From Science to Data Science

Stephane SENECAL – Orange Labs
CRI Data Science Club, 31/03/2017

Overview

- Academic background
- Professional background and activities
- Projects
- Working @ Orange Labs
- Machine Learning…

Academic background (1/3)

- Pre-PhD:
  - MS in mathematics:
    - obtained in 1998 @ Joseph Fourier University, Grenoble, France
    - thesis on holomorphic functions of several complex variables
  - MEng/MS in signal processing:
    - obtained in 1999 @ Grenoble INP, France
    - thesis on curvilinear component analysis for model order estimation

Academic background (2/3)

- PhD in statistical signal processsing for telecommunications:
  - obtained from Grenoble INP, France
  - conducted from 1999 to 2002 @ GIPSA-Lab
  - with MESR fellowship support

  - → **Statistical simulation methods:**
    - Markov Chain Monte Carlo (MCMC): Hastings-Metropolis, Gibbs sampling, reversible jumps MCMC
    - Sequential Monte Carlo/Particle Filtering

  - → Application to **Bayesian model estimation problems** in:
    - Independent Component Analysis/Blind Source Separation
    - Equalization of nonlinear system for satellite communications

Academic background (3/3)

- Post-PhD: Post-Doc in Computational Statistics
  - conducted @ Institute of Statistical Mathematics/Research Organization of Information and Systems (Tokyo, Japan) in 2003-2004
  - thanks to a JSPS fellowship support
  - **Design** of statistical **simulation** algorithms/methods/**techniques**:
    - Block/fixed-lag sampling strategies for Sequential Monte Carlo methods, applications in:
      - optimal filtering for bearing-only target tracking in radar
      - stochastic volatility in econometrics
    - Space alternating data augmentation, application to the estimation of finite mixtures of Gaussian distributions for speaker recognition

Professional background (1/7)

- Research Engineer/Scientist @ Orange Labs since 2005:
  - 2005-2006: Orange Labs Tokyo (Japan)
  - since 2007: Orange Labs Paris (now in Châtillon (92))

- Tackling problems and models/techniques/algorithms/methods in **Machine Learning** (statistical learning):
  - reinforcement learning
  - supervised learning
  - unsupervised learning

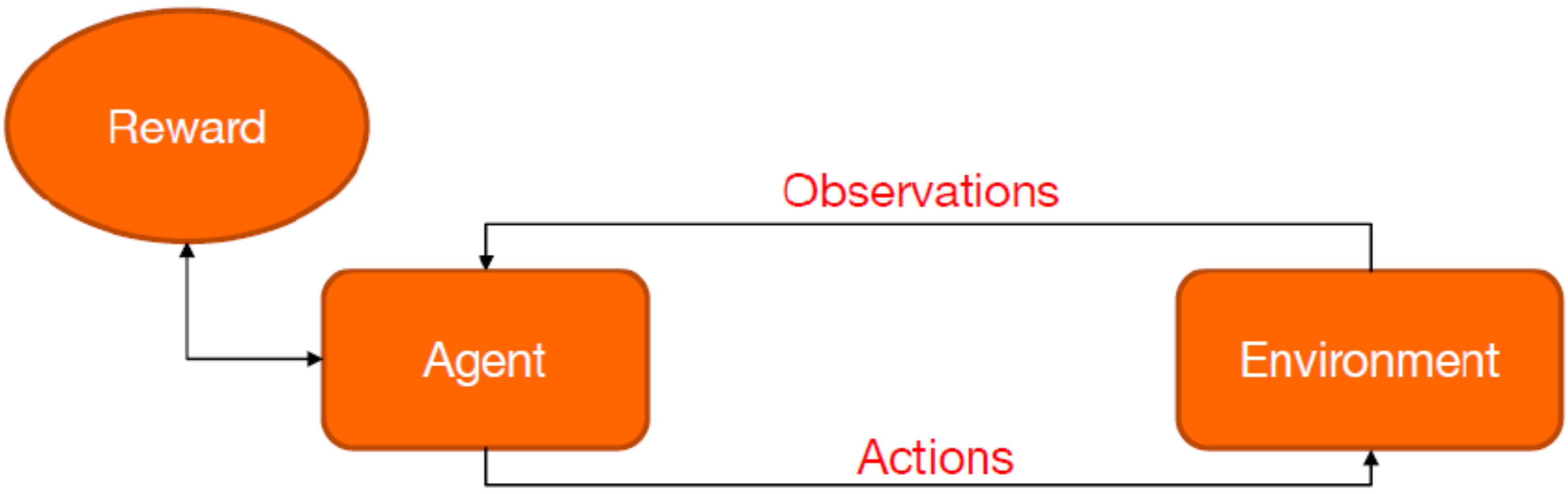
- Applications to telecommunications:
  - design of fixed and mobile networks optimization systems
  - design of traffic data processing systems

Professional background (2/7)
Orange Labs Tokyo 2005-2006

- Markov Decision Processes models and Reinforcement Learning techniques:
  - Dynamic programming
  - Temporal Differences (TD-lambda)
  - Q-Learning algorithm and its extensions (SARSA, eligibility traces)
  - Parametric approximation techniques (Policy Gradient, Least Squares Policy Iteration)
- Support Vector Machines (SVM) techniques and related extensions for regression and classification

Reinforcement Learning (1/2)

Reinforcement Learning (2/2)

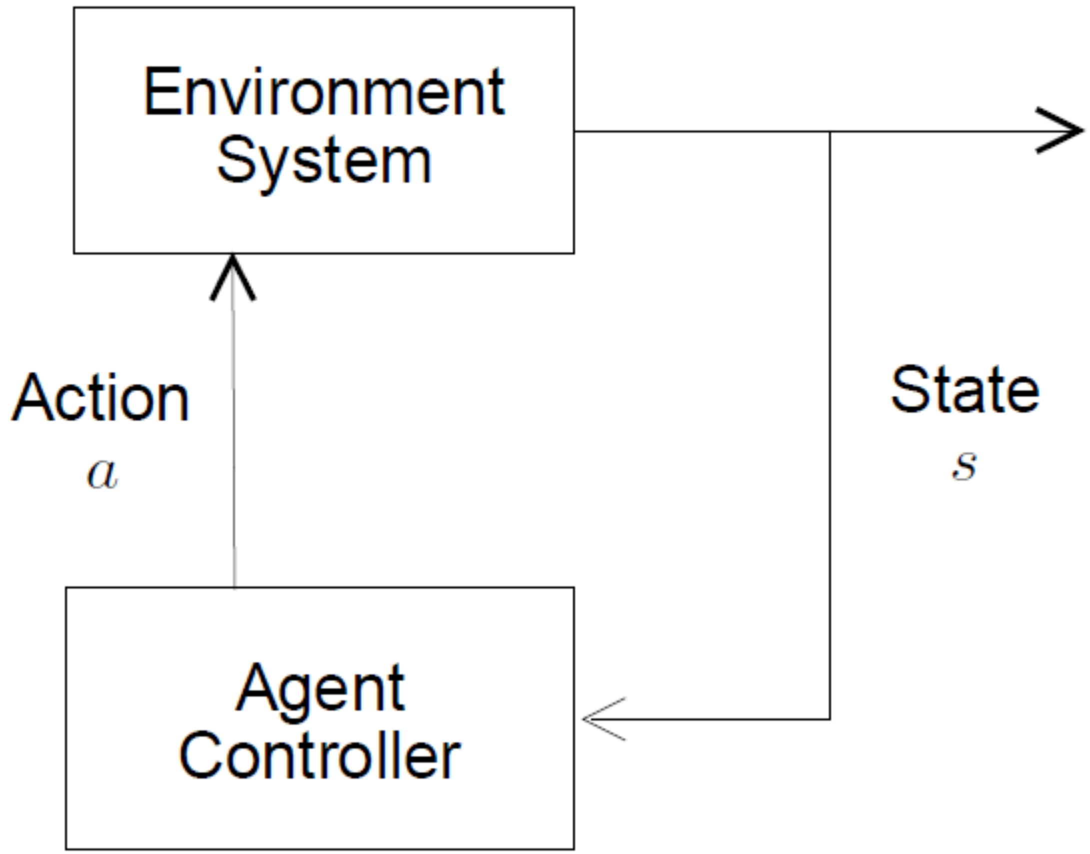


Reinforcement learning goal: **optimize** rewards by choosing adequately actions for given observations $\Rightarrow$ from **policies**

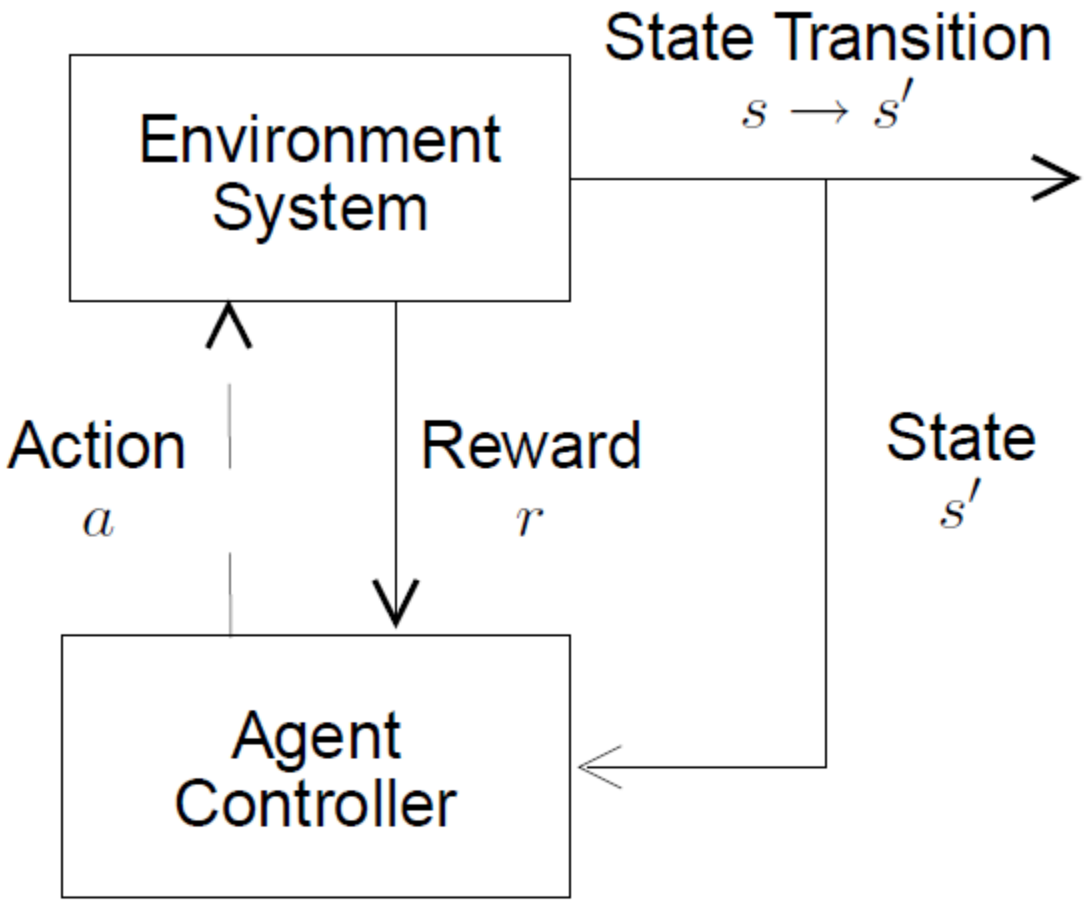
Professional background (3/7)
Orange Labs Tokyo 2005-2006

- Applications for telecommunication systems:
  - Radio Resource Management for mobile networks (via Reinforcement Learning techniques)
  - Automated selection of radio access networks (via Support Vector Machines techniques)

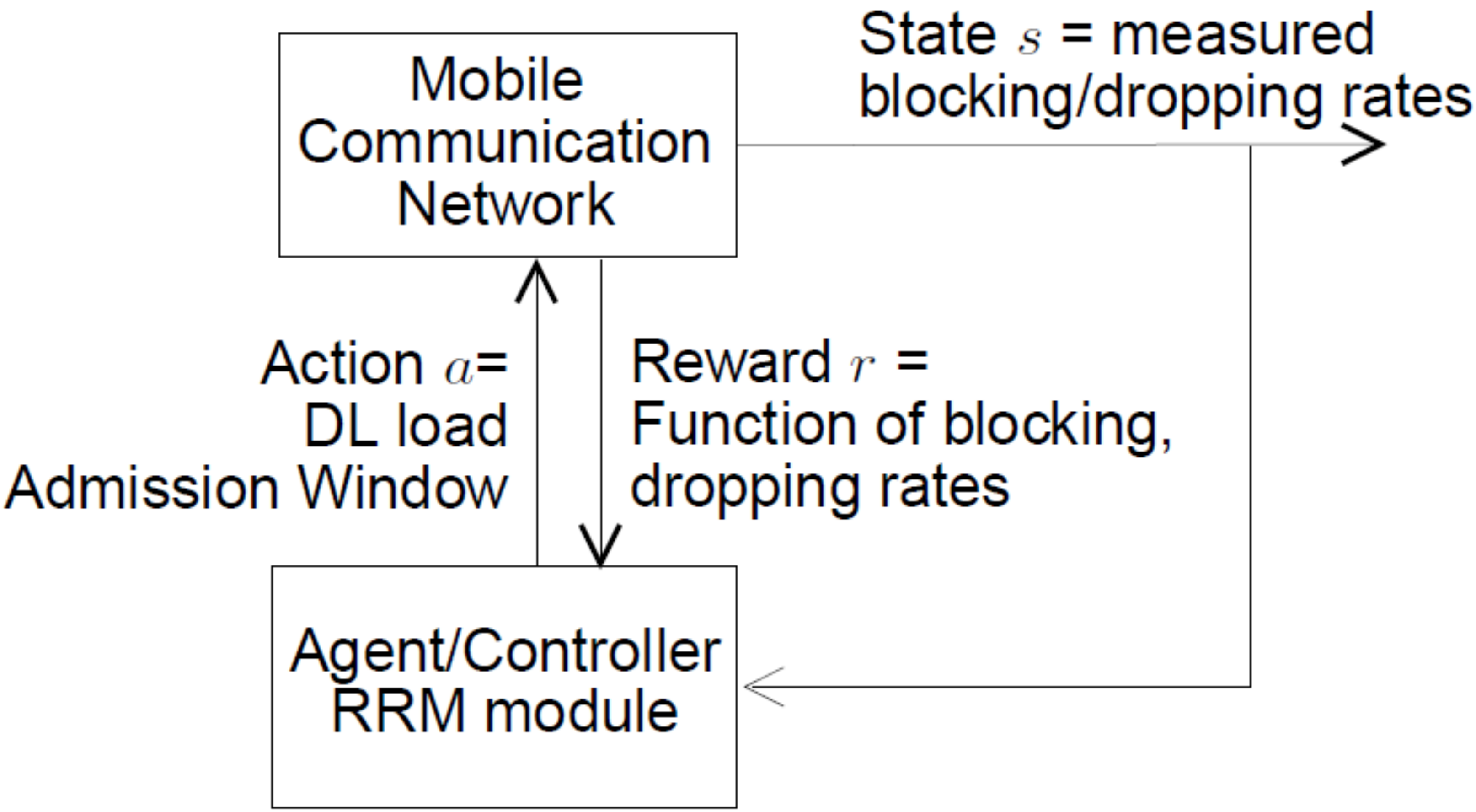
Learning & Control: Reinforcement Learning (1/3)



Learning & Control: Reinforcement Learning (2/3)



