Échantillonnage préférentiel pour le model checking statistique

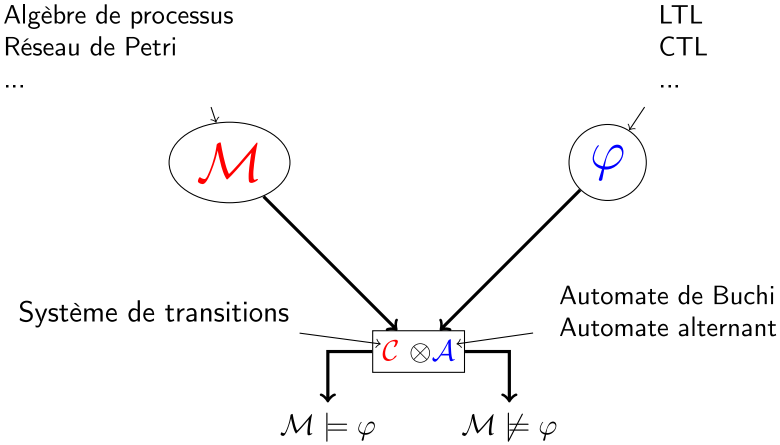
Benoît Barbot, Serge Haddad et Claudine Picaronny

LSV, ENS de Cachan & CNRS & INRIA

MSR'11, Vendredi 18 novembre 2011

Plan

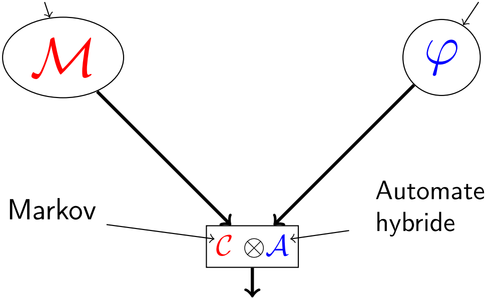
Model checking



Model checking de systèmes probabilistes

Algèbre de processus stochastique
Réseau de Petri stochastique
...

PCTL
HASL
...



Chaîne de Markov

Automate *déterministe* hybride

$$p = \Pr(\mathcal{M} \models \varphi)$$

Approches numérique et statistique

- Approche numérique
  - Logiques arborescentes (basées sur CTL)
  - Valeur exacte (aux approximations numériques près)
  - Implémenté efficacement dans de nombreux outils (PRISM, MRMC, GreatSPN)
  - Hypothèses probabilistes fortes
  - Taille de l'espace mémoire proportionnelle à la taille du processus stochastique

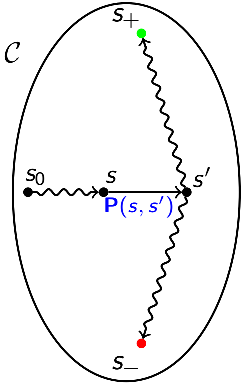
Approches numérique et statistique

- Approche numérique
  - Logiques arborescentes (basées sur CTL)
  - Valeur exacte (aux approximations numériques près)
  - Implémenté efficacement dans de nombreux outils (PRISM, MRMC, GreatSPN)
  - Hypothèses probabilistes fortes
  - Taille de l'espace mémoire proportionnelle à la taille du processus stochastique
- Approche statistique
  - Logiques linéaires (basées sur LTL)
  - Intervalle de confiance : encadrement probabiliste
  - Taille de l'espace mémoire très faible
  - Facilement parallélisable
  - Hypothèses probabilistes faibles
  - Inutilisable pour de très faibles probabilités (évènements rares)

Objectif

Développement d'une méthode pour traiter les évènements rares

Formalisation du problème



A partir de $\varphi = aUb$, $\mathcal{C}$ est transformé :
On agrège les états qui vérifient $\neg a$ en un état s_- et les états qui vérifient b en un état s_+.
Les états s_- et s_+ sont rendus absorbants.

On suppose que s_-, s_+ sont atteints avec probabilité 1

Soit $\sigma = s \to s_1 \to s_2 \to \cdots \to s_\pm$ une trajectoire aléatoire dans $\mathcal{C}$

$$V_s = \begin{cases} 1 \text{ si } \sigma \text{ finit dans l'état } s_+ \\ 0 \text{ si } \sigma \text{ finit dans l'état } s_- \end{cases}$$

Objectif : Calculer $\mathbf{E}(V_{s_0})$ lorsque $\mathbf{E}(V_{s_0}) \ll 1$

Difficulté : $\mathbf{V}(V_{s_0}) \approx \mathbf{E}(V_{s_0}) \gg \mathbf{E}^2(V_{s_0})$

