08 - Approximate Bayesian Computation - 01
Laplace Approximation

Master in Foundations of Data Science
Bayesian Statistics and Probabilistic Programming
Fall 2018-2019

Josep Fortiana

Monday, November 12, 2018

Universitat de Barcelona

08 - ABC - 01

Approximate Bayesian Computation *(ABC)*

Laplace approximation (univariate)

Bivariate normal distribution

Multivariate normal distribution

Laplace approximation (multivariate)

08 - ABC - 01

Approximate Bayesian Computation *(ABC)*

Laplace approximation (univariate)

Bivariate normal distribution

Multivariate normal distribution

Laplace approximation (multivariate)

Methods for approximate Bayesian inference

When a pdf, e.g., posterior, predictive, etc. cannot be expressed as a known function.

Approximate it,

- Deterministic methods: Laplace approximation, variational Bayes, expectation propagation.
- Stochastic methods: Monte Carlo (independent or MC)

Laplace approximation

Approximate target posterior pdf $h(\theta|x)$ with a gaussian (normal) pdf.

Tool: Taylor expansion of $h(\theta|x)$ up to second order.

Variational Bayesian inference

Choose a family of pdf's.

Find the member of this family that is closer to the target.

Closeness being measured by Kullback-Leibler divergence.

Symposium

Symposium on Advances in Approximate Bayesian Inference.

Montreal, December 2, 2018

08 - ABC - 01

Approximate Bayesian Computation *(ABC)*

Laplace approximation (univariate)

Bivariate normal distribution

Multivariate normal distribution

Laplace approximation (multivariate)

Univariate Laplace approximation

Target function $h(\theta)$, e.g. a pdf, with a maximum.

$$q(\theta) \stackrel{\text{def}}{=\!=} \log h(\theta).$$

Taylor expansion of $q(\theta)$ around a (global) maximum:

$$\theta_0 = \arg\max_\theta q(\theta).$$

For a posterior pdf, θ_0 is the MAP estimator.

Taylor expansion

$$q(\theta) = q(\theta_0) + (\theta - \theta_0)\cdot q'(\theta_0) + \tfrac{1}{2}\,(\theta-\theta_0)^2\cdot q''(\theta_0) + \cdots$$

$$\approx q(\theta_0) - \tfrac{1}{2}\,(\theta-\theta_0)^2\cdot |q''(\theta_0)|,$$

since $q'(\theta_0) = 0$ and $q''(\theta_0) < 0$ at the maximum.

$$h(\theta) = \exp(q(\theta)) \approx A\cdot\exp\left\{-\frac{1}{2}\left(\frac{\theta-\theta_0}{\sigma}\right)^2\right\},$$

a Gaussian pdf, $\mathrm{N}(\theta_0, \sigma^2)$, with $\sigma^2 = \dfrac{1}{|q''(\theta_0)|}$.

08 - ABC - 01

Approximate Bayesian Computation *(ABC)*

Laplace approximation (univariate)

Bivariate normal distribution

Multivariate normal distribution

Laplace approximation (multivariate)

Univariate normal distribution

A r.v. X has a normal (gaussian) distribution with parameters μ and σ^2, $X \sim \mathrm{N}(\mu, \sigma^2)$, if it is absolutely continuous, with pdf:

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

Bivariate normal with independent marginals

Two independent r.v. , $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathrm{N}(\mu_y, \sigma_y^2)$.

The random vector (X, Y), has pdf: $f_{(X,Y)}(x, y) =$

$$\frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

General bivariate normal pdf

(X, Y) is *nonsingular* bivariate normal if its distribution is absolutely continuous on $\mathbb{R}^2$ with joint pdf:

$$C \exp(-Q/2),$$

where Q is a positive definite quadratic form:

$$Q = a_{11}(x - \mu_x)^2 + 2a_{12}(x - \mu_x)(y - \mu_y) + a_{22}(y - \mu_y)^2.$$

Elementary algebra relates a_{11}, a_{12}, a_{22}, to known,

General bivariate normal pdf

Writing Q as a sum of squares:

$$
\begin{aligned}
Q &= a_{22}\left((y-\mu_y)+\frac{a_{12}}{a_{22}}(x-\mu_x)\right)^2+\left(a_{11}-\frac{a_{12}^2}{a_{22}}\right)(x-\mu_x)^2,\\
&= a_{22}\,u^2+\left(a_{11}-\frac{a_{12}^2}{a_{22}}\right)\,v^2.
\end{aligned}
$$

u and v appear as independent, centered normal variables, and:

$$
\operatorname{var}(u)=1/a_{22},\quad \operatorname{var}(v)=a_{22}/\Delta,\quad \Delta=a_{11}a_{22}-a_{12}^2.
$$

General bivariate normal pdf

As $v = x - \mu_x$, $\mathrm{E}(x) = \mu_x$ and $\sigma_x^2 \equiv \mathrm{var}(x) = \mathrm{var}(v) = a_{22}/\Delta$.

