Bayesian Inference in e-sport matches

Bayesian Computation Project

Presentation of the game

- 2 teams
- Assymetric game for the 2 teams
- No determined duration for a match

Presentation of the dataset

- All matches since 2014
- For each match is given:
  - Tournament
    - Regional
    - International
  - 2 opponent teams
  - Winner
  - Duration of the match
  - Year/Split of the match

Goal

- Given a match:
  - Predict winner
  - Predict duration
- Issues:
  - Is the duration predictable?
  - Do we have enough data?

Summary

1. Prediction of the winner of a match
    1. Logit Model
        1. Posterior
        2. Sampling Methods / Predictions
        3. Inflence of Time scale and Training Set
    2. Logit Model for International events
        1. Model
        2. Predictions
2. Prediction of the duration of a match
    1. Basic models
        1. Gaussian
        2. Gamma
        3. Skewed Gaussian
    2. Injection of Logit Model for duration prediction
        1. Model
        2. Predictions

Prediction of the winner of a match

Logit Parametric Model

Excluding international tournaments

- « Force » f_i for each team, Δ bias

- Probability of i winning against j:
  - $p_{ij} = logit^{-1}(f_i - f_j + \Delta)$

- Logit Model:
  - $p_{ij} = \frac{e^{f_i + \Delta}}{e^{f_i + \Delta} + e^{f_j}}$

Parameters

- m teams
- Parameter:
  - $\theta = \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \\ \Delta \end{pmatrix}$ « log-odds »
  - $(f_i)_{1 \leq i \leq m}$ « force » of each team
  - Δ bias for the blue side

Variables

- Pairs $(y_{i,j}, X_{i,j})$
  - $y_{i,j} \in \{0,1\}$ outcome of the match
  - $X_{i,j} = \left(\delta_{k,(i,j,m+1)}\right)_{(m+1)\times 1}$
  - ${X_{i,j}}_{m+1} = 1$ = Bias of Blue side

- Classification Task:
  - Predict outcome of matches

A Bayesian Linear Model

- $f(D|\theta) = \prod_{k=1}^{n} f(y_k|\theta, x_k)$

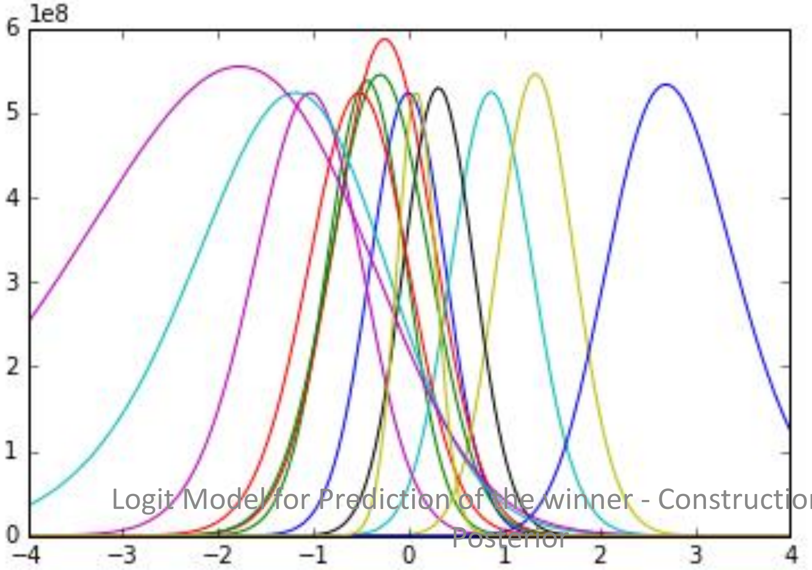
- We have linear dependency:
    - $y_k \sim B(logit^{-1}(\theta^T x_k))$; $y_k = 1 = blue\ side\ win$
    - $x_k = X_{i,j} = \left(\delta_{k,(i,j,m+1)}\right)_{k\times 1}$;team i vs j
- Then we have:
    - $p\left(y_k \middle| \theta, x_k = X_{i,j}\right) = \left(\frac{1}{1+e^{-(f_i - f_j + \Delta)}}\right)^{y_k} . \left(1 - \frac{1}{1+e^{-(f_i - f_j + \Delta)}}\right)^{1-y_k}$

Prior

- Gaussian Prior
  - $\theta_i \sim N\left(\mu_i, \sigma_i^2\right)$
  - Choice:
    - $\mu_{1\leq i\leq m} = \mu_0$; μ_0 not relevant
    - μ_{m+1}=0 ; game well-balanced
    - $\sigma_{1\leq i\leq m} = 4$; flexible model
    - σ_{m+1} =1 ; stronger intuition for the bias

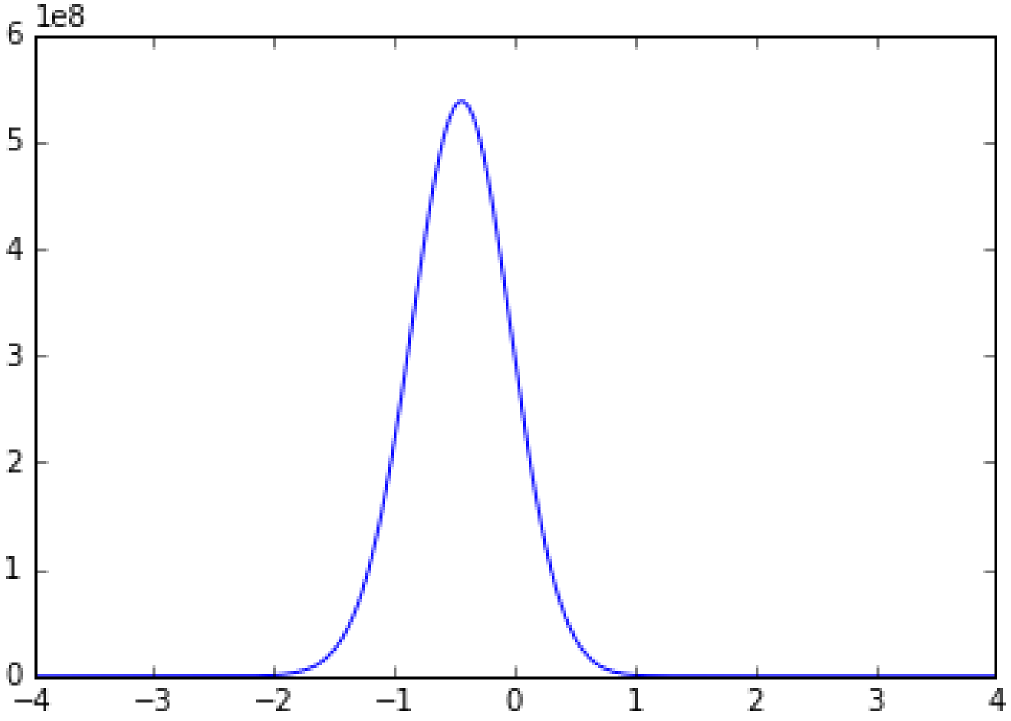
Posterior

- $f(\theta|D) = f(D|\theta)f(\theta)$
- $\theta_{bayes} = E(\theta|D) = \int_{\theta} f(\theta|D)d\theta$
- $\theta_{(m+1)\times 1}$
- Shape of the posterior:



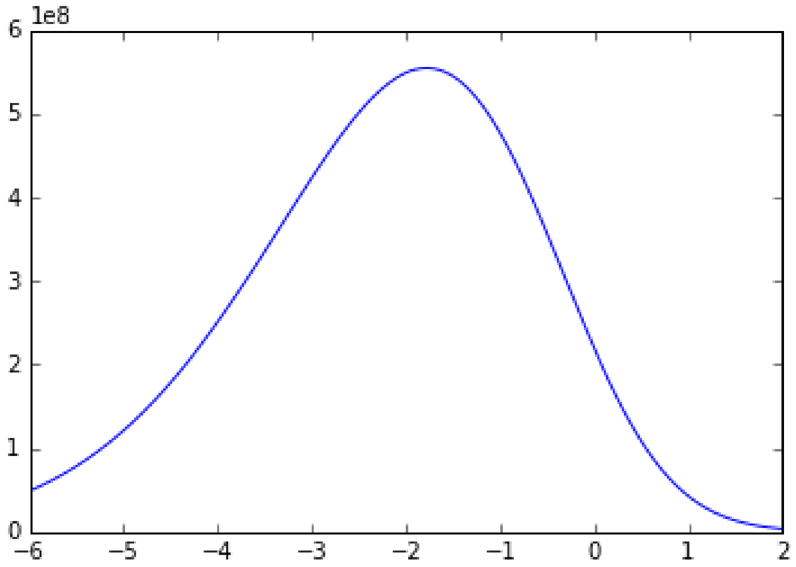
Shape of the posterior

- A lot of symmetric dimensions:



Shape of the posterior

- Few assymetric dimensions:



- Therefore, $\theta_{bayes} \neq \theta_{MLE}$

Fitting Methods used

- Laplace Approximation : $\theta = \theta_{MLE}$
- Importance Sampling : $\theta = \theta_{bayes}$
- Metropolis-Hastings : $\theta = \theta_{bayes}$
- Coded in Python

Predictor

- We get θ^* fitting our model

- Given $x_k = X_{i,j}$ we estimate y_k:
  - ${f_i}^* - {f_j}^* + \Delta^* \geq 0 \Longrightarrow \hat{y}_k = 1$
  - $\hat{y}_k = 0$ else

- « Accuracy » displayed by my algorithms for n matches:
  - $A_{\%} = \frac{\#\{k \in [\![1,n]\!] | \hat{y}_k = y_k\}}{n} .100 \quad ; \quad n = len(Validation\ Set)$

Laplace Approximation

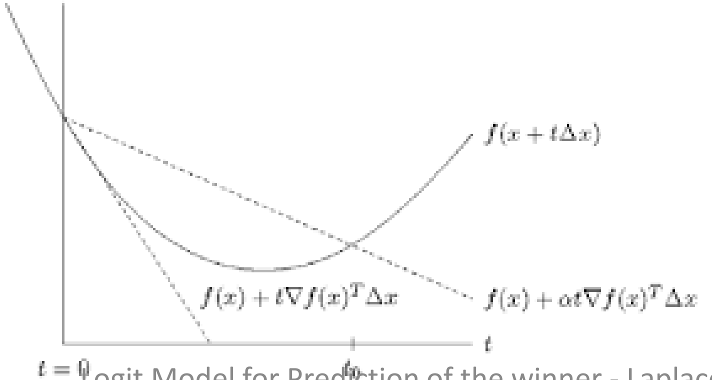
- Approximation of the curve by a Gaussian density

- Find $\theta^* = argmax_\theta f(\theta|D)$ by gradient descent
    - (also works with $\log(f(\theta|D))$
- Approximate Covariance Matrix by inverse of log-curvature:
    - $Cov^* = (\,-\nabla^2(\log(f))|_{\theta^*})^{-1}$