Échantillonnage préférentiel

Principe : Substituer W_s à V_s de même espérance mais de variance réduite

1. Substituer $\mathbf{P}'$ à $\mathbf{P}$ telle que $\mathbf{P}(s,s') > 0 \Rightarrow \mathbf{P}'(s,s') > 0 \vee s = s_-$
2. Pour chaque trajectoire $\sigma = s \to s_1 \to s_2 \cdots s_k \to s_\pm$
   On définit

$$W_s = \begin{cases} \frac{\mathbf{P}(s,s_1)}{\mathbf{P}'(s,s_1)} \cdot \frac{\mathbf{P}(s_1,s_2)}{\mathbf{P}'(s_1,s_2)} \cdot \ldots \cdot \frac{\mathbf{P}(s_k,s_+)}{\mathbf{P}'(s_k,s_+)} & \text{si } \sigma \text{ finit dans l'état } s_+ \\ 0 & \text{si } \sigma \text{ finit dans l'état } s_- \end{cases}$$

3. Estimer statistiquement $\mathbf{E}(W_{s_0})$

Échantillonnage préférentiel

Principe : Substituer W_s à V_s de même espérance mais de variance réduite

1. Substituer $\mathbf{P}'$ à $\mathbf{P}$ telle que $\mathbf{P}(s,s') > 0 \Rightarrow \mathbf{P}'(s,s') > 0 \vee s = s_-$
2. Pour chaque trajectoire $\sigma = s \to s_1 \to s_2 \cdots s_k \to s_\pm$
   On définit

$$W_s = \begin{cases} \frac{\mathbf{P}(s,s_1)}{\mathbf{P}'(s,s_1)} \cdot \frac{\mathbf{P}(s_1,s_2)}{\mathbf{P}'(s_1,s_2)} \cdot \ldots \cdot \frac{\mathbf{P}(s_k,s_+)}{\mathbf{P}'(s_k,s_+)} & \text{si } \sigma \text{ finit dans l'état } s_+ \\ 0 & \text{si } \sigma \text{ finit dans l'état } s_- \end{cases}$$

3. Estimer statistiquement $\mathbf{E}(W_{s_0})$

La méthode est non biaisée

$$\forall s \in S,\ \mathbf{E}(W_s) = \mathbf{E}(V_s)$$

Objectif

$$\mathbf{V}(W_{s_0}) \ll \mathbf{V}(V_{s_0})$$

Échantillonnage préférentiel optimal

Un résultat « non opérationnel »

Il existe un échantillonnage préférentiel de variance nulle

Soit $\mu(s) = \mathbf{E}(V_s)$
Soit $\mathbf{P}'(s,t) = \frac{\mu(t)}{\mu(s)} \cdot \mathbf{P}(s,t)$

$$W_s = \frac{\mathbf{P}(s,s_1)}{\mathbf{P}'(s,s_1)} \cdot \frac{\mathbf{P}(s_1,s_2)}{\mathbf{P}'(s_1,s_2)} \cdots \frac{\mathbf{P}(s_k,s_+)}{\mathbf{P}'(s_k,s_+)} = \frac{\mu(s)}{\mu(s_1)} \cdot \frac{\mu(s_1)}{\mu(s_2)} \cdots \frac{\mu(s_k)}{1} = \mu(s)$$

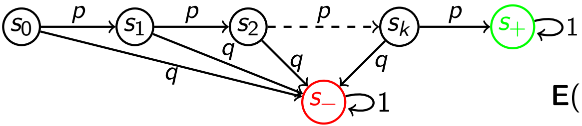
Donc $\mathbf{V}(W_s) = 0$

Problème

Nécessite de connaître μ, que l'on cherche à estimer !

Difficulté de mise en oeuvre

Cas simple (p' substitué à p)

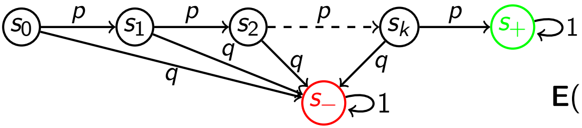


$$\mu(s_0) = p^{k+1}$$

$$\mathbf{E}(W_{s_0}) = \left(\frac{p}{p'}\right)^{k+1} p'^{k+1}$$

Difficulté de mise en oeuvre

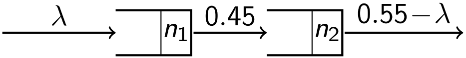
Cas simple (p' substitué à p)



$$\mu(s_0) = p^{k+1}$$

$$\mathbf{E}(W_{s_0}) = \left(\frac{p}{p'}\right)^{k+1} p'^{k+1}$$

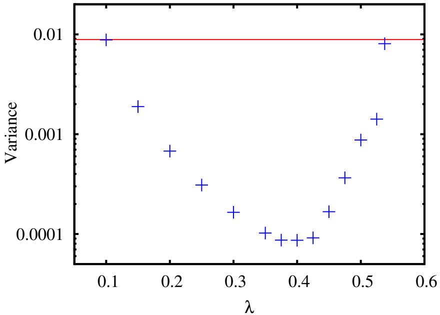
Cas plus complexe (λ' substitué à λ)



