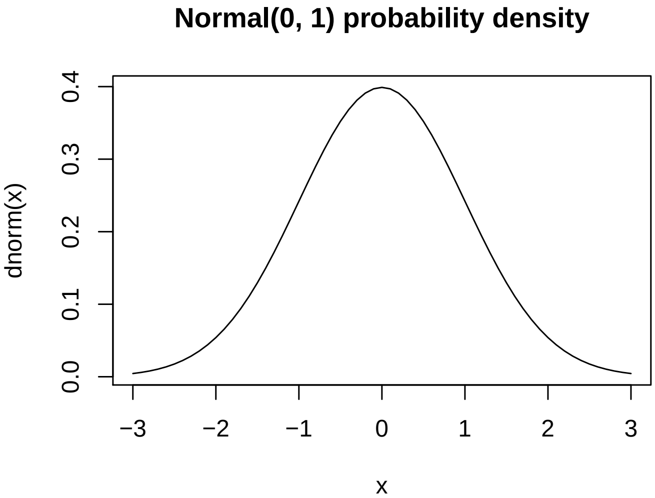
QDT: Gaussian process models

Max Joseph

October 21, 2015

Background: univariate normal

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



Background: multivariate normal

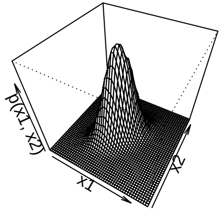
$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

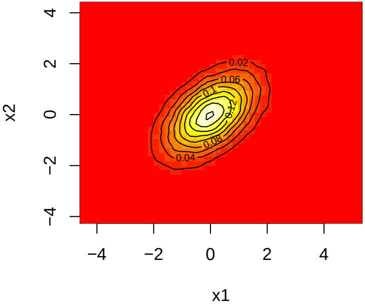
$\boldsymbol{\mu}$: vector of means

$\boldsymbol{\Sigma}$: covariance matrix

Bivariate normal probability density

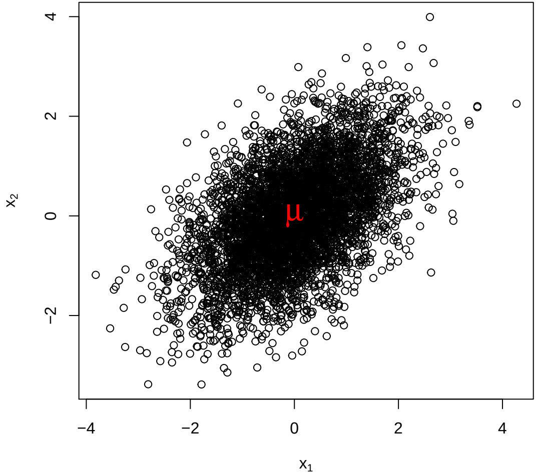
$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$





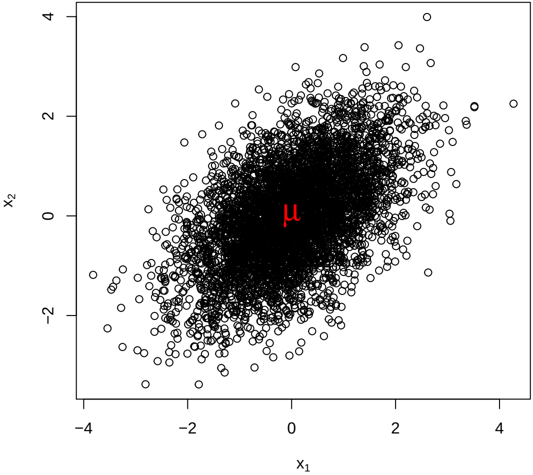
Bivariate normal parameters

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] \\ Cov[X_2, X_1] & Cov[X_2, X_2] \end{bmatrix}$$



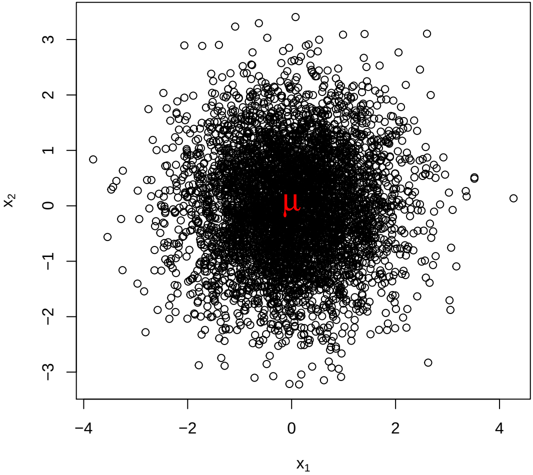
Bivariate normal parameters

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$



Uncorrelated bivariate normal

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



Common notation

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \; \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Common notation

