  Also: reparameterize Normal distrib. with precision matrix $\tau = \Sigma^{-1}$, $\mu \sim \mathcal{N}(\mu_0, \tau_0)$ and $\mu | x \sim \mathcal{N}(\mu_1, \tau_1)$, where $\mu_1 = \tau_1^{-1}(\tau_0\mu_0 + \tau x)$ and $\tau_1 = \tau_0 + \tau$.
- Also exist tractable families for unknown precision $\tau \sim \textit{Gamma}(\alpha, \beta)$.
- Other famous pairs: (Poisson, Gamma), (Exponential, Gamma), (Uniform, Pareto).

Overview

1. Uncertainty
2. Characterising Bayesian Updating
3. Properties of Bayesian Updating
4. More

More on Bayesian Learning

- **Merging of Opinions**: If people see similar data, (when) will their opinions tend to converge to the same thing?

  Blackwell & Dubins (1962), Kalai & Lehrer (1994 JMathEcon), and Acemoglu, Chernozhukov, & Yildiz (2016 TE) address these issues. Kalai & Lehrer (1993 Ecta) use this to study learning to play NE.

More on Bayesian Learning

- **Merging of Opinions**: If people see similar data, (when) will their opinions tend to converge to the same thing?

  Blackwell & Dubins (1962), Kalai & Lehrer (1994 JMathEcon), and Acemoglu, Chernozhukov, & Yildiz (2016 TE) address these issues. Kalai & Lehrer (1993 Ecta) use this to study learning to play NE.

- **Social Learning**: If everyone gets a signal, but we only learn from others' actions, (when) do we get to learning the true state?

  Applications: Innovation adoption, stock market, etc.

More on Bayesian Learning

- **Merging of Opinions**: If people see similar data, (when) will their opinions tend to converge to the same thing?

  Blackwell & Dubins (1962), Kalai & Lehrer (1994 JMathEcon), and Acemoglu, Chernozhukov, & Yildiz (2016 TE) address these issues. Kalai & Lehrer (1993 Ecta) use this to study learning to play NE.

- **Social Learning**: If everyone gets a signal, but we only learn from others' actions, (when) do we get to learning the true state?

  Applications: Innovation adoption, stock market, etc.

  Answer: it depends! See Bikhchandani, Hirshleifer, & Welch (1992 JPE), Smith & Sorensen (2000 Ecta) for classical refs.

More on Bayesian Learning

- **Common Learning**:

  Suppose group speculators observe signals about fundamentals. WT strike iff learn currency is weak with sufficiently high degree of certainty.

  Need to coordinate their efforts to strike. As they wait, they may learn perfectly whether or not the currency is weak, but also need to know that others have learned (to a prespecified sufficient degree of confidence) that the currency is weak. (and so on...)

More on Bayesian Learning

- **Common Learning**:

  Suppose group speculators observe signals about fundamentals. WT strike iff learn currency is weak with sufficiently high degree of certainty.

  Need to coordinate their efforts to strike. As they wait, they may learn perfectly whether or not the currency is weak, but also need to know that others have learned (to a prespecified sufficient degree of confidence) that the currency is weak. (and so on...)

  Common learning: have individuals learn the truth, and also learn that all learned the truth, and that all have learned that all have learned the truth, and so on.

More on Bayesian Learning

- **Common Learning**:

  Suppose group speculators observe signals about fundamentals. WT strike iff learn currency is weak with sufficiently high degree of certainty.

  Need to coordinate their efforts to strike. As they wait, they may learn perfectly whether or not the currency is weak, but also need to know that others have learned (to a prespecified sufficient degree of confidence) that the currency is weak. (and so on...)

  Common learning: have individuals learn the truth, and also learn that all learned the truth, and that all have learned that all have learned the truth, and so on.

  Cripps, Ely, Mailath, & Samuelson (2008 Ecta): common learning obtained if prior and likelihood are CK and have full support. But common learning may also fail!

More on Bayesian Updating at Large!

- **Methods to Elicit Beliefs:** important beyond just experimental and theory! E.g., *development and education* (Dizon-Ross 2019 AER), *macro* (Bordalo et al. 2020 AER), *health* (Miller, de Paula, & Valente, 2025 JEconometrics), *finance* (Giglio et al. 2021 AER), *political economy* (Ortoleva Snowberg 2015 AER)

More on Bayesian Updating at Large!

- **Methods to Elicit Beliefs:** important beyond just experimental and theory! E.g., *development and education* (Dizon-Ross 2019 AER), *macro* (Bordalo et al. 2020 AER), *health* (Miller, de Paula, & Valente, 2025 JEconometrics), *finance* (Giglio et al. 2021 AER), *political economy* (Ortoleva Snowberg 2015 AER)
- **Patterns in Belief Updating:** Turns out that in some contexts people important exhibit systematic deviations!

This motivates **Models of Belief Updating**: noisy belief updating, overconfidence, memory constraints, etc. (e.g., Benjamin 2019 Handbook, Cripps 2018 WP, Gonçalves et al. 2025 REStud)

More on Bayesian Updating at Large!

- **Methods to Elicit Beliefs:** important beyond just experimental and theory! E.g., *development and education* (Dizon-Ross 2019 AER), *macro* (Bordalo et al. 2020 AER), *health* (Miller, de Paula, & Valente, 2025 JEconometrics), *finance* (Giglio et al. 2021 AER), *political economy* (Ortoleva Snowberg 2015 AER)
- **Patterns in Belief Updating:** Turns out that in some contexts people important exhibit systematic deviations!

  This motivates **Models of Belief Updating**: noisy belief updating, overconfidence, memory constraints, etc. (e.g., Benjamin 2019 Handbook, Cripps 2018 WP, Gonçalves et al. 2025 REStud)

  Take + theory & experimental!

Where does this leave SEU and Bayes Rule?

Bayesian updating and SEU remain the main framework: very appealing principles and well-known virtues and vices.

Behaviourally: neither comes for free and it's important to know this.

Where does this leave SEU and Bayes Rule?

Bayesian updating and SEU remain the main framework: very appealing principles and well-known virtues and vices.

Behaviourally: neither comes for free and it's important to know this.

Implications for models: Understanding better deviations allows us to explicitly accommodate these in our model (even without letting go of Bayesian updating).

But unless some crucial element is missing, using BU and SEU allows better understanding of modelling innovations.