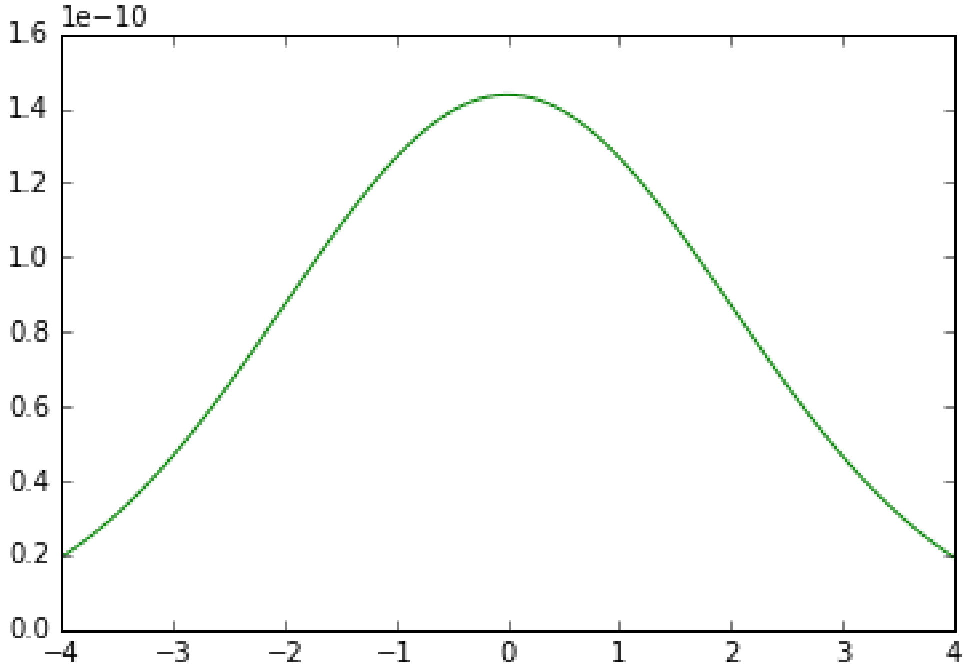
- Then $(\theta|D) \sim N(\theta^*, Cov^*)$

Backtracking Line-Search Gradient Descent

- Optimize choice of step by making sure we go down enough :
  - θ_n initial point ; λ_0 initial step-size ; E function to optimize ; (α, β)
  - At each step:
    - Initialize $\lambda = \lambda_0$
    - Candidate is $\theta_c = \theta_n - \lambda \nabla E(\theta_n)$
    - Accepted if $E(\theta_c) \leq E(\theta_n) - \lambda\alpha \|\nabla E(\theta_n)\|_2{}^2$; $\theta_{n+1} = \theta_c$
    - If rejected, $\lambda = \beta\lambda$

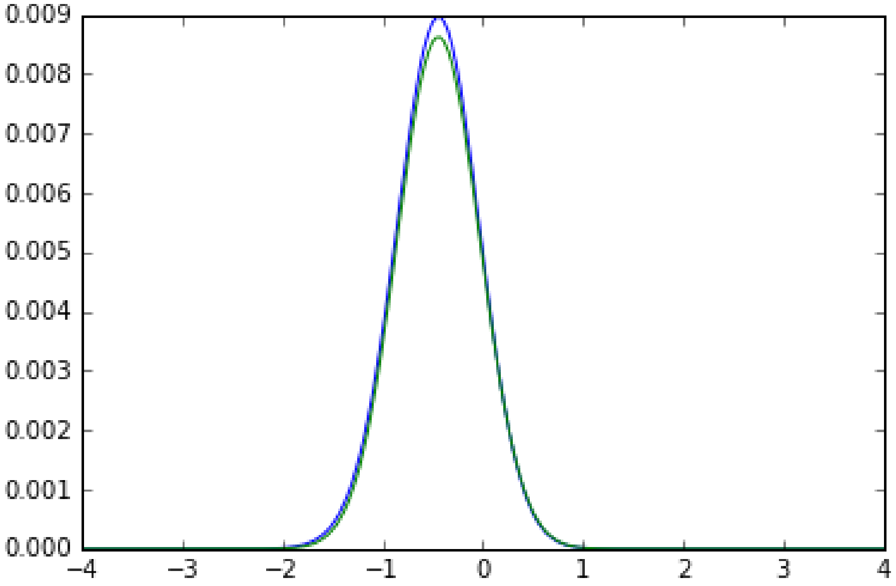


Laplace Approximation on Gaussian Example

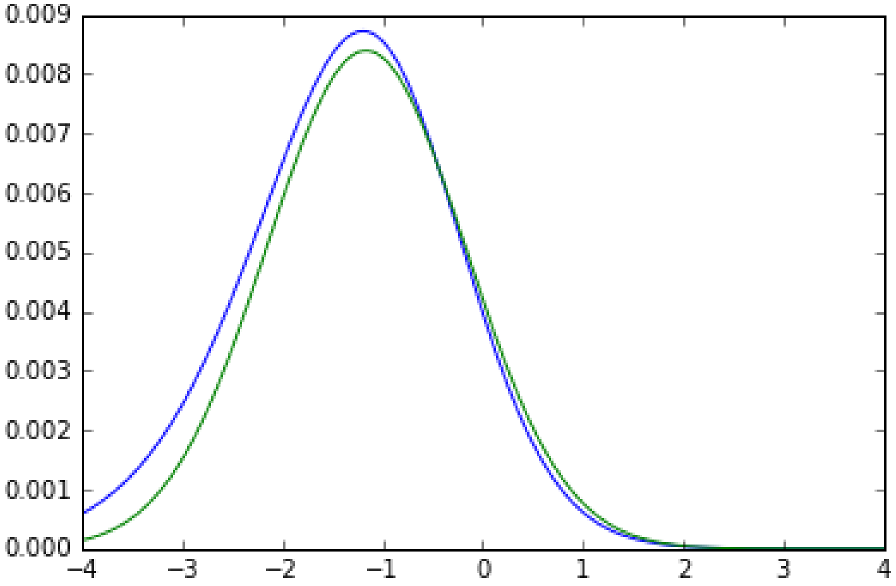


Gaussian and L.A.

Laplace Approximation on Posterior

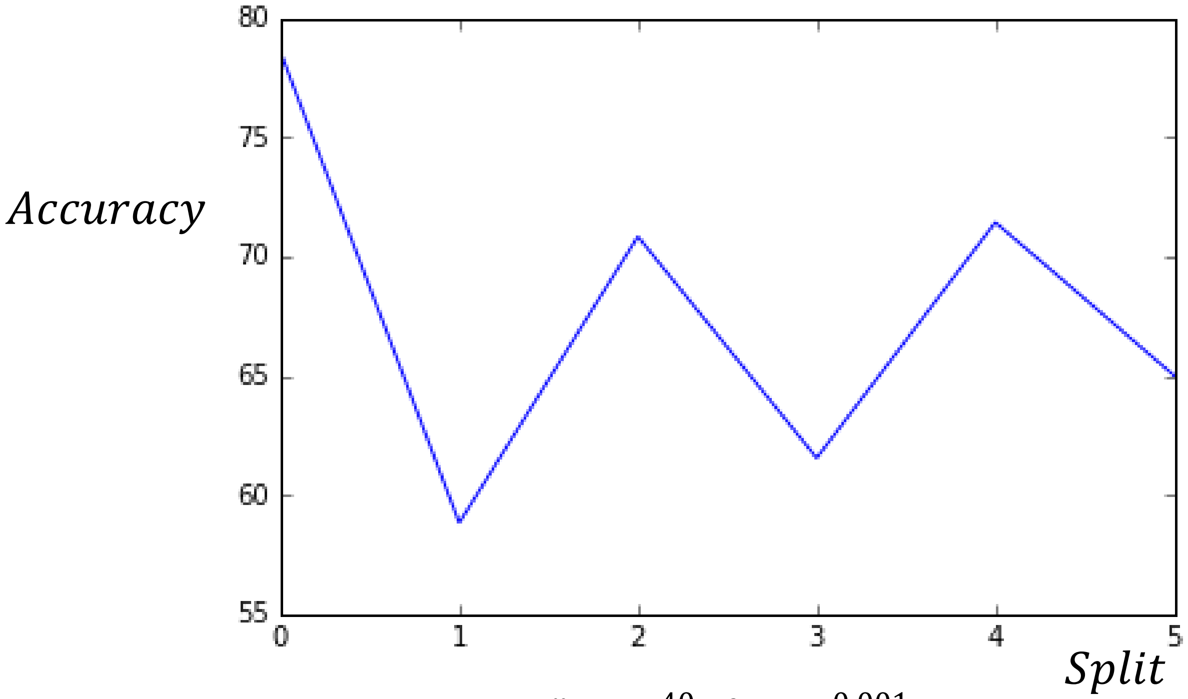


LA on symmetric dimension



LA on assymetric dimension

Laplace Predictions



$n_{\max} = 40\ ;\ e_{max} = 0{,}001$

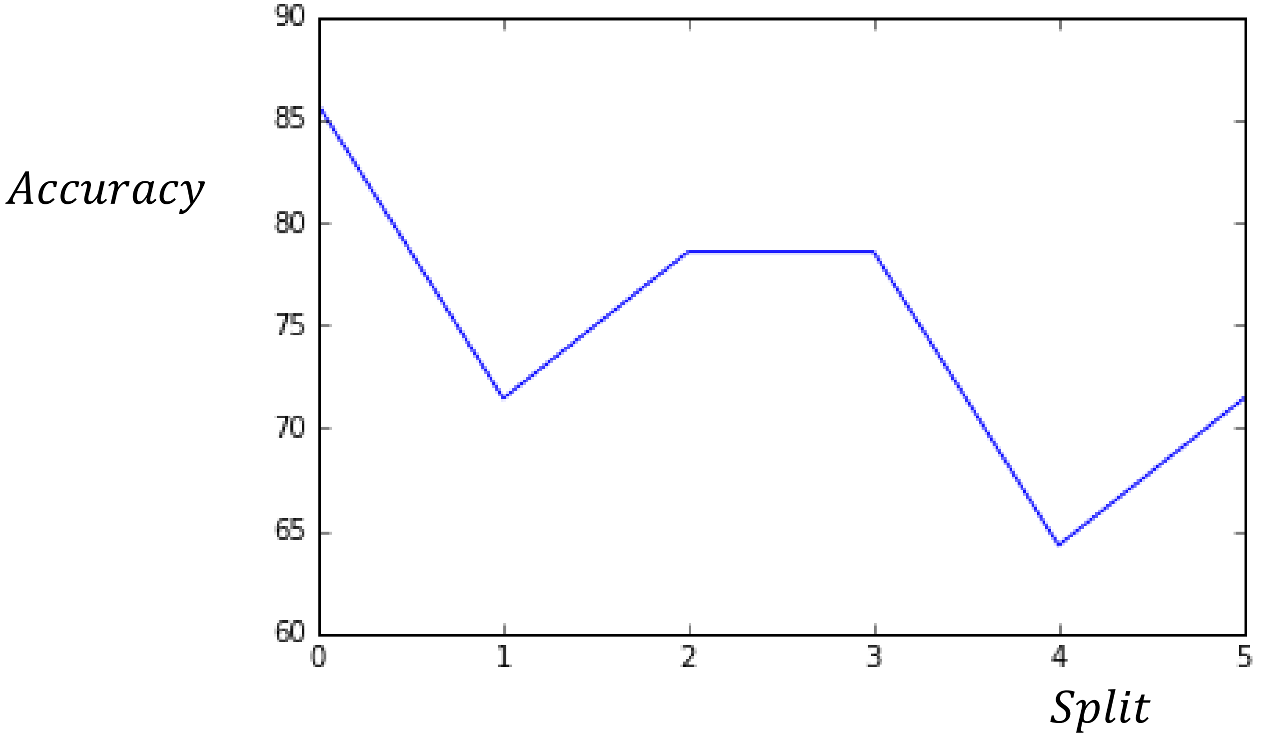
Predictions of winners by split in EULCS (2015-2017)

Training Set 1/2

Importance Sampling

- Goal: compute expectation of a random variable given its density
- Based on proposal distribution $h(\theta)$, like RS Sampling
- Principle:
  - $h(\theta)$ proposal we can sample from ; $f(\theta|D)$ our target density
  - Sample θ_i from h
  - Compute weight $p_i = \frac{f(\theta|D)}{h(\theta)}$
  - Estimator of Expectation:
    - $E(\theta \sim f(\theta|D)) \approx \frac{\sum_i p_i \theta_i}{\sum_i p_i}$

