Bayesian Inference in e-sport matches

Bayesian Computation Project

Etienne ROUANET-LABE

2018

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

Summary

1. <u>Prediction of the winner of a match</u>
    A. Logit Model
        a. Posterior
        b. Sampling Methods / Predictions
        c. Inflence of Time scale and Training Set
    B. Logit Model for International events
        a. Model
        b. Predictions

2. <u>Prediction of the duration of a match</u>
    A. Basic models
        a. Gaussian
        b. Gamma
        c. Skewed Gaussian
    B. Injection of Logit Model for duration prediction
        a. Model
        b. Predictions

Prediction
of the winner of a match

A Bayesian Linear Model

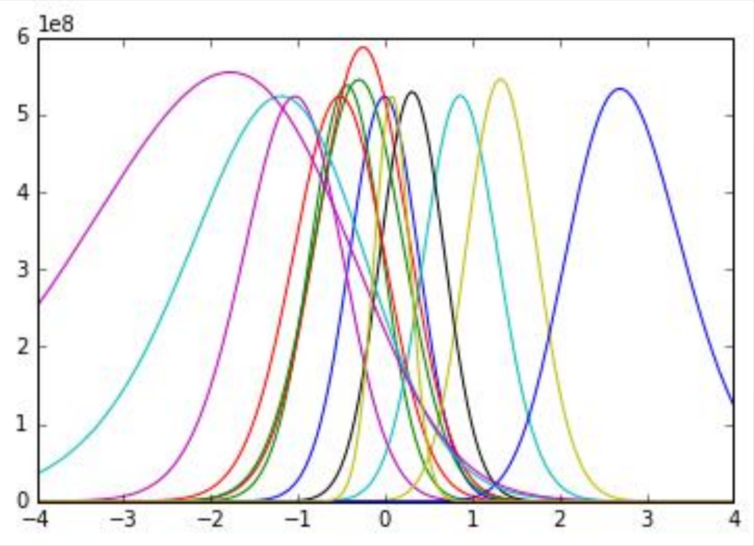
- $\theta = \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_m \\ \Delta \end{pmatrix}$
- $\boldsymbol{p_{ij} = logit^{-1}(f_i - f_j} + \Delta) = logit^{-1}(\theta^T x_k)$
- We have linear dependency:
  - $y_k \sim B(logit^{-1}(\theta^T x_k))$
- Then we have :
  - $p(y_k | \theta, x_k = X_{i,j}) = \left(\frac{1}{1+e^{-(f_i - f_j + \Delta)}}\right)^{y_k} . \left(1 - \frac{1}{1+e^{-(f_i - f_j + \Delta)}}\right)^{1-y_k}$

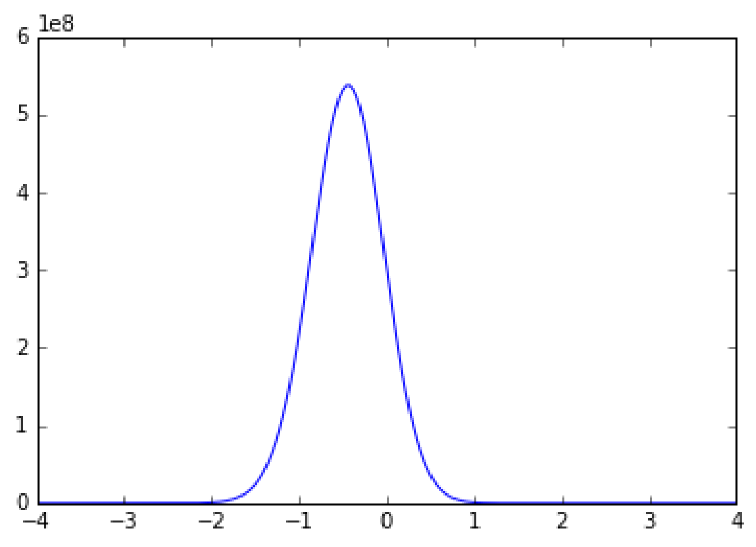
Gaussian Prior

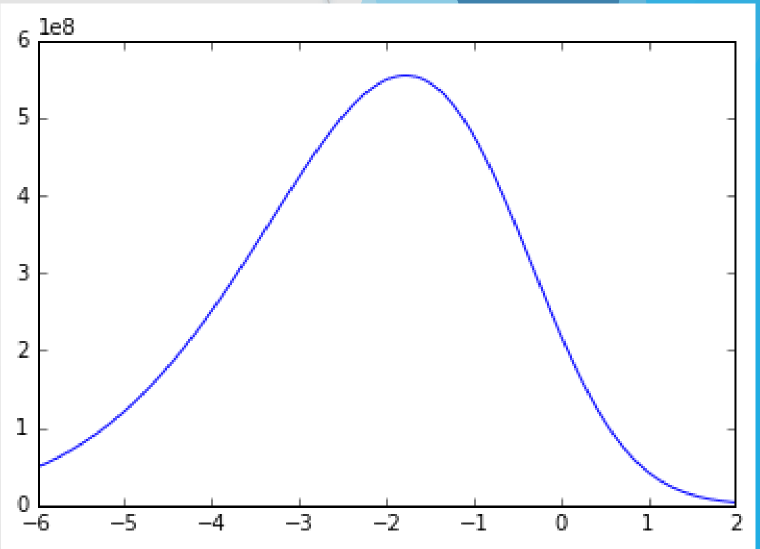
- $\boldsymbol{\theta_i \sim N\left(\mu_i, \sigma_i^2\right)}$
  - $\mu_{1 \le i \le m} = 0$; no intuition
  - μ_{m+1}=0 ; game well-balanced
  - $\sigma_{1 \le i \le m} = 4$; flexible model
  - σ_{m+1} =1 ; stronger intuition for the bias

Posterior

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







Predictor

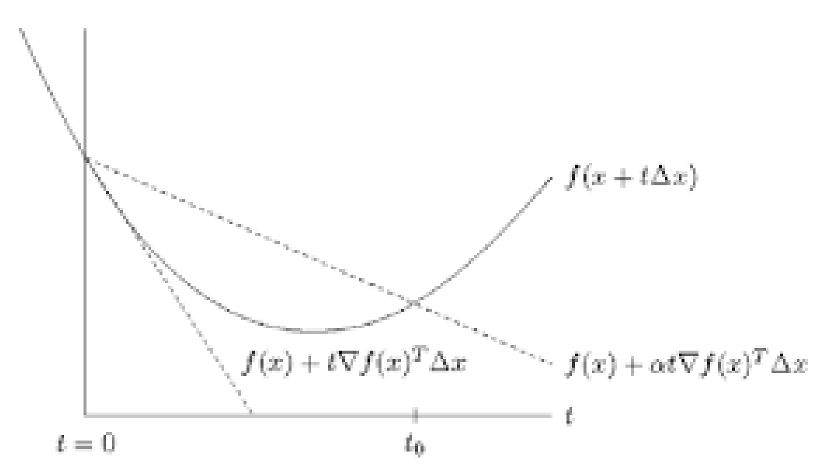
- Given $x_k = X_{i,j}$ we estimate y_k from θ^*:
  - $\boldsymbol{f_i}^* - \boldsymbol{f_j}^* + \Delta^* \geq \mathbf{0} \Longrightarrow \widehat{\boldsymbol{y}}_{\boldsymbol{k}} = \mathbf{1}$
- « Accuracy » displayed by my algorithms for n matches:
  - $A_{\%} = \frac{\#\{k \in [\![1,n]\!] \mid \hat{y}_k = y_k\}}{n} . 100 \quad ; \quad n = len(Validation\ Set)$