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \; \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$Cov[X_1, X_1] = Var[X_1] = \sigma_1^2$$

$$Cov[X_1, X_2] = \rho\sigma_1\sigma_2$$

Σ must be symmetric and positive semi-definite

Relevant properties of multivariate normals

1. Marginal distributions are normal
2. Conditional distributions are normal

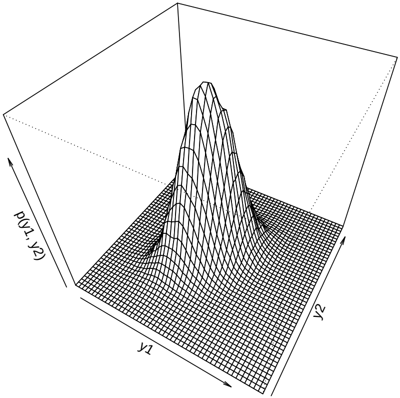
Marginals are normal

Joint distribution: $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

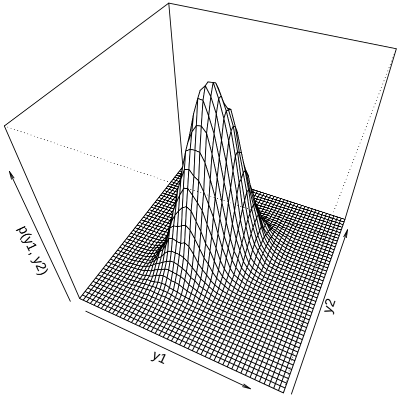
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Marginal distribution: $y_1 \sim N(\mu_1, \sigma_1^2)$