Atteindre 5 clients
versus
Revenir au repos

Pour le système originel ($\lambda = 0.1$)

$$\mathbf{E}(V_{s_0}) = 0.0089$$



État de l'art

Échantillonnage préférentiel asymptotiquement optimal
dans une classe d'échantillonnages
(P.Dupuis, A.D. Sezer, H. Wang 2007)

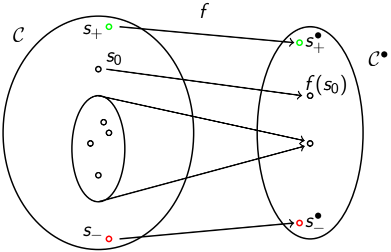
Utilisation d'heuristiques (e.g. apprentissage)
(P.E Heegaard, W. Sandmann 2007)

Analyse au cas par cas
(Rubino, Tuffin 2009)

Aucune de ces méthodes ne fournit
un intervalle de confiance du résultat
car la nature de W_s est inconnue.

Garantie de réduction de variance

Idée : substituer $\mu^\bullet(f(s))$ à $\mu(s)$ dans l'échantillonnage optimal



Quelle condition pour que W_{s_0} soit à valeurs dans $\{0; \mu^\bullet(f(s_0))\}$?

Une condition nécessaire

$\mathcal{C}^\bullet$ *est bornant :* $\mu(s_0) \leq \mu^\bullet(f(s_0))$

Le principe d'homogénéité conduit au renforcement suivant

$\mathcal{C}^\bullet$ *est uniformément bornant :* $\forall s \in S, \mu(s) \leq \mu^\bullet(f(s))$

Le principe de localité conduit à la condition suffisante suivante

$$\forall s \in S,\ \sum_{s' \in S} \mu^\bullet(f(s')) \cdot \mathbf{P}(s, s') \leq \mu^\bullet(f(s))$$

On dit alors que $(\mathcal{C}^\bullet, f)$ est une *réduction de $\mathcal{C}$ à variance garantie*

Échantillonnage préférentiel à variance garantie

Spécification de l'échantillonnage préférentiel

Pour tout $s' \neq s_-$, $\mathbf{P}'(s, s') = \frac{\mu^\bullet(f(s'))}{\mu^\bullet(f(s))}\mathbf{P}(s, s')$

$\mathbf{P}'(s, s_-) = 1 - \sum_{s' \neq s_-} \frac{\mu^\bullet(f(s'))}{\mu^\bullet(f(s))}\mathbf{P}(s, s')$ *(possible en vertu de la condition)*

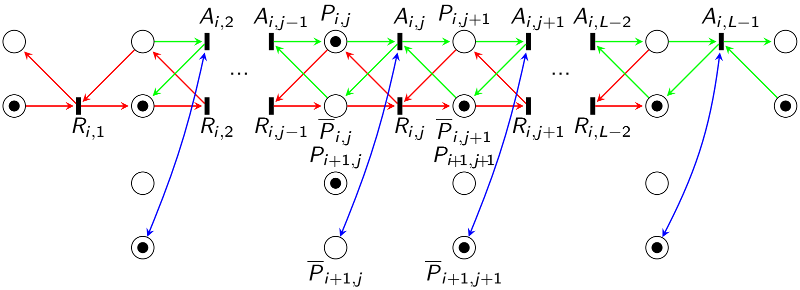
Résultats théoriques

Soit $(\mathcal{C}^\bullet, f)$ une réduction à variance garantie de $\mathcal{C}$.

- Pour tout s, W_s est une variable aléatoire à valeurs dans $\{0, \mu^\bullet(f(s))\}$
- Par conséquent, $\mathbf{V}(W_s) = \mu(s) \cdot \mu^\bullet(f(s)) - \mu^2(s) \approx \mathbf{V}(W_s) \cdot \mu^\bullet(f(s))$
- Comme la distribution (itérée) suit une loi « binomiale »,
  il est possible de calculer un intervalle de confiance.

