Example 1: Estimation of unknown Normal mean

Dimitris Korobilis
University of Essex

Barcelona GSE
25 June 2018

Example of Bayesian inference

In many simple cases calculation of the posterior is "easy"
$\rightarrow$ e.g. likelihood and prior in the exponential family

Mean of a Normal distribution

Assume you observe N observations y and you think they come from a normal distribution with variance 1 but unknown mean θ.

- Likelihood: $y|\mu \sim N(\theta, 1)$
- Prior: $\theta \sim N(\underline{\mu}, \underline{\tau}^2)$
- Posterior: ???

Parameters with an under-line are known with certainty: $\underline{\mu}, \underline{\tau}^2$ are *prior hyperparameters* which we need to choose.
From basic statistics you know that a consistent, unbiased point estimate of θ is $\widehat{\theta} = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$.

Example of Bayesian inference

The prior $\theta \sim N(\underline{\mu}, \underline{\tau}^2)$ is of the form

$$p(\theta) = \frac{1}{\sqrt{2\pi\underline{\tau}^2}} e^{-\frac{(\theta-\underline{\mu})^2}{\underline{\tau}^2}} \tag{1}$$

and the likelihood is of the form:

$$\mathcal{L}(\theta) = p(y|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\sum (y_i-\theta)^2} \tag{2}$$

Calculation of the posterior is straightforward given the property of the exponential function: $e^a \times e^b = e^{a+b}$.
This gives the posterior distribution

$$p(\theta|y) \sim N((1+\underline{\tau}^2)^{-1}(\underline{\mu}+\underline{\tau}^2\bar{y}), (1+\underline{\tau}^2)^{-1}\underline{\tau}^2) \tag{3}$$

where $\bar{y}$ is the sample mean of y.

Conjugate priors

The first decision we need to do is about the family of the prior distribution, the second decision is about the hyperparameters of this prior distribution

Choice of a well defined prior distribution is detrimental, however, applied scientists usually work with *conjugate priors*.

Conjugate Prior

A family $\mathcal{F}$ of probability distributions on Θ (the support of the parameters θ) is said to be conjugate (or closed under sampling) for a likelihood function $\mathcal{L}(y|\theta)$ if the posterior distribution $p(\theta|x)$ also belongs to $\mathcal{F}$.

Conjugate priors

Natural conjugate priors for some common exponential families

$p(y\|\theta)$	$p(\theta)$	$p(\theta\|y)$
Normal $N\left(\theta, \sigma^2\right)$	*Normal* $N\left(\mu, \tau^2\right)$	*Normal* $N\left(\xi\left(\sigma^2\mu + \tau^2\bar{y}\right)\xi\sigma^2\tau^2\right)$ $\xi^{-1} = \sigma^2 + \tau^2$
Gamma $gamma(\nu, \theta)$	*Gamma* $gamma(\alpha, \beta)$	*Gamma* $gamma(\nu + \alpha, \beta + \bar{y})$
Binomial $bin(n, \theta)$	*Beta* $beta(\alpha, \beta)$	*Beta* $beta(\alpha + \bar{y}, \beta + n - \bar{y})$
Normal $N(\mu, 1/\theta)$	*Gamma* $gamma(\alpha, \beta)$	*Gamma* $gamma\left(\alpha + \frac{1}{2}, \beta + \frac{(\mu-\bar{y})^2}{2}\right)$

How to choose prior hyperparameters?

Remember the example of the unknown Normal mean with known variance.
When no prior information is available, we can set $\theta \sim N(0, 10000)$ or $\theta \sim U(-\infty, +\infty)$

$$p(\theta|y) \sim N\left(\frac{10000}{10001}\bar{y}, \frac{10000}{10001}\right) \equiv N(\bar{y}, 1) \tag{4}$$

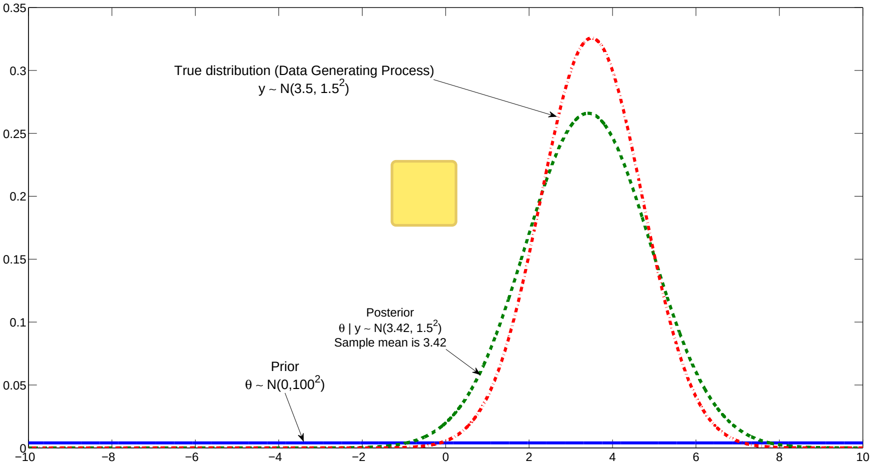
which means that the posterior is equal to the likelihood (the prior has no information for θ)
In general, for a uniform prior it holds:

$$p(\theta|X) \propto p(X|\theta)p(\theta) \equiv p(X|\theta) \times c \tag{5}$$

since $p(\theta) = U(-\infty, +\infty)$ means that each possible value θ in the support $(-\infty, +\infty)$ has equal prior probability of being the "true" value of θ.

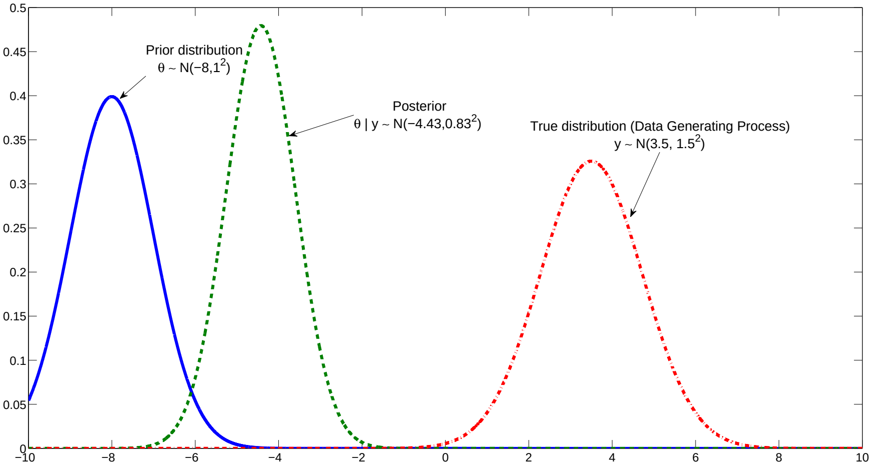
Noninformative prior

Note: True DGP $N(3.5, 1.5^2)$ is for data y, θ is "generated" from $N(3.5, 0)$



“Opinionated” informative prior

Note: True DGP $N(3.5, 1.5^2)$ is for data y, θ is “generated” from $N(3.5, 0)$



"Spot-on" informative prior

Note: True DGP $N(3.5, 1.5^2)$ is for data y, θ is "generated" from $N(3.5, 0)$