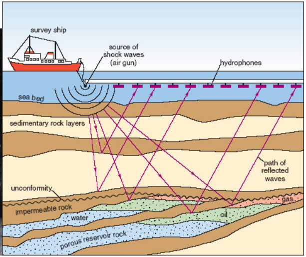
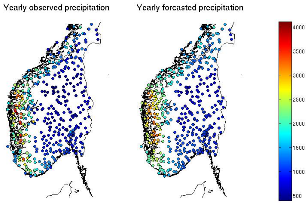
Topics - Schedule

- Jan: **Gaussian processes - applications and computations (optimization of function)**
- Jan: Gaussian Markov random fields (graphs and approximations of Gaussian processes)
- Jan: Latent Gaussian models, spatial Generalized linear mixed models
- Feb: Integrated nested laplace approximation - INLA (fast approximate Bayesian inference, examples of GLMMs)
- Feb: Template model builder (frequentist inference, examples of GLMMs)
- Feb: New Markov chain Monte Carlo methods (Bayesian inference, examples of GLMMs and complex function uncertainty quantifications)
- March: Discrete models: hidden Markov chains and Bayesian networks calculations (forward-backward /junction tree)
- March/April: Sequential Monte Carlo methods: particle filters, Ensemble Kalman filters.
- April: New clustering methods and dimension reduction techniques

Gaussian processes and applications

Gaussian processes are very commonly used in practice. Large spatial (spatio-temporal) datasets:





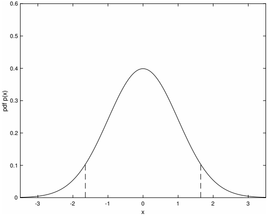
Gaussian processes and applications

Uncertainty quantification in a diverse range of applications:

- Genetic data (dependence via 'distance' in cells or in pedigrees)
- Functional data (dependence via 'distance' in covariates, use this to borrow information (smooth surface))
- Response surfaces modeling and optimization

Extremely common as building block in several machine learning applications.

Univariate Gaussian distribution



$Y \sim N(\mu, \sigma^2)$.

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right), \quad y \in \mathbb{R}.$$

$$Z = \frac{Y-\mu}{\sigma}, \qquad Y = \mu + \sigma Z.$$

Z is standard normal, mean 0 and variance 1.

Multivariate Gaussian distribution

Size $n \times 1$ vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$

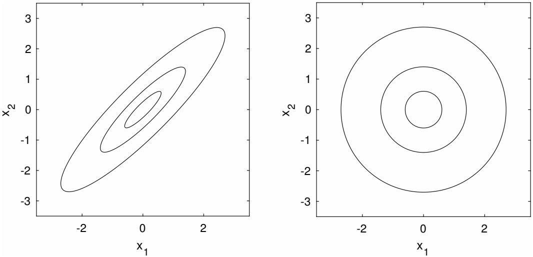
$$p(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right), \quad \boldsymbol{Y} \in \mathbb{R}^n.$$

Mean is $E(\boldsymbol{Y}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$. Positive-definite variance-covariance matrix is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{1,1} & \ldots & \Sigma_{1,n} \\ \ldots & \ldots & \ldots \\ \Sigma_{n,1} & \ldots & \Sigma_{n,n} \end{bmatrix},$$

$\Sigma_{i,i} = \sigma_i^2 = \mathrm{Var}(Y_i)$, $\Sigma_{i,j} = \mathrm{Cov}(Y_i, Y_j)$, $\mathrm{Corr}(Y_i, Y_j) = \Sigma_{i,j}/(\sigma_i\sigma_j)$.

Illustrations $n = 2$
- correlation 0.9 (left), independent (right).

