Institut National de Recherche en Informatique et Automatique

Institut de Recherche en Informatique et Systèmes Aléatoires

Robustesse des techniques de Monte Carlo dans l'analyse d'événements rares

H. Cancela, Univ. de la Republica, Montevideo

G. Rubino, B. Tuffin, IRISA-INRIA, Rennes

btuffin@irisa.fr

ACI Sure-Paths, Oct. 2005

Outline

1 - Generalities : ARMOR's work on Monte Carlo

- focus on the rare event case
- static and dynamic (Markov) models
- in the dynamic case, transient and stationary analysis
- algorithm design and generic analysis of Monte Carlo techniques

More precisely :

(i) (static) reliability analysis (network reliability area),

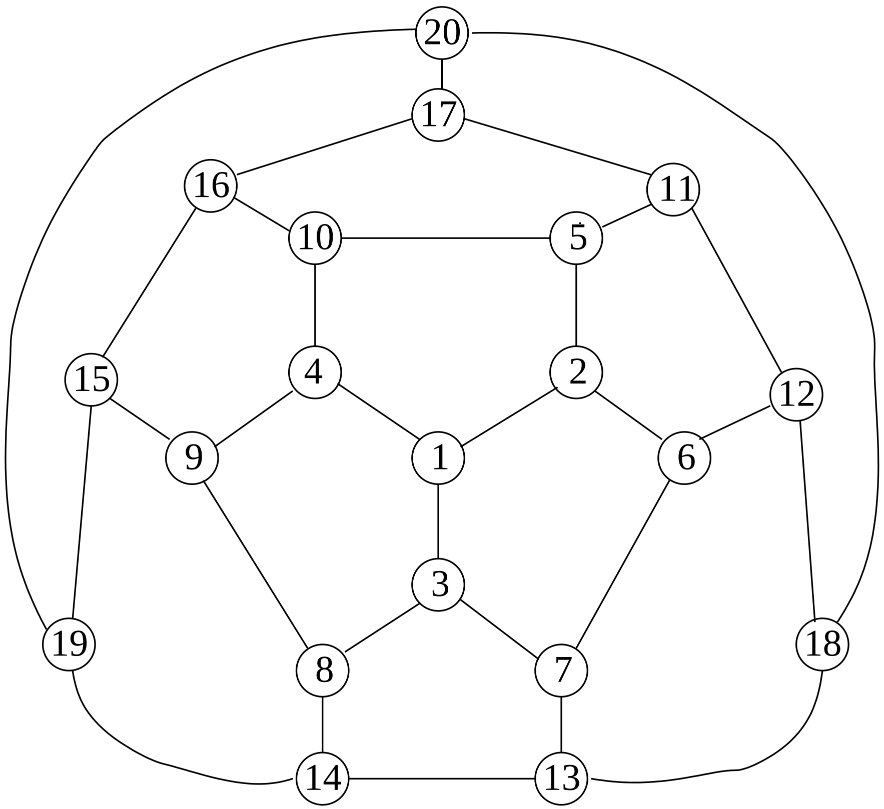
(ii) transient analysis of Markov models : estimation of the reliability at time t,

(iii) transient analysis of Markov models : Importance Sampling applied to estimating the MTTF of a system,

(iv) theoretical analysis of properties of Monte Carlo evaluation methods : illustration in the case of static models.

and variations around (performability, . . .).

Example of (i) :



– a communication network
– nodes are perfect, lines (edges) can fail
– lines are up or down, independently
– nodes 1 and 20 must talk to each other
– network reliability R is the probability that there is at least, a path up between 1 and 20.

Example of (ii), (iii) :

- a multi-component system (say, a database),
- K types of components (disks, power units, servers, . . .),
- type-k components fail with rate λ_k, possible failure propagations
- reparations following an exponential distribution
- system is up iff at least n_k type-k components are,
- we get a Markov model, homogeneous, finite, with states classed up or down.