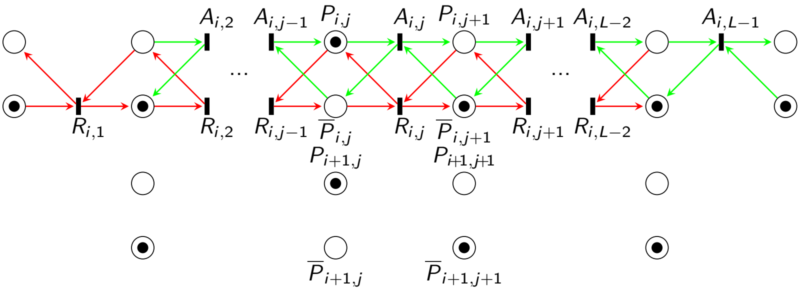
Exemple de la ruine des joueurs

N joueurs font une marche aléatoire de longueur L

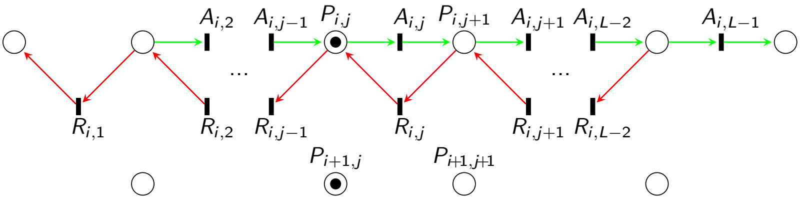


Taille du système L^N

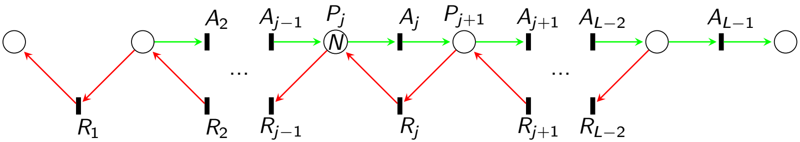
Exemple de la ruine des joueurs



Exemple de la ruine des joueurs



Exemple de la ruine des joueurs



Taille du système $\binom{N+L-1}{L-1}$

Comment vérifier la condition plus structurellement ?

On veut démontrer la condition de réduction sans analyse numérique.

$$\forall s \in S, \sum_{t \in S} \mu^{\bullet}(f(t)) \cdot \mathbf{P}(s,t) \leq \mu^{\bullet}(f(s))$$

Soient $\mathcal{C}$ une DTMC et $\mathcal{C}^{\bullet}$ une chaîne réduite. On suppose f surjective et localement injective. On note pour tout s :

- $Com(s)$, les transitions «communes» à $\mathcal{C}$ et à $\mathcal{C}^{\bullet}$:
  $Com(s) = \{(f(s), f(s')) \mid \mathbf{P}(s,s') > 0 \wedge \mathbf{P}^{\bullet}(f(s), f(s')) > 0\}$
- $Ext(s)$, les transitions de $\mathcal{C}$ «absentes» dans $\mathcal{C}^{\bullet}$:
  $Ext(s) = \{(f(s), f(s')) \mid \mathbf{P}(s,s') > 0 \wedge \mathbf{P}^{\bullet}(f(s), f(s')) = 0\}$
- $Inh(s)$, les transitions de $\mathcal{C}^{\bullet}$ «absentes» dans $\mathcal{C}$:
  $Inh(s) = \{(f(s), s^{\bullet}) \mid$
  $\forall s' \in f^{-1}(s) \;\; \mathbf{P}(s,s') = 0 \wedge \mathbf{P}^{\bullet}(f(s), s^{\bullet})) > 0\}$

Une nouvelle condition pour la variance garantie

Soit $h(s)$ (resp $h^\bullet(f(s))$), la probabilité d'emprunter une transition commune dans $\mathcal{C}$ (resp $\mathcal{C}^\bullet$).

Proposition

Si pour tout s non absorbant tel que $\mu(s) > 0$ on a :

- Les transitions communes ont des probabilités relatives identiques : $\forall (f(s), f(s')) \in Com(s)\ \mathbf{P}(s,s')/h(s) = \mathbf{P}^\bullet(f(s), f(s'))/h^\bullet(f(s))$
- Les transitions propres à $\mathcal{C}$ «diminuent» la probabilité d'atteindre $f(s_+)$ $\forall (f(s), f(s')) \in Ext(s)\ \mu^\bullet(f(s)) \geq \mu^\bullet(f(s'))$
- Les transitions propres à $\mathcal{C}^\bullet$ «augmentent» cette probabilité $\forall (f(s), s^\bullet) \in Inh(s)\ \mu^\bullet(f(s)) \leq \mu^\bullet(s^\bullet)$

Alors $\mathcal{C}^\bullet$ est une réduction à variance garantie.

