

# TP SdF N° 15

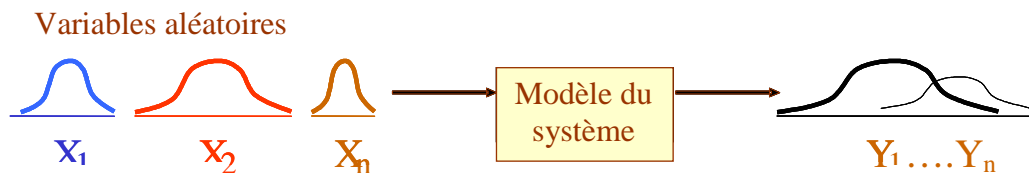
## Réduction de variance

Ce TP a pour objet d'améliorer la précision des résultats de simulations de Monte-Carlo (resserrement des intervalles de confiance) ; cette amélioration permettant en particulier d'estimer la probabilité d'apparition d'évènements rares ou de diminuer la durée des simulations.

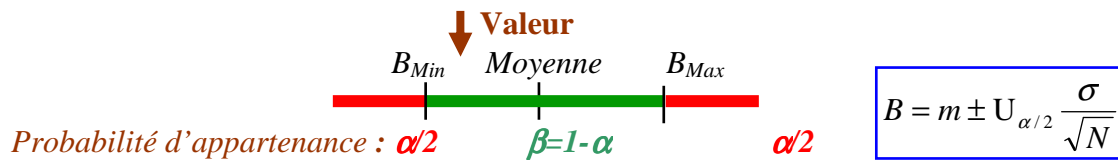
1. Expliquer les principes et techniques de réduction de variance dont l'échantillonnage stratifié et l'échantillonnage d'importance.
2. Montrer l'apport de ces techniques à l'estimation de fiabilité en comparant les résultats de simulation, avec ou sans réduction de variance, et ceux obtenus par calcul analytique dans les deux cas suivants :
  - Système mécanique évalué par la méthode Résistance / Contrainte ( $m_R=16$ ,  $\sigma_R=3$ ,  $m_C=8$ ,  $\sigma_C=2$ )
  - Système à vote entre 3 calculateurs identiques (fiabilité à 4500 heures avec  $MTTF_{\text{calculateur}} = 1000$  heures)
3. Montrer l'apport de ces techniques à l'estimation de disponibilité dans le cas d'une simple redondance active 1 parmi 2 ( $MTTF = 1000$  heures et  $MTTR = 500$  heures) puis dans le cas général en indiquant leurs limitations.

# 1. PRINCIPES ET TECHNIQUES DE REDUCTION DE VARIANCE

## A. RAPPEL SUR LA SIMULATION DE MONTE-CARLO

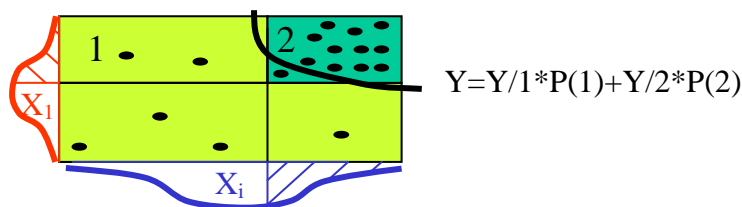


La méthode de simulation de Monte Carlo consiste à réaliser des tirages aléatoires des valeurs des données d'entrées d'un système. Pour chaque configuration de valeurs tirées, on évalue la valeur des données de sortie. La répétition des simulations permet d'obtenir des distributions de ces valeurs et, par là même, divers estimateurs dont notamment leur moyenne. Définie par l'intervalle de confiance, la précision des résultats obtenus est liée au nombre de simulations et à la variance des résultats (théorème central limite)



## B. REDUCTION DE VARIANCE

La réduction de variance consiste à privilégier un domaine d'intérêt au cours de la simulation puis de pondérer les résultats obtenus par application du théorème des probabilités totales.