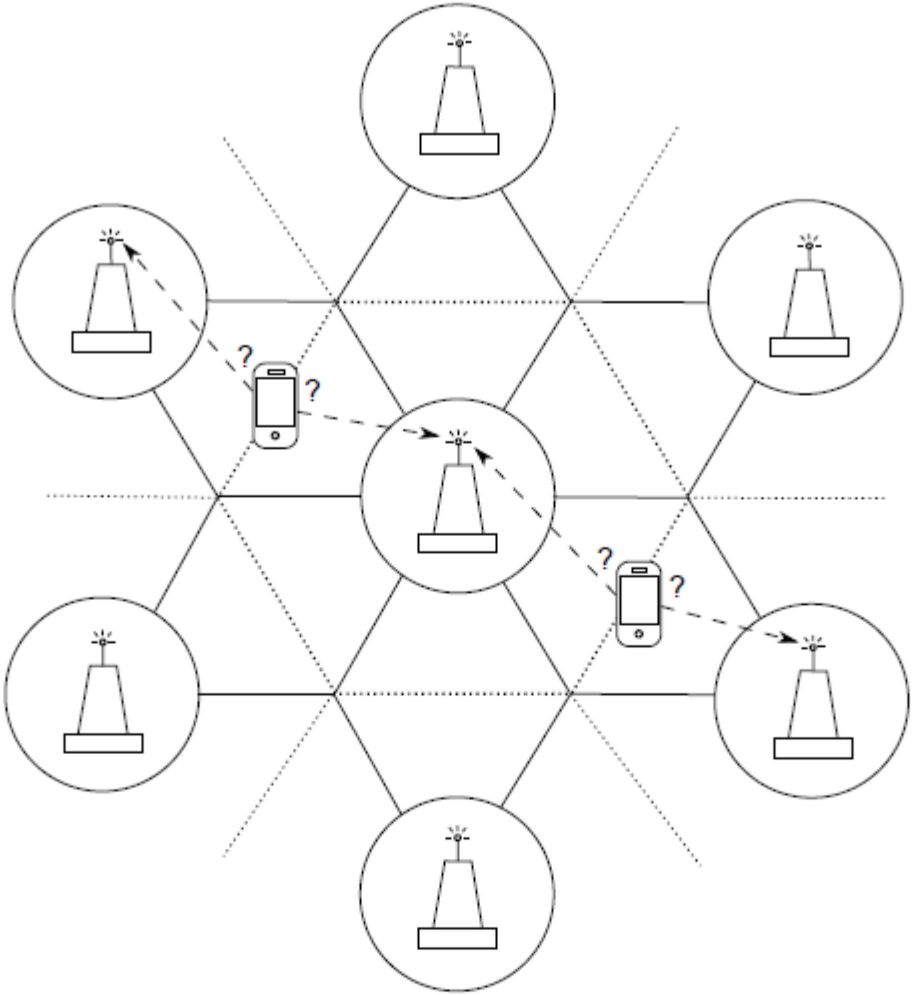
Learning & Control: Radio Resource Management for Mobile Networks (3/3)



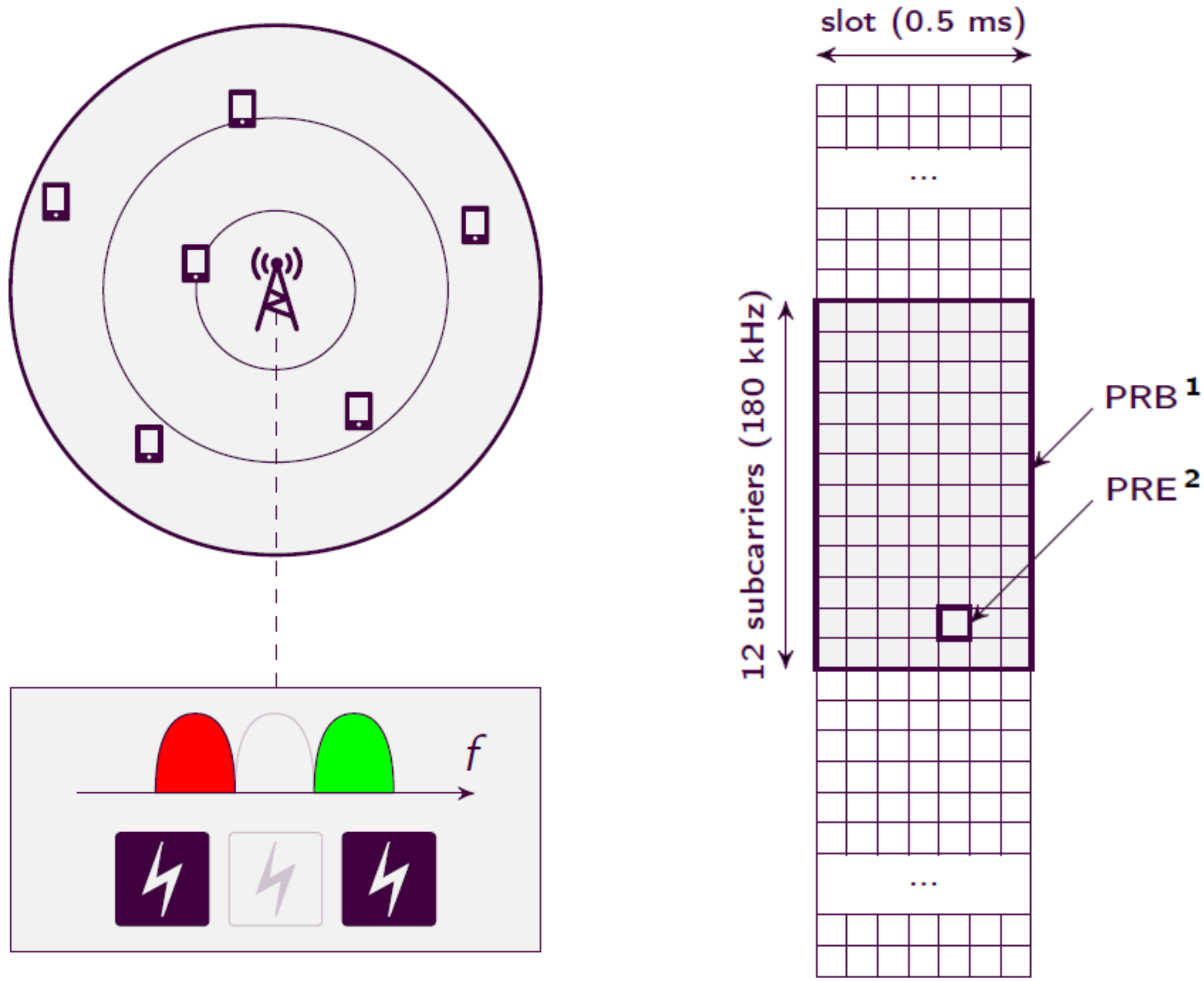
Professional background (4/7)
Orange Labs, since 2007

- **Reinforcement Learning:**
  - Partially observed models (POMDP) and dedicated learning techniques (Belief States → Monte Carlo POMDP, 2007-2009)
  - Policy Gradient (2011-2013):
    - Application to solution implementation for association problem of users to base stations for a mobile communication network
    - Design of variance reduction algorithms for Policy Gradient type estimation techniques in Reinforcement Learning
  - Dynamic Programming (2015):
    - application to joint QoS and energy consumption control for mobile communication networks

Association problem for mobile networks



Joint QoS and energy consumption control



Professional background (5/7)
Orange Labs, since 2007

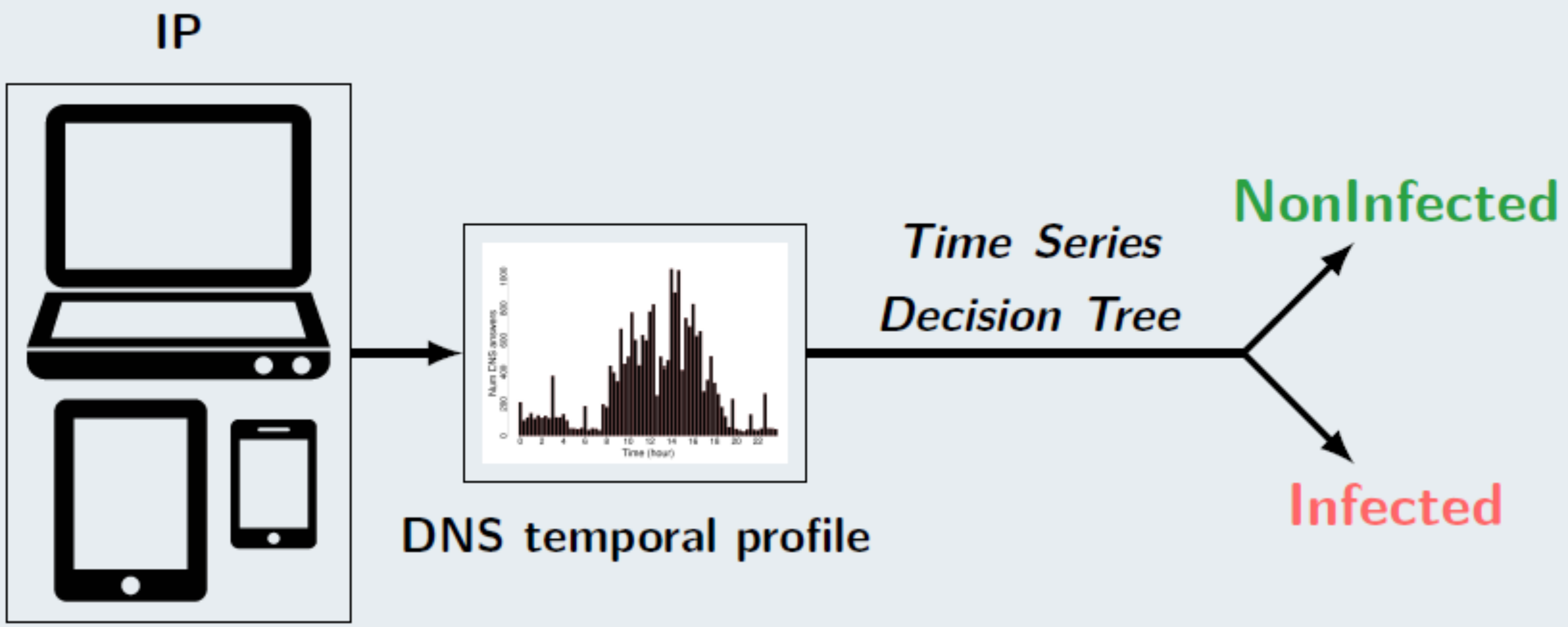
- **Unsupervised learning:**
  - Clustering: K-means algorithm and its extensions
  - Application to computer networks platforms optimization for handling, managing, processing Internet traffic
  - DNS, Internet traffic load balancing, DNSSEC implementation
  - 2010-2012

Professional background (6/7)
Orange Labs, since 2011

- **Supervised Learning:**
  - Methods for implementing traffic traces classification:
    - Kernel Learning (Multiple Kernel Learning, Support Vector Data Description, 2011-2014)
    - Artificial Neural Networks (Extreme Learning Machines, 2013-2014)
    - Decision Trees (with time series as inputs, 2014-2015)
  - Application to Internet traffic analysis for network security
    → Botnets detection

Botnets Detection

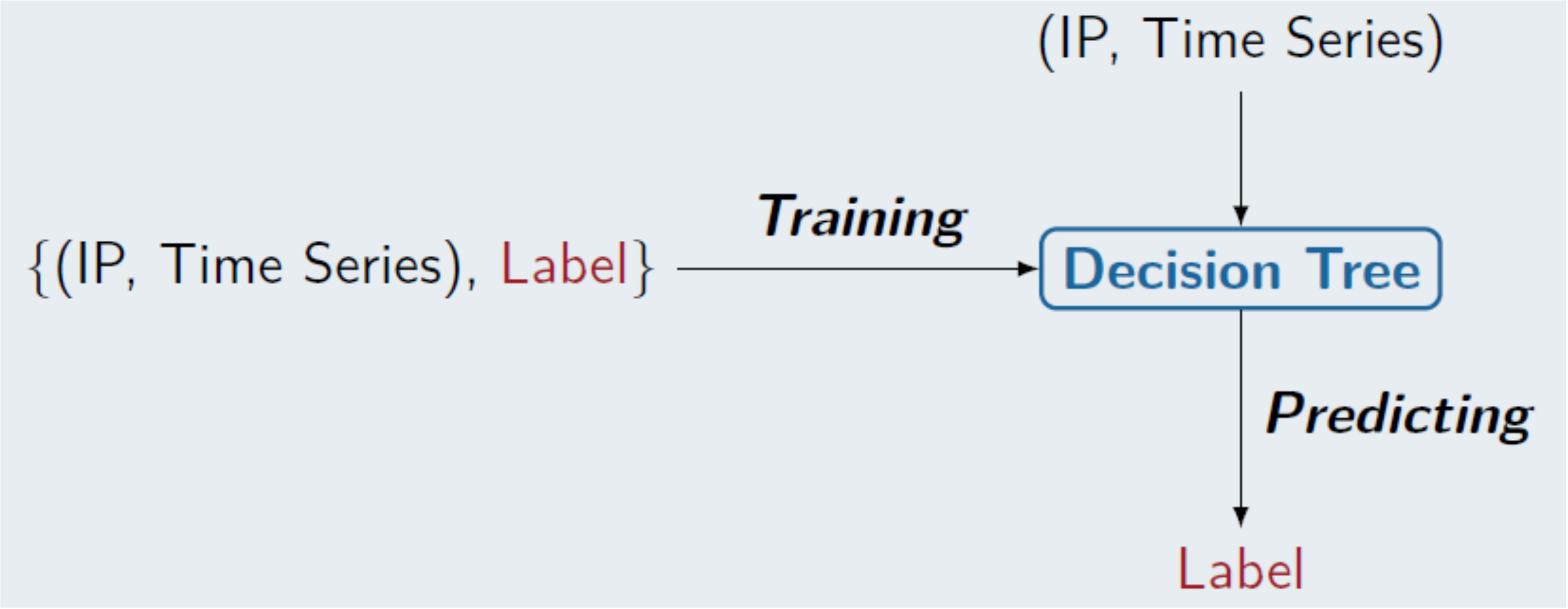
Detecting Infected IPs



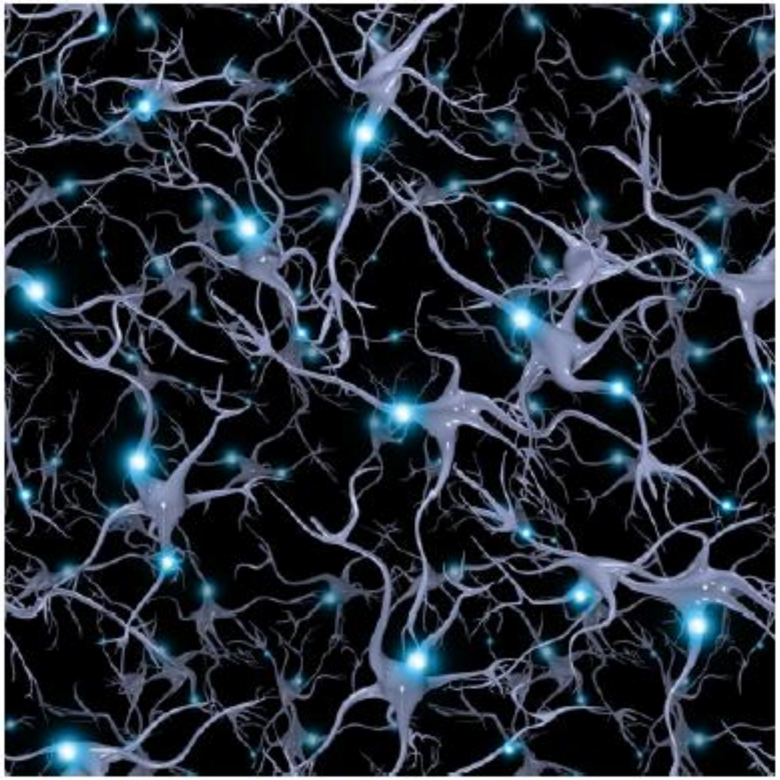
Supervised Learning (1/2): Training/Learning Phase

