Conjugate Priors

Case of Normal

Recall Normal density $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$E[x] = \mu \qquad \sigma_x^2 = \sigma^2 \qquad x \sim N(\mu, \sigma^2)$

If $x_1, x_2, \ldots, x_n$ are independent

and $x_i \sim N(\mu_i, \sigma_i^2)$

then $\sum_{i=1}^{n} a_i x_i \sim N\left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2\right)$

[sum of independent normals is normal]

Consider the problem

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

σ^2 is known

μ is unknown

that's why we are conditioning on it

$$x/\mu \sim N(\mu, \sigma^2)$$

When we don't know something we assign a prior to it.

Let $f(\mu)$ be our prior.

What can you say about μ if you are observing x?

$$\boxed{f(\mu|x) = \frac{f(x|\mu)\,f(\mu)}{f(x) = \int f(x|\mu) f(\mu)\, d\mu}}$$

Difficulty: Denominator !

From that point on, you can find

$P(\mu > \mu_0 | x)$, $E[\mu|x]$, etc...

How to compute the integral ? [it could be hard]

Solution: Choose $f(\mu)$ to make the math work !

Is this a limitation ?

Justification for solution above:

- Choice of prior does not matter in practice if we have a large amount of observations.

- [typical] Choose a prior that will produce a posterior of the same form as the prior. [Conjugate]

(update parameters without changing the model)

prior (original parameters)

observation

$$f(\mu|x) = \frac{f(x|\mu)\,f(\mu)}{\int f(x|\mu)\,f(\mu)\,d\mu}$$

Say $f(\mu) = \dfrac{1}{\pi\left[1+(\mu-\beta)^2\right]}$ Caushy

Denominator: $$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\pi\left(1+(\mu-\beta)^2\right)}\, d\mu$$

Observation: If $\mu \sim N(\beta, \tau^2)$, then $\mu | x \sim N(\sim, \sim)$

new parameters

Since posterior will have same form as prior, we say the prior is Conjugate

$$f(\mu|x) = \frac{f(x|\mu)\, f(\mu)}{?}$$

Normal

Form 2 ____________ = ____________ Form 1

prior is Conjugate to Normal

Normal is Conjugate to Normal

Conjugate prior means: You figured out the integral!

$$f(\mu|x) = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{(\mu-\beta)^2}{2\tau^2}}}{?}$$

Drop constants things that don't depend on μ

$$\propto e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot e^{-\frac{(\mu-\beta)^2}{2\tau^2}}$$

$$f(\mu|x) \propto f(x|\mu) f(\mu)$$

We say $f(x) \propto g(x)$ iff $f(x) = c \cdot f(x)$

"proportional to"

constant

$$f(\mu|x) \propto e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot e^{-\frac{(\mu-\beta)^2}{2\tau^2}}$$

$$= e^{-\frac{\tau^2(x-\mu)^2 + \sigma^2(\mu-\beta)^2}{2\sigma^2\tau^2}}$$

$$= e^{-\frac{\tau^2(x^2+\mu^2-2x\mu) + \sigma^2(\mu^2+\beta^2-2\mu\beta)}{2\sigma^2\tau^2}}$$

$$= e^{-\frac{(\sigma^2+\tau^2)\mu^2 - 2\mu(\tau^2 x + \sigma^2\beta) + }{2\sigma^2\tau^2}}$$

$$\propto e^{-\frac{(\sigma^2+\tau^2)\mu^2 - 2\mu(\tau^2 x + \sigma^2\beta)}{2\sigma^2\tau^2}}$$

$$= e^{-\frac{\mu^2 - 2\mu\left(\frac{\tau^2 x + \sigma^2 \beta}{\sigma^2+\tau^2}\right)}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}}$$

$$\propto e^{-\frac{\left[\mu - \frac{\tau^2 x + \sigma^2\beta}{\sigma^2+\tau^2}\right]^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}}$$

$$= e^{-\frac{(\mu - m)^2}{2V^2}}$$

where $m = \dfrac{\tau^2 x + \sigma^2\beta}{\sigma^2+\tau^2}$ $\qquad V^2 = \dfrac{\sigma^2\tau^2}{\sigma^2+\tau^2}$

$$\mu | x \sim N\left(\frac{\tau^2 x + \sigma^2\beta}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$$

More data: $x_1, x_2, \ldots, x_n \sim$ i.i.d. Conditioned on μ

$$f(\mu \mid x_1, \ldots, x_n) \propto f(x_1, \ldots, x_n \mid \mu)\, f(\mu)$$

$$= \prod_{i=1}^{n} f(x_i \mid \mu)\, f(\mu)$$

$$\propto \prod_{i=1}^{n} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \cdot e^{-\frac{(\mu-\beta)^2}{2\tau^2}}$$

What if we look at $\bar{x} = \sum_{i=1}^{n} x_i \big/ n$ (average)

$\bar{x} \mid \mu \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ [Sum of independent Normals]

$$f(\mu|\bar{x}) \propto \underbrace{f(\bar{x}|\mu)}_{N(\mu,\, \sigma^2/n)} \underbrace{f(\mu)}_{N(\beta,\, \tau^2)}$$

$$\mu|\bar{x} \sim N\left(\frac{\sigma^2\beta/n + \tau^2\bar{x}}{\sigma^2/n + \tau^2},\ \frac{\sigma^2\tau^2/n}{\sigma^2/n + \tau^2} \right)$$

If we work on the problem $f(\mu|x_1, \ldots, x_n)$ we will find $f(\mu|x_1, \ldots, x_n) = f(\mu|\bar{x})$.

$$n \to \infty : \ f(\mu|\bar{x}) \sim N\left(\bar{x}, \sigma^2/n\right)$$

Prior "doesn't matter"

Observe $\bar{x} = x_0$ and let y be such that

$$|y - x_0| \gg \varepsilon$$

$$\frac{f_\mu(x_0 \mid \bar{x} = x_0)}{f_\mu(y \mid \bar{x} = x_0)} = \frac{f(x_0 \mid \mu = x_0)\, f_\mu(x_0)}{f(x_0 \mid \mu = y)\, f_\mu(y)}$$

Assume $\dfrac{f_\mu(x_0)}{\max\limits_y f_\mu(y)} \geq c$

$$\geq c \cdot \frac{f(x_0 \mid \mu = x_0)}{f(x_0 \mid \mu = y)} \frac{2\varepsilon}{2\varepsilon} = c \cdot \frac{P(|\bar{x} - x_0| \leq \varepsilon \mid \mu = x_0)}{P(|\bar{x} - x_0| \leq \varepsilon \mid \mu = y)} \longrightarrow \frac{1}{0}$$

Chebychev $\quad P(|\bar{x} - \mu| \geq \varepsilon) \leq \dfrac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$