- An interesting metric is

MTTF = mean time to system down.

- A more detailed metric is the reliability at time t,

$$R(t) = \Pr(\text{system is up from time 0 to time } t).$$

2 - Standard Monte Carlo

- Consider a real random variable X and a real function $\psi()$.
- Let $\gamma = \mathsf{E}(\psi(X))$ and $\sigma^2 = \mathsf{Var}(\psi(X))$.
- The standard estimator of γ :
  - build n independent copies $X_1, \cdots, X_n$ of X
  - return $\gamma_n^{\text{STD}} = \dfrac{1}{n}\sum_{i=1}^{n}\psi(X_i)$.
- $\mathsf{E}(\gamma_n^{\text{STD}}) = \gamma, \quad \mathsf{Var}(\gamma_n^{\text{STD}}) = \sigma^2/n$

- A centered confidence interval for γ with (confidence) level δ is

$$\mathsf{C}_n^{\mathsf{STD}} = \left(\gamma_n^{\mathsf{STD}} \pm z_\delta \sqrt{\mathsf{Var}(\gamma_n^{\mathsf{STD}})}\right)$$

– the meaning is $\Pr(\mathsf{C}_n^{\mathsf{STD}} \ni \gamma) = \delta$

– factor z_δ is

$$z_\delta = \mathcal{N}^{-1}\left(\frac{1+\delta}{2}\right),$$

where

$$\mathcal{N}(x) = \frac{1}{2\pi}\int_0^x \exp(-u^2/2)d\,u.$$

(z_δ is the $1-\delta/2$ quantile of the std. normal distr.).

3 - Rarity

Rarity happens when $\gamma \ll 1$.

Examples :

- In the network reliability problem,

$$X = 1_{\{\text{chosen nodes can not communicate}\}},$$

and

$$\psi(x) = x.$$

(That is, $\gamma = 1 - R$).

When network components are very reliable, $\gamma \ll 1$.

- When estimating MTTF in a Markov model, we first write

$$\text{MTTF} = \frac{\mathsf{E}(\min\{U,V\})}{\Pr(U<V)},$$

where U is the r.v. “time to return to 0”, V is the r.v. “time to absorption”, and 0 is the initial state of the chain, assumed to be an “up” state.

- Focus on estimating the denominator, the only difficult component of the formula : $\gamma = \Pr(U < V)$ and we can put this into the general framework as before.
- If the system is highly reliable, again $\gamma \ll 1$.

Rarity parameter

- We assume that X depends on some real parameter ε (and, thus, γ, σ are also functions of ε).
- We call ε a *rarity parameter* because it verifies

$$\lim_{\varepsilon \to 0} \gamma = 0.$$

- Example : in the static model, we can assume that the reliability of line i, r_i, has the form

$$r_i = 1 - a_i \varepsilon^{b_i}.$$

where a_i and b_i are positive constants.

- In this case, we can prove that there exists some $r > 0$ such that

$$\gamma = \Theta(\varepsilon^r).$$

- If we look at the relative error when estimating γ using the standard estimator γ_n^{STD},

$$\lim_{\varepsilon \to 0} z_\delta \frac{\sqrt{\text{Var}(\gamma_n^{\text{STD}})}}{\gamma} = \frac{z_\delta}{\sqrt{n}} \lim_{\varepsilon \to 0} \sqrt{\frac{1-\gamma}{\gamma}} = \infty.$$

This is why, in general, we can not use γ_n^{STD} in case of rare events.