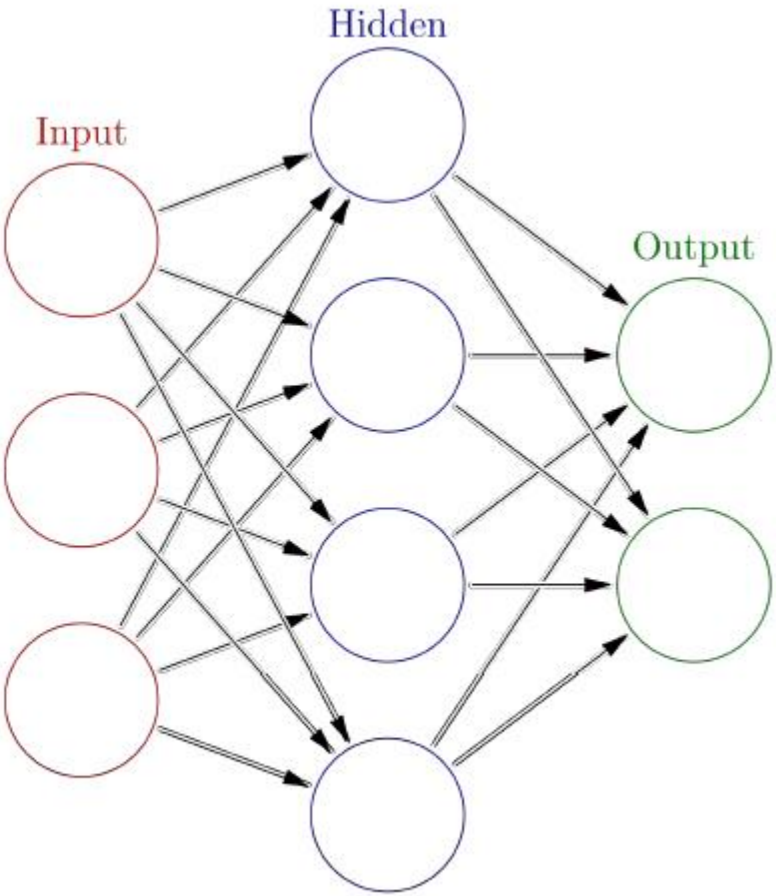
$$\{(\text{IP, Time Series}), \text{Label}\} \xrightarrow{\textit{Training}} \boxed{\text{Decision Tree}}$$

Supervised Learning (2/2): Prediction/Test Phase



Artificial Neural Networks





Decision Trees (1/4): Training Data

Id	Outlook (O)	Temperature (T)	Humidity (H)	Windy (W)	*Play*
a	overcast	$83°F$	86%	false	yes
b	overcast	$64°F$	65%	true	yes
c	overcast	$72°F$	90%	true	yes
d	overcast	$81°F$	75%	false	yes
e	rainy	$70°F$	96%	false	yes
f	rainy	$68°F$	80%	false	yes
g	rainy	$65°F$	70%	true	no
h	rainy	$75°F$	80%	false	yes
i	rainy	$71°F$	91%	true	no
j	sunny	$85°F$	85%	false	no
k	sunny	$80°F$	90%	true	no
l	sunny	$72°F$	95%	false	no
m	sunny	$69°F$	70%	false	yes
n	sunny	$75°F$	70%	true	yes

Decision Trees (2/4): Building the Model

$\mathtt{H} \geq 82.5\%$

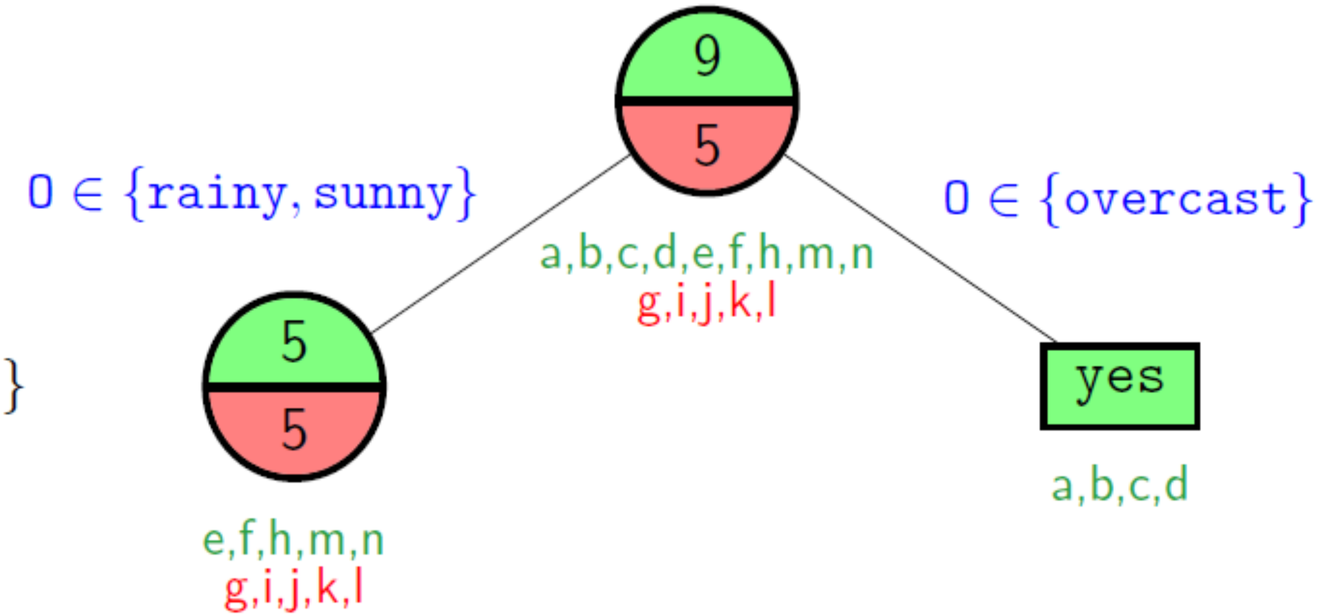
$\mathtt{H} \geq 75.5\%$

$\mathtt{O} \in \{\mathtt{rainy}, \mathtt{sunny}\}$

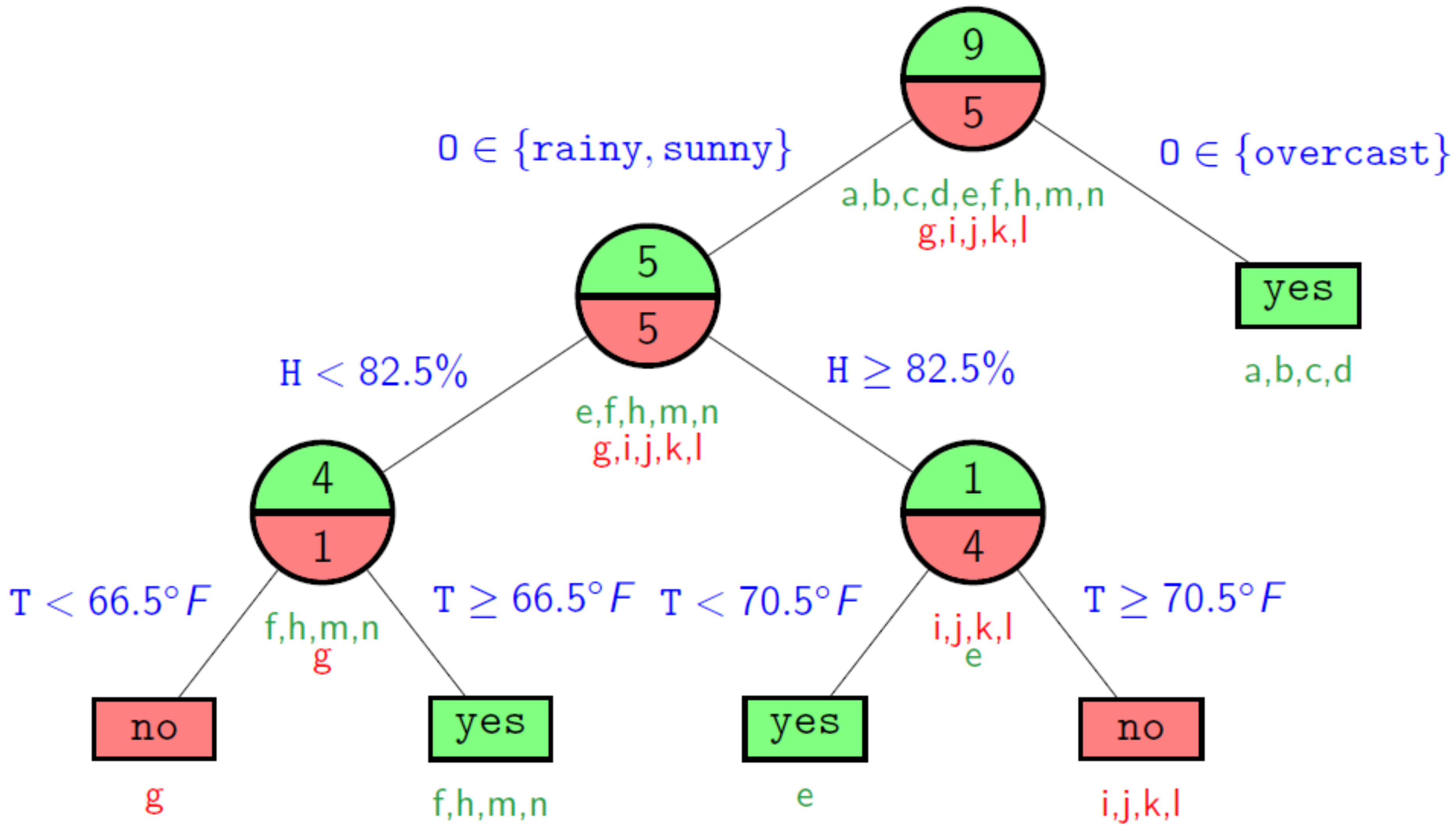
$\ldots$

$\mathtt{O} \in \{\mathtt{rainy}, \mathtt{overcast}\}$

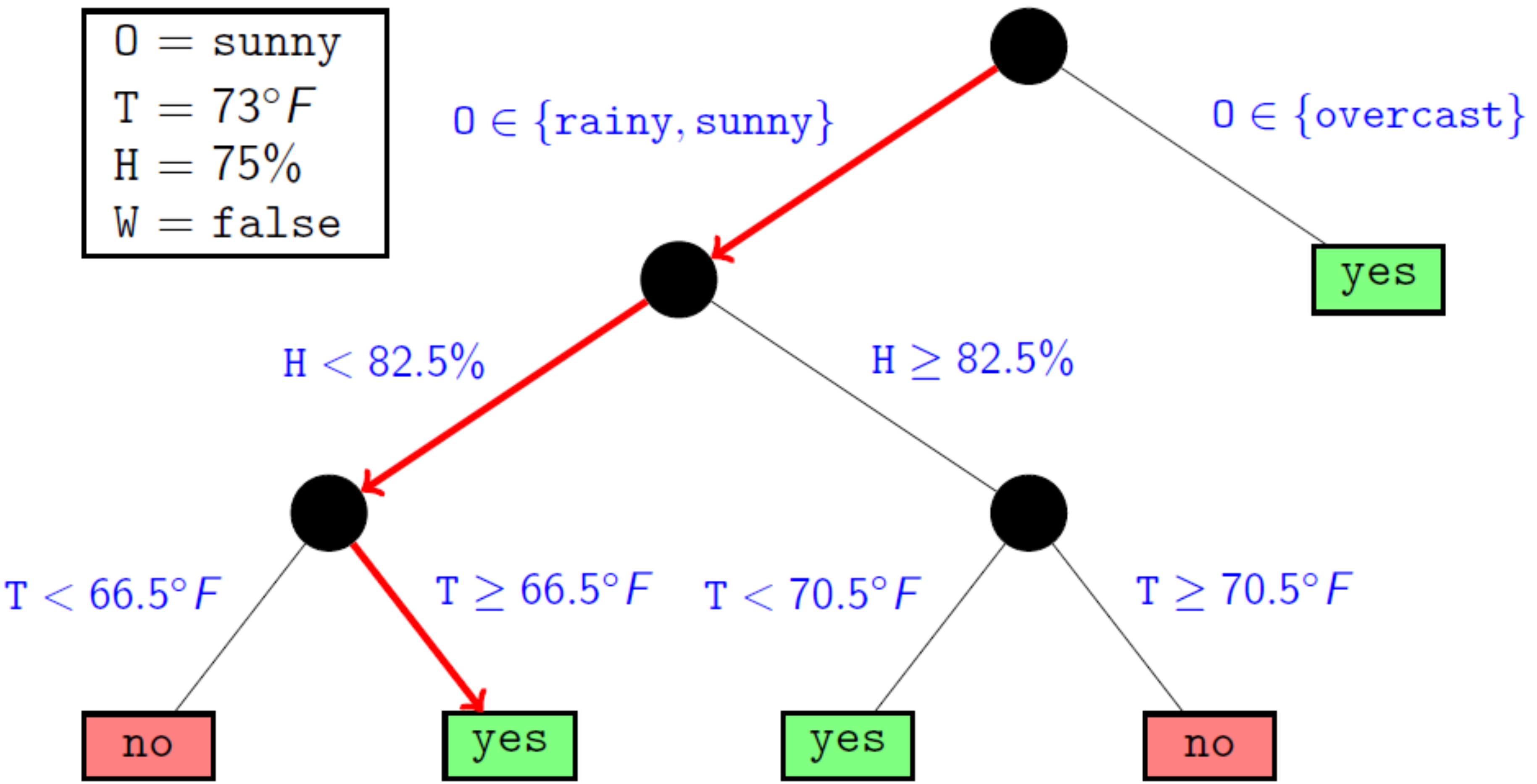
$\mathtt{W} = \mathtt{true}$



Decision Trees (3/4): Building the Model



Decision Trees (4/4): Prediction



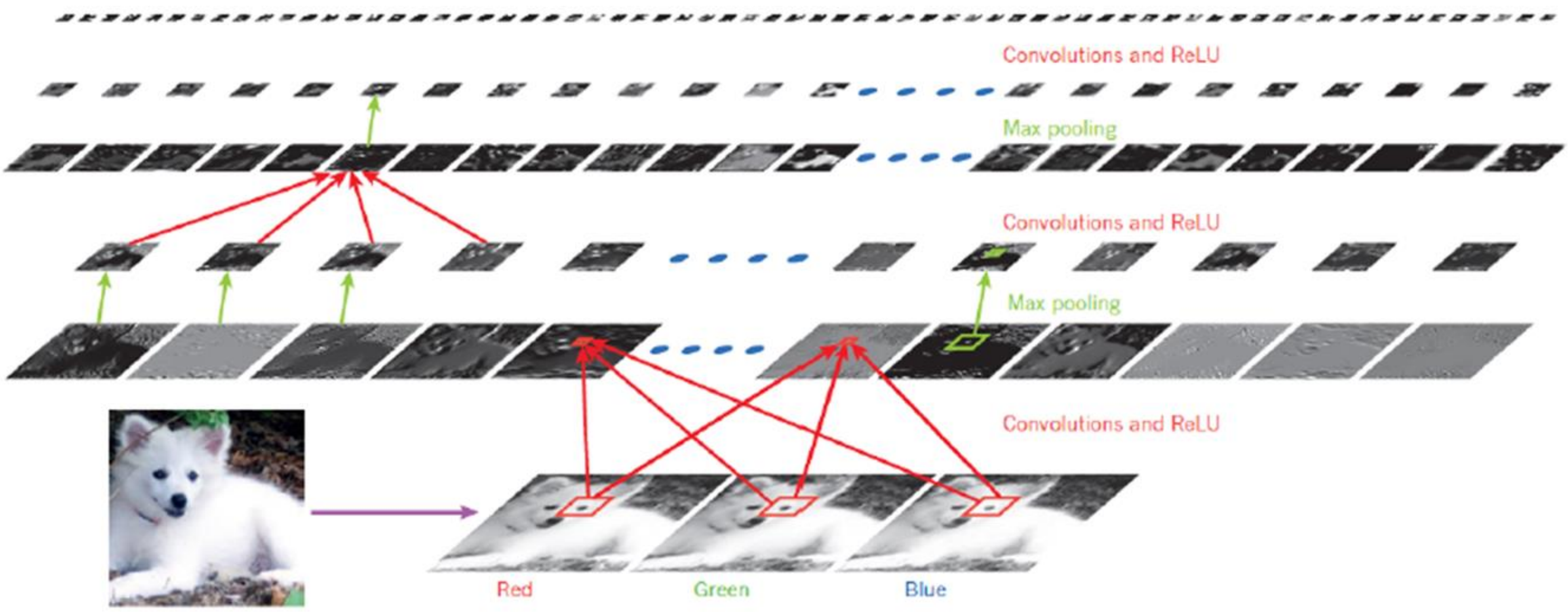
Professional background (7/7)
Orange Labs, currently

- **Deep Reinforcement Learning techniques:**
  - understanding of Google DeepMind AlphaGo system, 2016
  - application to resource allocation for mobile networks, 2017