Laplace Approximation

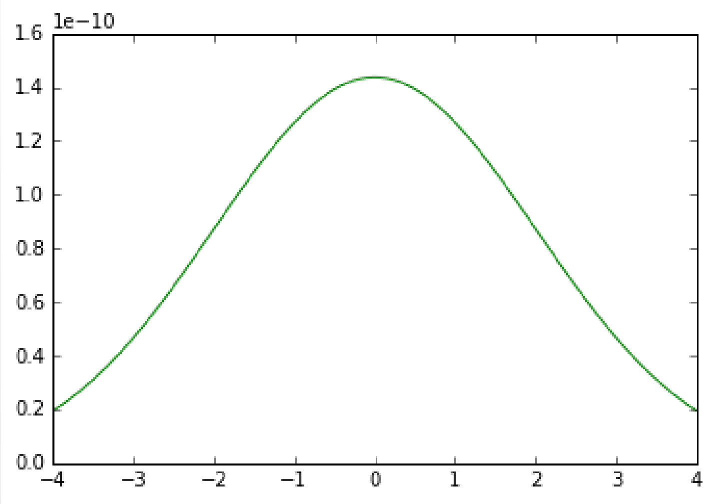
- Find $\boldsymbol{\theta}^* = \boldsymbol{argmax}_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{\theta}|\boldsymbol{D})$ by gradient descent
  - (also works with $\log(f(\theta|D))$
- Approximate Covariance Matrix by inverse of log-curvature:
  - $\boldsymbol{Cov}^* = \left(-\boldsymbol{\nabla}^2 (\mathbf{log}(\boldsymbol{f}))\big|_{\boldsymbol{\theta}^*} \right)^{-\mathbf{1}}$
- Then $(\boldsymbol{\theta}|\boldsymbol{D}) \sim \boldsymbol{N}(\boldsymbol{\theta}^*, \boldsymbol{Cov}^*)$

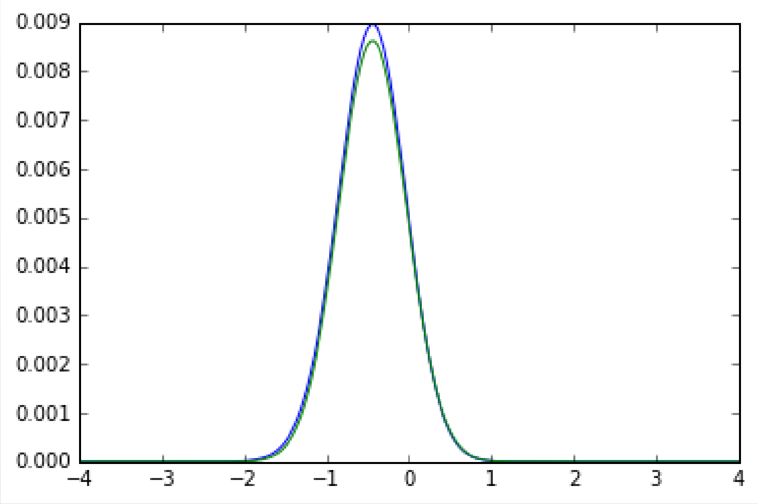
Backtracking Line-Search Gradient Descent

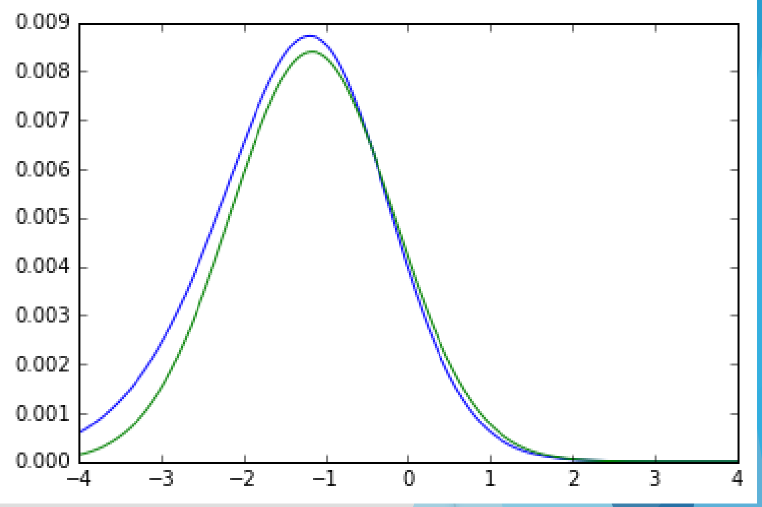
- Optimize choice of step by making sure we go down enough :
  - θ_n initial point ; λ_0 initial step-size ; E function to optimize ; (α, β)
  - At each step:
    - Initialize $\lambda = \lambda_0$
    - Candidate is $\theta_c = \theta_n - \lambda \nabla E(\theta_n)$
    - Accepted if $E(\theta_c) \leq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