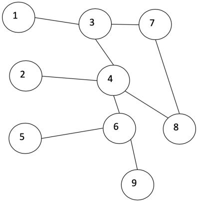


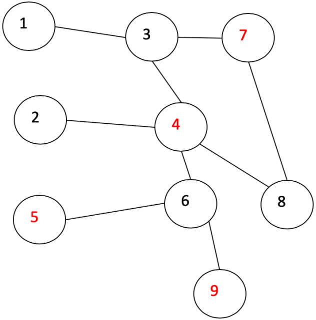
Joint for blocks

$\boldsymbol{Y}_A = (Y_{A,1}, \ldots, Y_{A,n_A})$, $\boldsymbol{Y}_B = (Y_{B,1}, \ldots, Y_{B,n_B})$, joint Gaussian with mean $(\boldsymbol{\mu}_A, \boldsymbol{\mu}_B)$, covariance:

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_A, \boldsymbol{\mu}_B), \quad \boldsymbol{\Sigma} = \left[\begin{array}{cc} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{A,B} \\ \boldsymbol{\Sigma}_{B,A} & \boldsymbol{\Sigma}_B \end{array} \right],$$

Observed set - non-observed set





$n = 9$, $n_A = 5$, $n_B = 4$.
A set is non-observed (black). B set is the observation set (red).
Because of the dependence (here illustrated by edges), the information on variables in set B will propagate to set A variables.

Conditioning

$$\begin{aligned} E(\boldsymbol{Y}_A|\boldsymbol{Y}_B) &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}(\boldsymbol{Y}_B - \boldsymbol{\mu}_B), \\ \mathrm{Var}(\boldsymbol{Y}_A|\boldsymbol{Y}_B) &= \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}\boldsymbol{\Sigma}_{B,A}. \end{aligned}$$

Mean is linear in conditioning variable (data).
Variance is not dependent on data.

Illustration for $n = 2$, mean 0, variance 1, correlation 0.9.

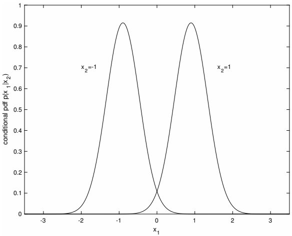


Figure: Conditional pdf for Y_1 when $Y_2 = 1$ or $Y_2 = -1$.

Transformation

$$\boldsymbol{Z} = \boldsymbol{L}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}), \quad \boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{Z}, \quad \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}'.$$

$\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ are independent standard normal, mean 0 and variance 1.

Cholesky factorization

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{1,1} & \dots & \Sigma_{1,n} \\ \dots & \dots & \dots \\ \Sigma_{n,1} & \dots & \Sigma_{n,n} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}',$$

Lower triangular matrix

$$\boldsymbol{L} = \begin{bmatrix} L_{1,1} & 0 & \dots & 0 \\ L_{2,1} & L_{2,2} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ L_{n,1} & L_{n,2} & \dots & L_{n,n} \end{bmatrix},$$

Cholesky - example

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \quad \boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 0.9 & 0.44 \end{bmatrix}.$$

Consider sampling from joint $p(y_1, y_2) = p(y_1)p(y_2|y_1)$:

- Draw independent standard Gaussian variables Z_1 and Z_2 (mean 0, variance 1).
- Sample from $p(y_1)$ by $Y_1 = \mu_1 + L_{1,1}Z_1$.
- Sample from $p(y_2|y_1)$ is constructed by $Y_2 = \mu_2 + L_{2,1}Z_1 + L_{2,2}Z_2 = \mu_2 + L_{2,1}\frac{Y_1-\mu_1}{L_{1,1}} + L_{2,2}Z_2$

Gaussian process

For any set of time locations $t_1, \ldots, t_n$. $Y(t_1), \ldots, Y(t_n)$ is jointly multivariate Gaussian.
Mean $\mu(t_i)$, $i = 1, \ldots, n$.

$$\mathbf{\Sigma} = \left[\begin{array}{ccc} \Sigma_{1,1} & \ldots & \Sigma_{1,n} \\ \ldots & \ldots & \ldots \\ \Sigma_{n,1} & \ldots & \Sigma_{n,n} \end{array} \right],$$