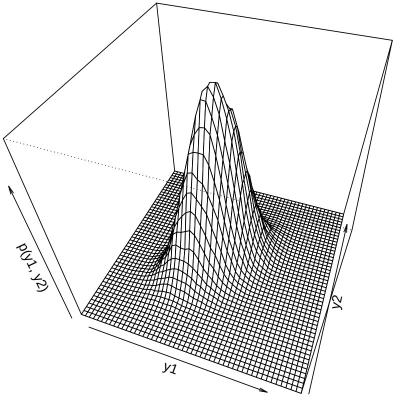
Graphical interpretation



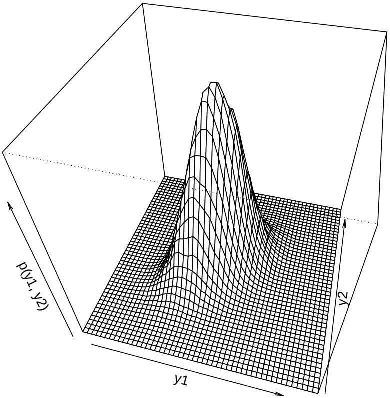
Graphical interpretation



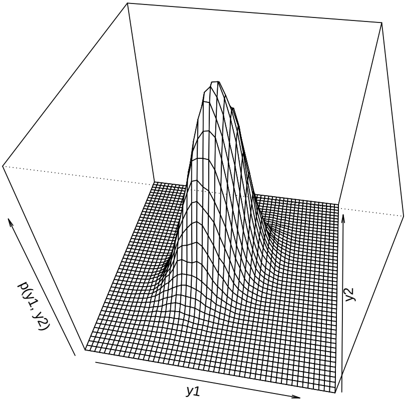
Graphical interpretation



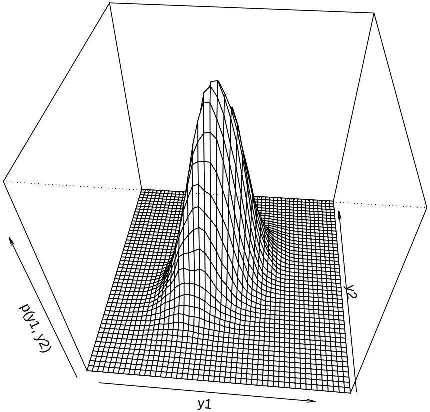
Graphical interpretation



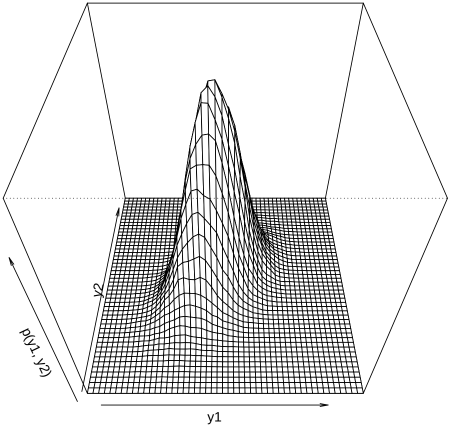
Graphical interpretation



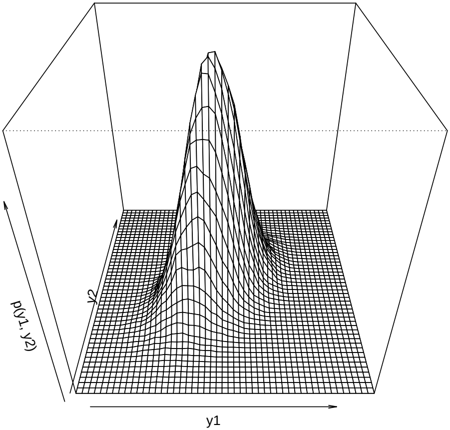
Graphical interpretation



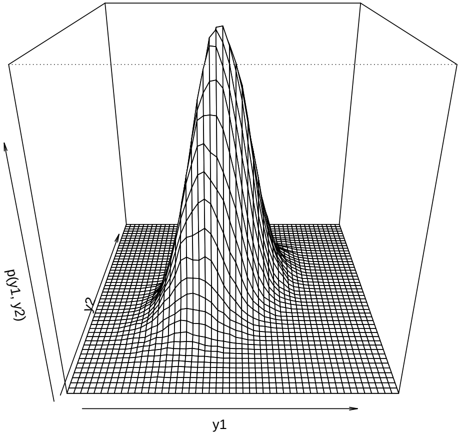
Graphical interpretation



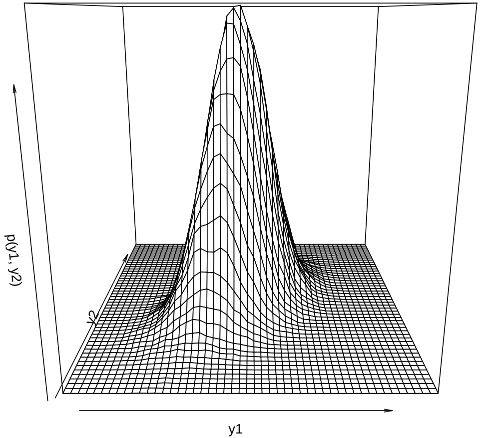
Graphical interpretation



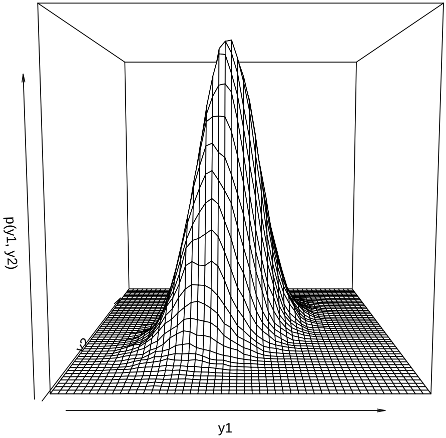
Graphical interpretation



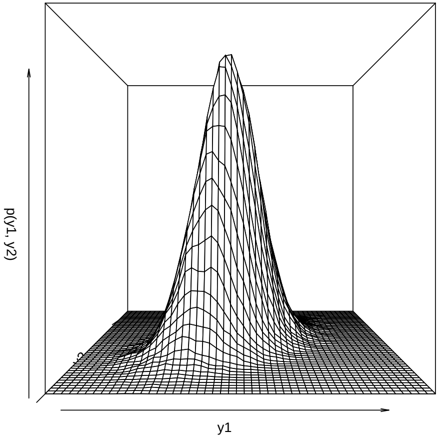
Graphical interpretation



Graphical interpretation

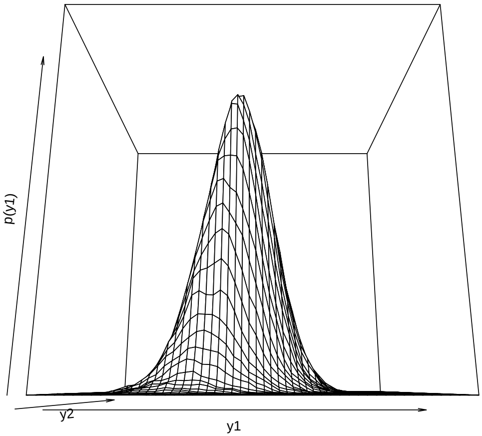


Graphical interpretation



Graphical interpretation: marginal of y_1

$N(\mu_1, \sigma_1^2)$



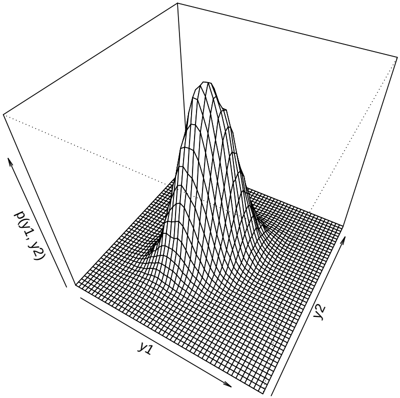
Conditional distributions are normal

Joint distribution: $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$