Laplace Approximation







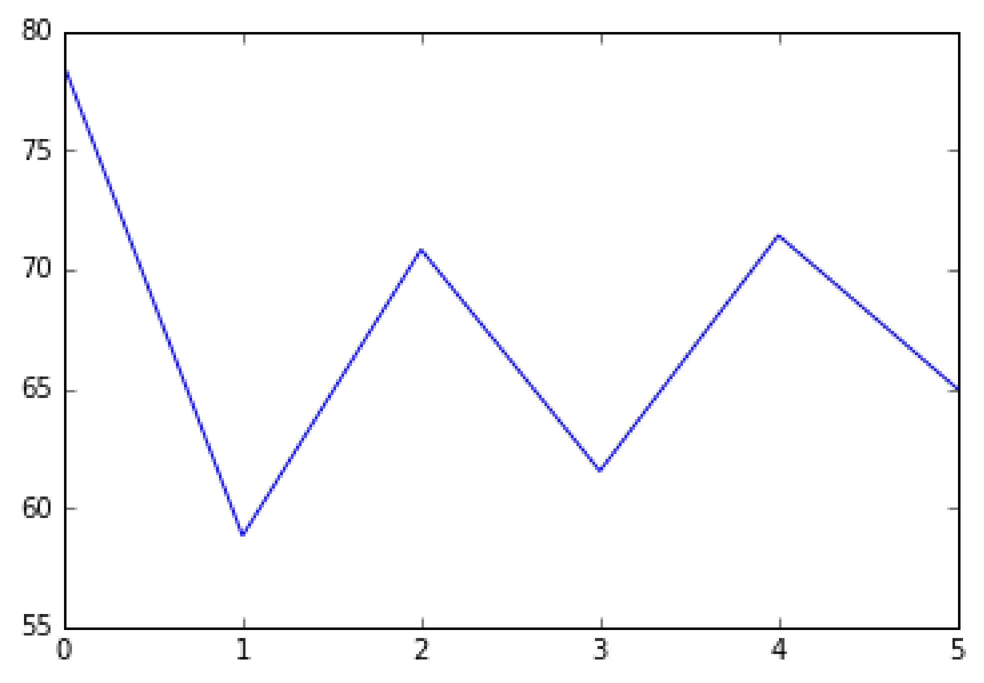
Gaussian and L.A.

LA on symmetric dimension

LA on assymetric dimension

Laplace Predictions

Accuracy



Split

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

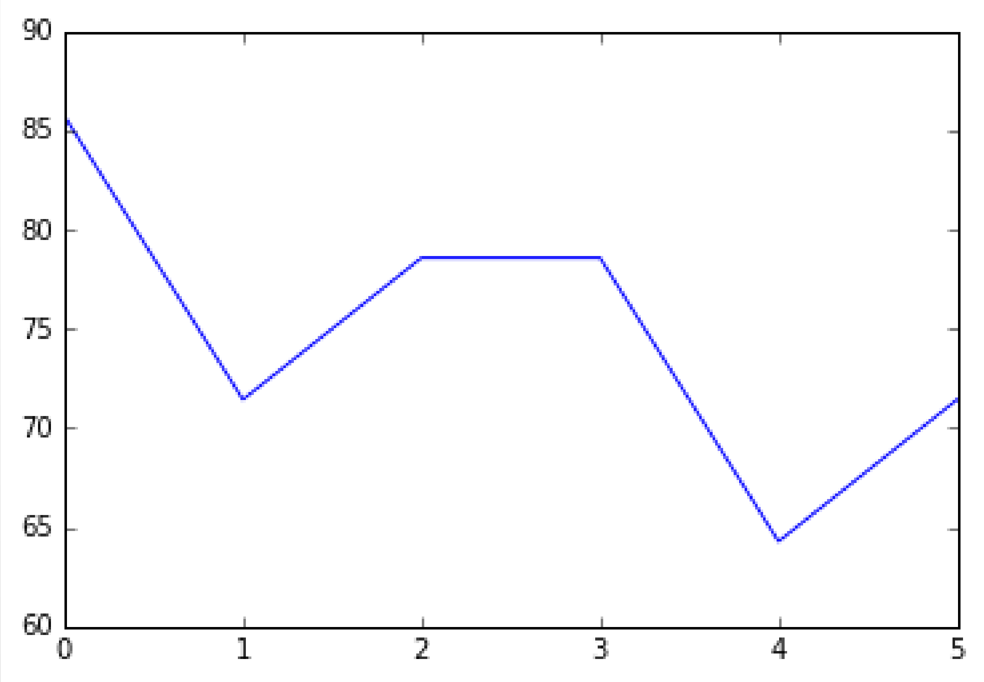
Predictions of winners by split in EULCS (2015-2017)

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:
    - $\boldsymbol{E(\theta \sim f(\theta|D)) \approx \frac{\sum_i p_i \theta_i}{\sum_i p_i}}$