IS Predictions



Predictions of winners by split in EULCS (2015-2017)

Training Set ½

$N = 15000$

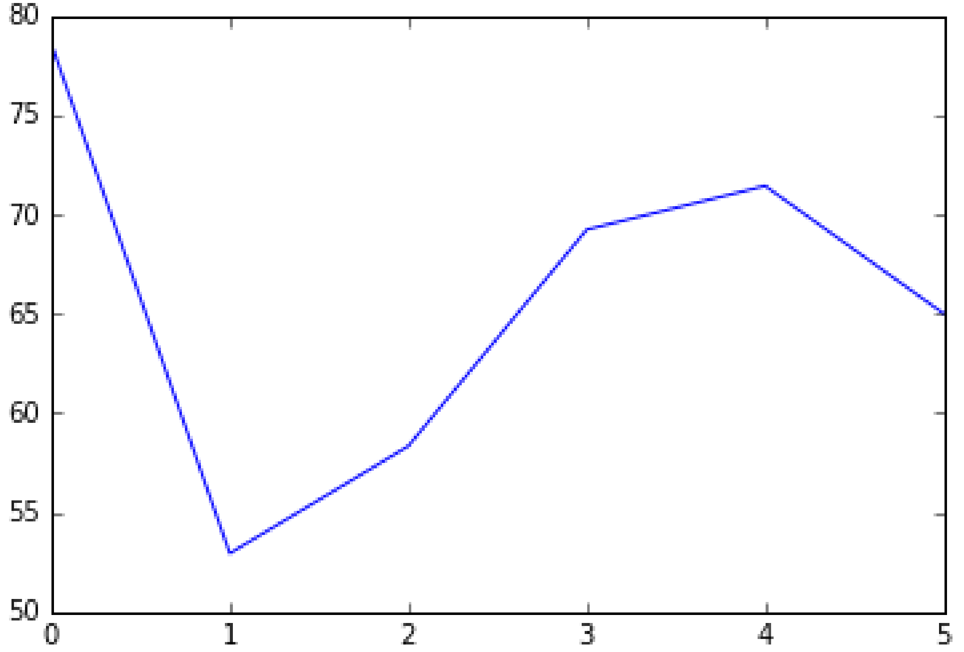
Metropolis-Hastings

- Goal : Compute expectation by sampling from a density $f(\theta)$
- Build Markov Chain with respect to density and a proposal
  - θ_0 initiation ; λ step-size
  - At each step n :
    - Compute $\theta_c = \theta_n + \lambda\eta_n$; $\eta_n \sim N(0,1)$ for example
    - Accept θ_c with probability $p_c = min\left\{\frac{f(\theta_c)}{f(\theta_n)}, 1\right\}$
      - If accepted $\theta_{n+1} = \theta_c$
      - Else $\theta_{n+1} = \theta_n$
  - We have sampled $\{\theta_n\}_{n_0 \leq n \leq N}$ from f
    - n_0 far enough, for independence

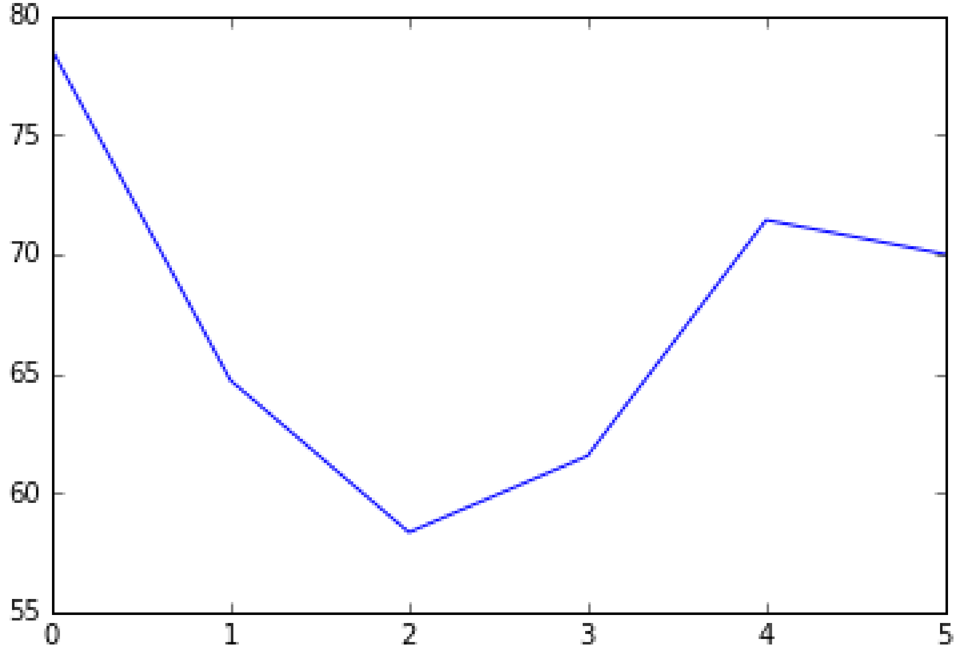
Some results with MH Sampling

- Sampling on a Gaussian of parameters ([0,1,4,9,16], identity) :
- $\theta_0 = [1,15,-3,-4,5]$; $n_0 = N/2$
- $\theta_{bayes}^1 = [0.8604, \ \ 14.8307, \ \ -3.1712, \ \ -3.7282, \ \ 5.0553]$
  - $\lambda = 0{,}001$; 40000 epochs
- $\theta_{bayes}^2 = [0.8660, \ \ 7.2102, \ \ 0.2559, \ \ 2.6136, \ \ 11.3700]$
  - $\lambda = 0{,}01$; 40000 epochs
- $\theta_{bayes}^3 = [-0.2776, \ \ 1.4215, \ \ 4.0391, \ \ 8.4282, \ \ 15.4884]$
  - $\lambda = 0{,}05$; 40000 epochs
- $\theta_{bayes}^4 = [-0.1525, \ \ 1.1469, \ \ 3.9139, \ \ 9.0190, \ \ 15.9634]$
  - $\lambda = 0{,}1$; 40000 epochs
- $\theta_{bayes}^5 = [-0.0008, \ \ 0.9551, \ \ 4.0527, \ \ 8.9214, \ \ 15.7864]$
  - $\lambda = 0{,}1$; 100000 epochs
- $\theta_{bayes}^6 = [-0.0693, \ \ 0.9115, \ \ 4.0223, \ \ 9{,}0129, \ \ 15.9835]$
  - $\lambda = 0{,}08$; 100000 epochs

MH Predictions



5000 epochs ; 0,1 step-size



5000 epochs ; 0,1 step-size

Predictions of winners by split in EULCS (2015-2017)

Training Set 1/2

Metropolis-adjusted Langevin

- Same idea than MH
- Difference of candidate and acceptance probability
  - $\theta_c = \theta_n + \frac{\lambda^2}{2} \nabla\left[\log\left(f(\theta)\right)\right] + \lambda\eta_n$
  - $p_c = min\left\{\frac{f(\theta_c)\pi(\theta_n|\theta_c)}{f(\theta_n)\pi(\theta_c|\theta_n)}, 1\right\}$;
  - with $\pi(\theta_1|\theta_2) = \alpha e^{-\frac{1}{\lambda^2}\left\|\theta_1 - \theta_2 - \frac{\lambda^2}{2}\nabla\left[\log\left(f(\theta_2)\right)\right]\right\|_2^2}$

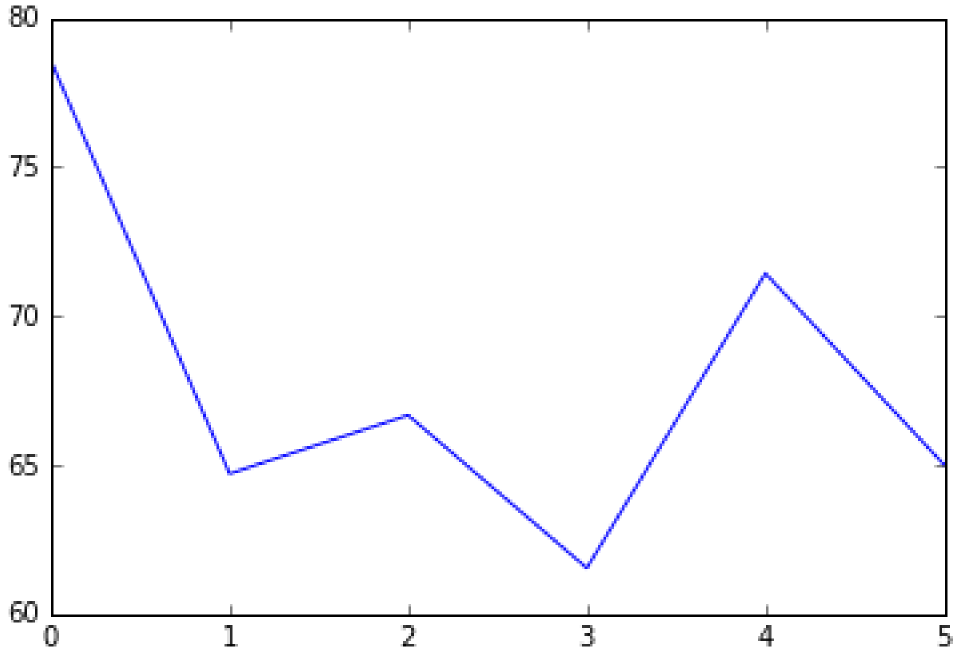
Some results with Langevin MH

- Sampling on a Gaussian of parameters ([0,1,4,9,16], identity) :
- $\theta_0 = [1,15,-3,-4,5]$; $n_0 = N/2$
- $\theta_{bayes}^1 = [0.8796, \;\; 12.1070, \;\; -1.6991, \;\; -1.0497, \;\; 7.9802]$
  - $\lambda = 0{,}1$; 100 epochs
- $\theta_{bayes}^2 = [-0.2568, \;\; 1.1059, \;\; 3.7440, \;\; 8.7557, \;\; 15.7393]$
  - $\lambda = 1$; 100 epochs
- $\theta_{bayes}^3 = [-0.0004, \;\; 1.0256, \;\; 4.0052, \;\; 8.9841, \;\; 15.9989]$
  - $\lambda = 1$; 10000 epochs
- $\theta_{bayes}^4 = [-0.1454, \;\; 1.1070, \;\; 3.9868, \;\; 8.9793, \;\; 15.9867]$
  - $\lambda = 0{,}2$; 20000 epochs
- $\theta_{bayes}^5 = [-0.0204, \;\; 0.9513, \;\; 4.0375, \;\; 8.9850, \;\; 15.9587]$
  - $\lambda = 0{,}5$; 20000 epochs
- $\theta_{bayes}^6 = [0.0003, \;\; 1.0045, \;\; 4.0029, \;\; 9{,}0069, \;\; 15.9955]$
  - $\lambda = 1$; 50000 epochs