Conditional distribution:

$Y_1 | Y_2 = y_2 \sim N(\mu, \Sigma)$

Graphical interpretation of conditioning



Conditional of Y_1

Joint distribution: $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \right)$

Conditional distribution:

$Y_1|Y_2 \sim N(E(Y_1|Y_2), Var(Y_1|Y_2))$

$E(Y_1|Y_2) = \mu_1 + \Omega_{12}\Omega_{22}^{-1}(Y_2 - \mu_2)$

$Var(Y_1|Y_2) = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$

Now the fun stuff

Classic linear modeling

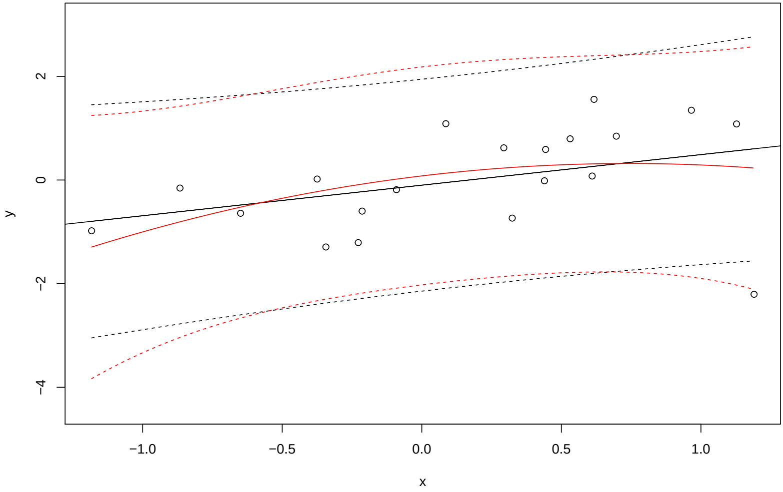
$$y = X\beta + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

Functional form determined by $X\beta$

Linear model functional forms

e.g. $y = \mu(x) + \epsilon$



Why not set a prior on $\mu(x)$?

Gaussian process as a prior for $\mu(x)$

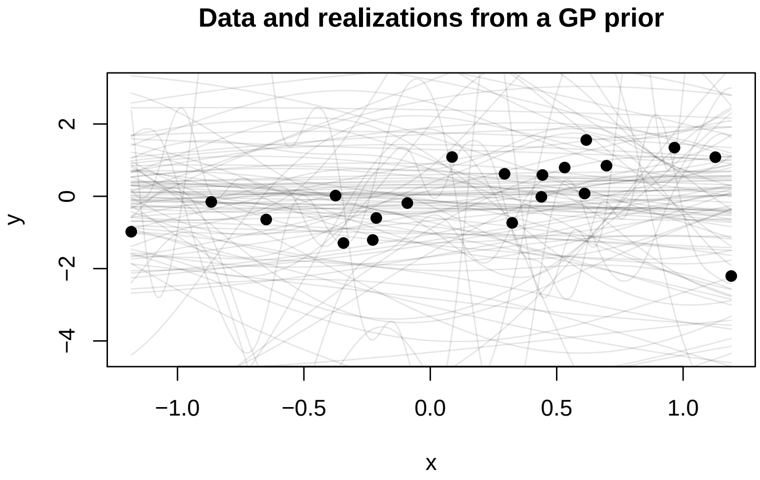
$$y \sim N(\mu(x), \sigma^2)$$

$$\mu(x) \sim GP(m, k)$$

GP prior for $\mu(x)$

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

$$\mu(x) \sim GP(m, k)$$



Wait, what's Gaussian about that?