IS Predictions



$Accuracy$

$Split$

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

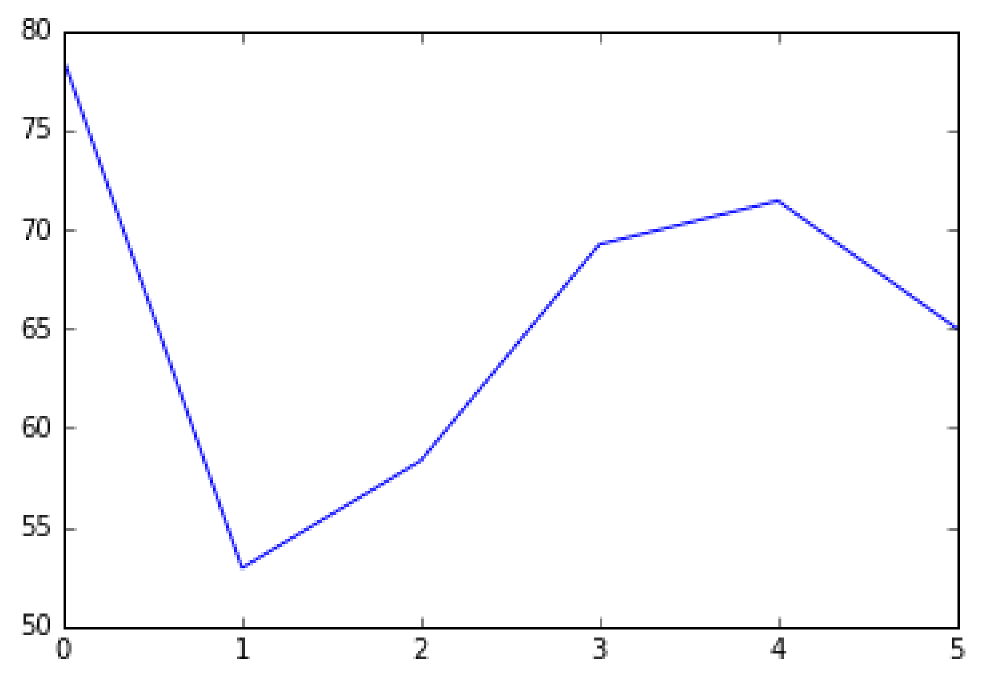
Training Set ½

$N = 15000$

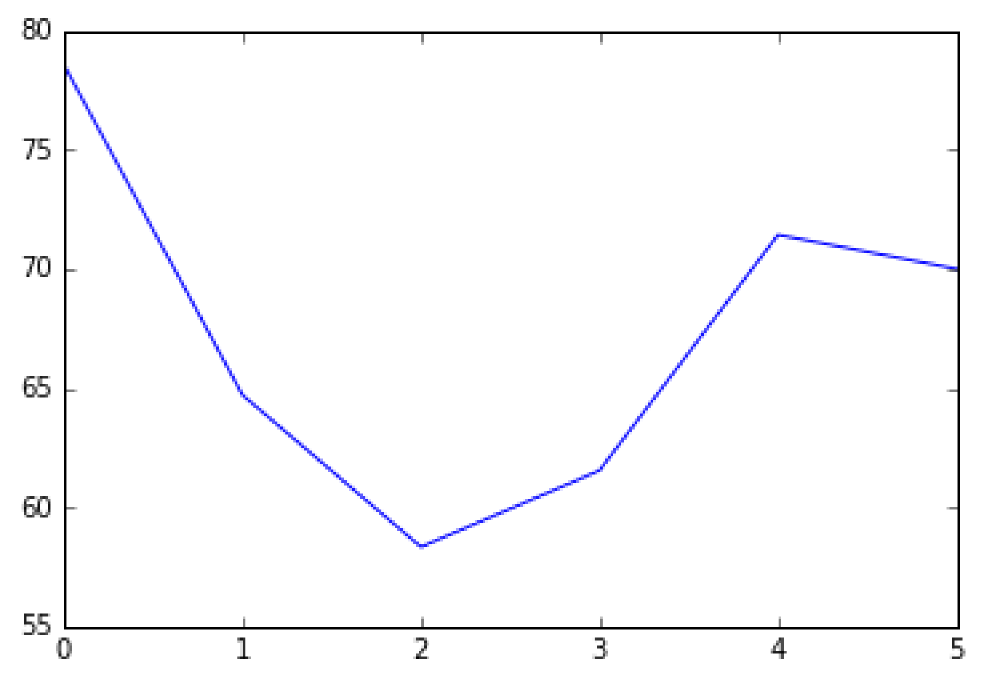
Metropolis-Hastings

- 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 $\boldsymbol{\theta_c = \theta_n + \lambda\eta_n}$; $\eta_n \sim N(0,1)$ for example
    - Accept θ_c with probability $\boldsymbol{p_c = min\left\{\frac{f(\theta_c)}{f(\theta_n)}, 1\right\}}$
      - 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