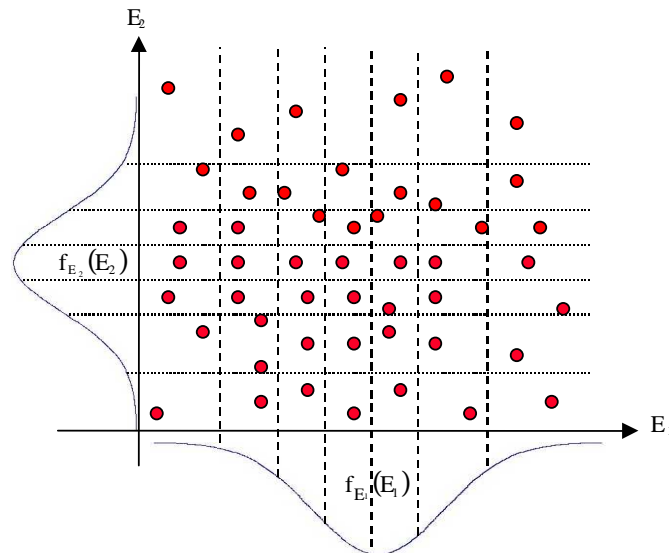


## C. MISE EN OEUVRE

Deux techniques de réduction de la variance sont présentées dans cette fiche : la méthode de l'échantillonnage stratifié et la méthode de l'échantillonnage d'importance.

### I. ECHANTILLONNAGE STRATIFIE

Cette méthode consiste à créer des **strates** (quadrillage) disjointes sur les espaces de départ et à effectuer des tirages dans chacune de ces strates. La valeur de sortie est alors évaluée dans chacune des strates. L'estimation finale du paramètre est la somme des estimations dans chaque strate pondérée par la probabilité que le tirage soit issu de la strate en question.



Cette technique permet ainsi d'explorer les zones généralement délaissées par les tirages classiques.

Trois solutions peuvent être appliquées lors de la partition du domaine :

- Echantillonnage stratifié **équiprobable** : la probabilité d'être dans chaque strate est la même. Cette méthode de stratification tend à « sous échantillonner » les queues de distribution des variables aléatoires d'entrée.
- Echantillonnage stratifié **équidistant** : les strates sont toutes de même taille.
- Echantillonnage stratifié **personnalisé** : lorsque l'on possède une connaissance a priori du système à étudier, il est possible de déterminer les strates en fonction des zones que l'on souhaite étudier en particulier. Il est également possible d'optimiser le nombre de tirage par strate afin d'augmenter la précision de la simulation dans les zones d'intérêt.

## II. ECHANTILLONNAGE D'IMPORTANCE

La méthode de Monte Carlo par échantillonnage d'importance consiste à concentrer les tirages sur les zones des données d'entrée jugées importantes. Pour ce faire, on utilise une fonction appelée **fonction d'importance** qui a pour but « d'encourager » les tirages dans les zones que l'on cherche à atteindre. L'utilisation de cette distribution biaisée conduira à un estimateur biaisé si nous l'appliquons directement aux simulations. Cependant, les différentes simulations peuvent être pondérées pour supprimer ce biais. Le poids qui est donné à chaque simulation est le ratio de la densité de probabilité des fonctions d'entrée par rapport à celle de la fonction d'importance.

Cette méthode est particulièrement adaptée à l'observation de « phénomènes locaux » comme par exemple l'estimation de probabilités extrêmes.

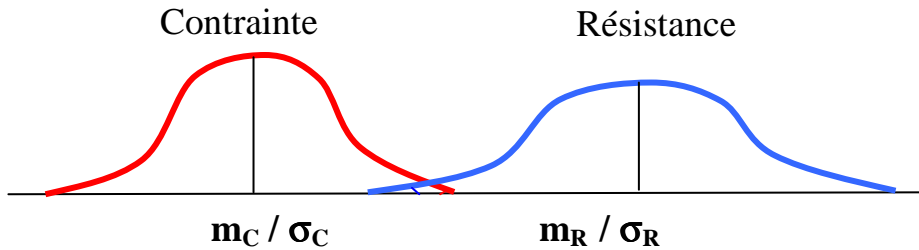
La principale difficulté lors de la mise en œuvre de cette méthode est le choix de la fonction d'importance. Elle est déterminée à partir d'une connaissance a priori de l'influence des paramètres d'entrée sur les données de sortie. Cette connaissance peut être obtenue par une première série d'essais.

## 2. APPORT A L'ESTIMATION DE FIABILITE

### a. ESTIMATION DE FIABILITE PAR LA METHODE RESISTANCE /CONTRAINTE

#### i. RESULTAT THEORIQUE

Le problème consiste à rechercher la probabilité de défaillance d'un matériel de résistance R et soumis à une contrainte C. Le matériel est défaillant si C est supérieur à R ( $R-C < 0$ ).