If $\mu(x) \sim GP(m, k)$, then

$\mu(x_1), ..., \mu(x_n) \sim N(m(x_1), ..., m(x_n), K(x_1, ..., x_n)$

m and k are functions!

Mean function: m

Classic example: $m(x) = X\beta$

e.g., $\mu(x) \sim GP(X\beta, k(x))$

But, the covariance function $k(x)$ is the real star.

Covariance functions

k specifies covariance between to x values

Squared exponential covariance:

$$k(x, x') = \tau^2 exp\Big(- \frac{|x - x'|^2}{l^2} \Big)$$

Lots of options: smooth, jaggety, periodic

Example of squared exponential

$$\boldsymbol{K} = \begin{bmatrix} \tau^2 exp(-\frac{|x_1 - x_1|^2}{l^2}) & \tau^2 exp(-\frac{|x_1 - x_2|^2}{l^2}) \\ \tau^2 exp(-\frac{|x_2 - x_1|^2}{l^2}) & \tau^2 exp(-\frac{|x_2 - x_2|^2}{l^2}) \end{bmatrix}$$

Example of squared exponential

$$\boldsymbol{K} = \begin{bmatrix} \tau^2 exp(-\frac{0^2}{l^2}) & \tau^2 exp(-\frac{|x_1 - x_2|^2}{l^2}) \\ \tau^2 exp(-\frac{|x_2 - x_1|^2}{l^2}) & \tau^2 exp(-\frac{0^2}{l^2}) \end{bmatrix}$$

Example of squared exponential

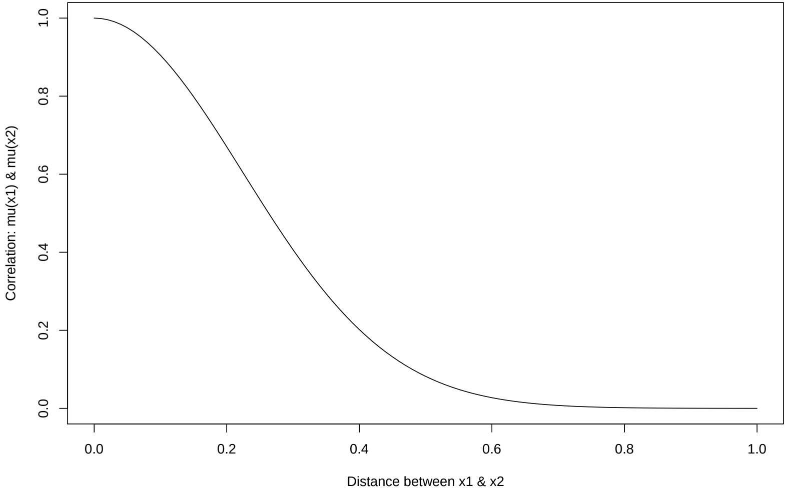
$$\boldsymbol{K} = \begin{bmatrix} \tau^2 exp(0) & \tau^2 exp(-\frac{|x_1 - x_2|^2}{l^2}) \\ \tau^2 exp(-\frac{|x_2 - x_1|^2}{l^2}) & \tau^2 exp(0) \end{bmatrix}$$

Example of squared exponential

$$\boldsymbol{K} = \begin{bmatrix} \tau^2 & \tau^2 exp(-\frac{|x_1 - x_2|^2}{l^2}) \\ \tau^2 exp(-\frac{|x_2 - x_1|^2}{l^2}) & \tau^2 \end{bmatrix}$$

$Cor(\mu(x_1), \mu(x_2)) = exp(-\frac{|x_1 - x_2|^2}{l^2})$.

Correlation function



Let's check it out

`sim.R`: GP simulations with 1d and 2d inputs

`estimate.R`: GP estimation with 1d inputs