MH Predictions



5000 epochs ; 0,1 step-size



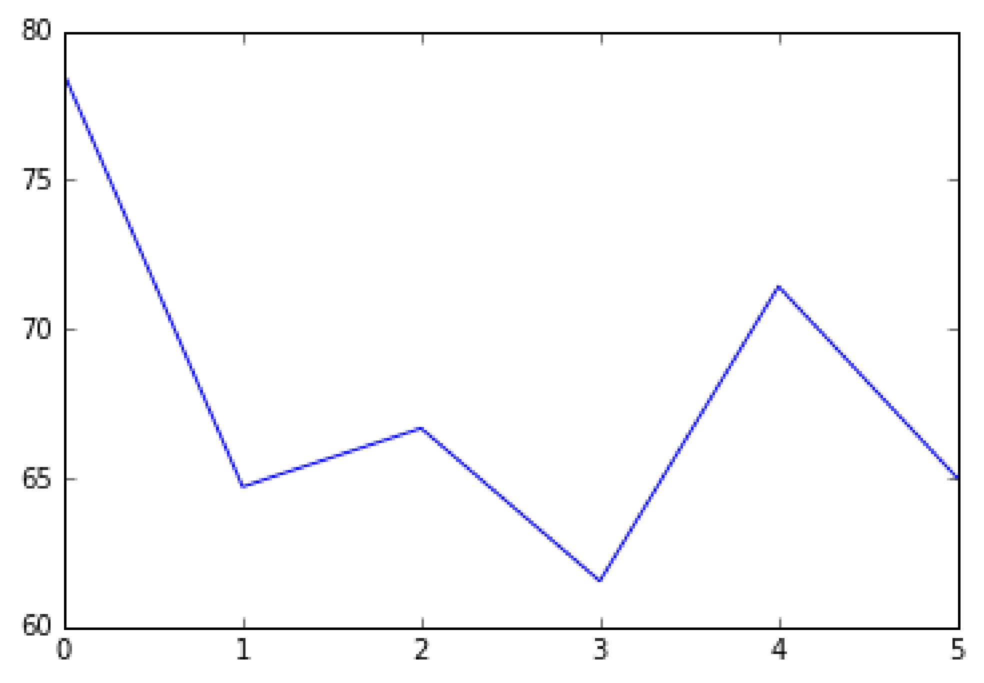
5000 epochs ; 0,1 step-size

Predictions of winners by split in EULCS (2015-2017)
Training Set 1/2

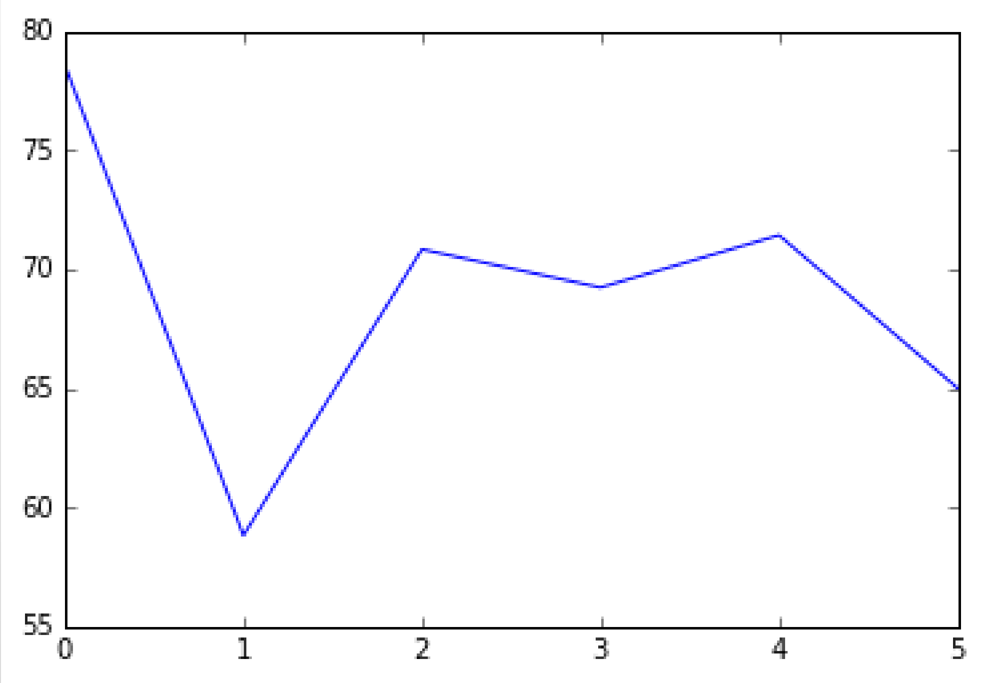
Metropolis-adjusted Langevin

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

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

- Estimated θ_{bayes} , EULCS 2015 :
  - Δ_{bayes}=0.41554176

- Interpretation: « force » for each team

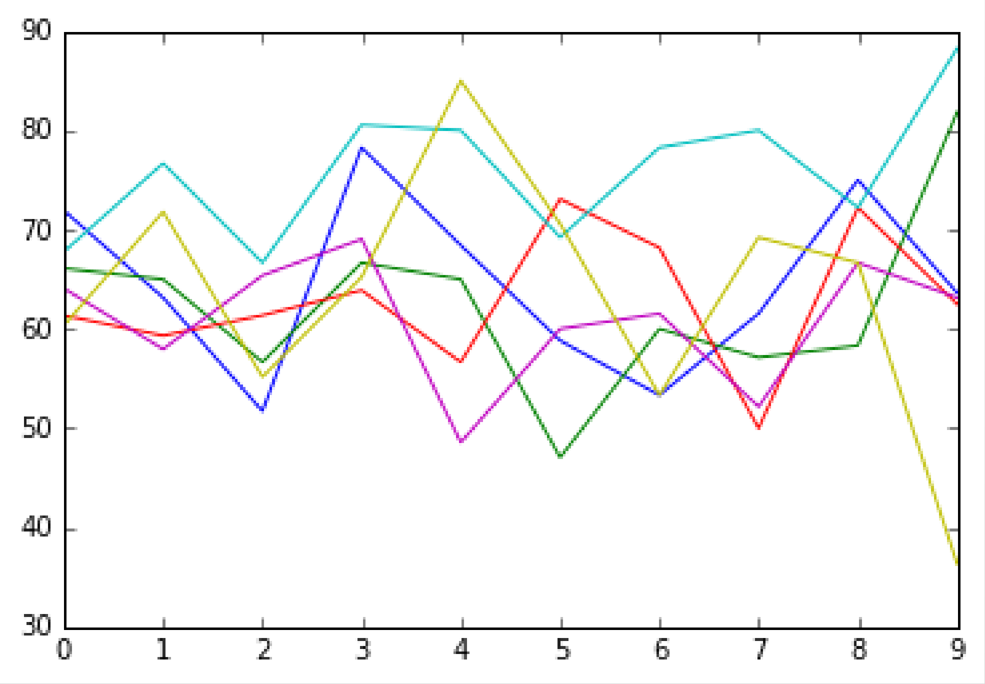
- The game is not well balanced : $p\left(y_{ij} = 1 \middle| f_i = f_j\right) \approx 0{,}6$