Méthodologie

1. Spécifier un modèle réduit $\mathcal{M}^\bullet$ de chaîne de Markov associée $\mathcal{C}^\bullet$ et une fonction f.

2. Etablir par examen de $\mathcal{M}$ et $\mathcal{M}^\bullet$ que la réduction est à variance garantie.

3. Calculer numériquement la fonction $\mu^\bullet$.

4. Calculer statistiquement $\mu(s_0)$
en utilisant l'échantillonnage préférentiel induit par $\mu^\bullet$.

Généralisation de la méthode

Spécification d'un échantillonnage préférentiel dans le cas général

Soient $\mathcal{C}$ une DTMC et $\mathcal{C}^\bullet$ une chaîne réduite par f.

Soit s tel que $\mu^\bullet(f(s)) < \sum_{t \in S} \mu^\bullet(f(t)) \cdot \mathsf{P}(s,t) \equiv g(s)$.

Alors :

$$\mathsf{P}'(s,s') = \frac{\mu^\bullet(f(s')) \cdot \mathsf{P}(s,s')}{g(s)}$$

Généralisation de la méthode

Spécification d'un échantillonnage préférentiel dans le cas général

Soient $\mathcal{C}$ une DTMC et $\mathcal{C}^\bullet$ une chaîne réduite par f.

Soit s tel que $\mu^\bullet(f(s)) < \sum_{t \in S} \mu^\bullet(f(t)) \cdot \mathbf{P}(s,t) \equiv g(s)$.

Alors :

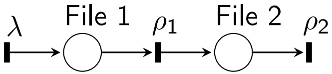
$$\mathbf{P}'(s,s') = \frac{\mu^\bullet(f(s')) \cdot \mathbf{P}(s,s')}{g(s)}$$

Observation

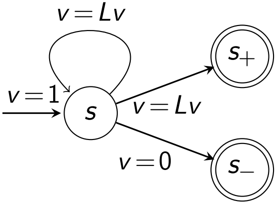
Pour tout s tel que $\mu(s) > 0$,
W_s est une variable aléatoire à valeurs dans $\{0\} \cup [\mu^\bullet(f(s)), \infty[$.

COSMOS un outil de model checking statistique

- Modèle sous forme de réseaux de Petri à distributions générales



- Spécification de formules par un automate hybride



- Model checking et évaluation de performance

- Vitesse de simulation supérieure à PRISM

Adaptation de COSMOS

Adaptation de la méthode aux modèles à temps continu

Calcul de la valeur des trajectoires par l'automate hybride

Encodage de la fonction f et de l'échantillonnage préférentiel.

Exemple de la ruine parallèle

- $p = 0.3$
- $L = 15$
- 300000 trajectoires

N	taille de $\mathcal{C}$	$\mu(s_0)$	T (s) numérique	taille de $\mathcal{C}^\bullet$	$\mu^\bullet(f(s_0))$	T $\mu^\bullet$ (sec)	$\mu(s_0)$ estimé	δ	T (sec) simulation
5	7.5E5	1.884E-9	20	11628	1.444E-8	≈ 0	1.902E-09	4.142E-11	1100
6	1.1E7	1.147E-12	435	38760	2.450E-11	≈ 0	1.157E-12	3.167E-14	930
7	1.7E8	#	#	116280	5.712E-11	3	2.934E-12	8.269E-14	1200
8	2.0E9	#	#	319770	1.033E-13	14	1.885E-15	9.724E-17	1000
9	3.8E10	#	#	817190	2.323E-13	47	4.693E-15	2.776E-16	1400
10	5.7E11	#	#	1961256	4.379E-16	153	3.209E-18	3.173E-19	1000
11	8.0E12	#	#	4457400	9.626E-16	481	7.959E-18	7.664E-19	1300
12	1.3E14	#	#	9657700	1.866E-18	1000	5.590E-21	1.030E-21	1100

Observation

Dés que $N \geq 7$ le résultat obtenu est à la fois hors de portée des model-checkers numériques et statistiques classiques.

Conclusion

- Contributions
  - Conception du premier échantillonnage préférentiel avec intervalle de confiance
  - Application au model checking statistique
  - Expérimentations variées
- Perspectives
  - Meilleure intégration dans l'outil
  - Meilleure condition structurelle (basée sur le couplage)
  - Prise en charge plus générale des systèmes infinis
  - Extension aux distributions non markoviennes