4 - Bounded Relative Error (BRErr)

- Consider now any estimator γ_n of γ (for instance, an estimator built using some Importance Sampling method).
- Assume E$(\gamma_n) = \gamma$ (that is, γ_n is unbiased) and denote $\sigma_n^2 =$ Var(γ_n).
- We can again build a centered confidence interval for γ from γ_n, writing

$$\mathrm{C}_n = (\gamma_n \pm z_\delta \sigma_n)$$

- The **relative error** is

$$\mathrm{RErr} = z_\delta \frac{\sigma_n}{\gamma}.$$

- We say that we have a bounded relative error (BRErr) if RErr remains bounded as $\varepsilon \longrightarrow 0$.
- This desirable property states that the relative size of the confidence interval remains bounded as $\varepsilon \longrightarrow 0$.
- In the case of the standard estimator, $\sigma_n^2 = \gamma(1-\gamma)/n$, and

$$\text{RErr} = z_\delta \frac{\sigma_n}{\gamma} = z_\delta \sqrt{\frac{1-\gamma}{n\gamma}} \longrightarrow \infty \quad \text{when } \varepsilon \longrightarrow 0.$$

4.1 - Asymptotic Optimality

- Often used in queuing theory when γ_n comes from the Importance Sampling method.
- $\gamma = E_f[g(X)] = E_{f^*}[g(X)L(X)]$ where $L(x) = f(x)/f^*(x)$.
- $\hat{\gamma}_n^{IS}$ is called asymptotically optimal, if $\lim_{\varepsilon\to 0} \frac{\ln E_{f^*}[g(X)^2 L(X)^2]}{\ln \gamma} = 2$.
- Bounded Relative Error implies Asymptotic Optimality (recently proved by Sandmann), the reciprocal being false.
- Bounded Relative Error appears to be the right property.

5 - Bounded Normal Approximation (BNA)

- **Berry-Esseen** : if $F_n()$ is the cdf of $(\gamma_n - \gamma)/\hat{\sigma}_n$, where $\hat{\sigma}_n^2$ is the standard estimator of σ^2, we have that for all real x,

$$|F_n(x) - \mathcal{N}(x)| \leq \frac{a\varrho}{\sigma_n^3\sqrt{n}},$$

  with $\varrho = \mathsf{E}(|\psi(X)|^3)$.
- The estimator γ_n has Bounded Normal Approximation iff the ratio ϱ/σ_n^3 remains bounded when $\varepsilon \to 0$
- Proved that BNA implies BRErr, the reciprocal being false. This means again that BNA appears as the right property to look for.

6 - Need for an extension of BRErr : BREff

- What is important in simulation ? the RErr *for a given simulation time*, not for a number n of replications.
- BRErr does not incorporate the second important characteristic of an estimator : *simulation time* (the computational cost).
- The average simulation time to get *one* replication can
  - *increase* with ε, or
  - *decrease* with ε.
- We will illustrate next this last case for a simulation method estimating the reliability of a static stochastic network.

6.1 - Illustration : simulation for static reliability analysis

- Consider the network reliability problem. Denote by G the undirected graph modelling the network, made of M links.
- Computing R, the probability that two fixed nodes s and t are connected, is an NP-hard problem.
- The method :
  - Let $\mathcal{P} = \{P_1, P_2, \cdots, P_H\}$ be a set of elementary (disjoint) paths connecting nodes s and t,
  - Let π_h be the event “all links of path P_h work” and $p_h = \Pr(\pi_h)$.
  - Assume an infinite sequence of independent copies of G is built.
  - Let F be the random variable “first element in the sequence where every path in $\mathcal{P}$ has at least one link that does not work”.

– Variable F is geometrically distributed with parameter $q = \prod_{h=1}^{H}(1 - p_h) = \Pr($no path in $\mathcal{P}$ “works”$)$:

$$\Pr(F = f) = (1 - q)^{f-1} q,$$

and, in particular,

$$\mathsf{E}(F) = \frac{1}{q}.$$

– Idea : sample first from the geometric distribution of F. The estimator is then built assuming that in the first $F - 1$ copies, nodes s and t are connected.

– Assume $F = f$. To know if they are connected in the fth copy, we must sample the network conditioning on the fact that at least one edge in each path is down.

– Consider a path P_h. Call W_h the r.v. giving the first failed edge of P_h.