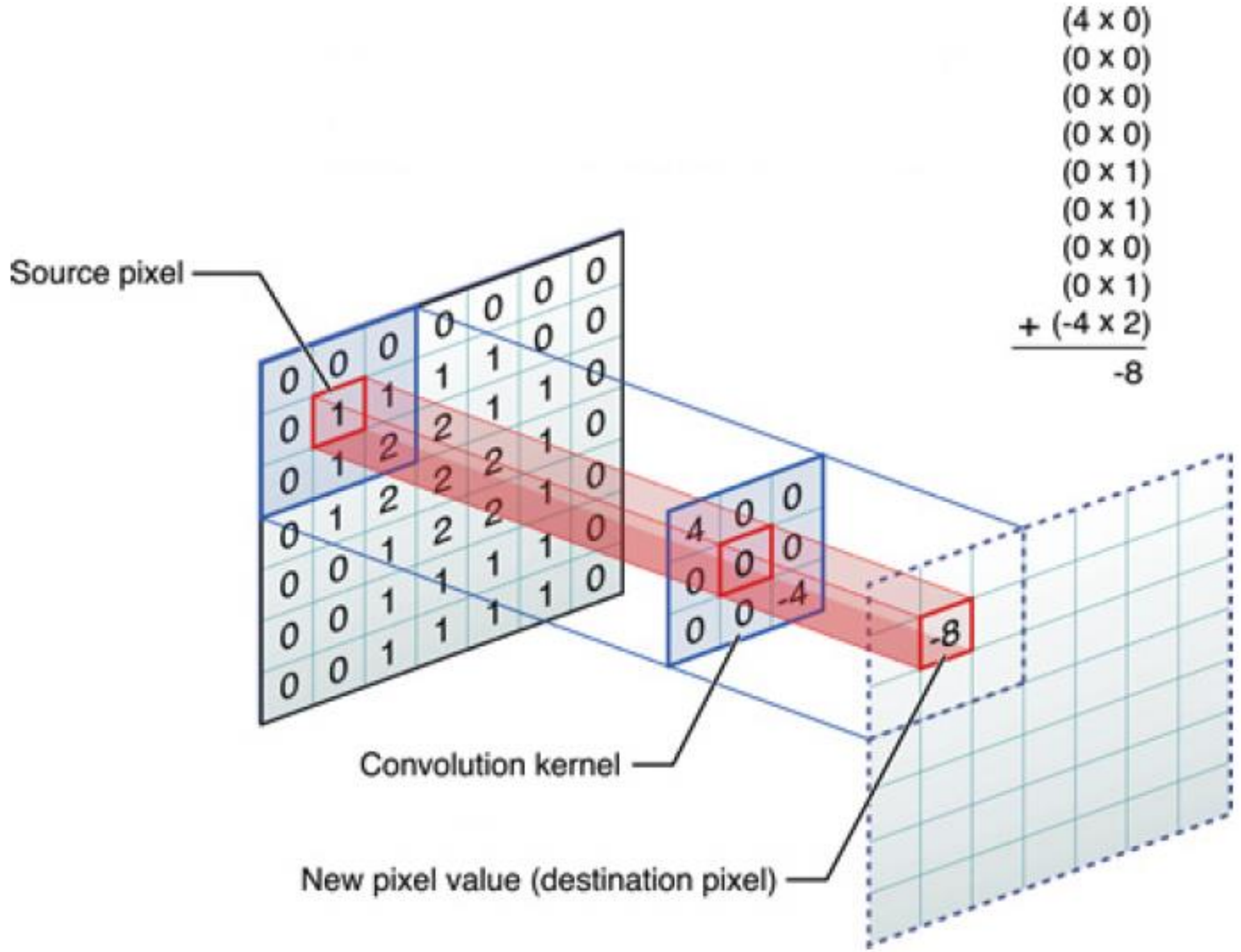
- **Markov Chain Monte Carlo (MCMC) simulation methods:**
  - Event-Chain based Monte Carlo techniques
  - Nonreversible Markov chains
  - 2016-2017

- Networks metrics **data prediction benchmark** for autonomic network management:
  - many Machine Learning models and techniques
  - essentially supervised learning
  - 2016-2017

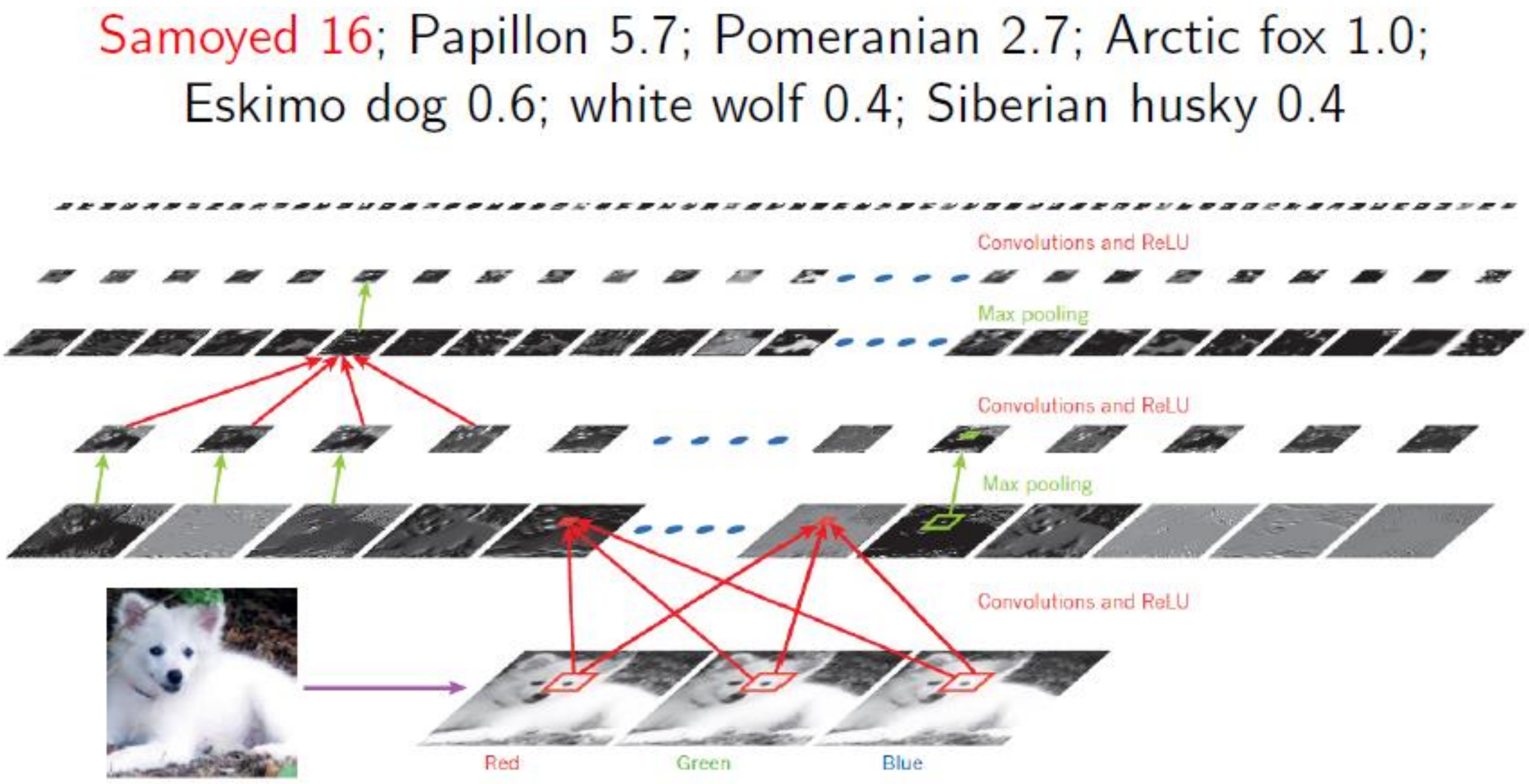
Convolutional Neural Networks: Modeling and Training/Learning (1/3)



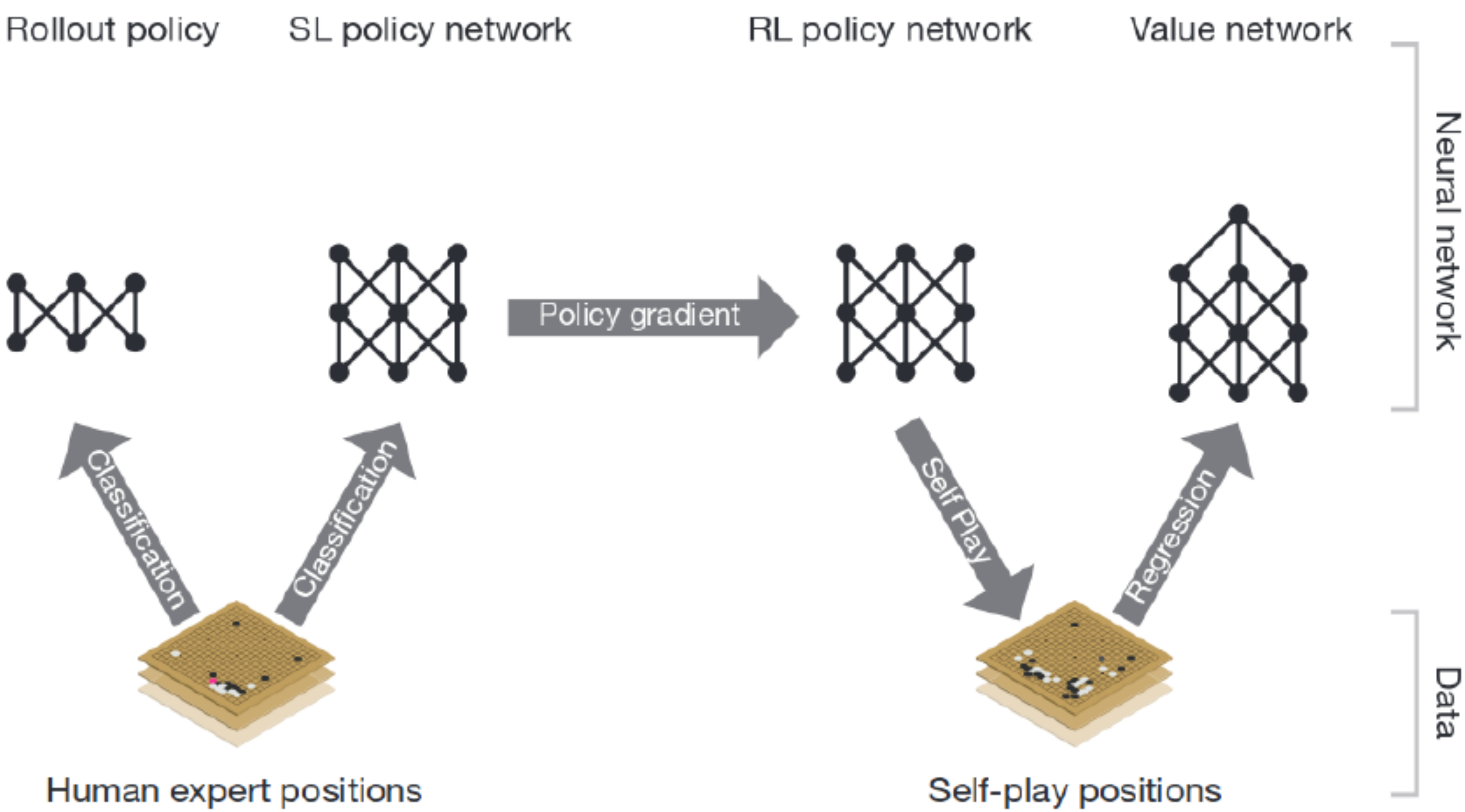
Convolutional Neural Networks: Convolutional Kernel (2/3)



Convolutional Neural Networks: Testing/Prediction (3/3)



AlphaGo Training/Learning Global Pipeline



Overview

- Academic background
- Professional background and activities
- → Projects
- Working @ Orange Labs
- Machine Learning…

Research activities → Projects

- Daily work organization in **Projects**, e.g. hosted by the "Applied Maths and Computer Science" Research Group @ Orange Labs:
  - Internal/Orange projects
  - Bilateral projects with Orange ("external research contracts")
  - Collaborative projects:
    - ANR "ECOSCELLS" 2009-2012
    - EU FP7 STREP "HARP" 2012-2015
    - ANR INFRA "NETLEARN" 2013-2017
    - EU H2020 5G-PPP "COGNET" 2015-2017

Research Projects: practical aspects (1/4)

- Working on a research theme in a "fixed-term" mode:
  - Work schedule: Project Management Plan (PMP), including Gantt charts, elaborated and submitted for validation before the launch of the project ("Kick-Off")
  - Costs management:
    - In human resources: People*Day, People*Month or People*Year with monthly follow-up/reporting of consumed resources
    - Financial: elaboration of an initial budget, then management of missions and material costs, with on-the-fly reporting

Research Projects: practical aspects (2/4)

- Organization of the works in “Work Packages” (WP) with specific Tasks with tasks and WPs leaders + 1coordinator (Project Head) and 1 technical coordinator/leader

- Working meetings (Face-to-Face, conf calls) and milestones meetings (plenary = for all partners, Face-to-Face)

- Scientific and technical skills indeed, but also good communication, relational skills, patience, resilience and a good sense of humor! ☺

Research Projects: practical aspects (3/4)

- Valorization of project works → "Deliverables":
  - Internal valorization:
    - Elaboration of technical reports and presentations
    - Reporting:
      - towards Orange: on the fly + semestrial official meetings with points with the hierarchy and with project entities (Project Head, Research Group)
      - towards the sponsor (ANR, EU) officially on a trimestrial, semestrial or annual basis
  - External valorization:
    - Publications in scientific and technical conferences (oral and poster presentations) and in scientific and technical journals/magazines: IEEE, ACM…
    - Organization of Workshops/Seminars in conferences or by ourselves

Research Projects: practical aspects (4/4)

- Valorization of project works → “Deliverables”:
  - Patent filling
  - Normalization/standardization activities: 3GPP, IETF, ETSI, ITU…
  - Development of technical solutions and industrialization:
    - internally: simulators et prototypes → development transfert → inclusion in the Information System and/or transfert towards technical and operational directions, even towards Business Units (BU)
    - externally: Open Source…

Work valorization example: publications

- Conferences:
  - 27 papers published (3 invited papers)
  - 1 submitted
  - 2 in preparation
- Journals:
  - 6 papers published
  - 2 in preparation
- Books:
  - 1 chapter in “Data Mining Applications with R”, Elsevier, 2013

Working as a team @ Orange Labs (1/2)

- “Department” (Team) “Modeling and Statistical Analysis”
- Activities on networks and traffic modeling for fixed (ADSL and Fiber Internet) et mobile (2G/3G/4G →5G) communications
- Currently 22 people, including:
  - 1 intern
  - 1 apprentice
  - 5 CIFRE PhD students
  - 1 post-doc

Working as a team @ Orange Labs (2/2)

- Goals:
  - Come up with a better understanding of communication trafic, related to terminals and usage evolution, regarding the very important increase of this trafic
  - Estimate and optimize networks infrastructure costs by geographical zones regarding the strategic choices for deploying new generation technologies (optical fiber, 4G mobile networks) with new architectures
  - Improve the Quality of Service (QoS) and performance for mobile networks
  - Provide analytical models for equipments energy consumption in order to estimate and predict the networks energy consumption

Infos about working @ Orange & contact details

- Opportunities for internships for MS/MEng students :
  - 4 to 6 months duration
  - Schedule: application from November for the forthcoming year
- Opportunities for apprenticeship, PhD programs (CIFRE) and post-docs
- Opportunities for permanent positions: Research Engineer/Scientist
- https://orange.jobs/site/en-home/
- Contact: Stephane SENECAL
  - email: stephane.senecal@orange.com
  - LinkedIn: http://fr.linkedin.com/in/stephanesenecal

More information on Machine Learning (1/2)

- Informal meetings and discussion groups (meet ups) in Paris/IDF:
  - Paris Machine Learning Applications Group:
  - http://www.meetup.com/Paris-Machine-learning-applications-group/
  - 1+ meeting(s) per month
  - + Groups on LinkedIn, Facebook, Google+, Twitter account, Nuit Blanche blog (including the meet ups archive)…

  - Deep Learning Paris:
  - http://www.meetup.com/Deep-Learning-Paris-Meetup/
  - + Workshops…

More information on Machine Learning (2/2)

