Information Theory (2)

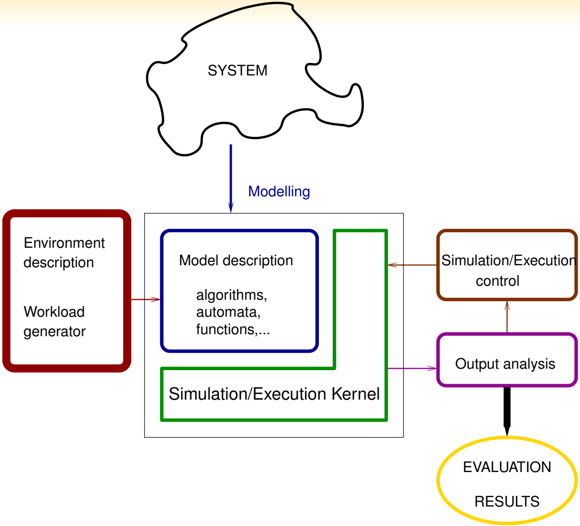
Some Applications in Computer Science

Jean-Marc Vincent

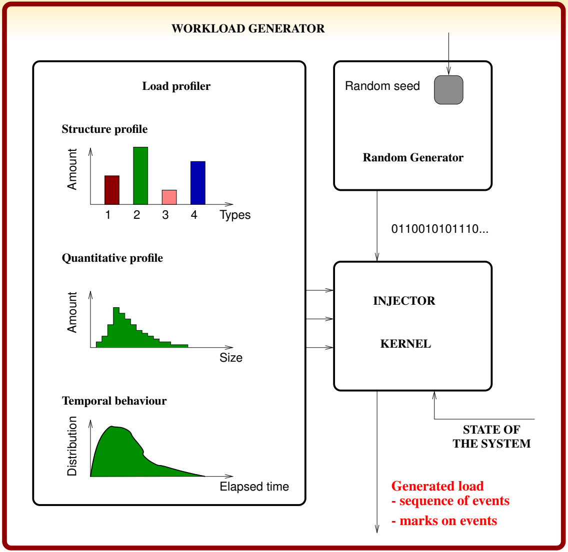
MESCAL-INRIA Project
Laboratoire d'Informatique de Grenoble
Universities of Grenoble, France
Jean-Marc.Vincent@imag.fr
LICIA - joint laboratory LIG -UFRGS

This work was partially supported by CAPES/COFECUB

Modeling Application



Workload generation problem (2)



Observations (example)

Consider a Web server, 4 types of requests A, B, C, D.

Basic Question

Build a stochastic model of the workload :
Without any other assumptions, we assume the workload be:
- a stochastic sequence of independent and uniformly distributed random variables on $\{A, B, C, D\}$

Some more assumptions

For each type of request

Type of request	A	B	C	D
Average processing time (s)	10	4	3	1

Observation :
The average processing time of N requests is NT with $T = 6$
Build a stochastic model of the workload

MaxEnt Principle

Observation

Partial information on the system state X :
Function Φ and we observe

$$\mathbb{E}\phi(X).$$

Examples :

- $\phi(x) = 1$ normalization;
- $\phi(x) = x$ average state;
- $\phi(x) = \log(x)$ order of magnitude;
- $\phi(x) = 1\!\!1_{X>\alpha}$ threshold constraint...

The total information on the system is given by

$$\mathcal{C} = \{(\phi_j, a_j), j = 0, 1, ..., m\},$$

$$\mathbb{E}\phi_j(x) = a_j,$$

Convention : $\phi_0 = 1$ and $a_0 = 1$

MaxEnt Principle

Principle

Under the set of constraints

$$\mathcal{C} = \{(\phi_j, a_j), j = 0, 1, ..., m\},$$

model the system by the distribution maximizing

$$\mathcal{H}(X) = \sum_i p_i(-log_2 p_i)$$

under $\mathcal{C}$.