– Write $P_h = (i_{h,1}, i_{h,2}, \cdots, i_{h,K_h})$.

We have

$$\Pr(W_h = w) = \frac{r_{i_{h,1}} r_{i_{h,2}} \cdots r_{i_{h,w-1}} (1 - r_{i_{h,w}})}{1 - r_{i_{h,1}} r_{i_{h,2}} \cdots r_{i_{h,K_h}}}.$$

– Assume $W_h = w$. Edges $i_{h,1}, i_{h,2}, \cdots, i_{h,w-1}$ are set to “up”, edge $i_{h,w}$ to “down”. Remaining edges in P_h are sampled from their *a priori* Bernoulli distributions.

- We have built (virtually) f copies of G by sampling F once and the network configuration (the states of the M links) once.

- Observe that
  - We have unbounded RErr (same variance than for crude Monte Carlo),
  - We sample F and the network, on the average, nq (= $n/(1/q)$) times, and we test the connectivity between s and t also nq times on the average.
  - The simulation time (or computing cost) is, on the average, $O(nq(M + K_1 + \cdots + K_H))$ (= $O(nqM)$ if we wish). It decreases with ε if $r_i = 1 - a_i\varepsilon^{b_i}$, since q decreases with ε.

6.2 - Bounded relative Efficiency

- It basically addresses the (relative) variance of an estimator obtained during a given simulation time.
- Consider an estimator γ_n of γ, with variance σ_n^2, built from n replications (possibly dependent), and denote by t_n the average simulation time needed to get these n replications.

  The relative efficiency of γ_n is

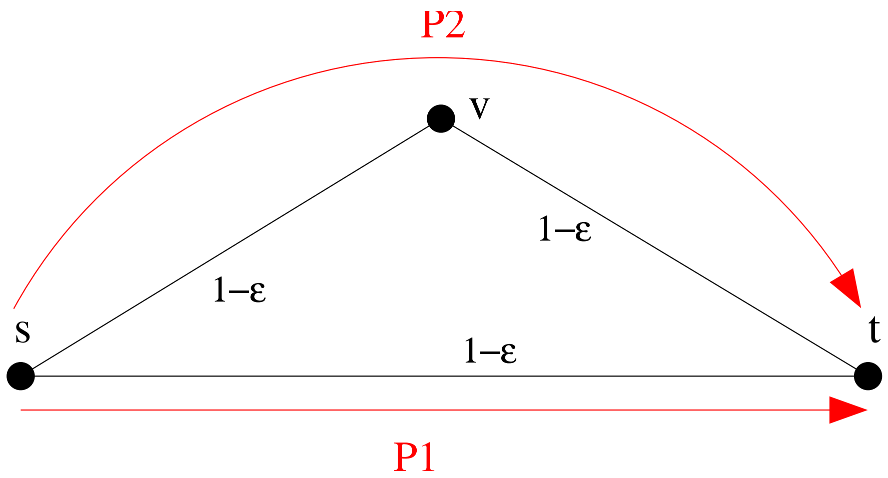
$$\text{REff} = \frac{\gamma^2}{\sigma_n^2 t_n}.$$

  The estimator γ_n has bounded relative efficiency (BREff) if there exists $d > 0$ such that REff is minored by d for all ε.

6.3 - Sufficient condition on the static reliability example

- Let $r > 0$ be the real such that $\gamma = \Theta(\varepsilon^r)$.
- $\forall P_h \in \mathcal{P}$, let $P_h = (i_{h,1}, \cdots, i_{h,K_h})$ and $b_h = \min_{1 \le k \le K_h} b_{i_{h,k}}$ (b_h is the exponent of ε in the most reliable edge of P_h (as $\varepsilon \to 0$)).
- Theorem : the estimator of the static unreliability described in previous section verifies Bounded Relative Efficiency if $\sum_{h=1}^{H} b_h \ge r$.