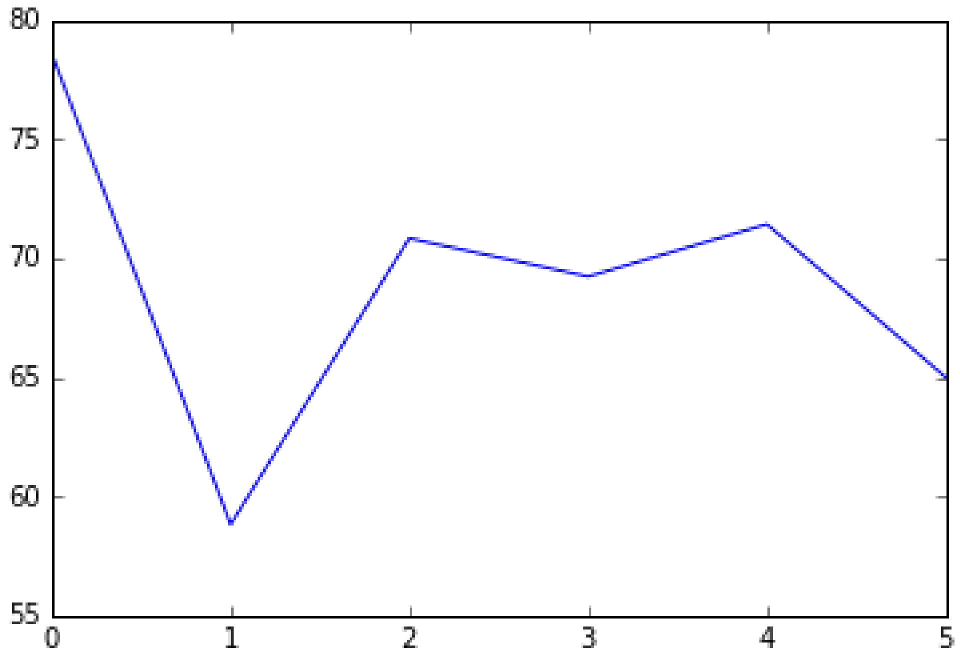
Advantages of Langevin

- Moves quicker to high density zones
- Good management of anisotropic zones
- Slows down and increases accuracy in flat zones / extrema
- <u>BUT</u> slower epochs than MH

Langevin MH Predictions



3000 epochs ; 0,5 step-size



3000 epochs ; 0,1 step-size

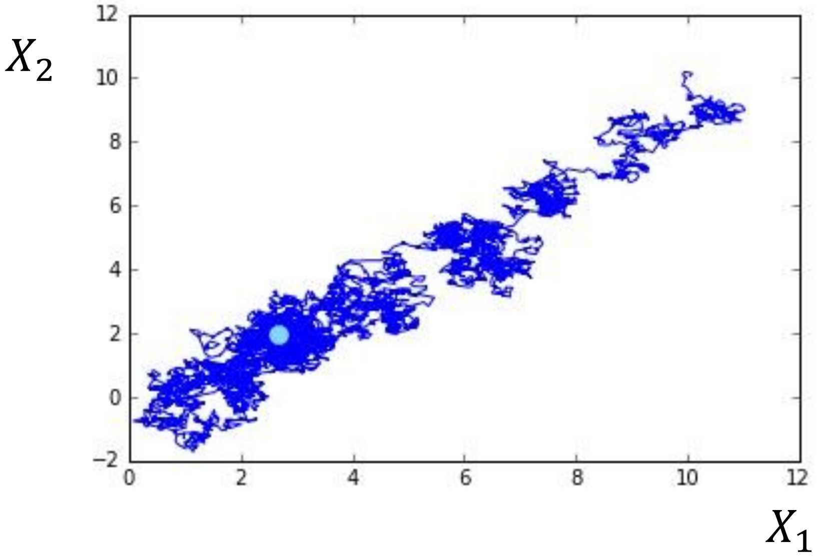
Predictions of winners by split in EULCS (2015-2017)

Training Set 1/2

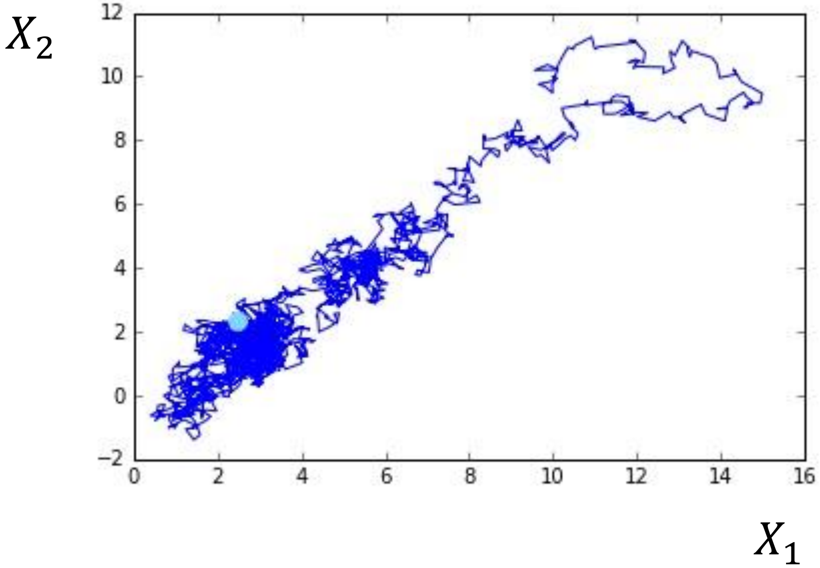
Interpretation of results

- EULCS 2015 Spring Split :
- Estimated θ_{bayes} :
  - $\left\{f_{i_{bayes}}\right\}_{1\leq i\leq m}$=[['EL', -0.06315812011347757], ['GIA', -0.5012171478884524], ['CW', 0.013228694659806372], ['SK', 1.6127797646398627], ['GMB', 0.654502158438717], ['MYM', -0.6741420566171826], ['ROC', -0.34513280133045615], ['FNC', 1.434772313332971], ['UOL', 0.9247568739344932], ['H2K', 1.219724556163784]]
  - Δ_{bayes}=0.41554176
- Interpretation: « force » for each team
- The game is not well balanced : $p\left(y_{ij}=1\middle|f_i=f_j\right)\approx 0{,}6$

Differences MH // Langevin



Langevin MH
Step-size $\lambda = 0{,}1$



MH
Step-size $\lambda = 0{,}2$

$\theta^* \approx (2{,}0)$
Initiation $\theta_0 = (10{,}10)$
5000 epochs

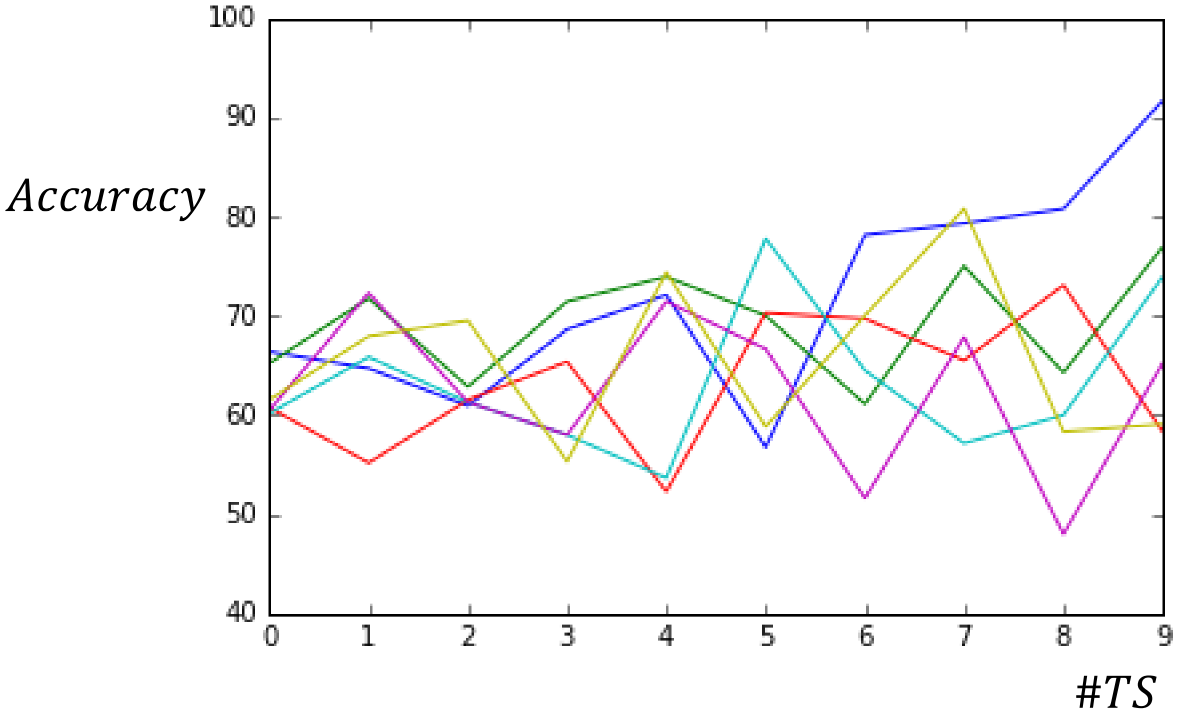
Comparison of algorithms

- Our posterior is :
  - Monomodal
  - Close to 0
- MH Methods are quite slow
- Laplace is quicker and more reliable
- IS is the fastest

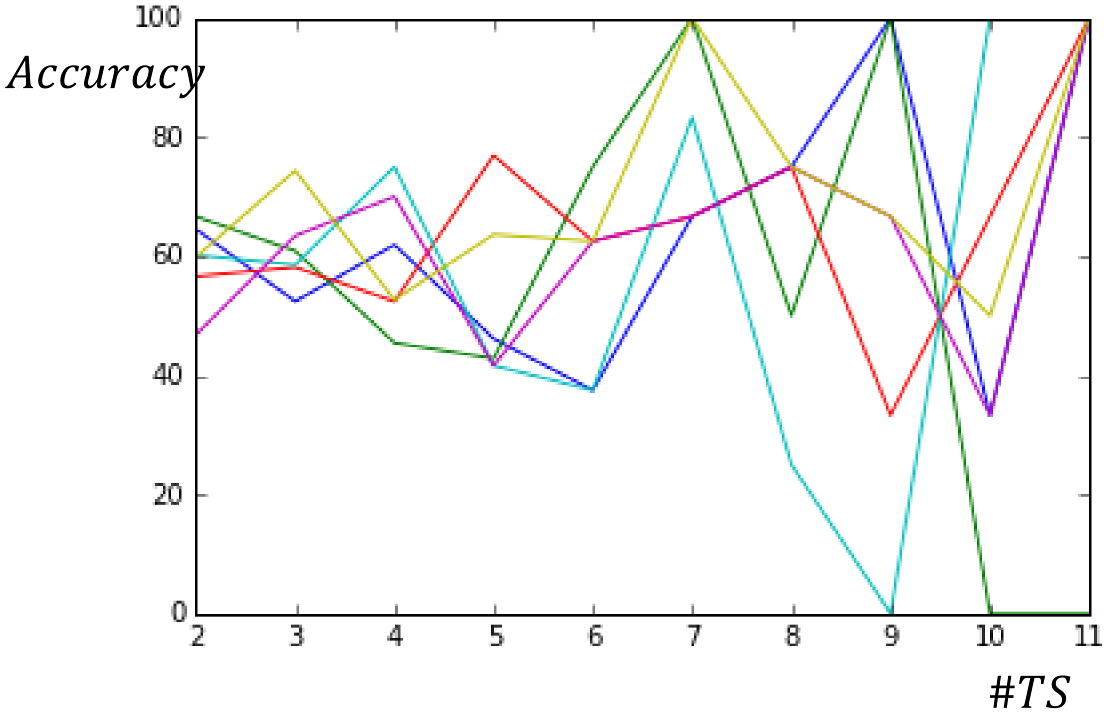
Influence of Training Set

- Fact : Data has regularity in time
- Is there a best size for training set?
- Periodic or random ?