By symmetry, $\mathrm{E}(y) = \mu_y$ and $\sigma_y^2 \equiv \mathrm{var}(y) = a_{11}/\Delta$.

By bilinearity of $\mathrm{cov}(\,\cdot\,,\,\cdot\,)$, from the definitions of u and v,

$$0 = \mathrm{cov}(u, v) = \mathrm{cov}(x, y) + \frac{a_{12}}{a_{22}}\,\sigma_x^2.$$

As a function of $\rho \equiv \mathrm{cov}(x, y)/(\sigma_x \sigma_y)$:

$$\mathrm{cov}(x, y) = \rho\sigma_x\sigma_y = -a_{12}/\Delta.$$

General bivariate normal pdf

Finally, we can express Q as a function of $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$:

$$a_{11} = \frac{1}{\sigma_x^2(1-\rho^2)}, \quad a_{22} = \frac{1}{\sigma_y^2(1-\rho^2)}, \quad a_{12} = -\frac{\rho}{\sigma_x\sigma_y(1-\rho^2)},$$

$$Q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]$$

General bivariate normal pdf

An absolutely continuous random vector (X, Y) is bivariate normal with parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ if its pdf is:

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \times\right.$$

$$\left.\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\}$$

Both marginal pdf's of such vector are univariate normal.

Standardization

As a function of the standardized vector (z_x, z_y),

$$z_x = \frac{x - \mu_x}{\sigma_x}, \qquad z_y = \frac{y - \mu_y}{\sigma_y},$$

the pdf is:

$$f(z_x, z_y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{z_x^2 + z_y^2 - 2\rho z_x z_y}{2(1-\rho^2)}\right\},$$

Contour levels of the pdf

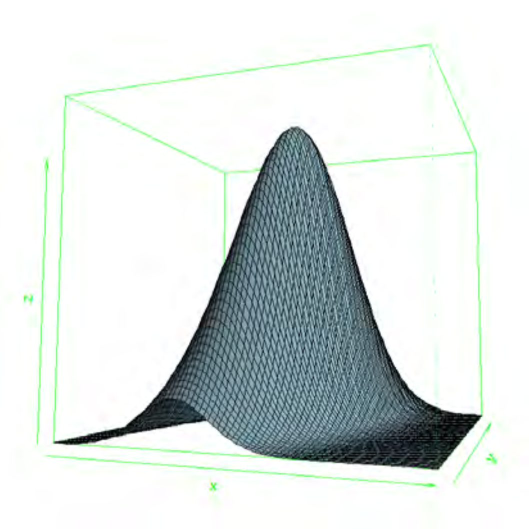
Contour lines in the bivariate normal pdf graph, i.e, curves along which the probability density is constant:

$$Q(x, y) = \text{const.}$$

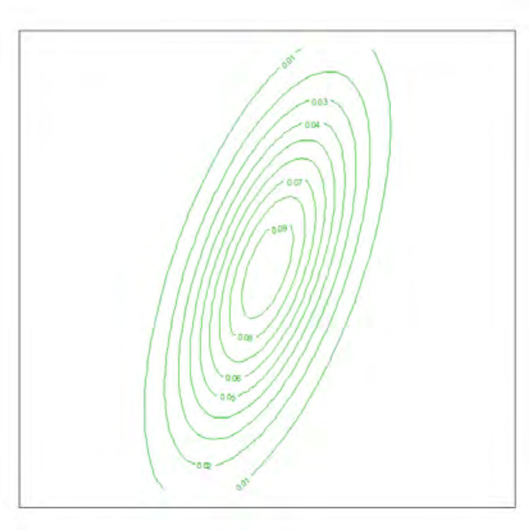
are ellipses.