7 - Illustrations 7.1 - Illustration on a (very) small problem



- $\gamma = 1 - R = \varepsilon^3 + 2\varepsilon^2(1-\varepsilon) \approx 2\varepsilon^2$
- $\sigma^2 = \gamma(1-\gamma) \approx 2\varepsilon^2$
- $\sigma/\gamma \approx 1/(\sqrt{2}\varepsilon)$: no BRErr
- $p_1 = 1 - \varepsilon, p_2 = (1-\varepsilon)^2$
- $q = (1-p_1)(1-p_2) \approx 2\varepsilon^2$;
- t_n proportional to q
- REff $\approx \dfrac{\gamma^2}{\sigma^2 q}$ bounded : BREff.

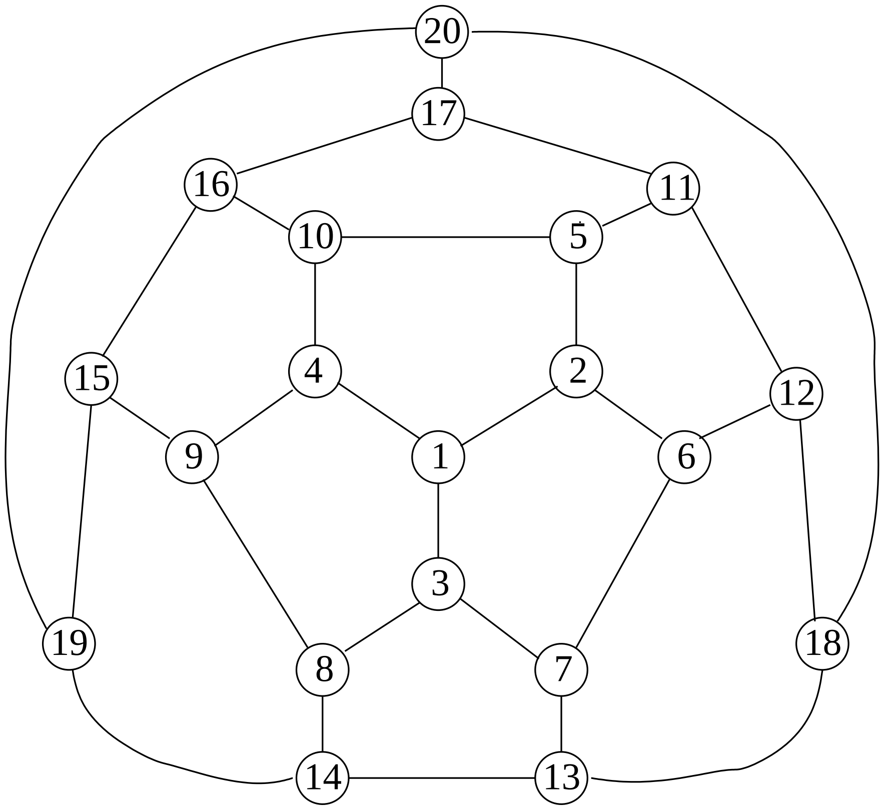
Results on the simple topology, with a number of replications fixed to $n = 10^4$:

$r_i \; \forall i$	Est.	Conf. interval	RErr	KS stat.
0.5	3.7790000e-01	(3.6839622e-01,3.8740378e-01)	2.51489e-02	5.6705e-02
0.9	1.9012678e-02	(1.8992726e-02,1.9032629e-02)	1.04937e-03	4.8610e-02
0.99	2.0000000e-04	(-7.7171997e-05,4.7717200e-04)	1.38586e+00	2.5436e-01
0.999	0	(0, 0)	—	1
0.9999	0	(0, 0)	—	1

Results for a simulation time $T = 10$ seconds :

$r_i \; \forall i$	Est.	Conf. interval	RErr
0.5	3.7498089e-01	(3.7466460e-01,3.7529718e-01)	8.43484e-04
0.9	1.9012678e-02	(1.8992726e-02,1.9032629e-02)	1.04937e-03
0.99	1.9905429e-04	1.9884527e-04,1.9926330e-04)	1.05005e-03
0.999	2.0003614e-06	(1.9982415e-06,2.0024813e-06)	1.05975e-03
0.9999	2.0014432e-08	1.9993102e-08,2.0035762e-08)	1.06572e-03