Training Set : Region LCK (different splits)

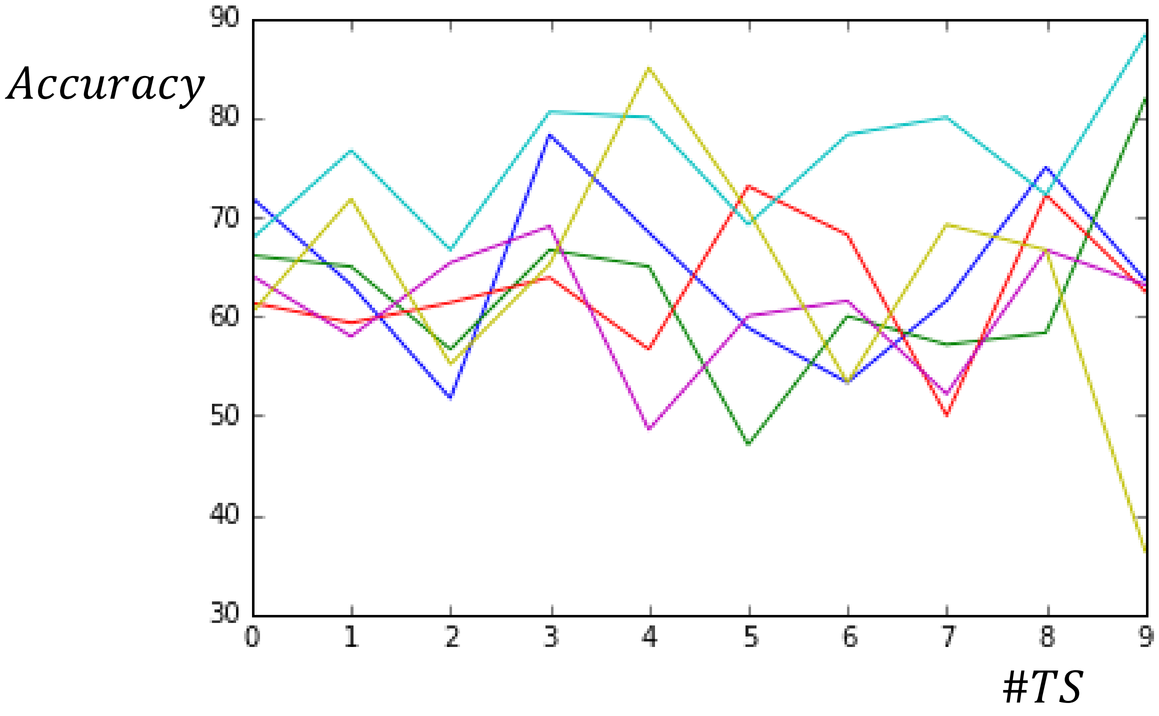


Periodic Training Set
$\{\#TS = n/i\}_{2 \le i \le 13}$
MH Sampling
500 epochs



Random Training Set
$\{\#TS = n/i\}_{2 \le i \le 13}$
MH Sampling
500 epochs

Training Set : Region EULCS (different splits)

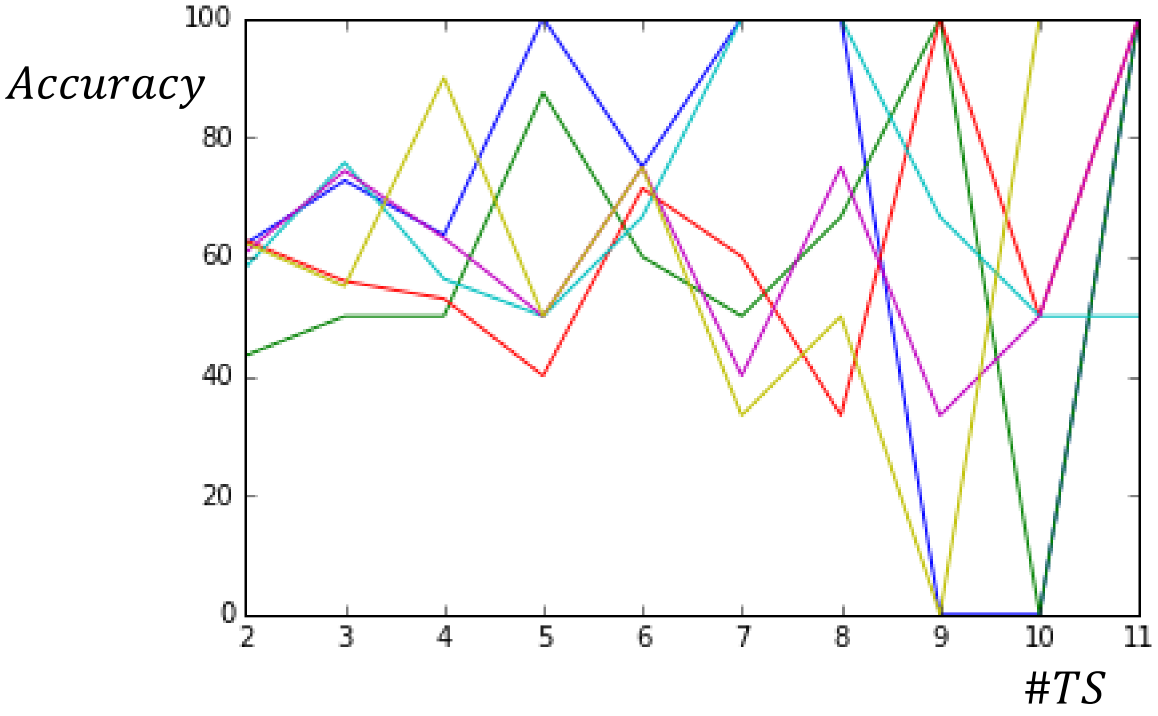


Periodic Training Set

$\{\#TS = n/i\}_{2 \le i \le 11}$

MH Sampling

500 epochs



Random Training Set

$\{\#TS = n/i\}_{2 \le i \le 13}$

MH Sampling

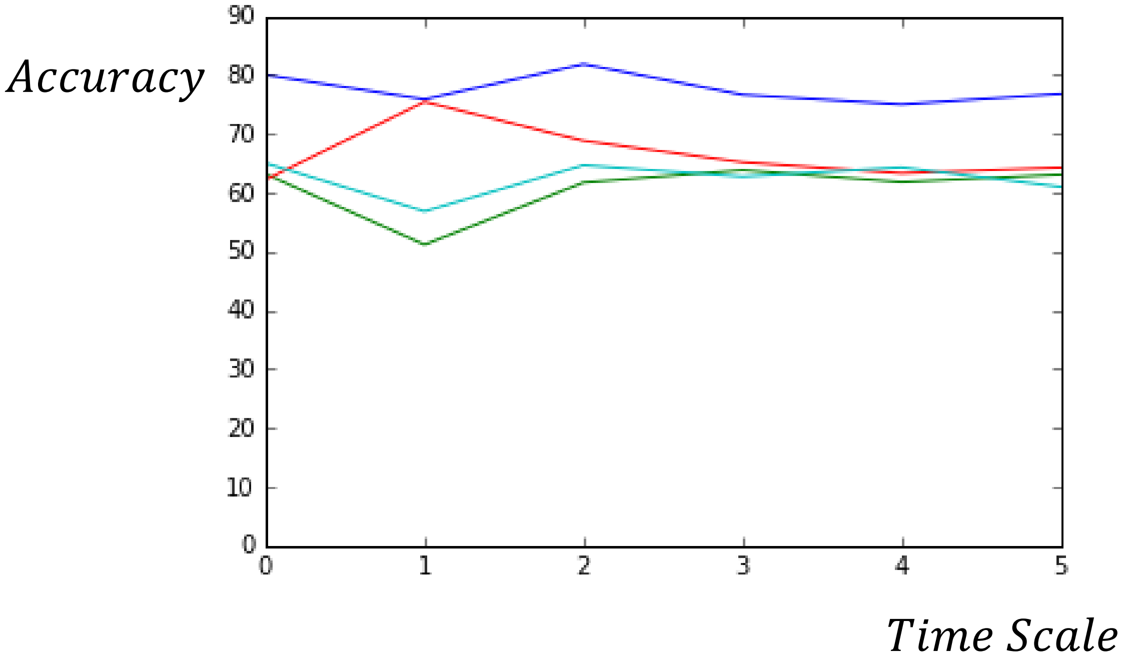
500 epochs

Influence of time scale

- Many time scales
  - Split
  - k Years ; $1 \leq k \leq 4$
- Not much data :
  - $10 \leq m \leq 15$ teams per split
  - $100 \leq n \leq 200$ matches per split ; 2 teams per match

- ⇒Temporal variation of data // Quantity of data ?

Time Scale Comparison



Different time scale predictions for 4 regions

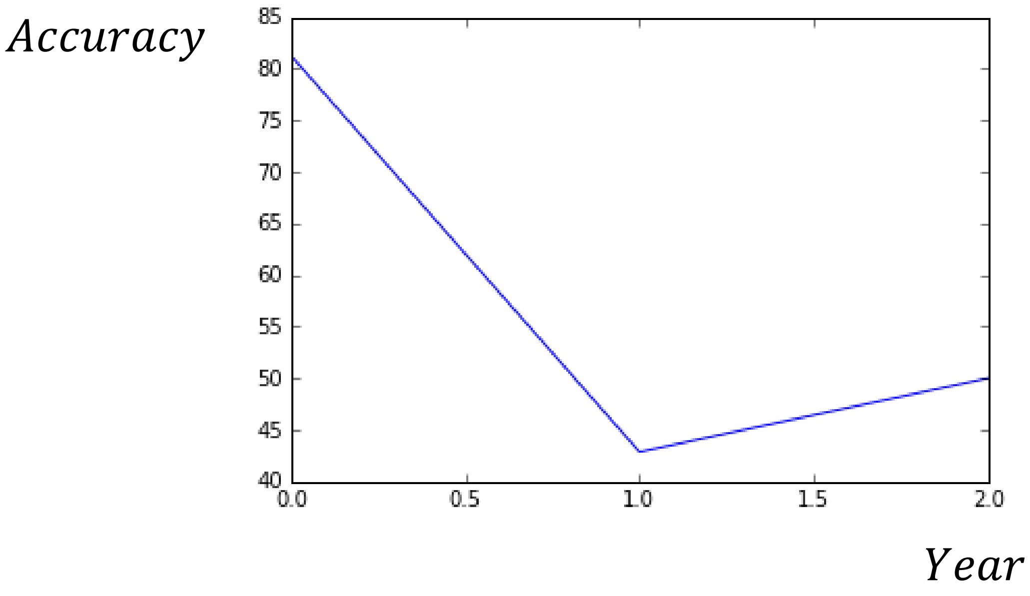
(1 split ; 1,2,3,4,5 years)

Training Set 1/10

IS Sampling ; $N = 15000$

Predictions for international tournaments

- Is our model relevant for international tournaments?