It is possible to write their equation in the canonical form: major and minor axes, excentricity, angle of the principal coordinate system with respect to the usual

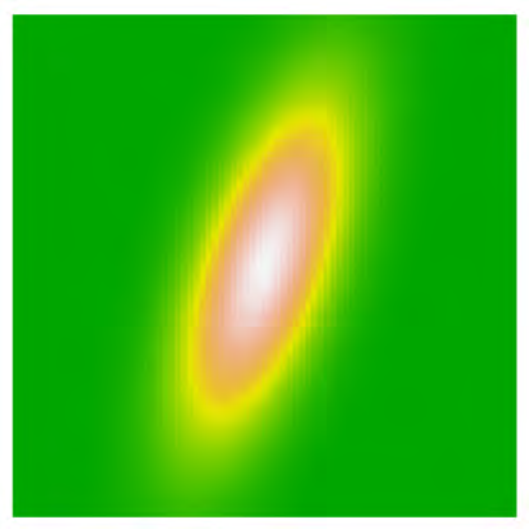
3D plot of a bivariate normal pdf



Contour level plot of a bivariate normal pdf



With another graphical design



Conditional pdf

The conditional pdf of $(Y|X = x)$ is normal:

$$c \ \exp\left\{-\frac{a_{22}}{2}\left((y-\mu_y)+\frac{a_{12}}{a_{22}}(x-\mu_x)\right)^2\right\}$$

$$= c \ \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left((y-\mu_y)+\frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)\right)^2\right\}.$$

Conditional pdf

The conditional expectation $\mu_{y|x} \equiv \mathrm{E}(Y|X=x)$ is:

$$\mu_{y|x} = \mu_y + \frac{\rho\,\sigma_y}{\sigma_x}\,(x - \mu_x),$$

which we can also write:

$$\mu_{y|x} = \beta_0 + \beta_1\,x, \quad \text{where } \beta_1 = \frac{\rho\,\sigma_y}{\sigma_x}, \quad \beta_0 = \mu_y - \beta_1\,\mu_x.$$

and the conditional variance:

Moments in matrix form

If (X, Y) is bivariate normal with parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$,

$$E(X, Y) = (\mu_x, \mu_y),$$

$$\mathrm{Var}(X, Y) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix},$$

$$\mathrm{cov}(X, Y) = \rho\sigma_x\sigma_y, \qquad \mathrm{cor}(X, Y) = \rho.$$

Inverse matrix of $\boldsymbol{\Sigma}$

$$\det \boldsymbol{\Sigma} = \sigma_x^2\,\sigma_y^2\,(1-\rho^2).$$

If $|\rho| \neq 1$, then $\det \boldsymbol{\Sigma} \neq 0$, and

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \dfrac{1}{\sigma_x^2} & -\dfrac{\rho}{\sigma_x\sigma_y} \\ -\dfrac{\rho}{\sigma_x\sigma_y} & \dfrac{1}{\sigma_y^2} \end{pmatrix},$$

The quadratic form in matrix notation

In the exponent of the pdf,

$$-\frac{1}{2}\,Q(x,y),$$

assuming $|\rho| \neq 1$, the quadratic form is:

$$Q(x,y) = \boldsymbol{u}' \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{u}, \quad \text{where } \boldsymbol{u} = \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}.$$

The pdf in matrix notation

The bivariate normal pdf becomes:

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{ -\frac{1}{2}\, \boldsymbol{u}' \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{u} \right\},$$

$$\text{where } \boldsymbol{u} = \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix}.$$

NB: Expression valid for any p-variate normal, $(p \geq 2)$.

Cautionary remarks

1. There are *singular* bivariate normal distributions. They are not absolutely continuous (with respect to the natural Lebesgue measure in $\mathbb{R}^2$).
2. In particular, if both F and G are univariate normal, the upper and lower Fréchet bounds, $H_+(F, G)$ and $H_-(F, G)$, are singular bivariate normal distributions, with the whole probability concentrated on a line –the regression line with, respectively, $\rho = +1$ and $\rho = -1$.
3. There exist bivariate distributions whose marginals are univariate normal but they themselves are not bivariate normal.

08 - ABC - 01

Approximate Bayesian Computation *(ABC)*

Laplace approximation (univariate)

Bivariate normal distribution

Multivariate normal distribution

Laplace approximation (multivariate)

A definition of the p-variate normal

A p-dimensional random vector $\boldsymbol{x}$, which we write as a $p \times 1$ column, follows a p-variate normal distribution if:

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{A} \cdot \boldsymbol{u},$$

where $\boldsymbol{u}$ is a a $q \times 1$ column random vector, $q \geq 1$, whose components are $\mathrm{N}(0, 1)$ r.v. , $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{A}$ is a $p \times q$ matrix of real numbers.

As a result, $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{\mu}$ and $\boldsymbol{\Sigma} \equiv \mathrm{Var}(\boldsymbol{x}) = \boldsymbol{A} \cdot \boldsymbol{A}'$.