Covariance function

The covariance tends to decay with 'distance':

$$\Sigma_{i,j} = \gamma(|t_i - t_j|),$$

for some covariance function γ.
Examples (distance $h = |t - t'|$):

$$\gamma_{\text{exp}}(h) = \sigma^2 \exp(-\phi h)$$

$$\gamma_{\text{mat}}(h) = \sigma^2 (1 + \phi h) \exp(-\phi h)$$
$$\gamma_{\text{gauss}}(h) = \sigma^2 \exp(-\phi h^2)$$

Illustration of covariance function

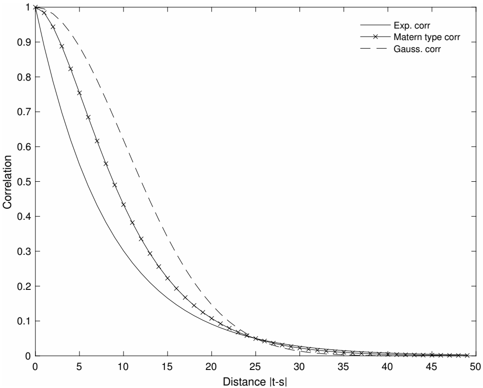
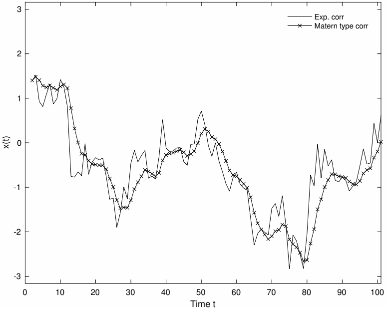


Illustration of samples of Gaussian processes

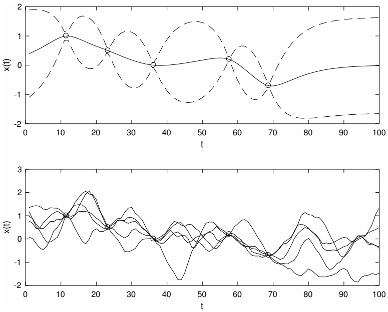


Conditional formula

$$
\begin{aligned}
E(\boldsymbol{Y}_A|\boldsymbol{Y}_B) &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}(\boldsymbol{Y}_B - \boldsymbol{\mu}_B), \\
\mathrm{Var}(\boldsymbol{Y}_A|\boldsymbol{Y}_B) &= \boldsymbol{\Sigma}_A - \boldsymbol{\Sigma}_{A,B}\boldsymbol{\Sigma}_B^{-1}\boldsymbol{\Sigma}_{B,A}.
\end{aligned}
$$

- Expectation linear in data.
- Variance only dependent on data locations, not data.
- Expectation close to conditioning variables near data locations, goes to $\boldsymbol{\mu}_A$ far from data.
- Variance small near data locations, goes to $\boldsymbol{\Sigma}_A$ far from data.
- Close data locations are not double data.

Illustration of conditioning in Gaussian processes



Application of Gaussian processes: function optimization

Several applications involve very time-demanding or costly experimentation, or tedious computer simulations.
Typical question: Which configurations of inputs give highest output?
Goal is to find the optimal input without too many trials / tests, because they cost so much.

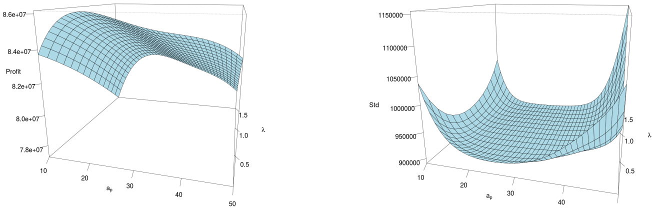
Approach: Fit a GP to the output function, with 'distance' between inputs. Train GP model from the evaluation points and results. This allows fast consideration of which evaluation points to choose next, based on large predictions or high uncertainty!