7.2 - Illustration : efficiency on the dodecahedron topology



with $s = 1$ and $t = 20$

$r_i \ \forall i$	Relative eff. w.r.t. crude MC
0.9	18.9
0.95	188.3
0.98	3800.2

$r_i \ \forall i$	Est.	Conf. interval	RErr	KS stat.
0.5	7.0820e-01	(6.99290e-01,7.17110e-01)	1.25818e-02	3.3896e-02
0.9	3.2000e-03	(2.09298e-03,4.30702e-03)	3.45945e-01	9.6531e-02
0.99	0	(0, 0)	—	8.2840e-01
0.999	0	(0, 0)	—	1

Results for a simulation time $T = 5$ seconds :

$r_i \ \forall i$	Est.	Conf. interval	RErr	KS stat.
0.5	7.1201e-01	(7.0918e-01 , 7.1485e-01)	3.98e-03	4.3126e-02
0.9	2.8894e-03	(2.7776e-03 , 3.0012e-03)	3.87e-02	7.0679e-02
0.99	2.0527e-06	(1.9211e-06 , 2.1844e-06)	6.41e-02	3.3475e-02
0.999	2.0047e-09	(1.8705e-09 , 2.1388e-09)	6.69e-02	6.5625e-02
0.9999	1.9812e-12	(1.8474e-12 , 2.1150e-12)	6.75e-02	5.4364e-02

8 - Generalized Bounded Normal Approximation

- Again, in BNA, time (cost) not taken into account.
- Fix time (computational budget) T and let $n = n(T)$ be the corresponding average number of of iterations.
- The estimator $\gamma_{n(T)}$ verifies GBNA if the ratio $\varrho_{n(T)}/(\sigma^3_{n(T)}\ n(T))$ remains bounded when $\varepsilon \longrightarrow 0$.
- It can be proved now that GNBA implies BREff and that in the case of the static reliability problem, both properties are equivalent.

