ECON0106: **Microeconomics**

8b. Bayesian Learning[*]

Duarte Gonçalves[†]

University College London

1. Overview

Suppose you have a model in which people are Bayesian and they learn from data. For instance, traders observing a signal about fundamentals, principals who can observe output but not effort, consumers who get some information on product quality. Would people learn the truth if they could get many signals? Alternatively, suppose that 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. These lecture notes take a closer look at these issues.

2. Bayesian Learning

We'll start by discussing Bayesian learning, which will be a central part of our career concerns model. Bayesian learning is nothing but the application of Bayes's rule to learning: we have a prior belief about the world, we observe some data, and we update our beliefs about the world. The decision-maker entertains a number of hypotheses about the world, summarised by a parameter θ taking values in Θ, and has some prior belief about which hypothesis is true, given by a probability measure μ on Θ.

The decision-maker observes data, a sequence of random variables X_n. We write $X^n = (X_1, ..., X_n)$ for the sequence of the first n observations. X^n is then distributed according to a probability P_θ^n, which depends on the parameter θ; such probability is called a likelihood, since it determines the likelihood of X^n given a parameter θ. Often, X_i are iid observations, but generally they need not be. Note that the set of all priors on Θ is denoted by $\Delta(\Theta)$, and each of such priors gives rise to a joint distribution of (θ, X^n) (think about what this statement means).

Upon observing the data, the decision-maker updates their beliefs about the world using

[*]Last updated: 23 September 2025.

[†] Department of Economics, University College London; duarte.goncalves@ucl.ac.uk. Please do not share these notes with people outside of this class.

Bayes's rule, forming a posterior belief $\mu_n \equiv \mu \mid X^n$, consisting of the conditional distribution of θ given X^n. Throughout, we'll consider the case where μ_n is well-defined for all n, in which case we have that 'posterior belief $\propto$ likelihood $\times$ prior belief'.

3. Consistency of Bayesian Learning

One important question about learning is whether the decision-maker will eventually learn the true model. This idea is captured by the notion of **consistency** of Bayesian learning: a prior μ is said to be consistent with the true model, θ_0, if it collapses to a mass point on θ_0 almost surely. In a way, consistency is the Bayesian counterpart to the law of large numbers. Formally,

Definition 1. The posterior belief μ_n is said to be consistent with the true model θ_0 if for every neighbourhood U of θ_0, $\mu_n(U) \to 1$ almost surely under the law determined by θ_0.

We say that a prior μ is **misspecified** if the true model θ_0 is not in the support of μ. Throughout, we'll assume that the true model is in the support of the prior, so that the posterior belief is well-defined for all n.[1] Furthermore, we'll also restrict to the case in which X_i are iid. Naturally, consistency requires that the parameter is identifiable from the data, that is, $P_{\theta_0} \neq P_\theta$ for any $\theta \neq \theta_0$. Since it would be bizarre to assume the above only for the true (unknown) parameter θ_0, we require $P_\theta \neq P_{\theta'}$ for any $\theta \neq \theta'$.

We'll now discuss some results on consistency of Bayesian learning. The first result is due to Doob (1948), who showed that consistency is obtained for almost all true models, where the almost all is with respect to the prior.

Theorem 1. ***(Doob (1948))*** *For any prior μ, the posterior μ_n is consistent at every θ except possibly on a set of μ-measure zero.*

While this result is nice, it is not very useful in practice, since we cannot be sure that the true model θ_0 the posterior is indeed consistent (θ_0 may be one of those μ-measure zero parameters for which consistency may fail). We can do better than that, however, if we impose some further conditions on the sample space of X_n.

Theorem 2. *If X_n are iid and can only take finitely many values, then for any prior μ, the posterior μ_n is consistent at every θ in the support of μ.*

Moreover, there are results that provide explicit convergence rates for the posterior belief and – what is also useful — bounds for how concentrated the posterior is around the empirical

[1]What if the prior is misspecified? We can still say something about convergence! See Berk (1966) for a classical reference and Esponda and Pouzo (2016) for an application to a game theory.

mean; see Diaconis and Freedman (1990).[2]

One may think that at least by obtaining humongous amount of iid data, Bayesian learning, fundamental as it is, is always consistent as long the prior belief is not misspecified. It turns out that, without further assumptions, when observations X_n can take infinitely many values (even if countably many), consistency may fail.

Example 1. (Inconsistent Bayesian Learning) Suppose that θ describes a probability mass function on the set of positive integers and θ_0, the true model, corresponds to a geometric distribution with parameter 1/4. Freedman (1963) shows that there is a prior μ that gives positive mass to every neighbourhood of θ_0 but the posterior belief concentrates on the neighbourhood of a geometric distribution with parameter 3/4.