First properties of a p-variate normal distribution

- $\boldsymbol{\Sigma}$ is a $p \times p$ symmetric, positive semidefinite matrix. It can be shown that $(\boldsymbol{\Sigma}, \boldsymbol{\mu})$ are uniquely determined, thus we write:

$$\boldsymbol{x} \sim \mathbf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

  But $\boldsymbol{A}$ is any $p \times q$ matrix such that $\boldsymbol{A} \cdot \boldsymbol{A}' = \boldsymbol{\Sigma}$. Also q is not unique.

- If $\boldsymbol{\Sigma}$ is non-singular, then necessarily $q \geq p$ and $\boldsymbol{x}$ has an absolutely continuous distribution. Otherwise it is singular.

p-variate normal pdf

In the non-singular case, the pdf is:

$$f(\boldsymbol{x}) = (2\,\pi)^{-p/2}\,|\det \boldsymbol{\Sigma}|^{-1/2}\,\exp\left\{-\tfrac{1}{2}\,Q(\boldsymbol{x})\right\}, \qquad \boldsymbol{x} \in \mathbb{R}^p,$$

where $Q(\boldsymbol{x})$ is the quadratic form:

$$Q(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu})' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{x} - \boldsymbol{\mu}).$$

Linear transformations

If $\boldsymbol{x} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, given $\boldsymbol{B}$, an $m \times p$ matrix of real numbers, the distribution of the vector:

$$\boldsymbol{y} = \boldsymbol{B} \cdot \boldsymbol{x}$$

is an m-dimensional Gaussian, with:

$$\operatorname{Var}(\boldsymbol{y}) = \boldsymbol{B} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{B}', \quad \operatorname{E}(\boldsymbol{y}) = \boldsymbol{B} \cdot \boldsymbol{\mu}.$$

Linear transformations

In particular, all marginals from a multivariate normal are normal.

That is, any vector formed by a subset of $\boldsymbol{x}$'s components is gaussian.

Its variances and covariances matrix is obtained by selecting the relevant rows and columns in $\boldsymbol{\Sigma}$.

Incorrelation and independence

In general, independent random variables or vectors are also incorrelated but the converse statement is false.

However, in the case of components of random vectors whose joint distribution is multivariate normal, incorrelation implies independence.

This follows from the fact that the covariances matrix completely determines the (joint) normal multivariate distribution.

08 - ABC - 01

Approximate Bayesian Computation *(ABC)*

Laplace approximation (univariate)

Bivariate normal distribution

Multivariate normal distribution

Laplace approximation (multivariate)

Multivariate Laplace approximation

Target function $h(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is now p-dimensional,

$$q(\boldsymbol{\theta}) \stackrel{\text{def}}{=\!=} \log h(\boldsymbol{\theta}).$$

Taylor expansion of $q(\boldsymbol{\theta})$ around a (global) maximum:

$$\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}} q(\boldsymbol{\theta}).$$

When $h(\,\cdot\,)$ is a posterior pdf, $\boldsymbol{\theta}_0$ is the MAP estimator.

Taylor expansion

$$
\begin{aligned}
q(\boldsymbol{\theta}) &= q(\boldsymbol{\theta}_0) + \dot{q}(\boldsymbol{\theta}_0)\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \\
&\qquad + \tfrac{1}{2}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\cdot\ddot{q}(\boldsymbol{\theta}_0)\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\cdots \\
&\approx q(\boldsymbol{\theta}_0) - \tfrac{1}{2}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\cdot(-\ddot{q}(\boldsymbol{\theta}_0))\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}_0),
\end{aligned}
$$

since the gradient $\dot{q}(\boldsymbol{\theta}_0)=0$ and the Hessian $\ddot{q}(\boldsymbol{\theta}_0)$ is negative definite at the maximum.

Notation: $\boldsymbol{\theta}$, etc., are $p\times 1$ column vectors. Derivatives are represented with dots, to avoid confusion with the prime used to indicate matrix transposition. The gradient $\dot{q}(\boldsymbol{\theta}_0)$ is a $1\times p$ row vector and the Hessian $\ddot{q}(\boldsymbol{\theta}_0)$ is a $p\times p$ symmetric psd matrix.

Approximate pdf

$$
\begin{aligned}
h(\boldsymbol{\theta}) &= \exp(q(\boldsymbol{\theta})) \\
&\approx A \cdot \exp\left\{-\tfrac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)\right\},
\end{aligned}
$$

a Gaussian pdf,

$$
\mathrm{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}),
$$

with $\boldsymbol{\Sigma} = -\ddot{q}(\boldsymbol{\theta}_0)^{-1}$.