Predictions by year in international tournaments (2015-2017)
Training Set 1/10
IS Sampling ; $N = 15000$

Parameters

- m_1 teams ; m_2 regions
- Parameter:
  - $\theta = \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_{m_1} \\ F_1 \\ \cdots \\ F_{m_2} \\ \Delta \end{pmatrix}$ « log-odds »
  - $(f_i)_{1 \leq i \leq m_1}$ « force » for each team
  - $(F_j)_{1 \leq j \leq m_2}$ « force » for each region
  - Δ bias for the blue side

Model

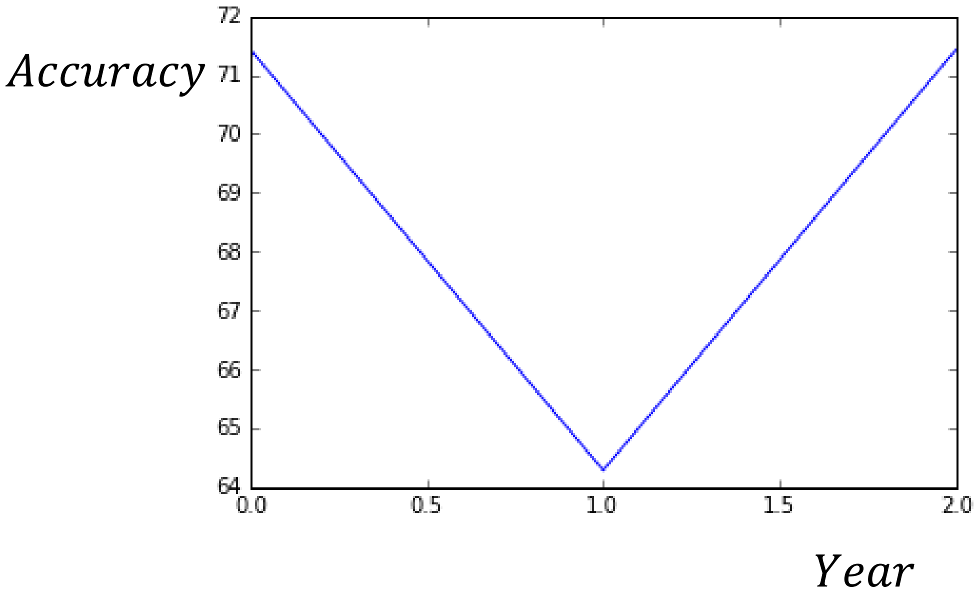
- $x_k = X_{i_1,j_1,i_2,j_2} = \left(\delta_{k,(i_1,j_1,i_2,j_2,m_1+m_2+1)}\right)_{(m_1+m_2+1)\times 1}$
    - (team i_1 from region j_1) vs (i_2 from j_2)

- Parametric : $y_k \sim B(logit^{-1}(\theta^T x_k))$

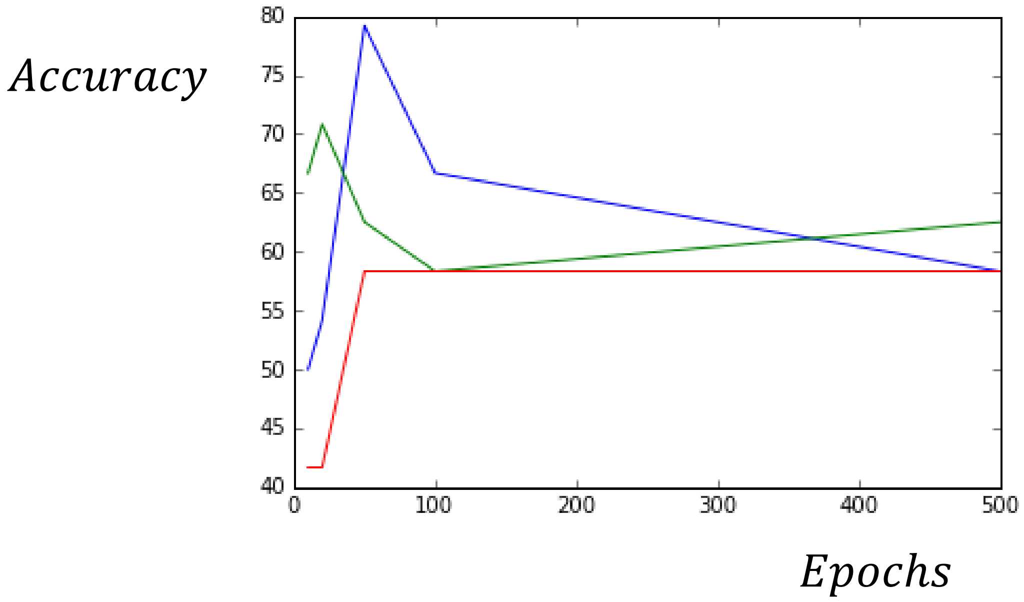
- $p_{i_1 j_1 i_2 j_2} = logit^{-1}\left(f_{i_1} - f_{i_2} + F_{j_1} - F_{j_2} + \Delta\right)$

- Regions quite heterogeneous ; model should fit

Predictions from new model



Predictions per year in international tournaments
(2015-2017)
Training Set 1/7
Langevin MH Sampling ; 500 Epochs



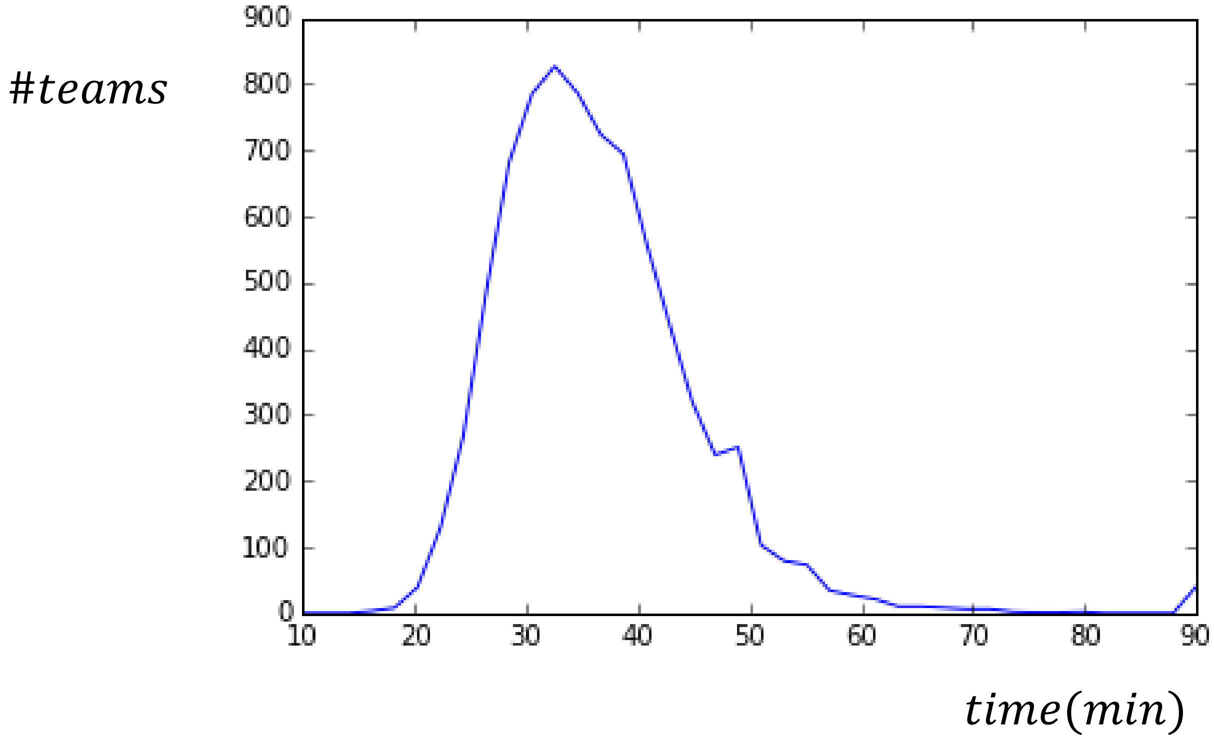
Accuracy per year // #Epochs
(10,20,50,100,500)
(2015-2017)
Training Set 1/7
Langevin MH Sampling

Prediction of the duration of a match

Prediction of match duration

- Useful for bets
- Possible to predict?
- Some teams play faster than others
- May depend on level difference between teams

Distribution of durations



Duration of matches

All years, all regions

Parametric model

- Expectation :
  - $\mu_{ij} = \alpha_i + \alpha_j + \mu_0$
  - $\{\alpha_i\}$ parameter for each team
  - μ_0 global expectation

- Parametric model :
  - $t_{ij} \sim L(\mu_{ij}, A)$

Variables

- Pairs $(t_{i,j}, X_{i,j})$
  - $t_{i,j} \in \mathbb{R}_+$ outcome : duration of the match
  - $X_{i,j} = \left(\delta_{k,(i,j,m+1,m+2)}\right)_{(m+2)\times 1}$
  - ${X_{i,j}}_{m+1} = 1$ $\sim$ Global mean
  - ${X_{i,j}}_{m+2} = 1$ $\sim$ Variance

  - For skewed Normal :
    - ${X_{i,j}}_{m+3} = 1$ = Skewness

Predictor

- Once fitted, we estimate :

- $\hat{t}_{ij} = \alpha_{i_{bayes}} + \alpha_{j_{bayes}} + \mu_{0_{bayes}}$

- Error:
    - $MSE = 1/n \sum_1^n (t_{ij} - \hat{t}_{ij})^2$
    - Compared to variance

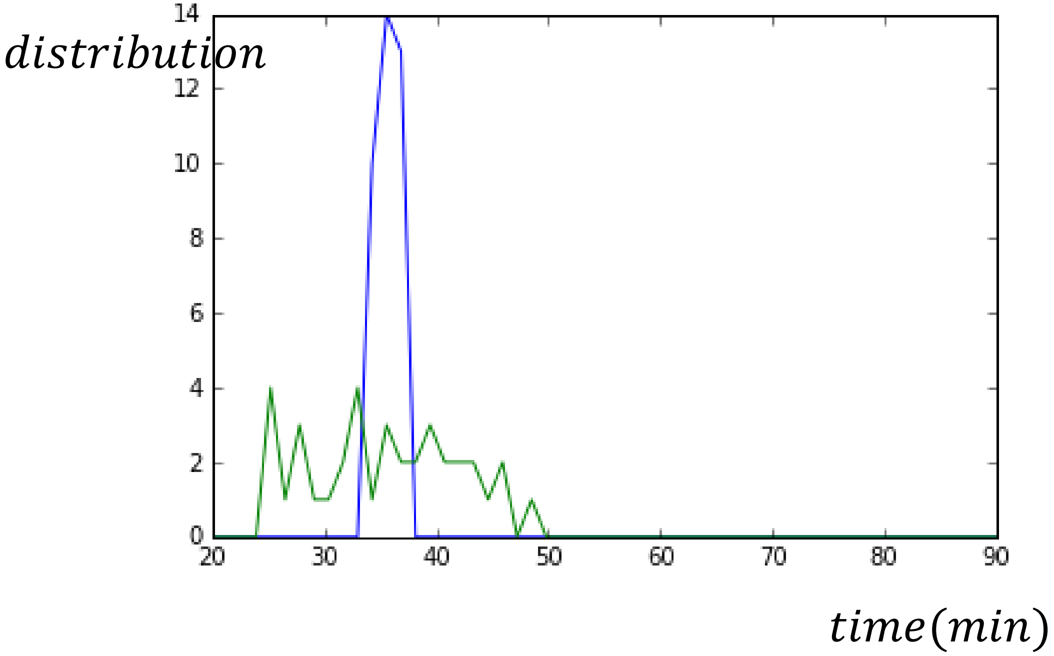
Gaussian Model

- $t_{ij} \sim N\left(\alpha_i + \alpha_j + \mu_0, \sigma\right)$; σ identical for each team

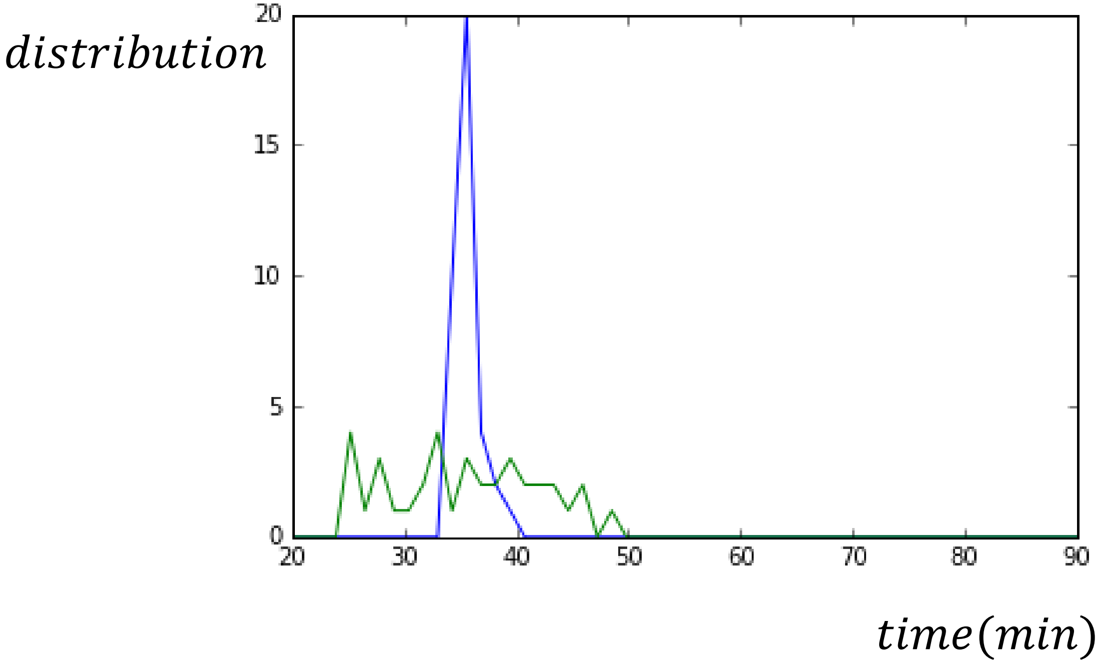
- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_m \\ \mu_0 \\ \sigma \end{pmatrix}$