- Academic seminar on Machine Learning in Paris:
  - Statistical Machine Learning “SMILE in Paris”
  - https://sites.google.com/site/smileinparis/
  - Organized by ENS and Mines-ParisTech

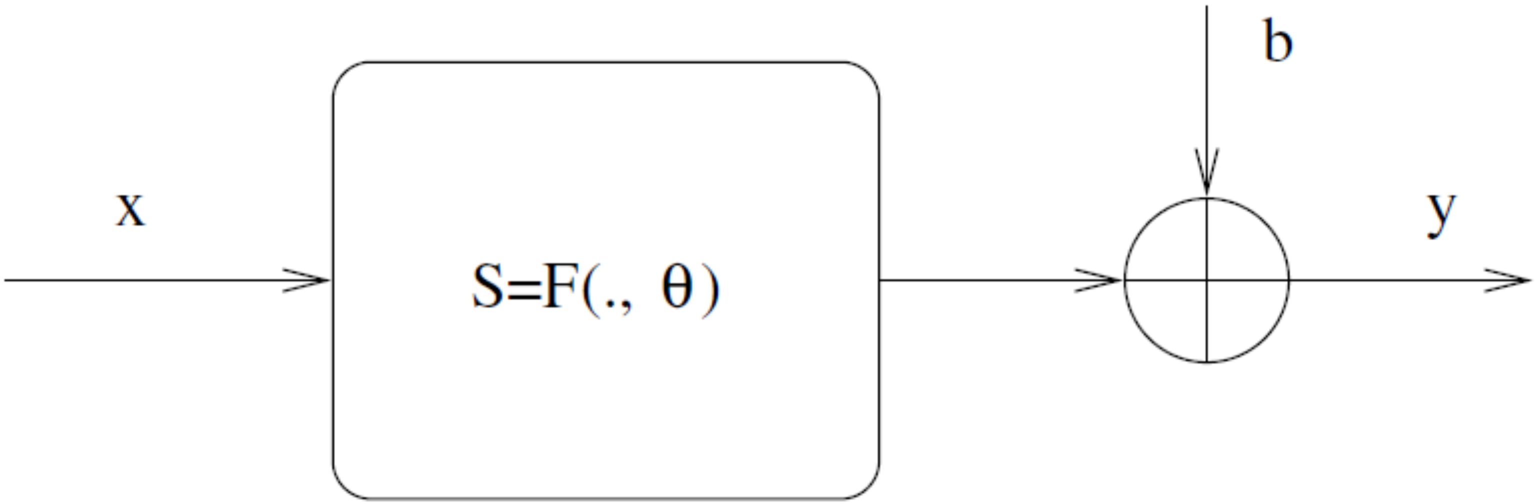
- Internet Groups/Forums (worldwide audience → in English):
  - Google Group: Machine Learning News
  - CFP, job offers, …
  - https://groups.google.com/forum/#!forum/ML-news

Thank you!☺

Questions?

Appendix

Bayesian Estimation



$$\text{Information on } (x, \theta)\text{: distribution of probability}$$

$$p(x, \theta | y, F, prior\ ) \propto p(y | x, \theta, F, prior\ ) \times p(x, \theta | prior\ )$$

$$\Rightarrow \ \text{Estimates } (\widehat{x}, \widehat{\theta})$$

Bayesian Estimates

- *Maximum a posteriori* (MAP)

$$(\widehat{x}, \widehat{\theta}) = \arg\max_{x,\theta} p(x, \theta | y, prior\ )$$

- Expectation: *posterior* mean $\mathrm{E}\{x, \theta | y, prior\}$

$$\mathrm{E}_{p(.|y,prior\ )}\{f(x,\theta)\} = \int f(x,\theta) p(x,\theta|y, prior\ ) d(x,\theta)$$

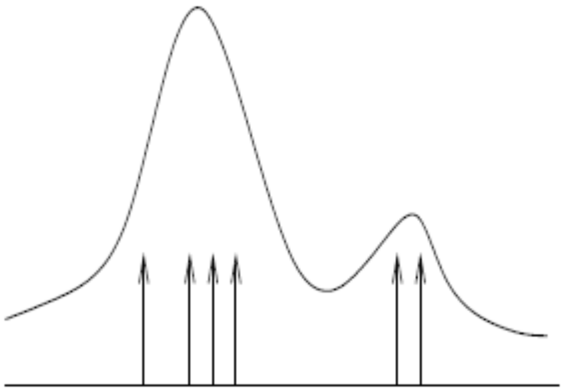
Computation : asymptotic, numerical, stochastic methods

$\Rightarrow$ Monte Carlo simulation methods

Monte Carlo Estimates

$$x_1, \ldots, x_N \sim \pi$$

$$\Rightarrow \widehat{\pi}_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_n}$$



$$\widehat{S}_N(f) = \frac{1}{N} \sum_{n=1}^{N} f(x_n) \longrightarrow \int f(x)\pi(x)dx = \mathrm{E}_\pi \{f\}$$

$$\widehat{x}_{max} = \arg\max_{x_n} \widehat{\pi}_N \text{ approximates } x_{max} = \arg\max_{x} \pi(x)$$

$$\Rightarrow \text{generate samples } x_\ell \sim \pi \text{ ?}$$

$\rightarrow$ Markov chain and sequential Monte Carlo

Simulation Techniques

- Classical distributions : cumulated density function
  $\rightarrow$ transformation of uniform random variable
- Non-standard distributions, $\mathbb{R}^n$, known up to a normalizing constant $\rightarrow$ usage of instrumental distribution:

Accept-reject, importance sampling $\rightarrow$ sequential/recursive
$\Rightarrow$ SMC aka particle filtering, condensation algorithm

$\Rightarrow$ MCMC : distribution = fixed point of an operator

$$\pi = K\pi$$

$\rightarrow$ simulation schemes with Markov chain: Hastings-Metropolis, Gibbs sampling

Markov Chain

Definition:

$$X_n | X_{n-1}, X_{n-2}, \ldots, X_0 \stackrel{d}{=} X_n | X_{n-1}$$

$$\text{homogeneity} : X_n | X_{n-1} \text{ independent of } n$$

Realization:

$$X_0 \sim \pi_0(x_0)$$

$$p.d.f. \text{ of } X_n | X_{n-1} = \text{transition kernel } K(x_n | x_{n-1})$$

Simulation of a Markov Chain

Convergence: $X_n \sim \pi$ asymptotically ?

$$\pi\text{-invariance} : \pi(.) = K\pi(.)$$

$$\int_A \pi(x)dx = \int_{y \in A} \int K(y|x)\pi(x)dxdy$$

$$\Leftarrow \pi\text{-reversibility} : Pr(A \to B) = Pr(B \to A)$$

$$\int_{y \in B} \int_{x \in A} K(y|x)\pi(x)dxdy = \int_{y \in A} \int_{x \in B} K(y|x)\pi(x)dxdy$$

Construct kernels $K(.|.)$ such that the chain is π-invariant

- Hastings-Metropolis algorithm
- Gibbs sampling

Hastings-Metropolis algorithm (1/2): scheme

Draw $\boldsymbol{x}$ from $\pi(.)$

1. initialize $\boldsymbol{x}_0 \sim \pi_0(\boldsymbol{x})$
2. Iteration ℓ
   - propose candidate $\boldsymbol{x}^\star$ for $\boldsymbol{x}_{\ell+1}$ $\quad \rightarrow \boldsymbol{x}^\star \sim q(\boldsymbol{x}|\boldsymbol{x}_\ell)$
   - accept it with prob $\alpha = \min\{1, r\}$
3. $\ell \leftarrow \ell + 1$ and go to (2)

$$r = \frac{\pi(\boldsymbol{x}^\star)q(\boldsymbol{x}_\ell|\boldsymbol{x}^\star)}{q(\boldsymbol{x}^\star|\boldsymbol{x}_\ell)\pi(\boldsymbol{x}_\ell)} \rightarrow \pi(x)K(y|x) = \pi(y)K(x|y)$$

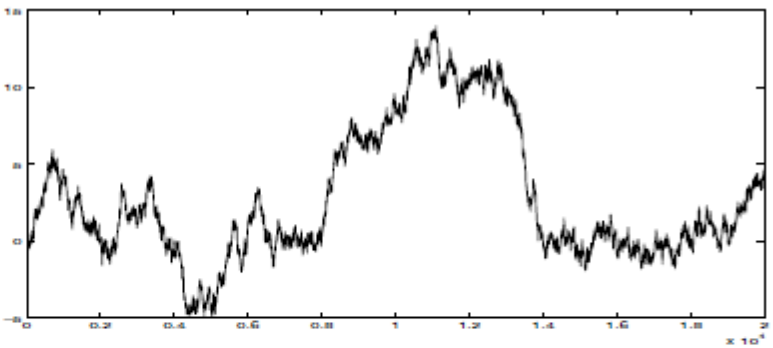
$$\pi(x)q(y|x)\min\left\{1, \frac{\pi(y)q(x|y)}{q(y|x)\pi(x)}\right\} = \min\left\{\pi(x)q(y|x), \pi(y)q(x|y)\right\}$$

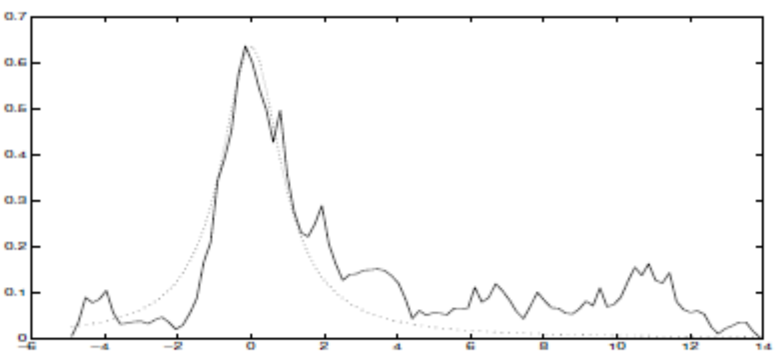
$$q(\boldsymbol{x}^\star|\boldsymbol{x}_\ell) = q(\boldsymbol{x}^\star) \quad q(\boldsymbol{x}^\star|\boldsymbol{x}_\ell) = q(|\boldsymbol{x}^\star - \boldsymbol{x}_\ell|)$$

Hastings-Metropolis algorithm (2/2): example

sample $x \sim p(x) \propto \frac{1}{1+x^2}$ 20,000 iterations

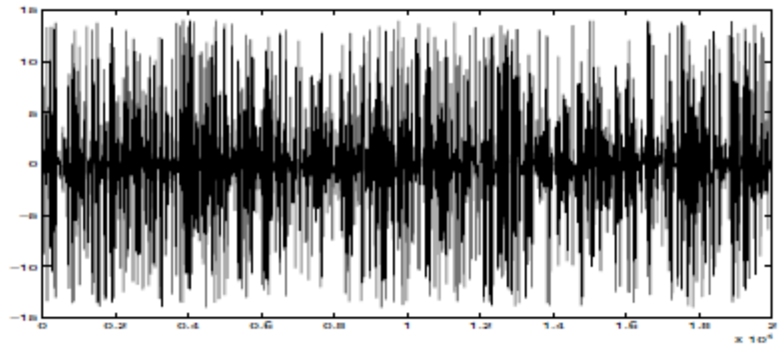
$x^\star \sim \mathcal{N}(x_\ell, 0.1^2)$

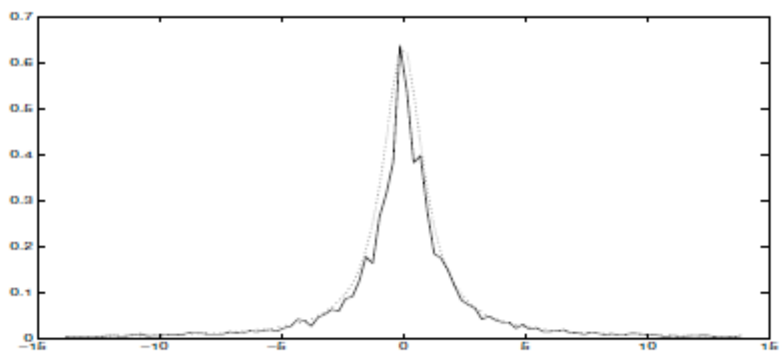




acc. rate = 97%

$x^\star \sim \mathcal{U}_{[a,b]}$





acc. rate = 26%

Gibbs Sampling algorithm (1/2): scheme

Sample $\boldsymbol{x} = (x_1, ...x_p) \sim \pi(x_1, ...x_p)$

1. initialize $\boldsymbol{x}^{(0)} \sim \pi_0(\boldsymbol{x})$, $\ell = 0$

2. iteration ℓ : Sample

$$\begin{aligned}
x_1^{(\ell+1)} &\sim \pi_1(x_1 | x_2^{(\ell)}, \ldots, x_p^{(\ell)}) \\
x_2^{(\ell+1)} &\sim \pi_2(x_2 | x_1^{(\ell+1)}, x_3^{(\ell)}, \ldots, x_p^{(\ell)}) \\
&\vdots \\
x_p^{(\ell+1)} &\sim \pi_p(x_p | x_1^{(\ell+1)}, \ldots, x_{p-1}^{(\ell+1)})
\end{aligned}$$

3. $\ell \leftarrow \ell + 1$ and go to (2)

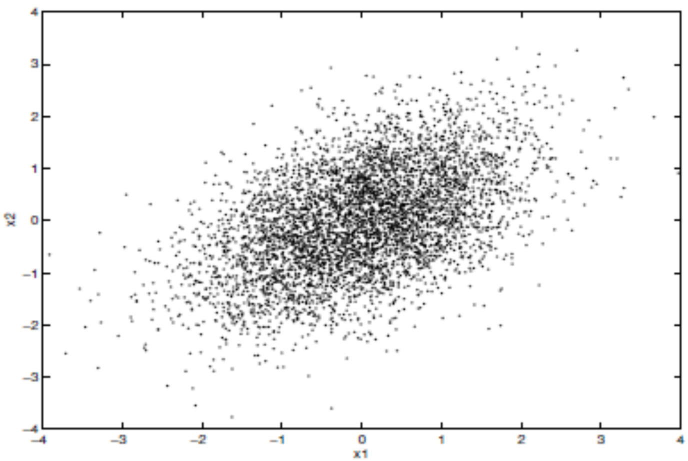
$\longrightarrow$ no rejection, reversible kernel

Gibbs Sampling algorithm (2/2): example

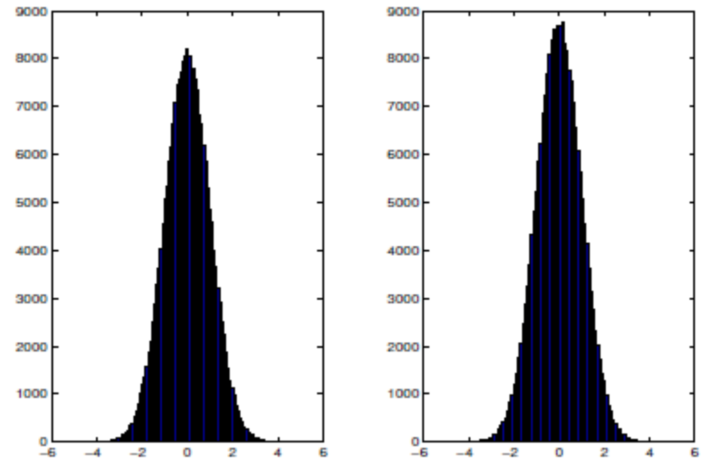
$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$x_1^{(\ell+1)} | x_2^{(\ell)} \sim \mathcal{N}\left(\rho x_2^{(\ell)}, 1 - \rho^2\right)$$

$$x_2^{(\ell+1)} | x_1^{(\ell+1)} \sim \mathcal{N}\left(\rho x_1^{(\ell+1)}, 1 - \rho^2\right)$$



5,000 samples, ρ=0.5



histograms (x_1^ℓ, x_2^ℓ)

Improving convergence of simulation techniques

How to obtain fast converging simulation scheme ?