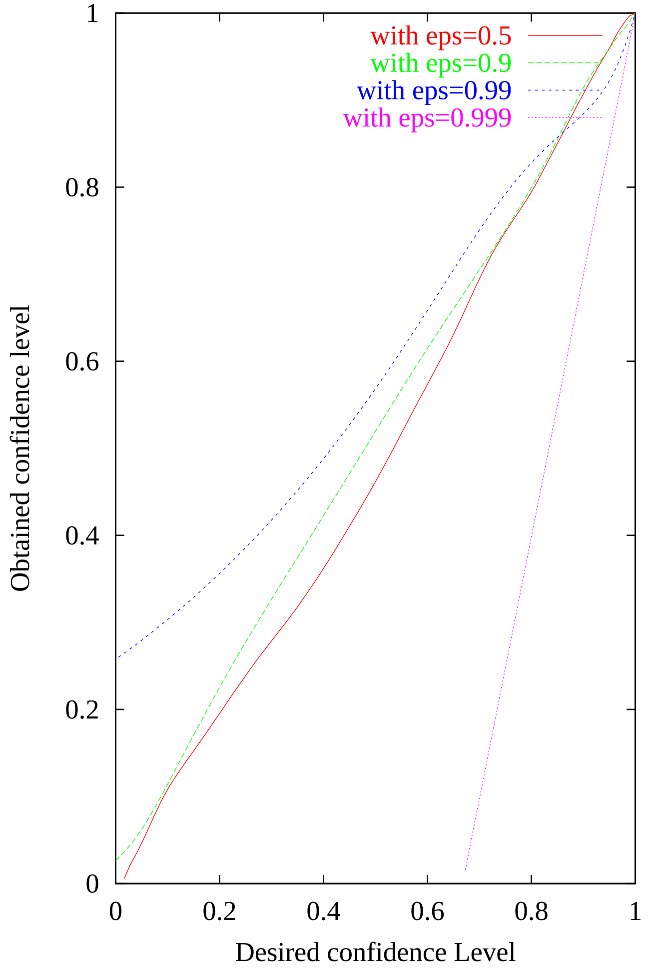
Coverage function

- Confidence interval $R(\eta, \mathbb{X})$ for γ, at confidence level η with (random) data $\mathbb{X}$.
- We should have $\Pr[\gamma \in R(\eta, \mathbb{X})] = \eta$.
- If $\eta^* = \inf\{\eta \in [0,1] : \gamma \in R(\eta, \mathbb{X})\}$, then, η^* should be uniformly distributed :

$$F_{\eta^*}(\eta) = \Pr[\eta^* \leq \eta] = \eta.$$

- $F_{\eta^*}(\eta)$ is the *actual* coverage level : empirical estimation.