Application : optimal vessel fleet mix for offshore wind-farm operation

Application : optimal vessel fleet mix for wind-farm operation

- More vessels is better because easier to do repair and maintenance.
- More vessels is worse because it costs more to operate.
- The 'price' of each combination of large/small vessels takes a long time to evaluate (using a simulator model with various weather and waves inputs along with Poisson process failures).
- Approach: Fit a GP to the profit function, for different input variables (number of different ships available, number of various personnel) based on some evaluation points and results. Run batches of simulations, to iteratively find the optimum giving the best combinations of inputs.

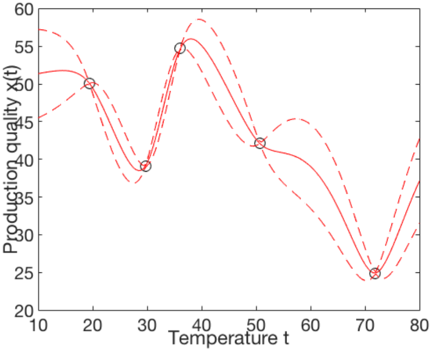
Application : optimal vessel fleet mix



Fitted mean (left) and variance (right) after some batches of iterations. The selection of evaluation points is done by **Expected Improvement**, which has a closed form expression for Gaussian process models.

Simplified example - Production quality

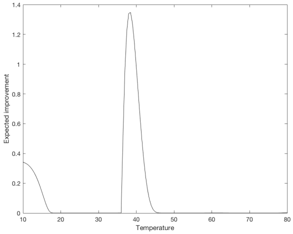
Goal: Find the best temperature input, to give optimal production. Which temperature to evaluate next? Experiment is costly, want to do few evaluations.



Expected improvement

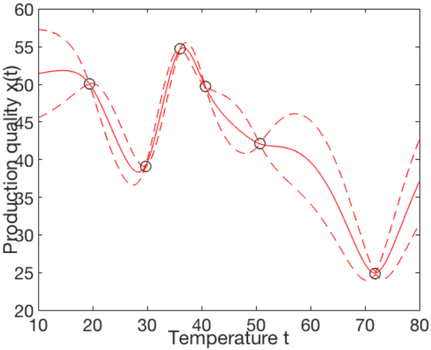
Maximum so far $Y^* = \max(\boldsymbol{Y})$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$.

$$EI(t) = E(max(Y(t) - Y^*, 0)|\boldsymbol{Y})$$



Sequential uncertainty reduction

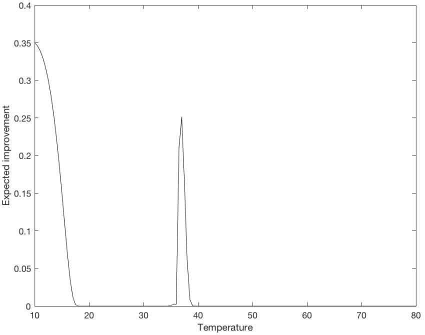
Perform test at $t = 40$, result is $Y(40) = 50$.



Sequential uncertainty reduction and optimization

Maximum so far $Y^* = \max(\boldsymbol{Y}, Y_{n+1})$.

$$EI_{n+1}(t) = E(max(Y(s) - Y^*, 0)|\boldsymbol{Y}, Y_{n+1})$$



Project on function optimization using EI next week

Analytical solutions to parts of computational challenge:

$$
\begin{aligned}
\text{EI} &= \int_{-\infty}^{\infty} \max\{0, v - v^*\}\, p(v) dv = \int_{v^*}^{\infty} (v - v^*) p(v) dv \\
&= \int_{v^*}^{\infty} v p(v) dv - v^* \int_{v^*}^{\infty} p(v) dv \\
&= \int_{\frac{v^* - m}{s}}^{\infty} (m + sz) p(z) dz - v^* \int_{\frac{v^* - m}{s}}^{\infty} p(z) dz \\
&= s\phi(z) + (m - v^*)\Phi(z),
\end{aligned}
$$

Relies on Gaussian standard density ($\phi(z)$) and cumulative function ($\Phi(z)$).