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 : Region EULCS

$Accuracy$



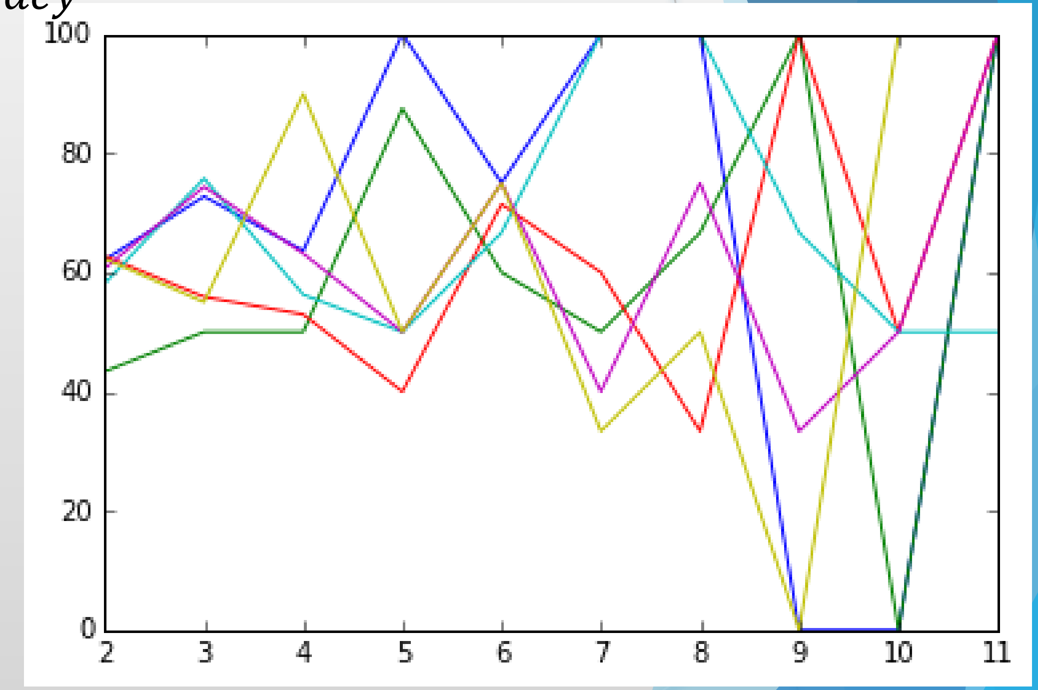
$\#TS$

Periodic Training Set

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

MH Sampling
500 epochs

$Accuracy$



$\#TS$

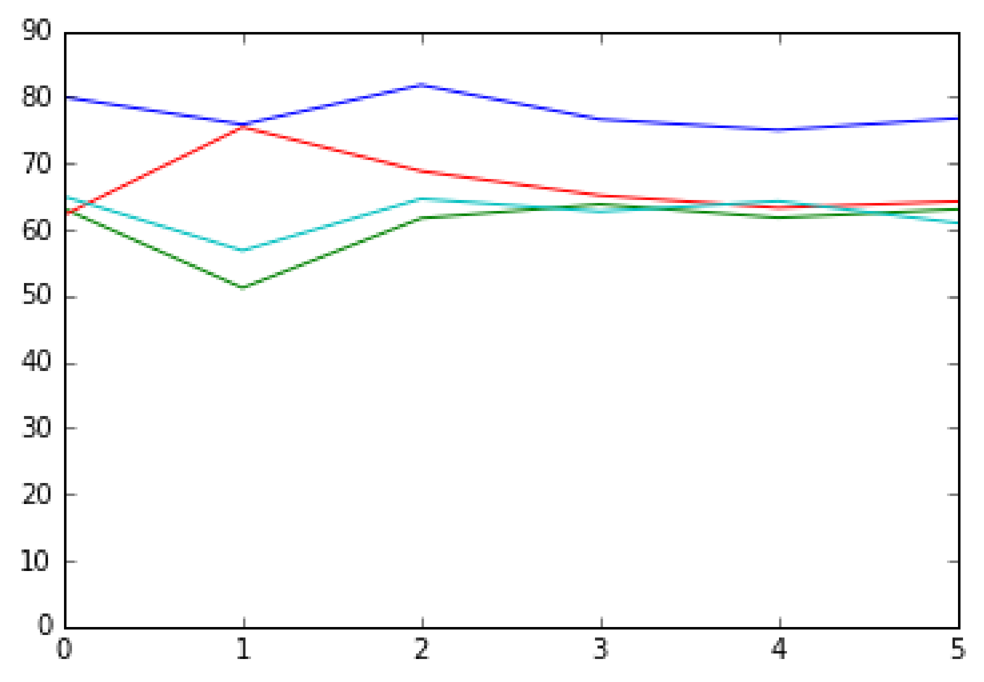
Random Training Set

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

MH Sampling
500 epochs

Influence of Time Scale

$Accuracy$



$Time\ Scale$

Different time scale predictions for 4 regions

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

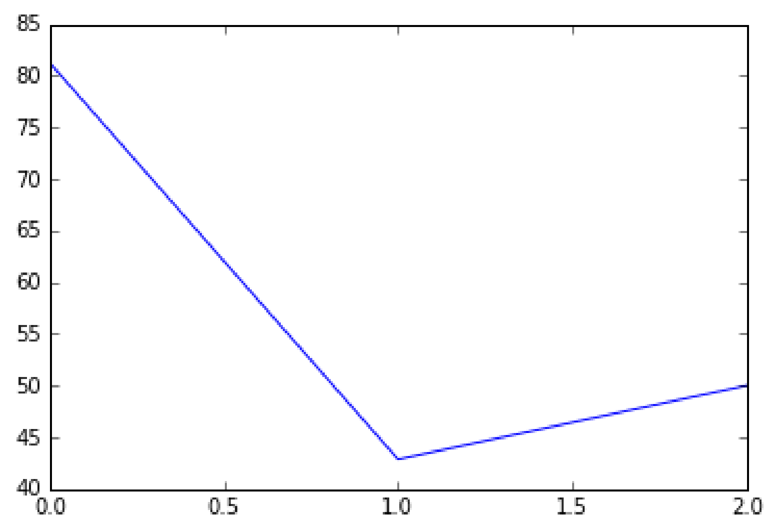
Training Set 1/10

IS Sampling ; $N = 15000$

Predictions for international tournaments

- Is our model relevant for international tournaments?

$Accuracy$



$Year$

Predictions by year in international tournaments (2015-2017)

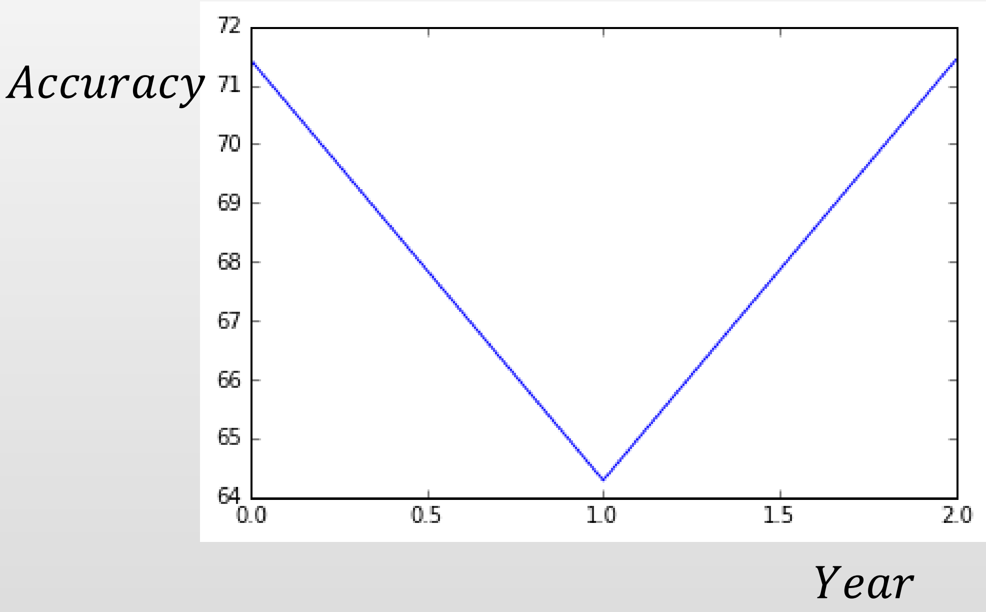
Training Set 1/10

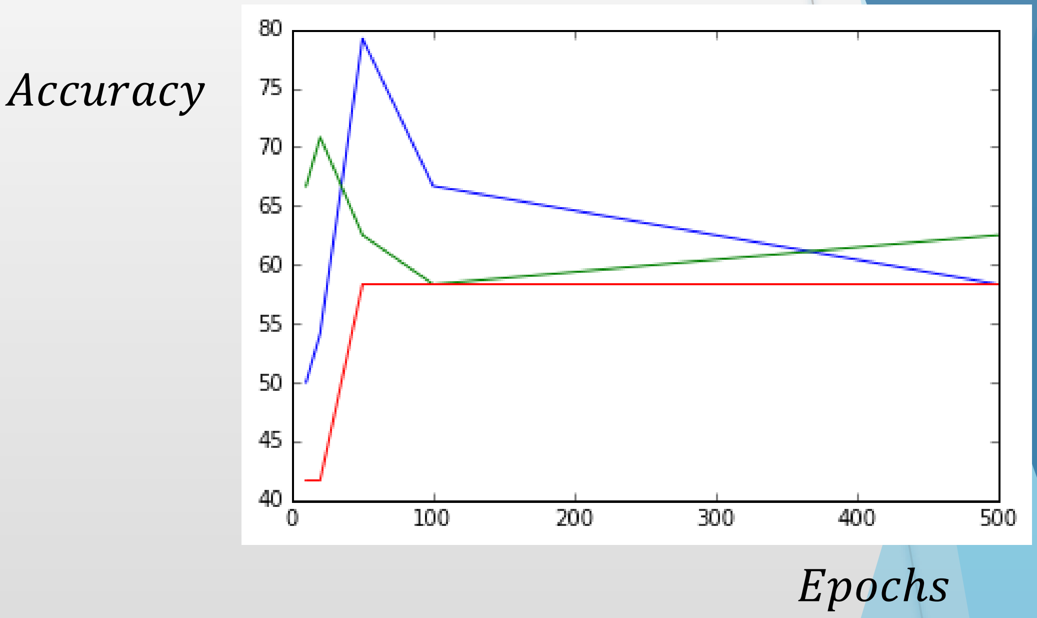
IS Sampling ; $N = 15000$

Parametric Model for International Matches

- $\theta = \begin{pmatrix} f_1 \\ f_2 \\ \cdots \\ f_{m_1} \\ F_1 \\ \cdots \\ F_{m_2} \\ \Delta \end{pmatrix}$
- Parametric : $y_k \sim B(logit^{-1}(\theta^T x_k))$
- $\boldsymbol{p_{i_1 j_1 i_2 j_2} = logit^{-1}(f_{i_1} - f_{i_2} + F_{j_1} - F_{j_2} + \Delta)}$
- Regions quite heterogeneous ; model should fit