Illustration on the simple topology



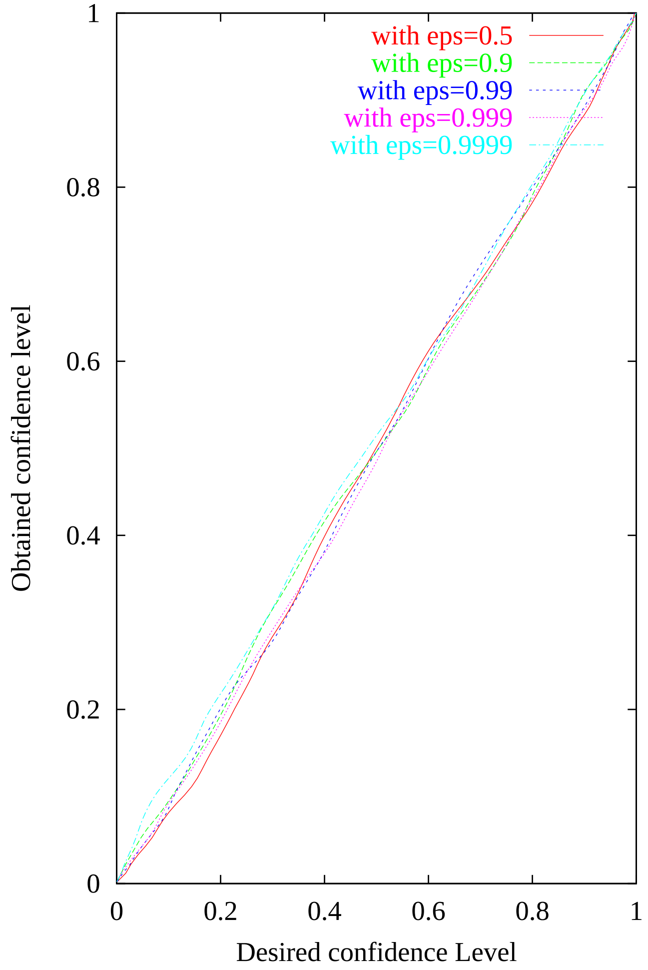
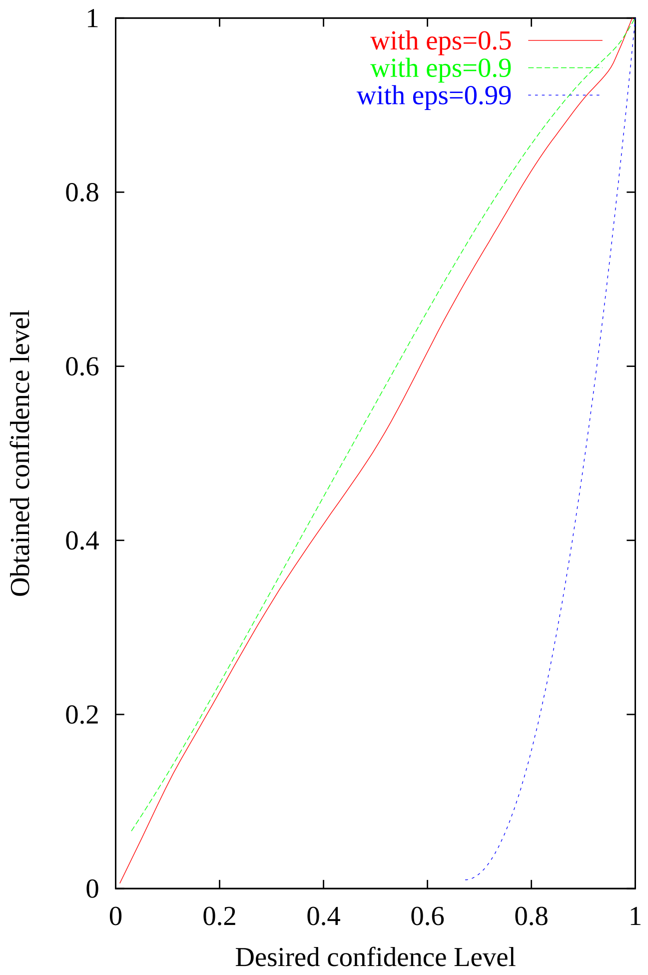
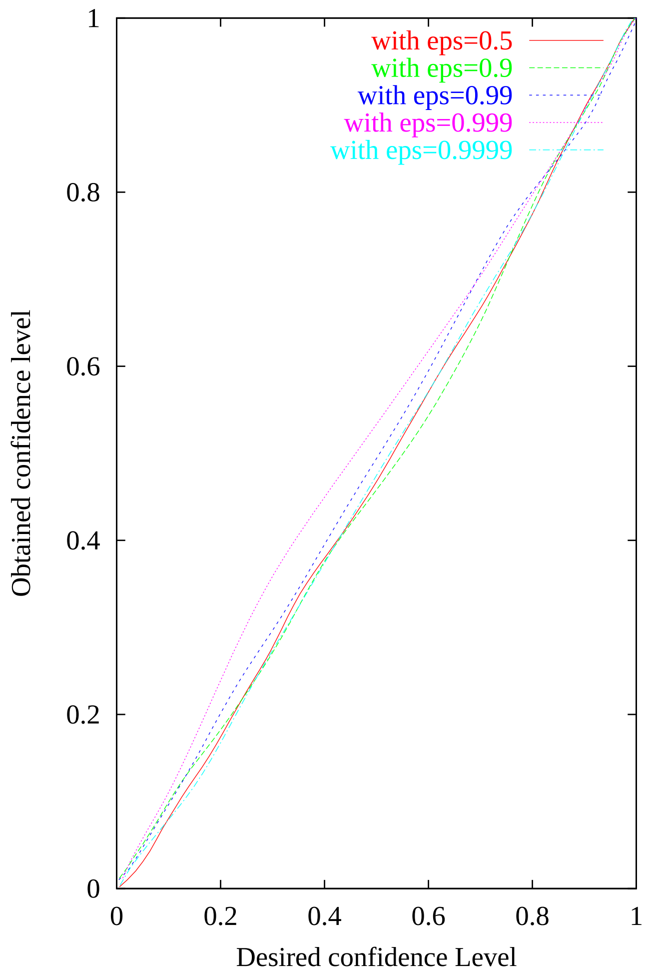


Illustration on the Dodecahedron topology





9 - Conclusions

- Rare event simulation requires sophisticated techniques with robustness properties.
- Bounded Relative Error does not take into account the estimator's full information.
- We have
  - defined Bounded Relative Efficiency to cope with this problem
  - defined Generalized Bounded Normal Approximation (and previously BNA)
  - Illustrated the validity of the approach on a reliability analysis problem.
  - deep investigation of the coverage of the confidence interval