Quantifying Uncertainty

Sai Ravela

M. I. T

2012

Parametric forms

If we can produce a (posterior) distribution and sample from it, we are done (more or less). In certain cases, where the prior is a conjugate prior, then sampling the posterior is nothing but sampling an updated prior.

We are interested in parameterized distributions as a start. Then we will move on to mixture and nonparametric densities thereafter.

There are numerously parameterized distributions, we focus on the versatile exponential family

Examples

Type	Form	Domain
Gaussian	$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$	$x \in \mathbb{R}$
Bernoulli	$p(x) = \alpha^x (1-\alpha)^{1-x}$	$x \in \{0, 1\}$
Binomial	$\binom{n}{x} \alpha^x (1-\alpha)^{n-x}$	$x \in \{0, 1, 2, \ldots, n\}$
Exponential	$p(x) = \lambda e^{-\lambda} x$	$x \in \mathbb{R}^+$
Poison	$p(x) = \frac{e^{-\lambda}}{x!} \lambda^x$	$x \in \{0, 1, 2, \ldots\}$
Dirichlet	$p(x) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i x_i^{\alpha_i - 1}$	$x_i \in [0, 1],\ \sum_i x_i = 1$

Also includes Beta, Gamma, χ^2, Weibull, Wishart, Inverse Wishart and many others ...

What is Special?

$$p(x) = h(x)e^{\theta^T T(x) - A(\theta)}$$

Where,

θ is a vector of parameters (the natural parameter)

T(x) is a vector of sufficient statistics

$A(\theta)$ is cumulant

There is only one cross-term relating parameters and sufficient statistics, which will come in handy.

Cumulant

$$\int p(x)dx = e^{-A(\theta)} \int h(x)e^{\theta^T T(x)}dx = 1$$

$$A(\theta) = \ln\left(\int h(x)e^{\theta^T T(x)}dx\right)$$

The log of the Laplace transform of $h(x)$!

Bernoulli

$$\theta = \ln \frac{\alpha}{1-\alpha}$$
$$T(x) = x$$
$$A(\theta) = \ln(1 + e^{\theta})$$

$$\begin{aligned}
p(x) &= h(x)e^{\theta^T T(x) - A(\theta)} \\
&= e^{\left(\ln \frac{\alpha}{1-\alpha} x - \ln\left(1 + \frac{\alpha}{1-\alpha}\right)\right)} \\
&= e^{\left(\ln \frac{\alpha}{1-\alpha} x + \ln(1-\alpha)\right)} \\
&= e^{\ln(\alpha^x (1-\alpha)^{1-x})}
\end{aligned}$$

Gaussian

$$\theta = \begin{pmatrix} \sum^{-1} \mu \\ -\frac{1}{2}\Sigma^{-1} \end{pmatrix}$$

$$T(x) = \begin{pmatrix} x \\ xx^T \end{pmatrix}$$

$$h(x) = (2\pi)^{-D/2}$$

How to fit?

$$\frac{dA\theta}{d\theta} = E_\theta[T(x)]$$
$$\frac{d^2A(\theta)}{d\theta^2} = Cov(T(x)) \geq 0$$

Therefore, A(θ) is convex in θ.

Mechanism

$$I(\theta) = \sum \ln p(x_i|\theta) = \sum \ln h(x_i) + \theta^T T(x_i) - A(\theta)$$

$$\frac{dI}{d\theta} = 0 \rightarrow \frac{1}{N} \sum_i T(x_i) = \frac{dA}{d\theta}$$

There is a unique global solution for θ and, further, it may be calculated from sample statistics!

This normal equation will have to be solved iteratively; e.g. newton-raphson, and other approaches.

Density Estimation

The products of exponential families are in exponential families? (Yes, but not easy)

It is useful when prior and posterior have same form in density estimation.

$$P(\theta|x) \propto p(x|\theta)p(\theta)$$

Then, for example, one might iteratively update the distribution's parameters as new measurements come in. Sampling from posterior just as sampling from prior.

Conjugate Priors

$$
\begin{aligned}
P(\theta|x) \leftarrow & \; p(x|\theta) & \leftarrow p(\theta) \\
Beta \leftarrow & \; bernoulli & \leftarrow Beta \\
Bernoulli \leftarrow & \; Beta & \leftarrow Bernoulli \\
Poisson \leftarrow & \; Gamma & \leftarrow Gamma
\end{aligned}
$$

$$
P(\alpha) = \frac{1}{B(\gamma, \Delta)} \alpha^{\gamma-1}(1-\alpha)^{\Delta-1}
$$

$$
P(x|\alpha) = \alpha^{x}(1-\alpha)^{1-x}
$$

Derivation for $p(\alpha|x)$? (Hint: you'll need $p(x) = \sum p(x|\alpha)p(\alpha)$)

Kernel Density

We have a collection of (high dimensional) points, and want to create a density function.

Kernel methods to the rescue! (tongue in cheek)

$$p(x) = \frac{1}{N}\sum_{i=1}^{N} K(x - x_i)$$

x_i are basis or collocation points.
K is an interpolation function, e.g. (no!)

What type of K?

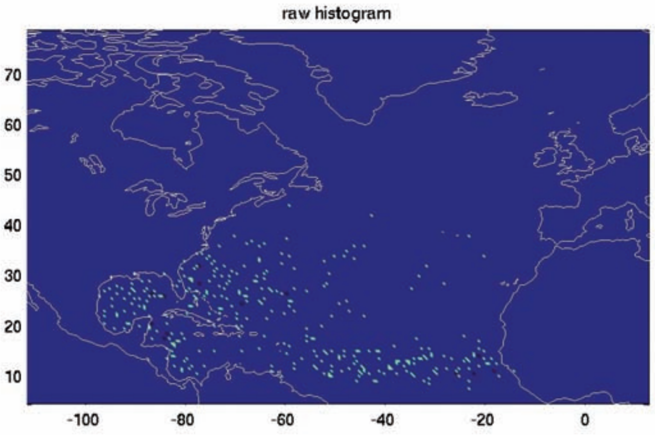
Natural scale parameterization

$$K(x - x_i; h) = exp(-||x - x_i||/2h^2)$$

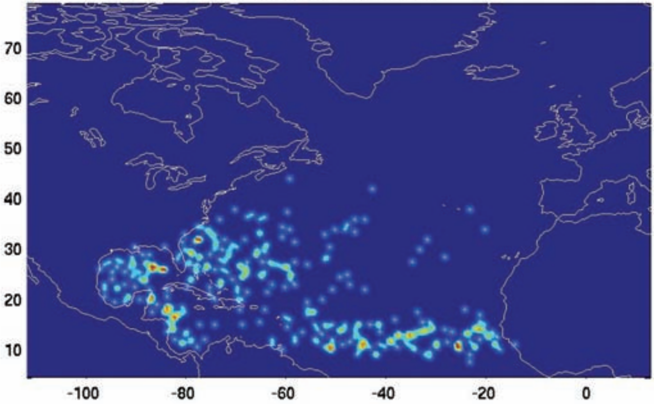
adaptive

$$h \sim h(x)$$

Example



Vanilla KDE



“Balloon’d” KDE