Predictions from new model





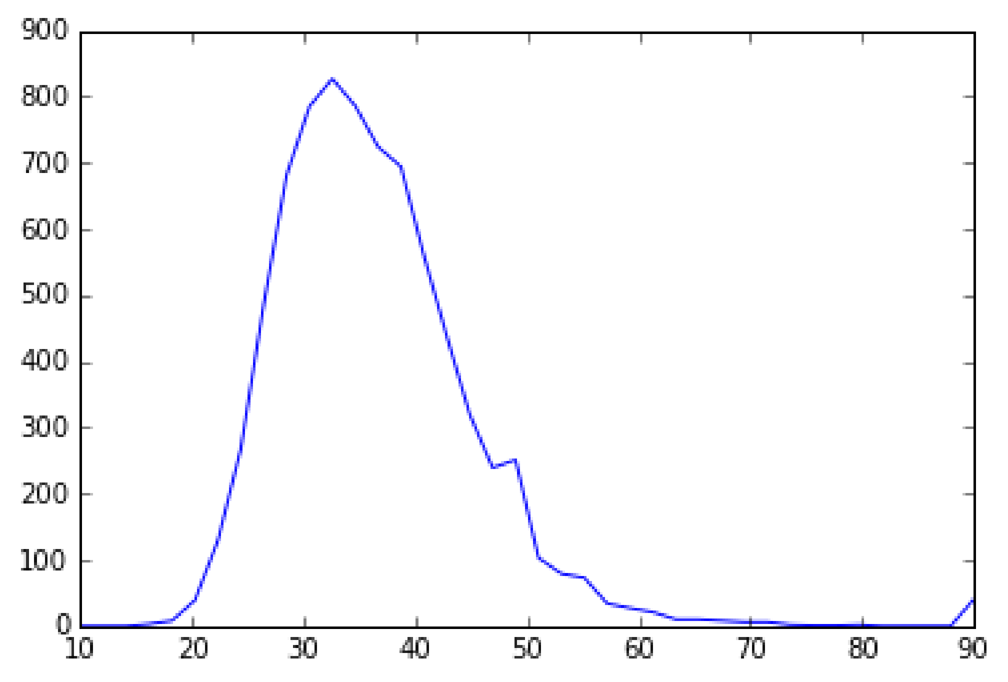
Predictions per year in international tournaments
(2015-2017)
Training Set 1/7
Langevin MH Sampling ; 500 Epochs

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

Distribution of durations

$\#teams$



$time(min)$

Duration of matches
All years, all regions

Parametric model

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

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

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

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

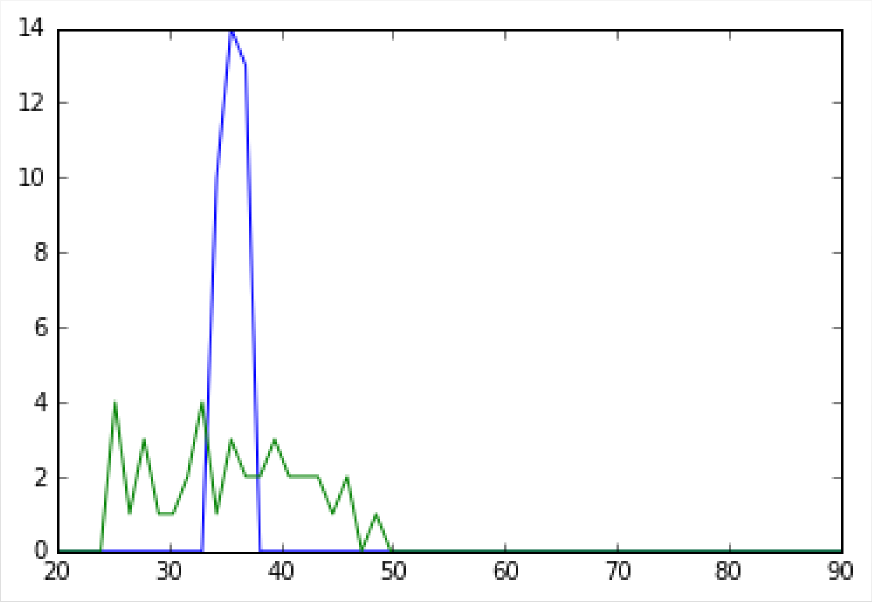
- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_m \\ \mu_0 \\ \sigma \end{pmatrix}$

- **Gaussian Prior** for θ

- Fitted using Langevin MH Sampling or IS

Gaussian predictions

$distribution$



$time(min)$

- $MSE = 45.12$
- $Var = 43.34$

EULCS, 2015

Distribution of durations
Langevin MH Sampled

Gamma Model

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

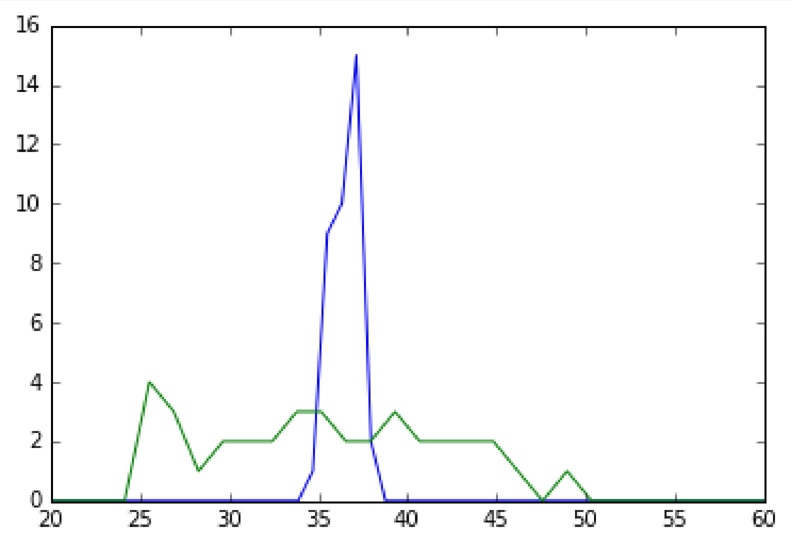
- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_m \\ \mu_0 \\ k \end{pmatrix}$

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

- Fitted by Langevin MH Sampling

Gamma predictions

$distribution$



$time(min)$

EULCS, 2015

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

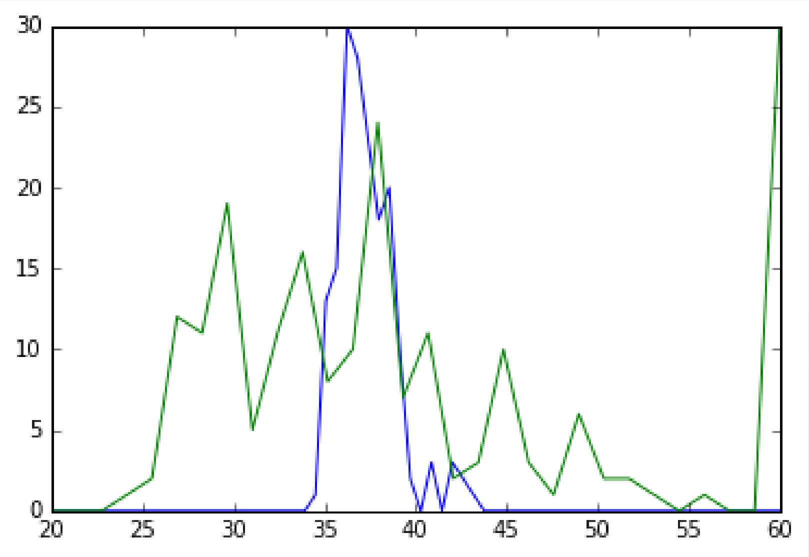
Distribution of durations
Langevin MH Sampled

