Lecture 2: Bayes Estimation

David A. Levin

DAIICT

January 2020

Bayesian Estimation

- In this set-up, we assume the parameter(s) θ has a **prior** distribution $\pi(\theta)$.
- Given θ, the random variables $(X_1, \ldots, X_n)$ have a distribution $f(x_1 \ldots, x_n \mid \theta)$.
- If the distributions of X_i's are conditionally independent, then

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

- Inference about θ is made using the **posterior** distribution

$$f(\theta \mid x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n \mid \theta)\pi(\theta)}{\int f(x_1, \ldots, x_n \mid \theta)\pi(\theta)\, d\theta}$$

- The issue is often computing the normalizing constant in the posterior:
$$\int f(x_1, \ldots, x_n \mid \theta)\pi(\theta)\, d\theta$$
- If θ is high-dimensional, this especially can be difficult. Modern Bayesian statistics uses many methods including Markov Chain Monte Carlo to evaluate this constant.
- Often the prior is chosen so that the posterior is easy to determine; if the posterior has the same parametric form as the prior, the distribution is called **conjugate**

Normal Example

- Suppose that

$$\pi(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2\tau^2}(\mu - \nu)^2\right)$$

  That is $\theta \sim N(\nu, \tau^2)$.

- Suppose that $X_1, \ldots, X_n$ are i.i.d.

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right),$$

  that is, X_i are $N(\mu, \sigma^2)$.

- Here ν, τ^2, σ^2 are assumed known. They are called **hyperparameters**.
- Calculating the posterior seems messy but it all works out because the prior is conjugate.

$$
\begin{aligned}
f(\mu \mid x_1,\ldots,x_n) &\propto f(x_1,\ldots,x_n \mid \mu)\pi(\mu;\nu,\tau) \\
&= c(\sigma,\tau)\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\tau^2}(\mu-\nu)^2\right)\right) \\
&= c(\sigma,\tau,\boldsymbol{x})\exp\left(\mu\left(\frac{S_n}{2\sigma^2}+\frac{\nu}{2\tau^2}\right)-\mu^2\left(\frac{n}{2\sigma^2}+\frac{1}{2\tau^2}\right)\right) \\
&= c(\sigma,\tau,\boldsymbol{x})\exp\left[-\frac{1}{2}\frac{n\tau^2+\sigma^2}{\sigma^2\tau^2}\left(\mu^2-\mu\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}\left(\frac{\bar{x}n\tau^2+\nu\sigma^2}{\sigma^2\tau^2}\right)\right)\right] \\
&= c(\sigma,\tau,\boldsymbol{x})\exp\left[-\frac{1}{2v(\sigma,\tau)}\left(\mu-\frac{n\bar{x}\tau^2+\nu\sigma^2}{n\tau^2+\sigma^2}\right)^2\right]
\end{aligned}
$$

All the exponent that does not depend on μ is thrown into the multiplicative constant $c(\sigma,\tau,\boldsymbol{x})$.

But the only distribution this can be is Normal, with variance $v(\sigma,\tau)=\sigma^2\tau^2/(n\tau^2+\sigma^2)$, and with mean

$$
\bar{x}\frac{\tau^2}{\tau^2+\sigma^2/n}+\nu\frac{\sigma^2/n}{\tau^2+\sigma^2/n}.
$$

- If the goal is minimize

$$\mathbb{E}[(\mu - T)^2 \mid \boldsymbol{x}]$$

  among statistics T depending on $\boldsymbol{x}$, then the minimizer is $T = \mathbb{E}[\mu \mid \boldsymbol{x}]$.
- In this case, the Bayes estimator is

$$\bar{x}\frac{\tau^2}{\tau^2 + \sigma^2/n} + \nu\frac{\sigma^2/n}{\tau^2 + \sigma^2/n}$$

- This is a convex combination of the data-only estimator $\bar{X}$ and the prior mean ν. The weight of $\bar{X}$ tend to 1 as $n \to \infty$.
- Other inferences are possible, e.g. *credible intervals*, so we can find a and b so that

$$\mathbb{P}(a < \mu < b \mid \boldsymbol{x}) = 0.95\,.$$