Minimize the a priori information inside the model description
Uniform distribution is the 0-knowledge hypothesis
Computable form of the distribution
Danger : such a distribution may not exist

Application of MaxEnt Principle

Maximize entropy of the model under constraints

$$(\Phi_0) \quad p_A + p_B + p_C + p_D = 1,$$

$$(\Phi_1) \quad 10p_A + 4p_B + 3p_C + p_D = 6.$$

Using Lagrange multipliers

$$g(p, \lambda_0, \lambda_1) = \sum p_i(-\log p_i) + \lambda_0(\sum p_i - 1) + \lambda_1(\sum T_i p_i - T).$$

$$\frac{\partial g}{\partial p_i} = -\log p_i - 1 + \lambda_0 + \lambda_1 T_i,$$

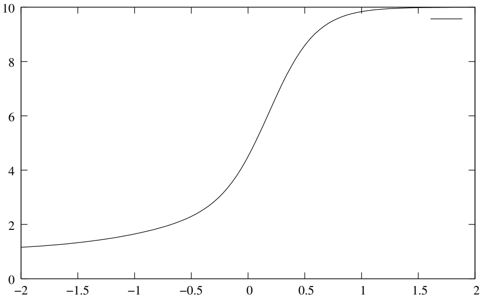
and

$$p_i = 2^{\lambda_0 - 1 + \lambda_1 T_i}.$$

Application of MaxEnt Principle

Using constraints

$$\frac{\sum T_i 2^{\lambda_1 T_i}}{\sum 2^{\lambda_1 T_i}} = T.$$



$\lambda_1 = 0.1712$

Type of request	A	B	C	D
Average processing time (s)	10	4	3	1
Probability	0.44	0.22	0.19	0,15

Classical laws

Integer valued variable on $\{0, 1, 2, \cdots, n\}$

- No constraints : uniform distribution
- $\phi_1(x) = x$ average constraint

$$p_i = \frac{1-\rho}{1-\rho^{n+1}}\rho^i,$$

Geometric distribution

Extends to $\mathbb{N}$

Continuous Variables

Continuous Variable Entropy

For X random variable with density f_X we define

$$\mathcal{H}(X) = \int (-\log f_X(x)) f_X(x) dx.$$

The properties of $\mathcal{H}$ are almost the same as for discrete distributions (positivity fails)

Gibbs distributions

A random variable has a Gibbs distribution when the density has the form

$$f(x) = \exp\left\{ \sum_{j=0}^{m} \lambda_j \phi_j(x) \right\},$$

play a central role in statistics
exponential, Gaussian,... are Gibbs distributions

Optimal distributions

Gibbs densities are MaxEnt distributions

Suppose that there is a Gibbs distribution of X^* satisfying the constraints

$$\mathcal{C} = \{(\phi_j, a_j), j = 1, ..., m\}.$$

then X^* has the maximum entropy distribution under $\mathcal{C}$.

- Uniform distribution is MaxEnt under constraint $X \in [a, b]$
- Exponential distribution is MaxEnt under constraint $\mathbb{E}X = m$
- Normal distribution is MaxEnt under constraint $\mathbb{E}X = m$ and $Var\ X = \sigma^2$
- Poisson process, Markov process, etc.

Synthesis

For modeling systems

- establish the knowledge on your system parameters (fixed or variable)
- establish what is really random
- establish the knowledge on the random part (put the constraints)
- apply the MaxEnt principle (use independence if there are no correlations)
- generate/analyse your workload

Generalization

- Combinatorial structures
- Non uniforme reference distribution (Gaussian)
- ...

References

Cover, T. and Thomas, J. (2006), *Elements of Information Theory* 2nd Edition, Wiley-Interscience

E. T. Jaynes, “Information Theory and Statistical Mechanics,” Physical Review, vol. 106, no. 4, pp. 620-630; May 15, 1957.