Skewed Gaussian Model

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

Skewed Gaussian Predictions

$distribution$



$time(min)$

EULCS, 2015

Distribution of durations
MH Sampled

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

Injection of Logit Results in Skewed Gaussian

- Intuition : Difference of « force » has an influence
- Model:
  - $\boldsymbol{\mu_{ij}} = \boldsymbol{\alpha_i} + \boldsymbol{\alpha_j} + \boldsymbol{\mu_0} + \boldsymbol{\kappa}\left|\boldsymbol{f_i}^* - \boldsymbol{f_j}^* + \Delta^*\right|$
- We expect $\kappa < 0$

Skewed Gaussian with forces

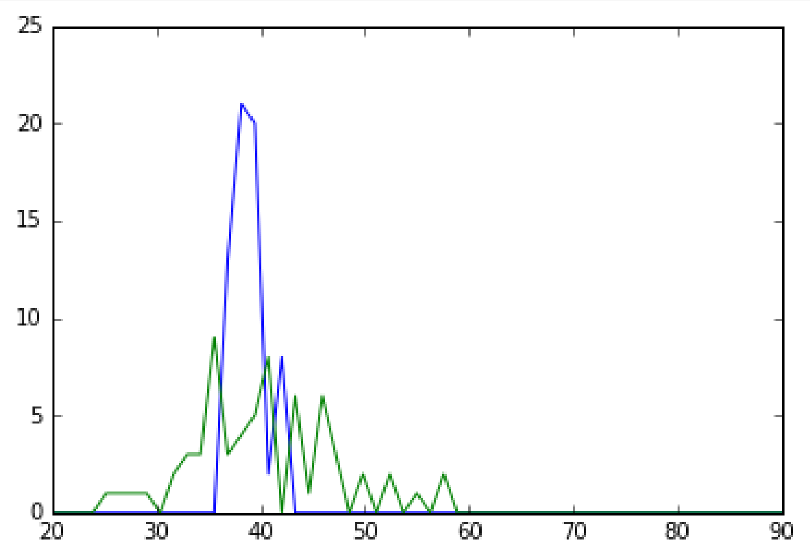
- $\boldsymbol{\mu_{ij} = \alpha_i + \alpha_j + \mu_0 + \kappa\left|f_i{}^* - f_j{}^* + \Delta^*\right|}$
- $t_{ij} \sim SkewedGaussian(\mu_{ij}, \omega, \varepsilon)$
- $\alpha_i + \alpha_j + \mu_0$ is max likelihood and **not expectation**
- $\theta = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdots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

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

Fitted by Langevin MH Sampling

Skewed Gaussian with forces Predictions

$distribution$



$time(min)$

EULCS, 2015

Error on training set :

$MSE_{1,2}$ = 58.90 , 61.78

$Var_{1,2}$ = 63.85 , 66.96

Error on Validation Set :

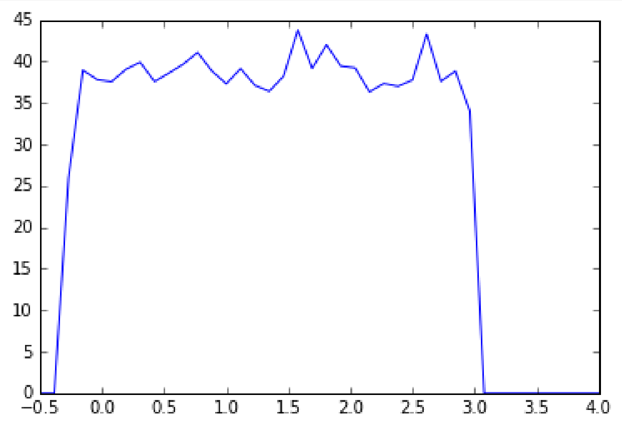
$MSE_{1,2}$ = 59.07 , 59.16

$Var_{1,2}$ =63.86 , 64.70

Distribution of durations
MH Sampled

Influence of Force Parameters

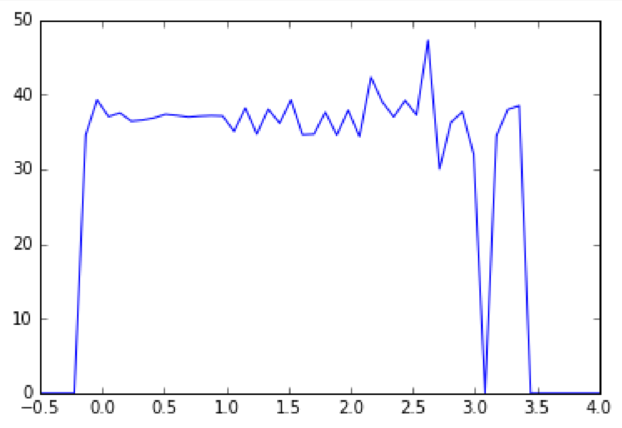
$\overline{t_{ij}}$



$|\Delta f|$

LCK, 2014-2018

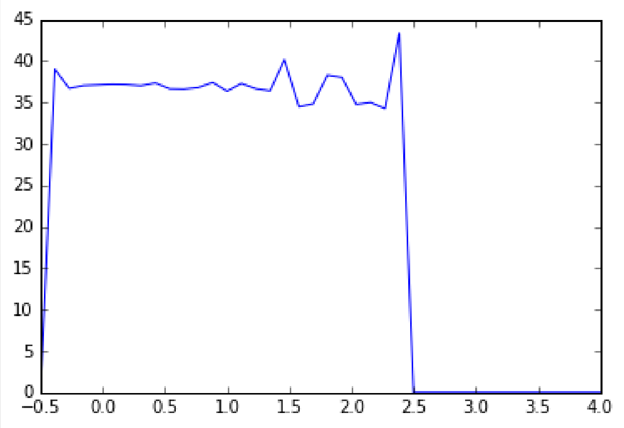
$\overline{t_{ij}}$



$|\Delta f|$

EULCS, 2014-2018

$\overline{t_{ij}}$



$|\Delta f|$

All regions, 2014-2018

Questions ?