In fact, the above example is not pathological: Freedman (1965) showed that inconsistency is the rule rather than the exception when no further conditions are imposed on Θ or the likelihood. This serves as a cautionary tale: *Bayesian learning is not always consistent.*

To obtain consistency, we need restrictions. A famous condition was obtained by Schwartz (1965):

Theorem 3. ***(Schwartz (1965))*** *Let Θ be a class of densities and let X_n be iid with density θ_0, where $\theta_0 \in \Theta$. Let μ be a prior on Θ such that for every $\epsilon > 0$, $\mu\left(\{\theta \in \Theta \mid \int \theta_0 \ln(\theta_0/\theta) < \epsilon\}\right) > 0$. Then, the posterior belief μ_n is consistent at θ_0.*

Here's a **practical condition** that is sufficient for consistency: if $\{f_\theta, \theta \in \Theta\}$ is a family of densities smoothly parametrised by a real or vector valued parameter θ, and $X_n \overset{iid}{\sim} f_{\theta_0}$, then consistency is obtained if and only if θ_0 lies in the support of the prior.

Now, if consistency is akin to a LLN, do we have a Bayesian CLT? The answer is yes, but we need to impose some further conditions – see Ghosal (1997) and references therein for details.

4. Conjugate Priors

A particularly useful concept is that of **conjugate priors**. Given a random variable X with likelihood P_θ, $\theta \in \Theta$ and taking values in $\mathscr{X}$, a prior $\mu \in M$ is said to be conjugate if the posterior $\mu \mid X$ is in the same family M as the prior μ for any value of X. Note that there is no such thing as *the* conjugate prior. For instance, if Θ is finite, then the set of all probability measures on Θ is a conjugate prior no matter the likelihood – a true statement, but not a very useful one.

Some nice examples of conjugate priors are the following:

[2]Fun fact: Persi Diaconis was a professional magician before becoming a statistician.

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

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

- If the likelihood is Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$, then the Normal distribution family is a conjugate prior: $\mu \sim \mathcal{N}(\mu_0, \Sigma_0) \Longrightarrow \mu|X = x \sim \mathcal{N}(\mu_1, \Sigma_1)$, where $\mu_1 = \Sigma_1(\Sigma_0^{-1}\mu_0 + \Sigma^{-1}x)$ and $\Sigma_1 = (\Sigma_0^{-1} + \Sigma^{-1})^{-1}$.
  Another way to think about this, is to reparametrise the Normal distribution in terms of its precision matrix $\tau = \Sigma^{-1}$, so that $\mu \sim \mathcal{N}(\mu_0, \tau_0)$ and $\mu|x \sim \mathcal{N}(\mu_1, \tau_1)$, where $\mu_1 = \tau_1^{-1}(\tau_0\mu_0 + \tau x)$ and $\tau_1 = \tau_0 + \tau$.
  There are also tractable results for unknown precision $\tau \sim Gamma(\alpha, \beta)$.

- Other famous pairs include (Poisson, Gamma), (Exponential, Gamma), (Uniform, Pareto).

5. Related Topics

5.1. Merging of Opinions

If (at least under some conditions) individuals learn the truth, then at some point they will agree on the true model. This is the idea behind a classical result by Blackwell and Dubins (1962) and extended by Kalai and Lehrer (1994): two individuals with two different beliefs (mutually absolutely continuous) will tend to have similar posterior beliefs as they observe more and more data. This has important implications, namely for showing that Bayesian players eventually learn to play Nash equilibrium in a repeated game (Kalai and Lehrer, 1993). Acemoglu et al. (2016) discuss the limits of this merging of opinions.

5.2. Common Learning

Suppose a group of speculators observe signals about fundamentals and they want to strike if they learn the currency is weak with sufficiently high degree of certainty. However, they need to coordinate their efforts to strike, and while, as they wait, they may learn perfectly whether or not the currency is weak, they do need to also know that others have learned (to a prespecified sufficient degree of confidence) that the currency is weak. This is the idea behind **common learning**: not only do individuals learn the truth, but they also learn that they have learned the truth, and that they have learned that they have learned the truth, and so on. When does common learning actually occur? This is the question addressed by Cripps et al. (2008), who show that common learning is obtained, e.g., if the prior is common knowledge and the likelihood is common knowledge and has full support. But, as most we've seen, common learning may also fail (see examples in Cripps et al. (2008)).

5.3. Other Related Topics

Learning is also a central part of the literature on social learning – referenced below – which studies how individuals learn from the actions of others (e.g. online reviews, fads and fashions, neighbours' crops' success, etc), as well as in game theory, being it a fundamental idea underlying the very concept of equilibrium behaviour – as emerging from learning and experience of players (references also below).