$\rightarrow$ **Missing Data, Data Augmentation, Latent Variables**

Idea : extend sampling space $x \rightarrow (x, z)$ and distribution $\pi(x) \rightarrow \widetilde{\pi}(x, z)$ with constraint

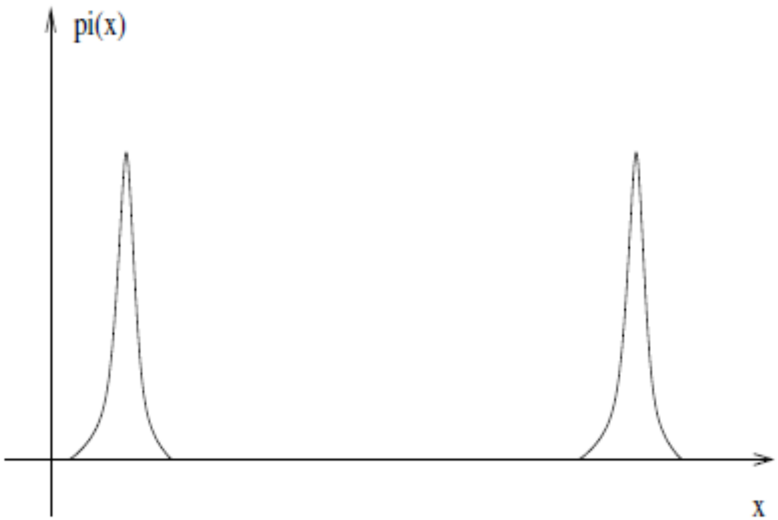
$$\int \widetilde{\pi}(x, z) dz = \pi(x)$$

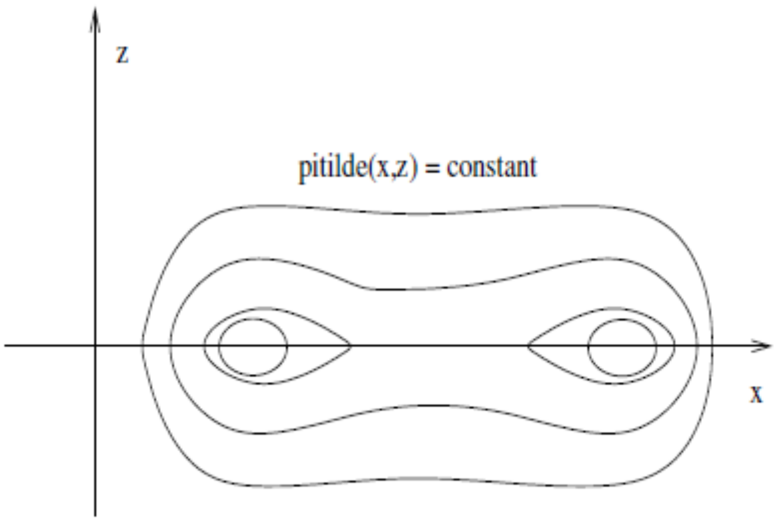
such that Markov chain $(x^{(i)}, z^{(i)}) \sim \widetilde{\pi}$ faster

- Optimization : Expectation-Maximization (EM) algorithm
- Simulation : Data Augmentation, Gibbs sampling

Efficient Data Augmentation Schemes

Idea: construct missing data space as less informative as possible





$x \sim \pi(x)$ $\qquad$ $(x, z) \sim \widetilde{\pi}(x, z)$

Information introduced in missing data $\searrow$: convergence $\nearrow$

Estimation of State Space Models

$$x_t = f_t(x_{t-1}, u_t) \qquad y_t = g_t(x_t, v_t)$$

$$p(x_{0:t}|y_{1:t}) \quad \longrightarrow \quad p(x_t|y_{1:t}) = \int p(x_{0:t}|y_{1:t})dx_{0:t-1}$$

distribution of $x_{0:t} \Rightarrow$ computation of estimate $\widehat{x}_{0:t}$:

$$\widehat{x}_{0:t} = \int x_{0:t}p(x_{0:t}|y_{1:t})dx_{0:t} \rightarrow \mathrm{E}_{p(.|y_{1:t})}\{f(x_{0:t})\}$$

$$\widehat{x}_{0:t} = \arg\max_{x_{0:t}} p(x_{0:t}|y_{1:t})$$

Computation of the estimates

$p(x_{0:t}|y_{1:t}) \Rightarrow$ multidimensionnal, non-standard distributions:

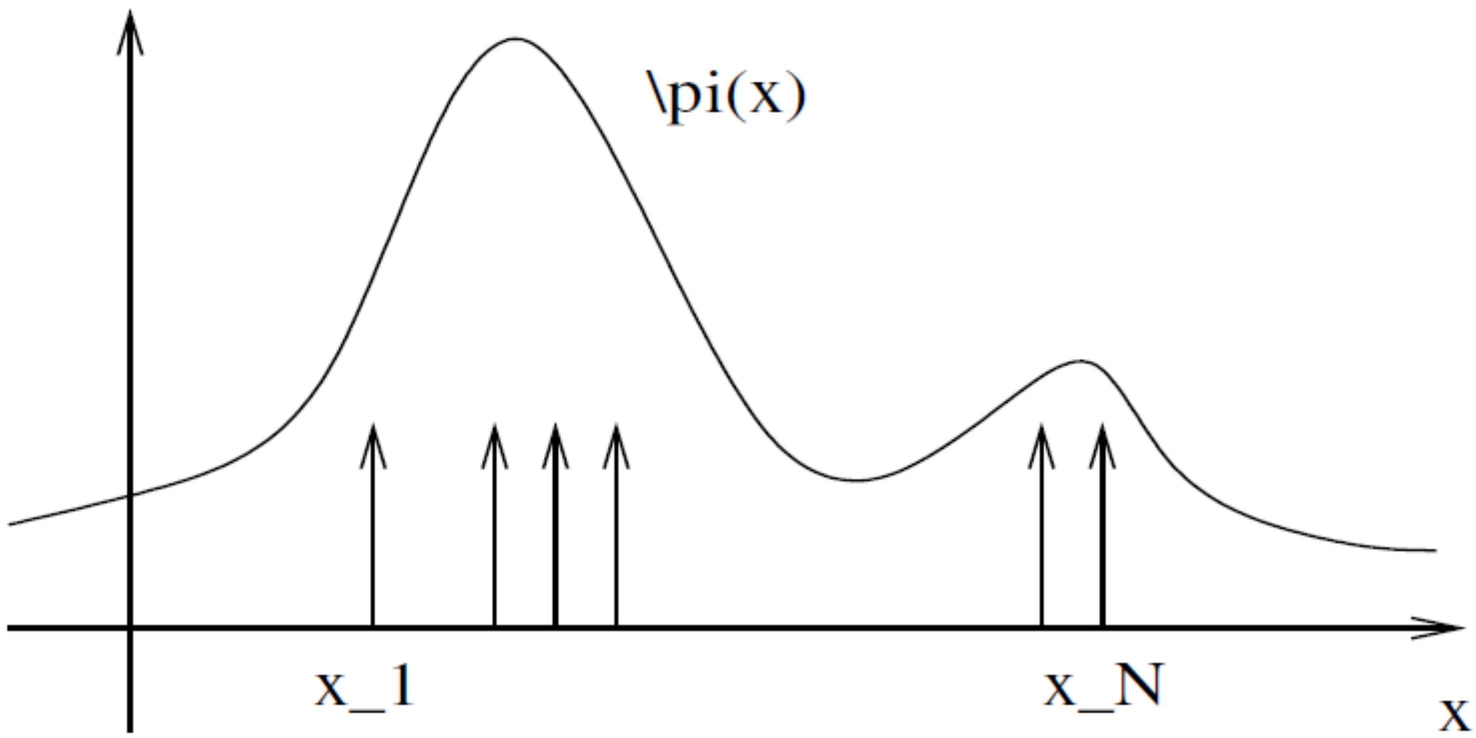
$\rightarrow$ analytical, numerical approximations

$\rightarrow$ integration, optimisation methods

$\Rightarrow$ Monte Carlo techniques

Monte Carlo Approach

compute estimates for distribution $\pi(.) \rightarrow$ samples $x_1, \ldots, x_N \sim \pi$



$\Rightarrow$ distribution $\widehat{\pi}_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ approximates $\pi(.)$

Monte Carlo Estimates

$$\widehat{S}_N(f) = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \longrightarrow \int f(x)\pi(x)dx = \mathrm{E}_\pi\{f(x)\}$$

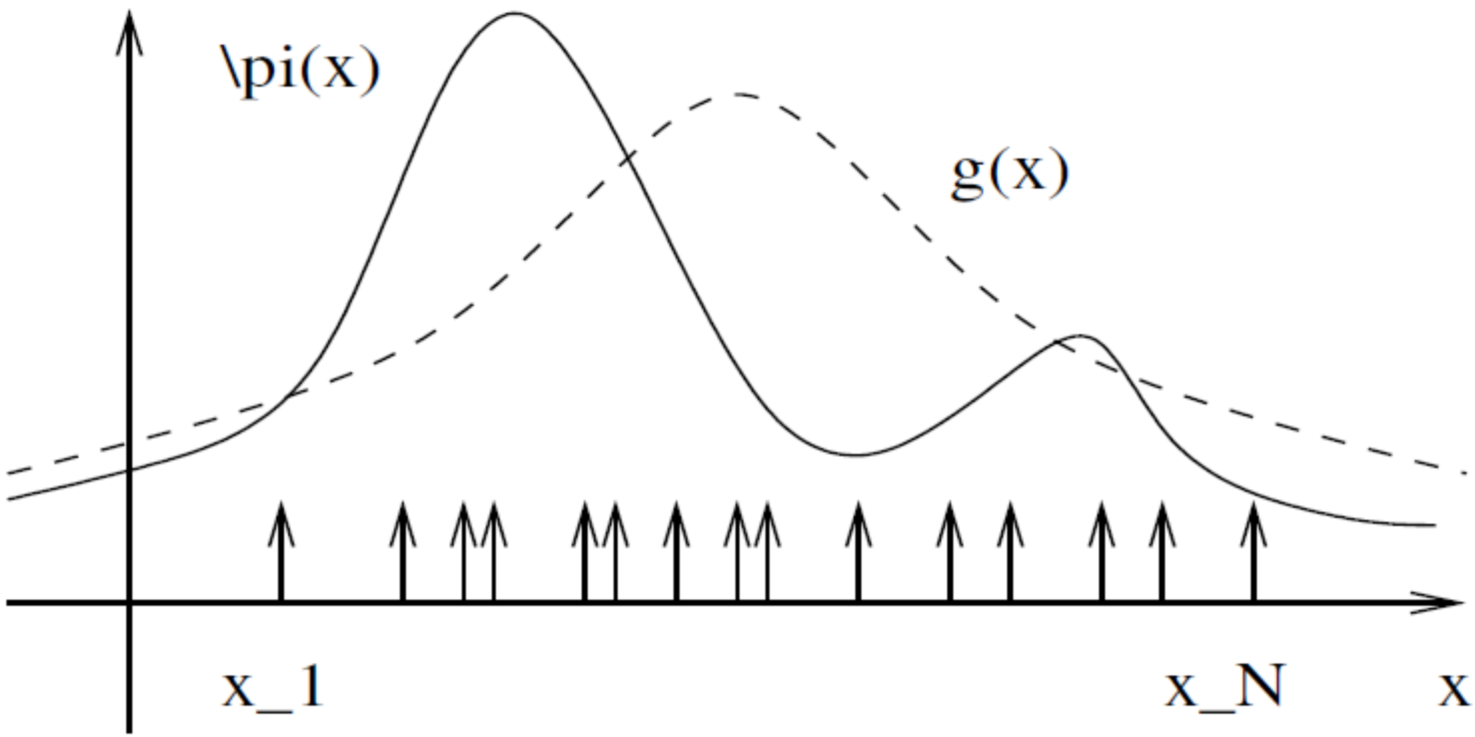
$\arg\max_{(x_i)_{1\leq i\leq N}} \widehat{\pi}_N(x_i)$ approximates $\arg\max_x \pi(x)$

$\Rightarrow$ sampling $x_i \sim \pi$ difficult

$\rightarrow$ importance sampling techniques

Importance Sampling (1/2)

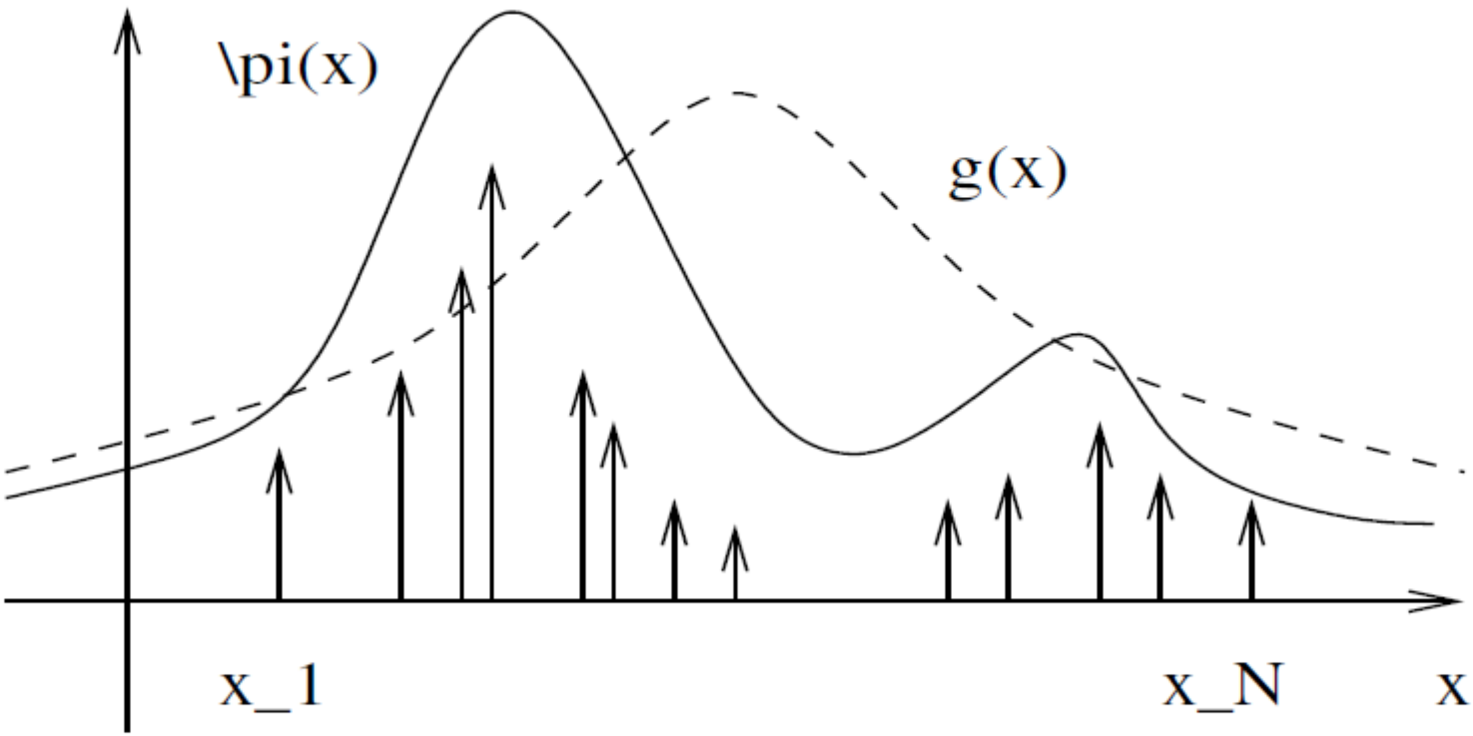
$x_i \sim \pi \rightarrow$ candidate/proposal distribution $x_i \sim g$



Importance Sampling (2/2)

$$x_i \sim g \neq \pi \rightarrow (x_i, w_i) \text{ weighted sample}$$

$$\Rightarrow \text{ weight } w_i = \frac{\pi(x_i)}{g(x_i)}$$



Estimation

importance sampling $\rightarrow$ computation of Monte Carlo estimates

e. g. expectations $\mathrm{E}_\pi\{f(x)\}$:

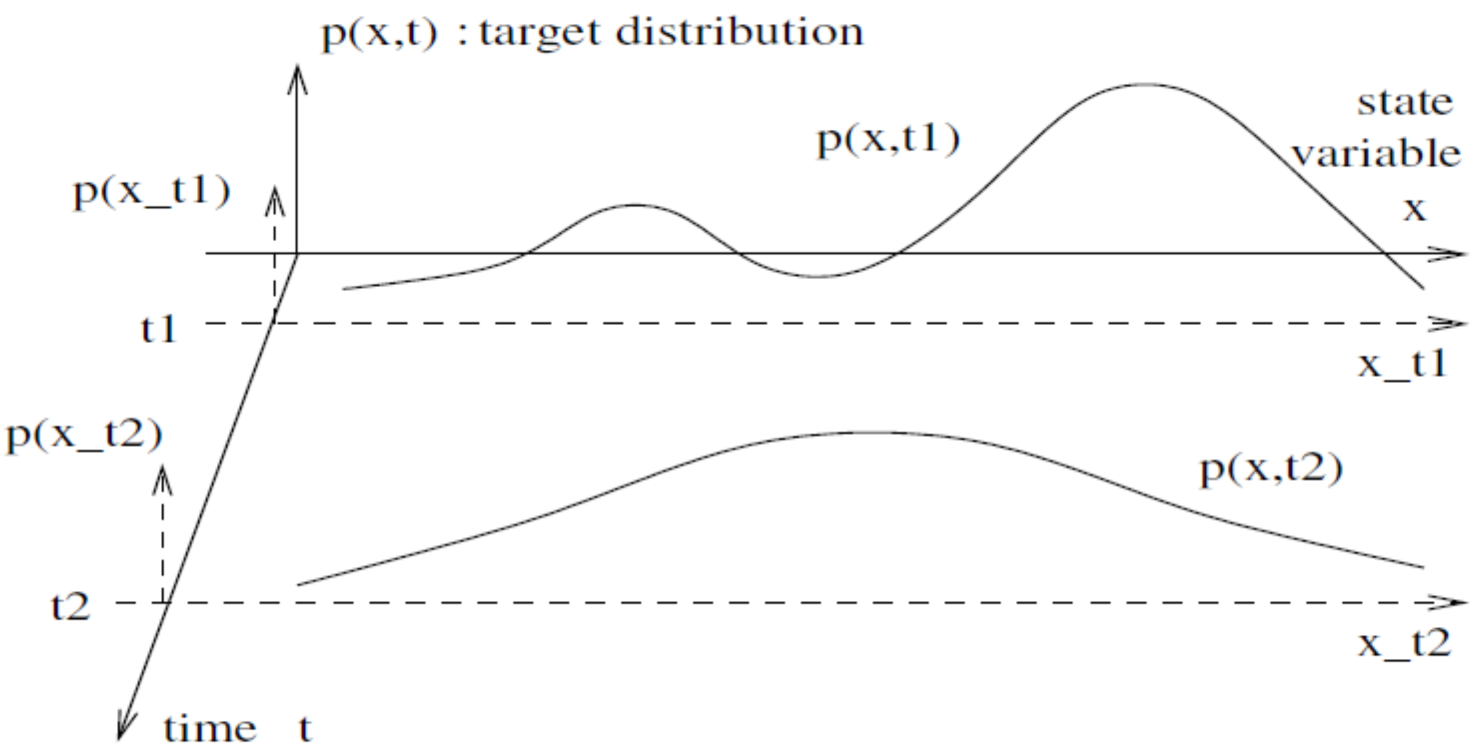
$$\begin{aligned}\int f(x)\frac{\pi(x)}{g(x)}g(x)dx &= \int f(x)\pi(x)dx \\ \sum_{i=1}^{N} w_i f(x_i) &\rightarrow \int f(x)\pi(x)dx = \mathrm{E}_\pi\{f(x)\}\end{aligned}$$

dynamic model $(x_t, y_t) \Rightarrow$ recursive estimation $\widehat{x}_{0:t-1} \rightarrow \widehat{x}_{0:t}$

Monte Carlo techniques $\Rightarrow$ sampling sequences $x^{(i)}_{0:t-1} \rightarrow x^{(i)}_{0:t}$

Sequential Simulation

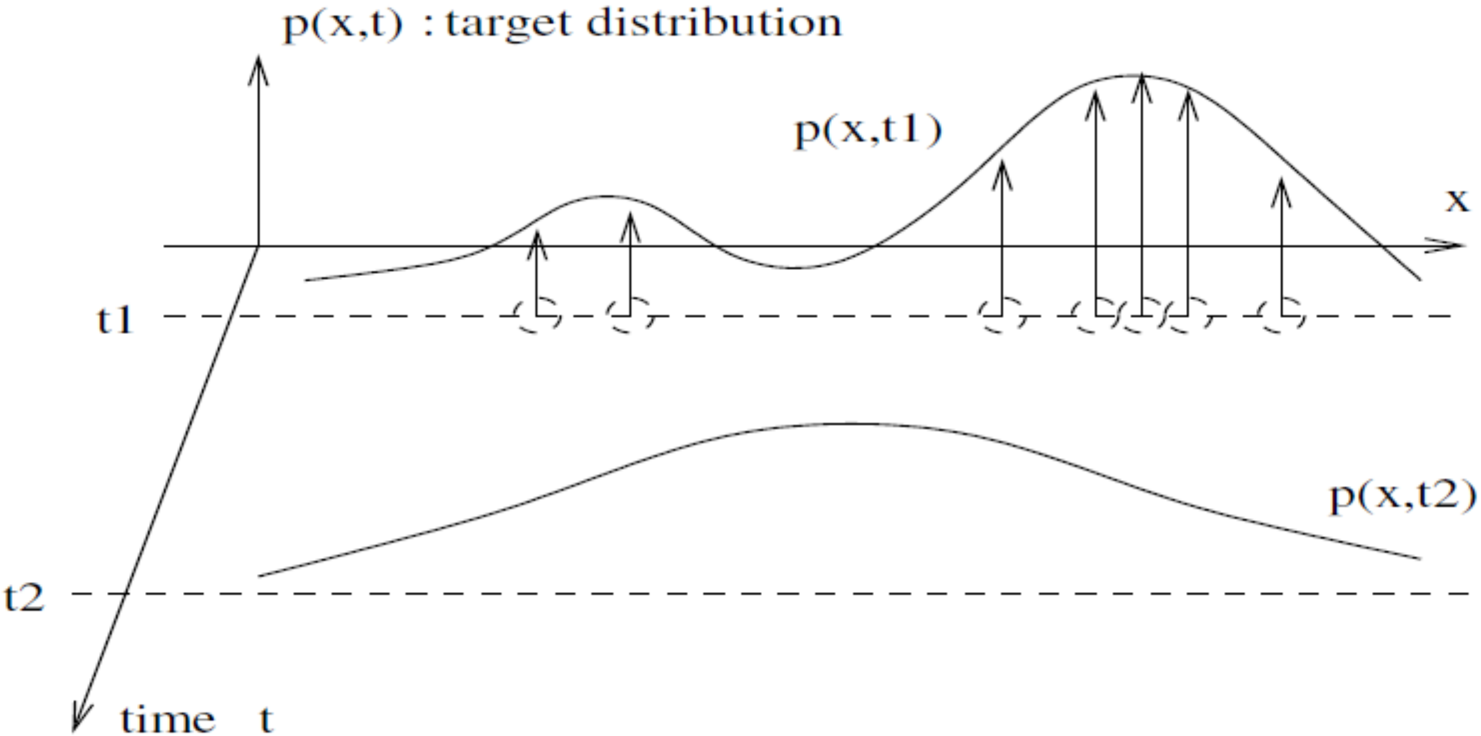
sampling sequences $x_{0:t}^{(i)} \sim \pi_t(x_{0:t})$ recursively:



Sequential Simulation: Importance Sampling

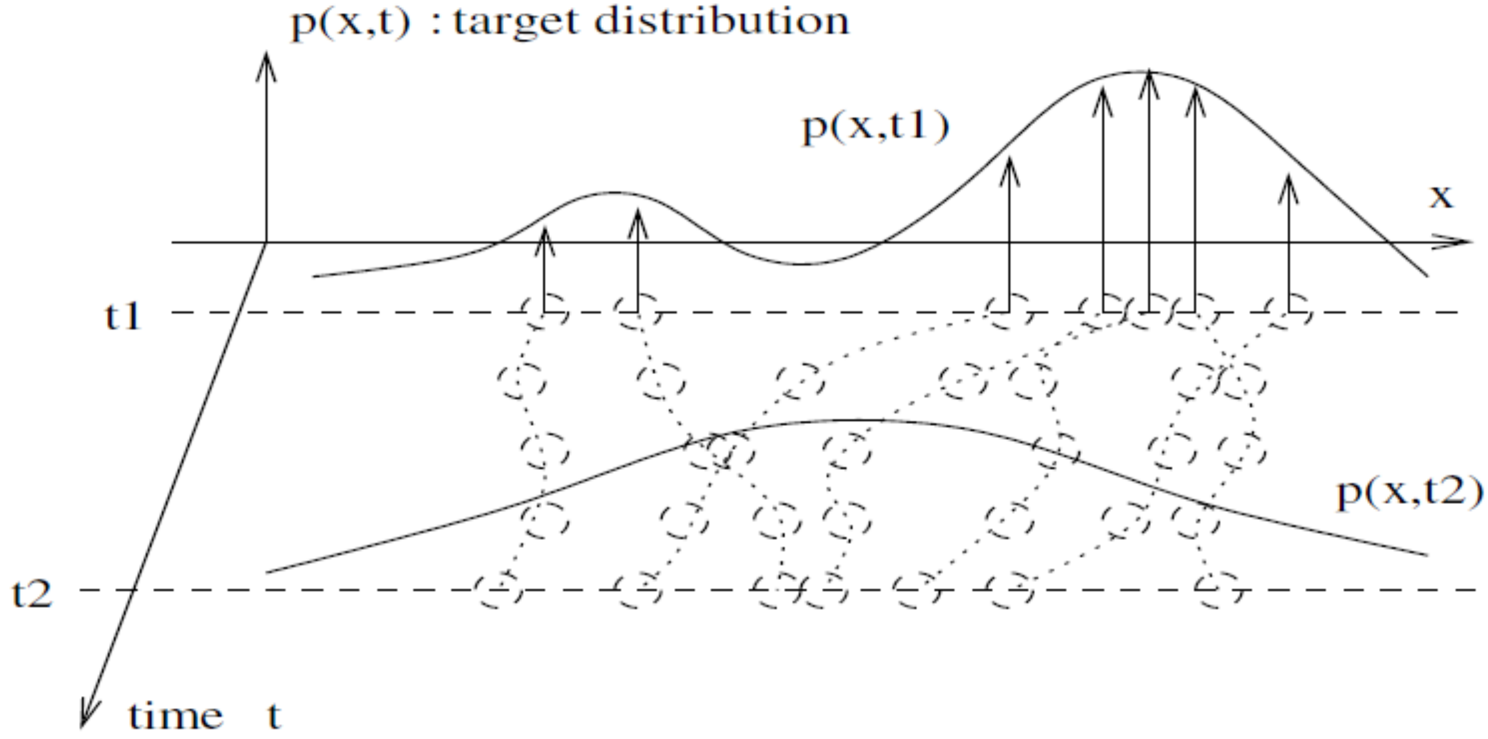
samples $x_{0:t}^{(i)} \sim \pi_t(x_{0:t})$ approximated by weighted particles

$$(x_{0:t}^{(i)}, w_t^{(i)})_{1 \leq i \leq N}$$



Sequential Importance Sampling (1/2)

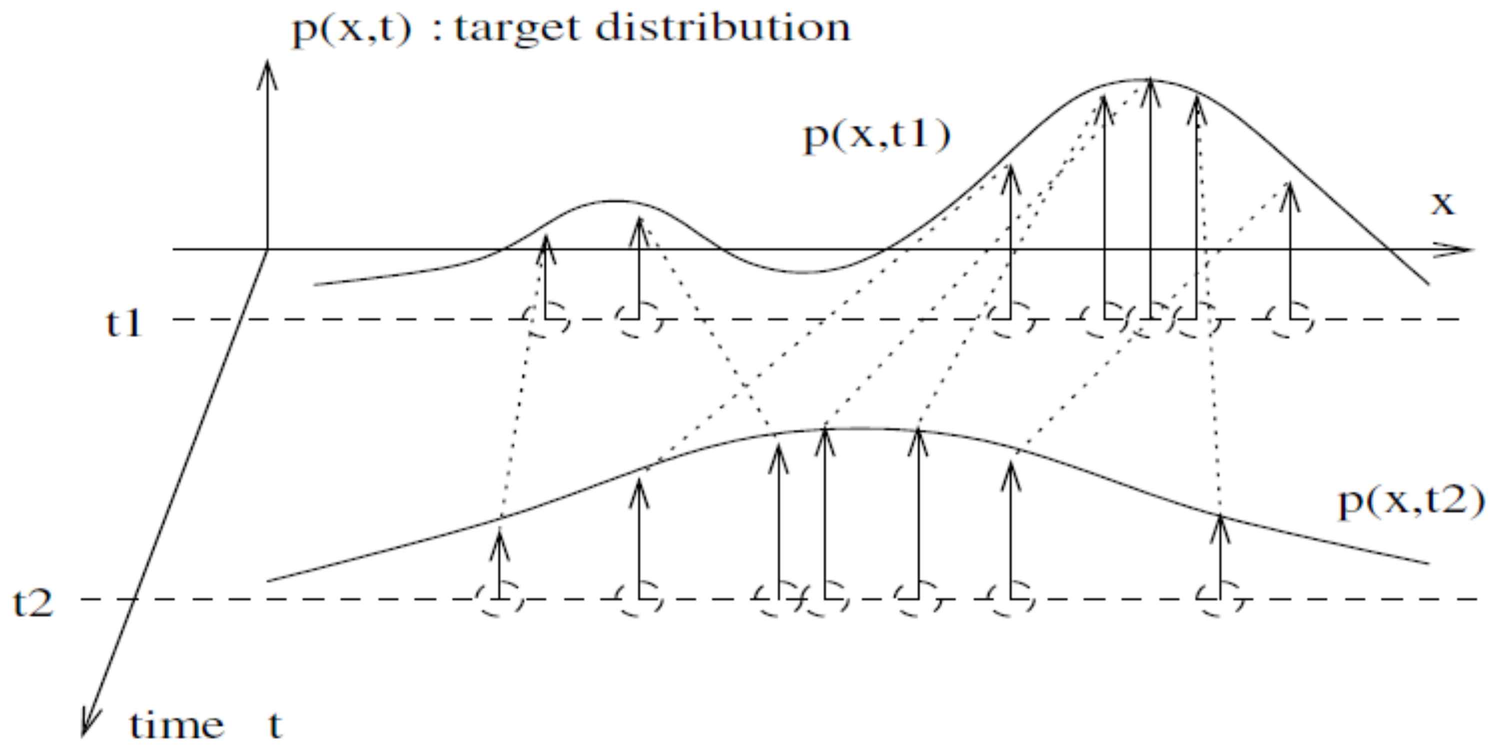
diffusing particles $x_{0:t_1}^{(i)} \longrightarrow x_{0:t_2}^{(i)}$



$\Rightarrow$ sampling scheme $x_{0:t-1}^{(i)} \longrightarrow x_{0:t}^{(i)}$

Sequential Importance Sampling (2/2)

updating weights $w_{t_1}^{(i)} \rightarrow w_{t_2}^{(i)}$



$$\Rightarrow \text{ updating rule } w_{t-1}^{(i)} \rightarrow w_t^{(i)}$$

Sequential Importance Sampling Scheme

$$x_{0:t} \sim \pi_t(x_{0:t}) \Rightarrow (x_{0:t}^{(i)}, w_t^{(i)})_{1 \leq i \leq N}$$

Simulation scheme $t-1 \rightarrow t$:

- Sampling step $x_t^{(i)} \sim q_t(x_t | x_{0:t-1}^{(i)})$

- Updating weights

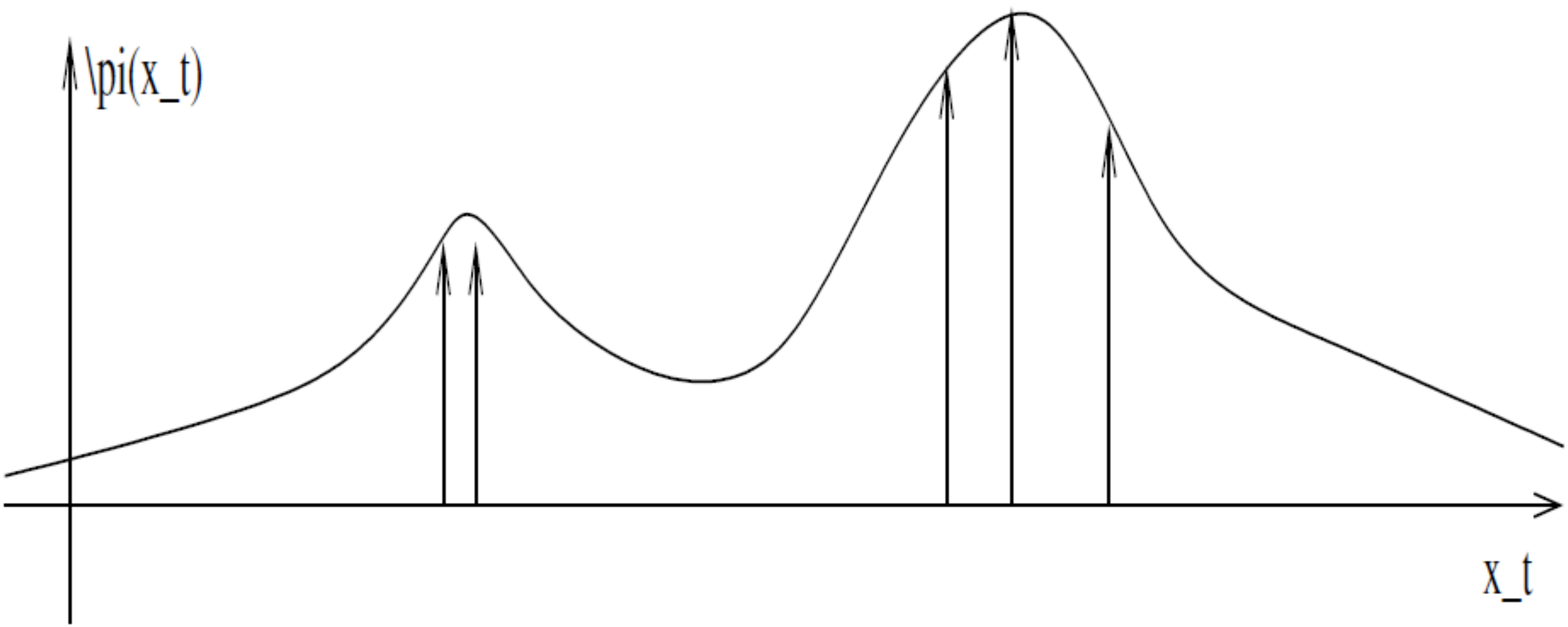
$$w_t^{(i)} \propto w_{t-1}^{(i)} \times \underbrace{\frac{\pi_t(x_{0:t-1}^{(i)}, x_t^{(i)})}{\pi_{t-1}(x_{0:t-1}^{(i)}) q_t(x_t^{(i)} | x_{0:t-1}^{(i)})}}_{\text{incremental weight (iw)}}$$

normalizing $\sum_{i=1}^{N} w_t^{(i)} = 1$

Sequential Importance Sampling Issue (1/2)

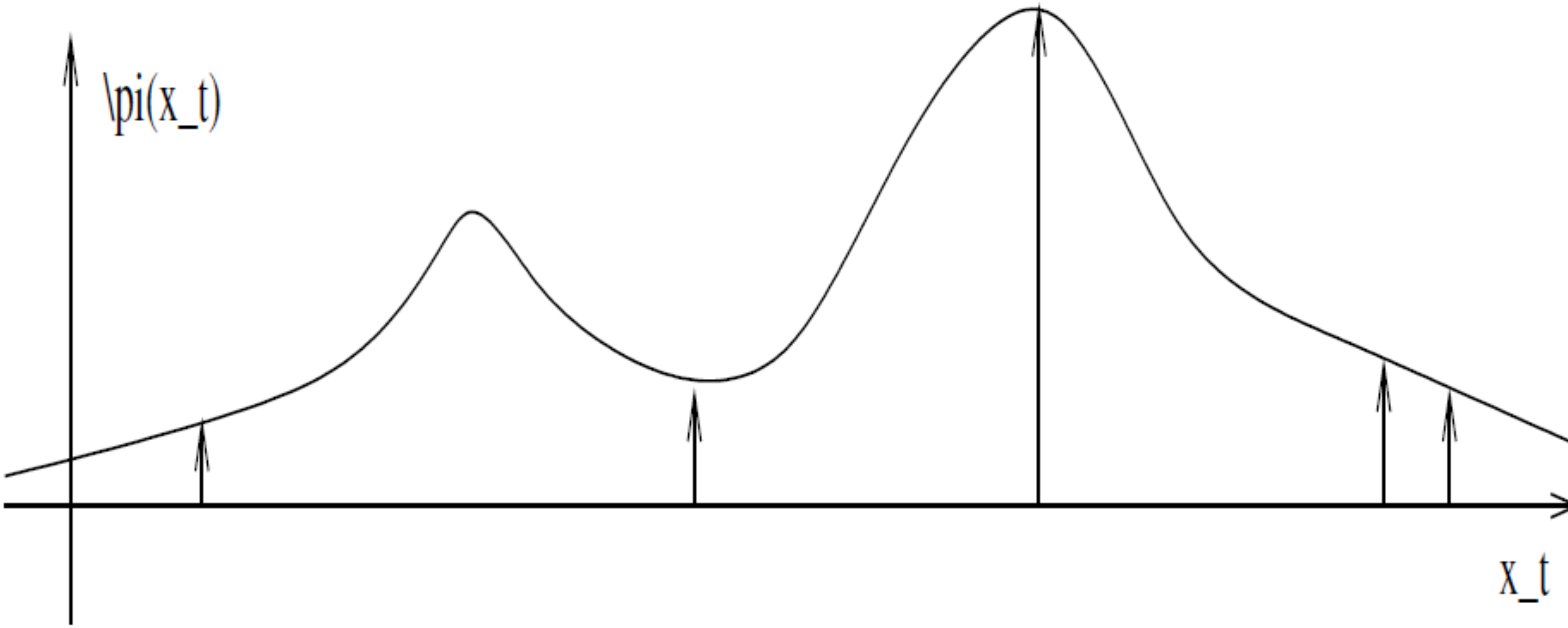
$$x_{0:t} \sim \pi_t(x_{0:t}) \Rightarrow (x_{0:t}^{(i)}, w_t^{(i)})_{1 \leq i \leq N}$$

proposal + reweighting $\rightarrow$



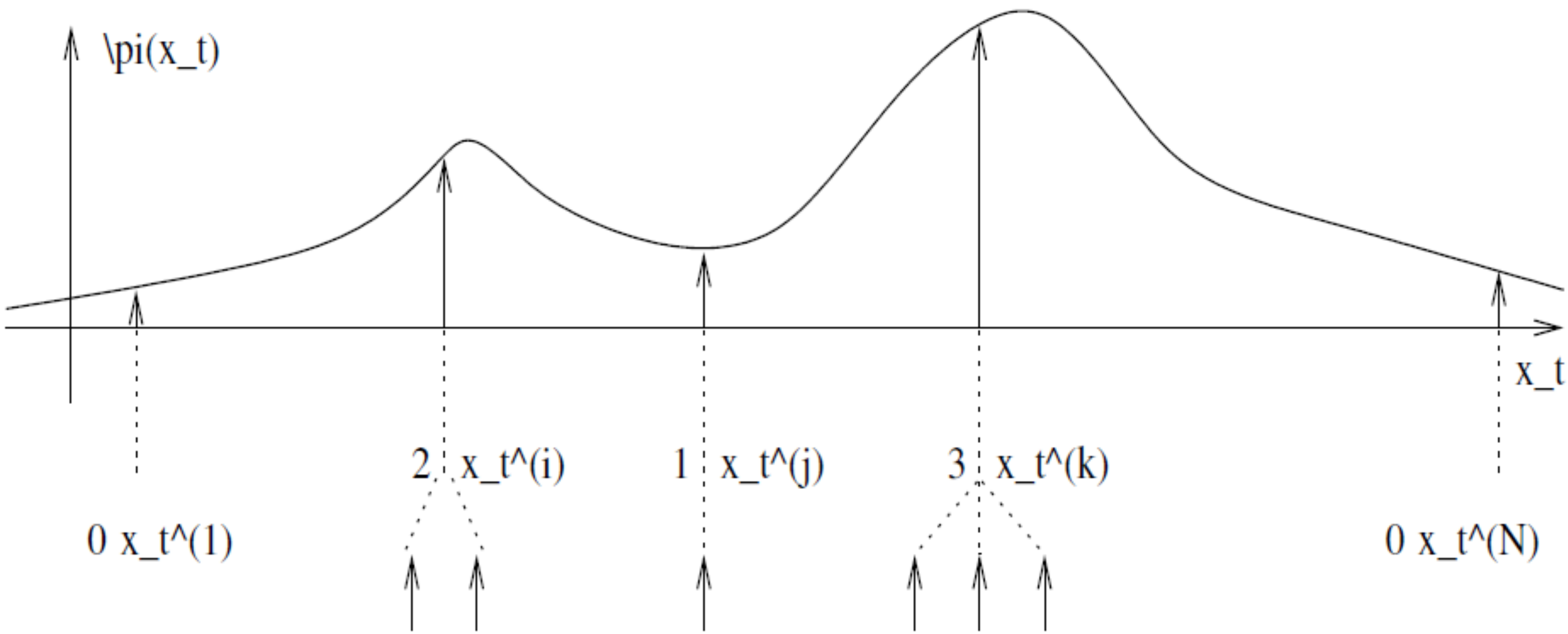
Sequential Importance Sampling Issue (2/2)

$$\text{proposal} + \text{reweighting} \rightarrow \text{var}\{(w_t^{(i)})_{1\leq i\leq N}\} \nearrow \text{with } t$$



$$\rightarrow w_t^{(i)} \approx 0 \text{ for all } i \text{ except one}$$

→Resampling



$$\rightarrow \text{draw } N \text{ particles paths from the set } (x_{0:t}^{(i)})_{1\leq i\leq N}$$
$$\text{with probability } (w_t^{(i)})_{1\leq i\leq N}$$

Sequential Importance Sampling/Resampling Scheme

Simulation scheme $t-1 \to t$:

- Sampling step $x_t^{\prime(i)} \sim q_t(x_t^{\prime}|x_{0:t-1}^{(i)})$

- Updating weights $w_t^{(i)} \propto w_{t-1}^{(i)} \times \frac{\pi_t(x_{0:t-1}^{(i)}, x_t^{\prime(i)})}{\pi_{t-1}(x_{0:t-1}^{(i)}) q_t(x_t^{\prime(i)}|x_{0:t-1}^{(i)})}$

  $\to$ parallel computing

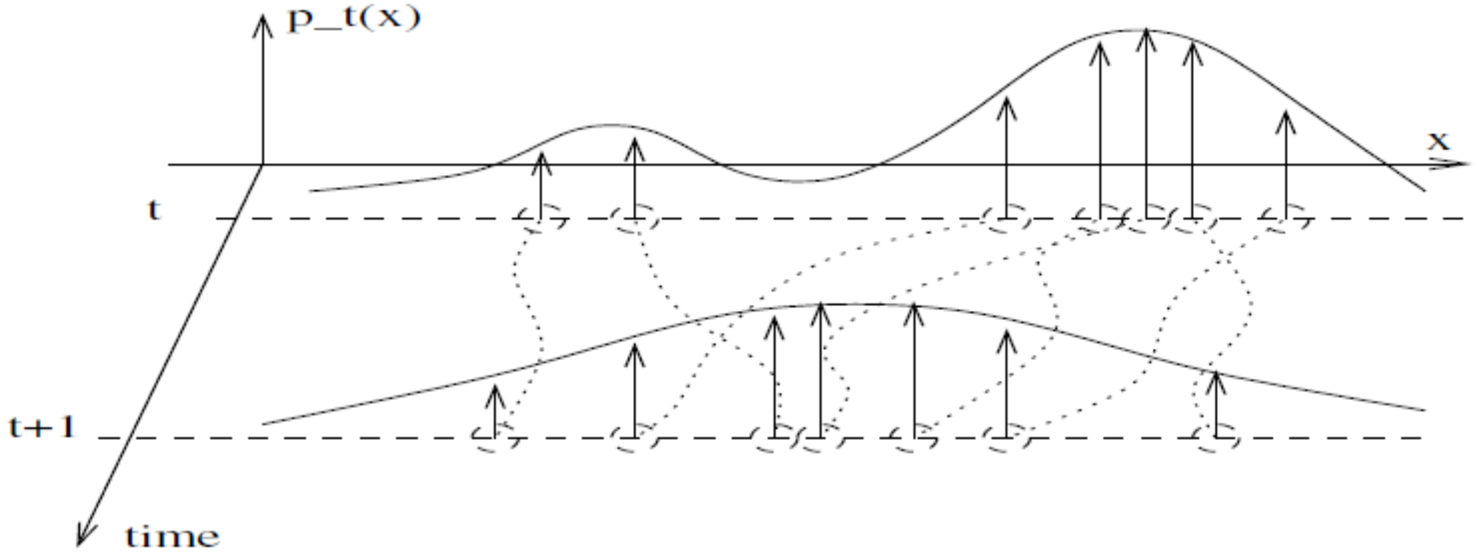
- $\Rightarrow$ **Resampling step**: sample N paths from $(x_{0:t-1}^{(i)}, x_t^{\prime(i)})_{1 \leq i \leq N}$

  $\to$ particles interacting : computation at least $O(N)$

Sequential simulation: SISR

Recursive estimation of state space models.

Approximation with particles, importance sampling.



Bootstrap, particle filtering

Gordon et al. 1993, Kitagawa 1996, Doucet et al. 2001

$\rightarrow$ time series, tracking.

Sequential Importance Sampling Resampling (SISR)

Samples $x_{0:t}^{(i)} \sim \pi_t(x_{0:t})$ approximated by
weighted particles $(x_{0:t}^{(i)}, w_t^{(i)})_{1 \leq i \leq N}$

Simulation scheme $t-1 \rightarrow t$:

- Sampling step $x_t'^{(i)} \sim q_t(x_t' | x_{0:t-1}^{(i)})$

- Updating weights $w_t^{(i)} \propto w_{t-1}^{(i)} \times \underbrace{\frac{\pi_t(x_{0:t-1}^{(i)}, x_t'^{(i)})}{\pi_{t-1}(x_{0:t-1}^{(i)}) q_t(x_t'^{(i)} | x_{0:t-1}^{(i)})}}_{\text{incremental weight (iw)}}$

- Resampling step: sample N paths from $(x_{0:t-1}^{(i)}, x_t'^{(i)})_{1 \leq i \leq N}$

SISR for Recursive Estimation of State Space Models

$$x_t = f_t(x_{t-1}, u_t) \quad \rightarrow \quad p(x_t|x_{t-1})$$
$$y_t = g_t(x_t, v_t) \quad \rightarrow \quad p(y_t|x_t)$$

Usual SISR: Bootstrap filter (Gordon et al. 93, Kitagawa 96):

- Sampling step $x_t^{(i)} \sim p(x_t|x_{t-1}^{(i)})$
- Updating weights : incremental weight $w_t^{(i)} \propto w_{t-1}^{(i)} \times iw$

$$iw \propto p(y_t|x_t^{(i)})$$

- Stratified/Deterministic resampling

efficient, easy, fast for a wide class of models

tracking, time series $\rightarrow$ nonlinear non-Gaussian state spaces

Improving Simulation

Optimal proposal distribution $q_t(x_t|x_{0:t-1}^{(i)})$

$\rightarrow$ mimimizing variance of incremental weight $(w_t^{(i)} \propto w_{t-1}^{(i)} \times iw)$

$$iw = \frac{\pi_t(x_{0:t-1}^{(i)}, x_t^{(i)})}{\pi_{t-1}(x_{0:t-1}^{(i)})q_t(x_t^{(i)}|x_{0:t-1}^{(i)})}$$

$\Rightarrow$ 1-step ahead predictive:

$$\pi_t(x_t|x_{0:t-1}) = p(x_t|x_{t-1}, y_t)$$

$\Rightarrow$ incremental weight:

$$\begin{aligned} iw \rightarrow \frac{\pi_t(x_{0:t-1})}{\pi_{t-1}(x_{0:t-1})} &= \frac{p(x_{0:t-1}|y_{1:t})}{p(x_{0:t-1}|y_{1:t-1})} \\ &\propto p(y_t|x_{t-1}) = \int p(y_t|x_t)p(x_t|x_{t-1})dx_t \end{aligned}$$