- Fitted using Langevin MH Sampling or IS

Gaussian predictions



EULCS, 2015



EULCS, 2014-2018

Distribution of durations
Langevin MH Sampled

Fitted Estimator

- θ_{bayes}= [1.57 -0.239 0.038 -0.423 1.71 -1.33 -0.47 0.28 0.005 -0.392 0.342 -1.44 -0.946 2.08 1.020 0.235 35.3 6.38]

- $MSE = 45.12$

- $Var = 43.34$

- For EULCS, 2015 ; IS Sampled

Interpretation

- $|\alpha_i| < 2$; $\mu_0 \approx 35$
- No correlation between paramerers α_i and t_{ij}
- Distribution peak at μ_0

Gamma Model

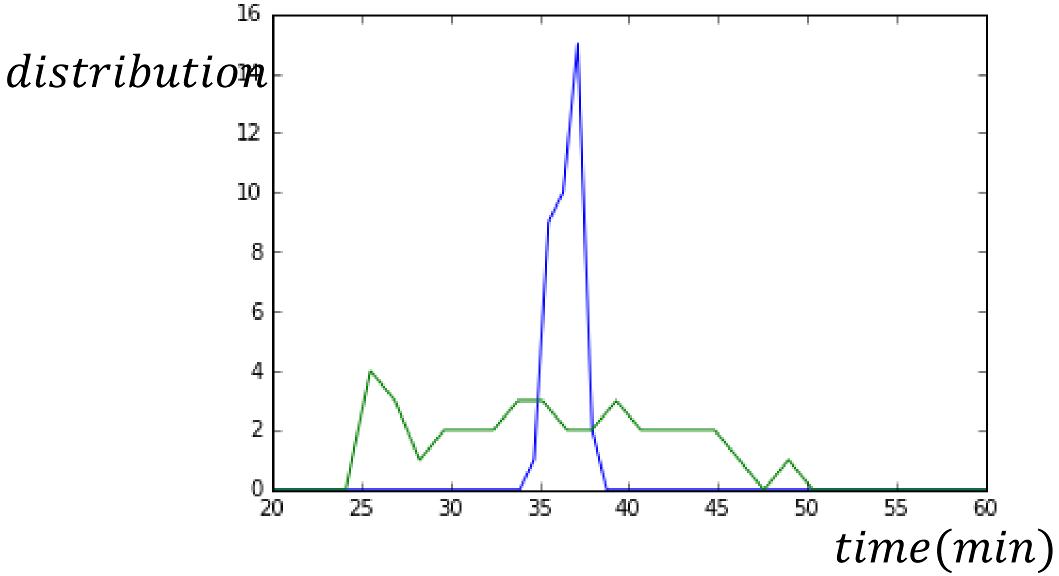
- $t_{ij} \sim Gamma\left(k, \frac{\alpha_i + \alpha_j + \mu_0}{k}\right)$

- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_m \\ \mu_0 \\ k \end{pmatrix}$

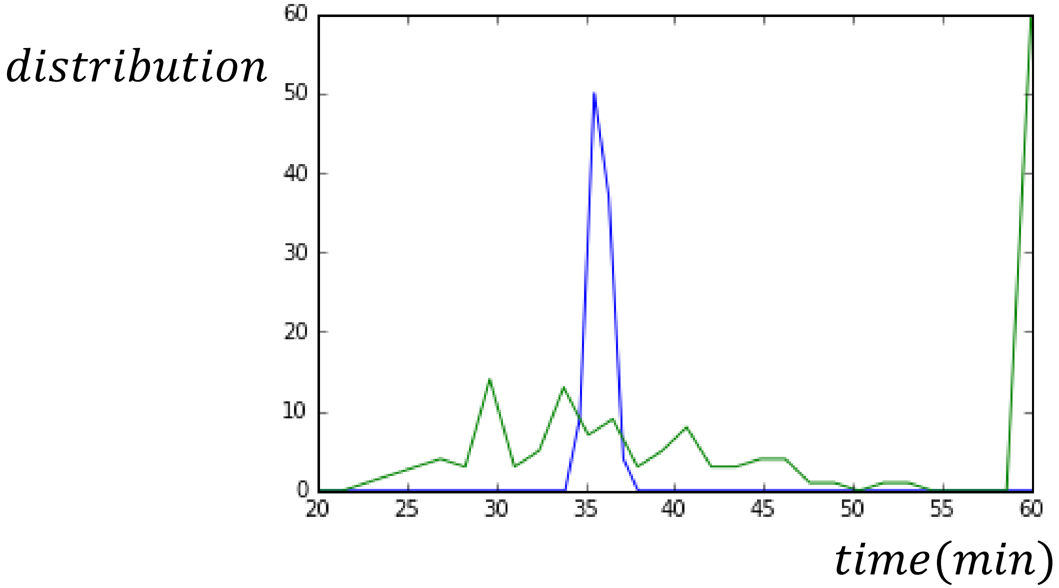
- $\mu_{ij} = k. \frac{\alpha_i + \alpha_j + \mu_0}{k} = \alpha_i + \alpha_j + \mu_0$

- Fitted by Langevin MH Sampling

Gamma predictions



EULCS, 2015



EULCS, 2014-2018

Distribution of durations
Langevin MH Sampled

Errors on 2 Sets

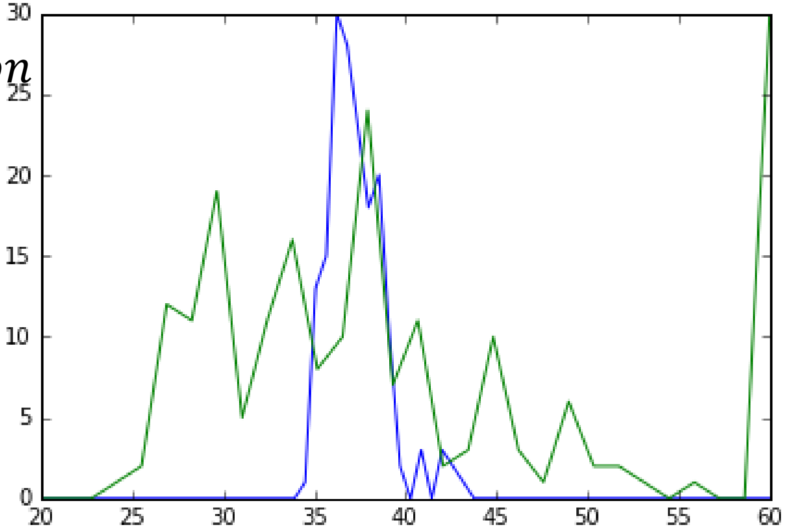
- Error on training set:
  - $MSE_{1,2} = 53.43 , 43.34$
  - $Var_{1,2} = 53.91 , 43.37$
- Error on validation set:
  - $MSE_{1,2} = 57.21 , 68.16$
  - $Var_{1,2} = 57.23 , 68.77$

- No correlation between parameters and estimated quantity

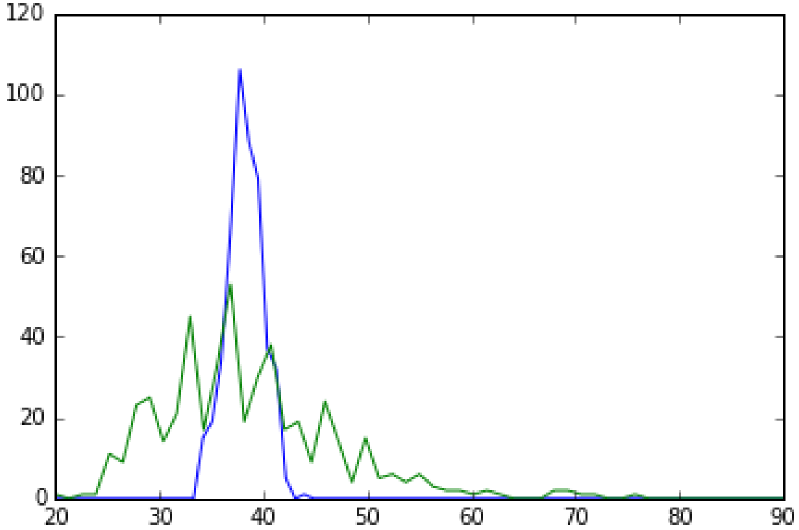
Skewed Gaussian Model

- $t_{ij} \sim SkewedGaussian\left(\alpha_i + \alpha_j + \mu_0, \omega, \varepsilon\right)$
- $\alpha_i + \alpha_j + \mu_0$ is max likelihood and not expectation
- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_m \\ \mu_0 \\ \omega \\ \varepsilon \end{pmatrix}$
- $\mu_{ij} = \alpha_i + \alpha_j + \mu_0 + \omega \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} \sqrt{\frac{2}{\pi}}$, Expectation
- Fitted by Langevin MH Sampling, MH Sampling