6. Study and Further Reading

6.1. Further Reading

- Learning in games: Fudenberg and Levine (1998), Fudenberg and Levine (2009), Fudenberg and Levine (2016)
- Misspecified Learning: Berk (1966), Fudenberg et al. (2023); in games: Esponda and Pouzo (2016), Fudenberg et al. (2021).
- Social learning: Bikhchandani et al. (1992), Smith and Sørensen (2000), a survey (including empirical references) Mobius and Rosenblat (2014), a textbook Chamley (2010)
- Bayesian Learning in Macro: Baley and Veldkamp (2021)
- Merging of opinions: Acemoglu et al. (2016)
- Experiments: belief updating Charness and Levin (2005), learning about political facts Hill (2017), learning in games Camerer (2003, ch. 6)

7. References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2016. "Fragility of asymptotic agreement under Bayesian learning." *Theoretical Economics* 11 (1): 187–225. 10.3982/TE436.

Baley, Isaac, and Laura Veldkamp. 2021. "Bayesian Learning." *NBER Working Paper* (29338): . 10.3386/w29338.

Berk, Robert H. 1966. "Limiting Behavior of Posterior Distributions when the Model is Incorrect." *Annals of Mathematical Statistics* 37 (1): 51–58. 10.1214/aoms/1177699597.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026. 10.1086/261849.

Blackwell, David, and Lester Dubins. 1962. "Merging of opinions with increasing information." *The Annals of Mathematical Statistics* 33 (3): 882–886. 10.1214/aoms/1177704456.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press.
Chamley, Christophe P. 2010. *Rational Herds: Economic Models of Social Learning.* Cambridge University Press.
Charness, Gary, and Dan Levin. 2005. "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect." *American Economic Review* 95 (4): 1300–1309. 10.1257/0002828054825583.
Cripps, Martin, Jeffrey C. Ely, George J. Mailath, and Larry Samuelson. 2008. "Common Learning." *Econometrica* 76 (4): 909–933. 10.1111/j.1468-0262.2008.00862.x.
Diaconis, Persi, and David Freedman. 1990. "On the uniform consistency of Bayes estimates for multinomial probabilities." *Annals of Statistics* 18 (3): 1317–1327.
Doob, Joseph L. 1948. "Application of the theory of martingales." *Colloques Internationaux du CNRS* 22–28.
Esponda, Ignacio, and Demian Pouzo. 2016. "Berk-Nash Equilibrium: A Framework for Modeling Misspecified Models." *Econometrica* 84 (3): 1093–1130. 10.3982/ECTA12609.
Freedman, David. 1963. "On the asymptotic behavior of Bayes estimates in the discrete case I.." *Annals of Mathematical Statistics* 34 1386–1403.
Freedman, David. 1965. "On the asymptotic behavior of Bayes estimates in the discrete case II.." *Annals of Mathematical Statistics* 36 454–456.
Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack. 2021. "Limit Points of Endogenous Misspecified Learning." *Econometrica* 89 (3): 1065–1098. 10.3982/ECTA18508.
Fudenberg, Drew, Giacomo Lanzani, and Philipp Strack. 2023. "Pathwise concentration bounds for Bayesian beliefs." *Theoretical Economics* 18 (4): 1585–1622. 10.3982/TE5206.
Fudenberg, Drew, and David K. Levine. 1998. *The Theory of Learning in Games.* MIT Press.
Fudenberg, Drew, and David K. Levine. 2009. "Learning and Equilibrium." *Annual Review of Economics* 1 385–420. 10.1146/annurev.economics.050708.142930.
Fudenberg, Drew, and David K. Levine. 2016. "Whither Game Theory? Towards a Theory of Learning in Games." *Journal of Economic Perspectives* 30 (4): 151–70. 10.1257/jep.30.4.151.
Ghosal, Subhashis. 1997. "A Review of Consistency and Convergence of Posterior Distribution." *Manuscript* 1–10, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3dac008f296da5ee41d9f54b032479615df53a3a.
Hill, Seth J. 2017. "Learning Together Slowly: Bayesian Learning about Political Facts." *The Journal of Politics* 79 (4): 1201–1236. 10.1086/692739.
Kalai, Ehud, and Ehud Lehrer. 1993. "Rational Learning Leads to Nash Equilibrium." *Econometrica* 61 (5): 1019–1045. 10.2307/2951492.
Kalai, Ehud, and Ehud Lehrer. 1994. "Weak and strong merging of opinions." *Journal of Mathematical Economics* 23 (1): 73–86. 10.1016/0304-4068(94)90037-X.
Mobius, Markus, and Tanya Rosenblat. 2014. "Social Learning in Economics." *Annual Review of Economics* 6 827–847. 10.1146/annurev-economics-120213-012609.
Schwartz, Lorraine. 1965. "On Bayes procedures." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 4 10–26. 10.1007/BF00535479.
Smith, Lones, and Peter Sørensen. 2000. "Pathological Outcomes of Observational Learning." *March 2000* 68 (2): 371–398. 10.1111/1468-0262.00113.