Section 4 – Global games

Juan Dodyk

Gov 2006 – Harvard

February 18, 2020

Bayesian updating with continuous distributions

I will tell you a trick that people doing Bayesian statistics use a lot.

Let's revisit something we did in class.

- There is a parameter θ_0 we don't know.
- We have a flat improper prior: $p(\theta_0) = c > 0$.
- We observe a signal $\theta_1|\theta_0 \sim N(\theta_0, \sigma^2)$.
- We update our belief about θ_0.

So, we want to know the distribution $\theta_0|\theta_1$.

Bayes rule:

$$p(\theta_0|\theta_1) = \frac{p(\theta_1|\theta_0)p(\theta_0)}{p(\theta_1)}.$$

The numerator is easy: we are given $p(\theta_1|\theta_0)$ and $p(\theta_0)$.

The denominator is hard to calculate.

The trick

Key observation: it's just a constant.

Key observation 2: if you know a PDF $p(x)$ up to a constant, then you know it, because the constant is determined by the formula $\int p(x)\,dx = 1$.

If you know that $p(x) = kf(x)$ but you don't know k, you can calculate it:

$$1 = \int p(x)\,dx = \int kf(x)\,dx = k\int f(x)\,dx,$$

so

$$k = \frac{1}{\int f(x)\,dx}.$$

We can ignore constant factors.

Notation: $p(x) \propto f(x)$ means that $p(x) = kf(x)$, where $k > 0$ is a constant.

Bayes rule is $p(\theta_0|\theta_1) = \frac{p(\theta_1|\theta_0)p(\theta_0)}{p(\theta_1)}$, but, since the denominator is a constant, we can write it simply as

$$p(\theta_0|\theta_1) \propto p(\theta_1|\theta_0)p(\theta_0).$$

(If you read Bayesian stats papers, they do this all the time.)

Going back to our example, we have

Bayes rule is $p(\theta_0|\theta_1) = \frac{p(\theta_1|\theta_0)p(\theta_0)}{p(\theta_1)}$, but, since the denominator is a constant, we can write it simply as

$$p(\theta_0|\theta_1) \propto p(\theta_1|\theta_0)p(\theta_0).$$

(If you read Bayesian stats papers, they do this all the time.)

Going back to our example, we have

$$\begin{aligned} p(\theta_0|\theta_1) \propto p(\theta_1|\theta_0)p(\theta_0) = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\theta_1 - \theta_0)^2\right] c \propto \\ &\propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\theta_0 - \theta_1)^2\right]. \end{aligned}$$

Now, this function of θ_0 integrates to 1, so it must be exactly the PDF. It is clearly the PDF of a normal $N(\theta_1, \sigma^2)$.

Hence, $\theta_0|\theta_1 \sim N(\theta_1, \sigma^2)$, which we knew already.

Adding a second signal

Suppose that we observe another signal $\theta_2|\theta_0 \sim N(\theta_0, \tau^2)$, independent of $\theta_1|\theta_0$. We want the posterior $\theta_0|\theta_1, \theta_2$.

Bayes rule:

$$
\begin{aligned}
p(\theta_0|\theta_1,\theta_2) &\propto p(\theta_1,\theta_2|\theta_0)p(\theta_0) = \\
&= p(\theta_1|\theta_0)p(\theta_2|\theta_0)p(\theta_0) \propto \\
&\propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\theta_1-\theta_0)^2\right] \frac{1}{\sqrt{2\pi}\tau} \exp\left[-\frac{1}{2\tau^2}(\theta_2-\theta_0)^2\right] \propto \\
&\propto \exp\left[-\frac{1}{2\sigma^2}(\theta_1-\theta_0)^2 - \frac{1}{2\tau^2}(\theta_2-\theta_0)^2\right] = \\
&= \exp\left[-\frac{1}{2\sigma^2}(\theta_0^2 - 2\theta_1\theta_0 + \theta_1^2) - \frac{1}{2\tau^2}(\theta_0^2 - 2\theta_2\theta_0 + \theta_2^2)\right] = \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\theta_0^2 - \left(\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}\right)\theta_0 - \frac{\theta_1^2}{\sigma^2} - \frac{\theta_2^2}{\tau^2}\right] \propto \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\theta_0^2 - \left(\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}\right)\theta_0\right]
\end{aligned}
$$

$$
\begin{aligned}
p(\theta_0|\theta_1, \theta_2) \propto & \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\theta_0^2 - \left(\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}\right)\theta_0\right] = \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\left(\theta_0 - \frac{\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)^2 + \frac{\left(\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}\right)^2}{2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)}\right] \propto \\
&\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\left(\theta_0 - \frac{\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)^2\right] \propto \\
&\propto \frac{1}{\sqrt{2\pi\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}}} \exp\left[-\frac{1}{2\frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}}\left(\theta_0 - \frac{\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)^2\right]
\end{aligned}
$$

So...

$$
\theta_0|\theta_1, \theta_2 \sim N\left(\frac{\frac{\theta_1}{\sigma^2} + \frac{\theta_2}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right).
$$

This is a lot more intuitive in terms of precisions $h_1 = 1/\sigma^2$ and $h_2 = 1/\tau^2$.

The precision is the inverse of the variance: more noise, less precision.

Written in terms of precisions, the above result is

$$\theta_0|\theta_1, \theta_2 \sim N\left(\frac{h_1\theta_1 + h_2\theta_2}{h_1 + h_2}, \frac{1}{h_1 + h_2}\right).$$

So, what happened?

- We got a signal θ_1 with precision h_1.
- We got a signal θ_2 with precision h_2.
- Then, our estimate of θ_0 is a weighted average of θ_1 and θ_2, where we weight the signals according to their precision.
- And the precision of our estimate is the sum of the precisions of the signals.

This generalizes to n signals, in which case we get DeGroot's formula

$$\theta_0|\theta_1,\ldots,\theta_n \sim N\left(\frac{h_1\theta_1+\cdots+h_n\theta_n}{h_1+\cdots+h_n},\frac{1}{h_1+\cdots+h_n}\right).$$

This formula is key in models of career concerns (Holmström 1999, ReStud, originally published in an obscure book in 1982). Those models are the basis of models in electoral accountability (see Scott Ashworth's papers). So, it's a useful formula.

The normal distribution is very special in this regard. In general distributions do not behave as nicely.

Look up "conjugate prior" if you want to know what happens with other distributions.

Correlated signals

There is a state $\theta \in \mathbb{R}$ that nobody knows.

Two leaders, 1 and 2, communicate signals s_{1i} and s_{2i} to a group of agents indexed by i. Each agent i knows her signals s_{1i} and s_{2i} but doesn't know the signals of other agents.

For every leader j and agent i we have

$$s_{ji} = \theta + \eta_j + \epsilon_{ij},$$

with $\eta_j \sim N(0, \sigma_j^2)$ and $\sigma_{ij} \sim N(0, \tau_j^2)$, all jointly independent.

Two questions:

- Calculate $\mathbb{E}_i[\theta \mid s_{1i}, s_{2i}]$.
- Calculate $\mathbb{E}_i[s_{1k} \mid s_{1i}, s_{2i}]$ for $k \neq i$.