Skewed Gaussian Predictions



EULCS, 2015



EULCS, 2014-2018

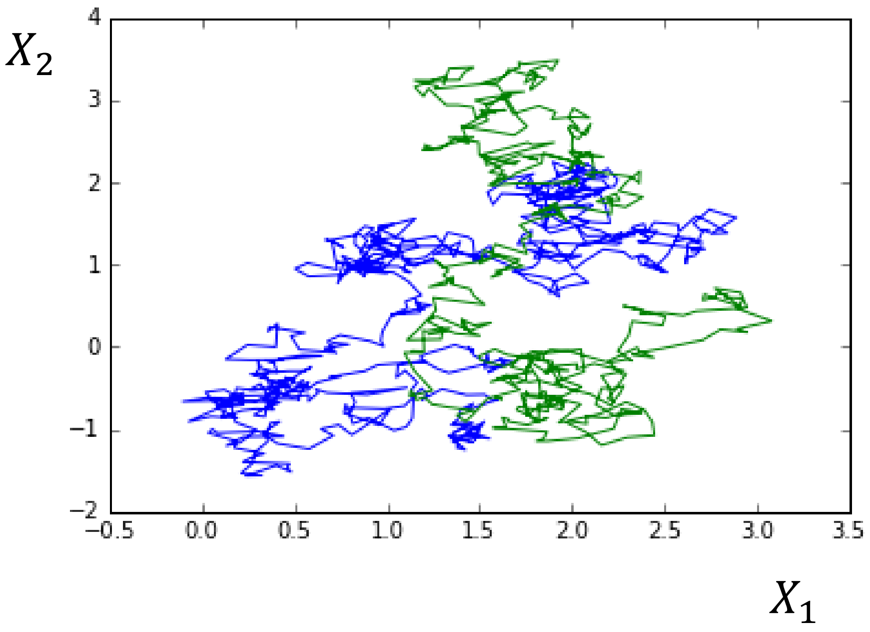
Distribution of durations
MH Sampled

Errors on 2 Sets

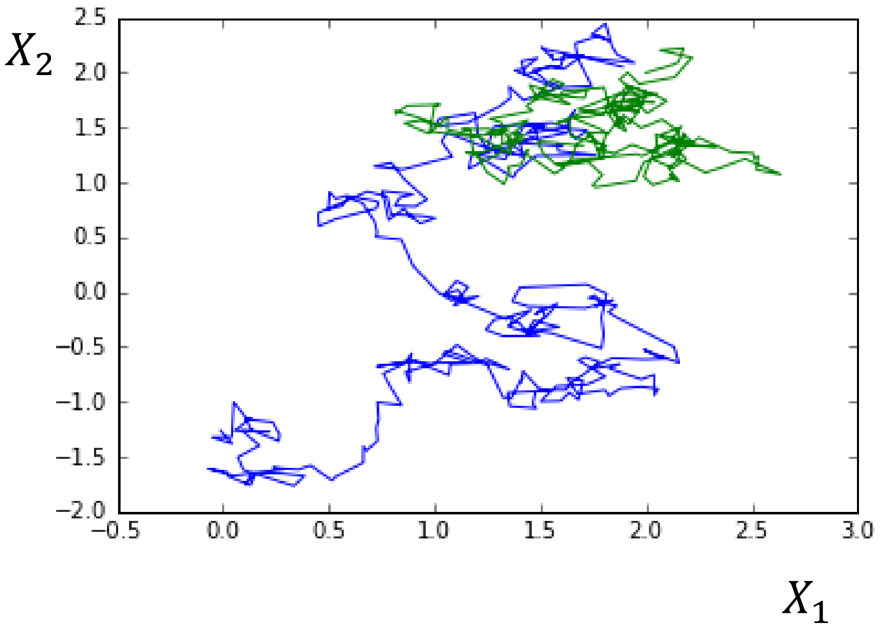
- Error on training set :
  - $MSE_{1,2} = 67.81 , 68.11$
  - $Var_{1,2} = 71.21, 72.70$
- Error on Validation Set :
  - $MSE_{1,2} = 47.57 , 72.28$
  - $Var_{1,2} = 48.32 , 72.84$

- Best model so far

Langevin // MH Sampling



Step-size : 0,1 ; 0,1
600 Epochs



Step-size : 0,1 ; 0,1
300 Epochs

Injection of Logit Results in Skewed Gaussian

- Intuition : Difference of « force » has an influence
- Model:
  - $\mu_{ij} = \alpha_i + \alpha_j + \mu_0 + \kappa \left| f_{i_{bayes}} - f_{j_{bayes}} + \Delta_{bayes} \right|$
- We expect $\kappa < 0$

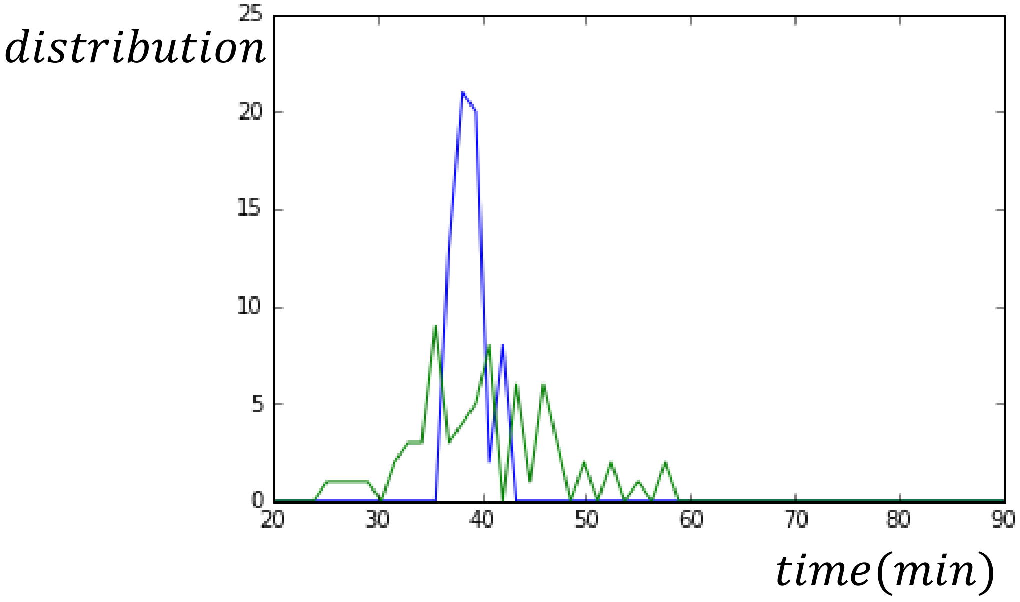
Skewed Gaussian with forces

- $\mu_{ij} = \alpha_i + \alpha_j + \mu_0 + \kappa\left|{f_i}^* - {f_j}^* + \Delta^*\right|$
- $t_{ij} \sim SkewedGaussian\left(\mu_{ij}, \omega, \varepsilon\right)$
- $\alpha_i + \alpha_j + \mu_0$ is max likelihood and not expectation
- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_m \\ \mu_0 \\ \omega \\ \varepsilon \\ \kappa \end{pmatrix}$
- Expectation :

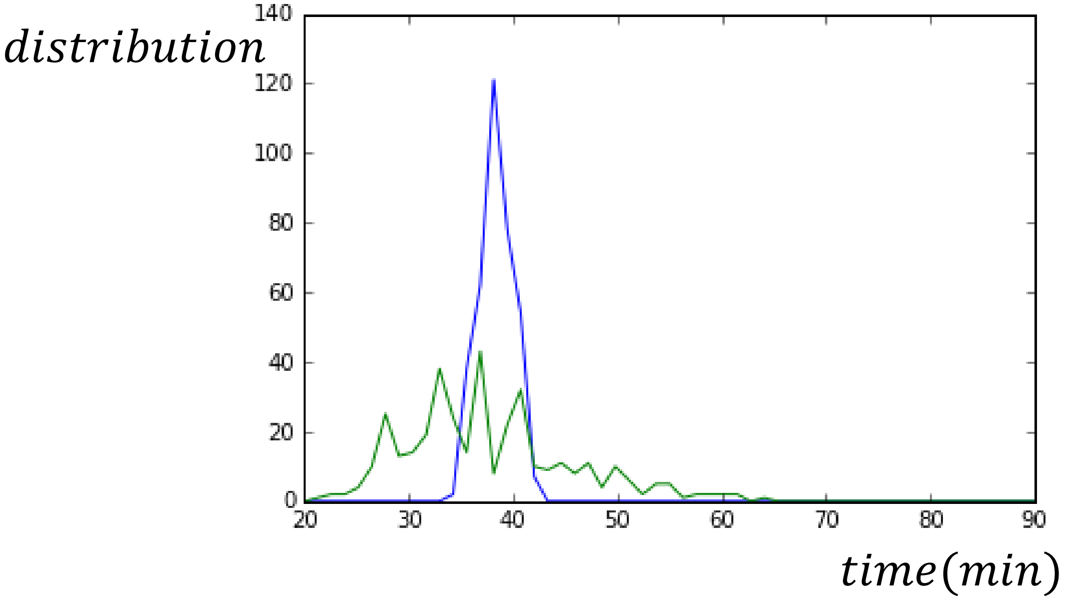
$$\theta_{bayes} = \alpha_{i_{bayes}} + \alpha_{j_{bayes}} + \mu_{0_{bayes}} + \kappa_{bayes}\left|{f_i}^* - {f_j}^* + \Delta^*\right| + \omega_{bayes} \frac{\varepsilon_{bayes}}{\sqrt{1 + {\varepsilon_{bayes}}^2}} \sqrt{\frac{2}{\pi}}$$

- Fitted by Langevin MH Sampling

Skewed Gaussian with forces Predictions



EULCS, 2015



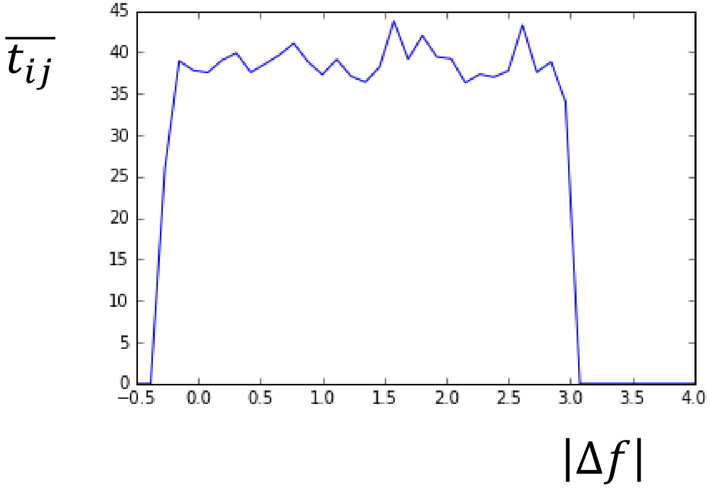
EULCS, 2014-2018

Errors on 2 Sets

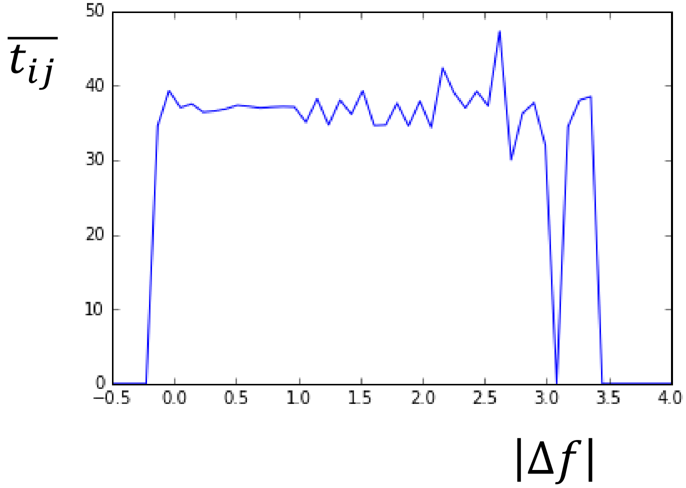
- Error on training set :
  - $MSE_{1,2} = 58.90 , 61.78$
  - $Var_{1,2} = 63.85 , 66.96$
- Error on Validation Set :
  - $MSE_{1,2} = 59.07 , 59.16$
  - $Var_{1,2} = 63.86 , 64.70$

- Best model, but not much influence of force parameters

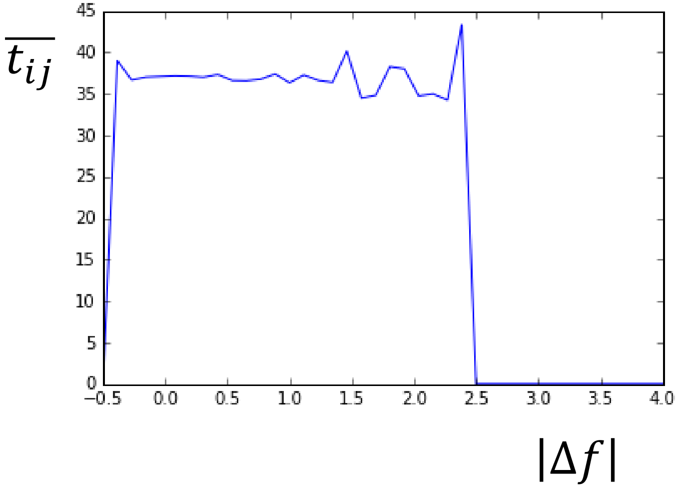
Influence of Force Parameters



LCK, 2014-2018



EULCS, 2014-2018



All regions, 2014-2018