**Résistance (R)** : moyenne  $m_R=16$

écart type :  $\sigma_R=3$

**Contrainte (C)** : moyenne  $m_C=8$

écart type :  $\sigma_C=2$

Le résultat théorique peut être calculé dans le cas de lois normales car la somme de deux variables gaussiennes **indépendantes** est elle-même une variable gaussienne.

Soit la variable aléatoire  $Z=R-C$

$$m_Z = m_R - m_C$$

et

$$\sigma_Z^2 = \sigma_R^2 + \sigma_C^2$$

$$m_Z = 8$$

$$\sigma_Z^2 = 13$$

La probabilité de défaillance est donnée par  $P(Z \leq 0) = 0,013250095$  soit  $R = 0,986749905$

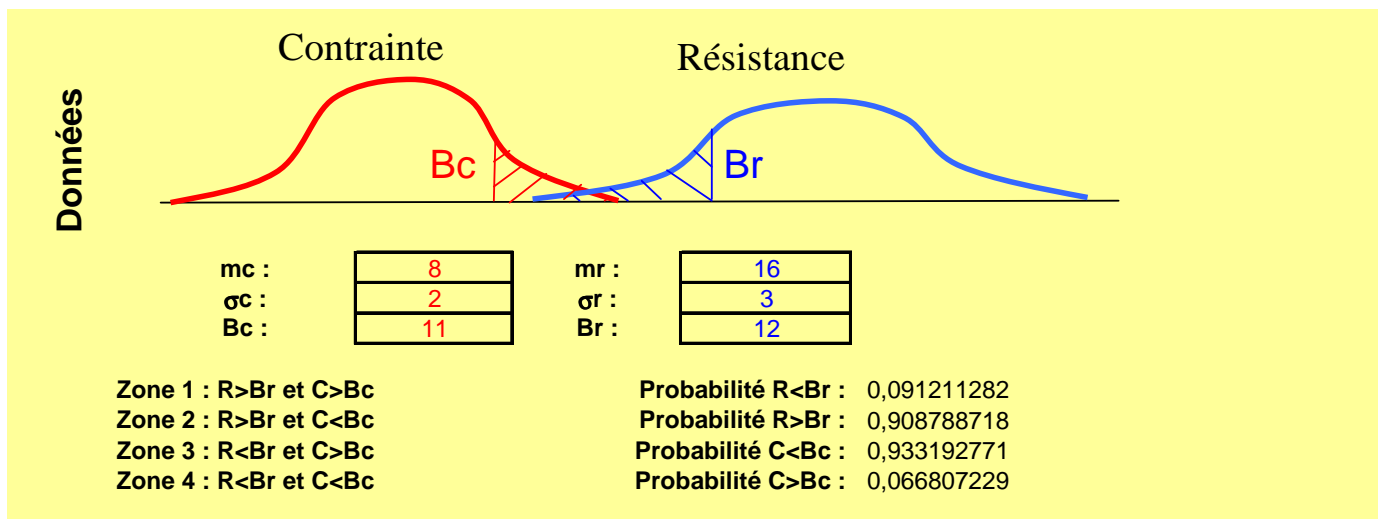
## ii. METHODE DE MONTE CARLO PAR ECHANTILLONNAGE STRATIFIE

Dans le cas du problème de résistance/contrainte l'espace d'état que l'on souhaite étudier plus particulièrement est celui dans lequel on a les plus fortes probabilités de défaillance, c'est-à-dire lorsque la contrainte est importante et la résistance est faible. On peut alors se limiter à deux domaines sur chacun des espaces de départ, et donc à quatre zones différentes.

Le découpage des strates (Br = 12 et Bc = 11) est d'abord choisi approximativement, puis il peut être affiné à partir des résultats de simulation.

Traité par l'outil SIMCAB, les résultats obtenus sont présentés ci-après (le fichier Excel correspondant qui est incrusté dans le document Word peut s'ouvrir par double clic de la souris).

On notera l'utilisation des fonction de tirage des lois de probabilité avec bornes min et max, proposées par l'outil, tel que « L\_Bin(m;σ;;min;max) » dans cet exemple.

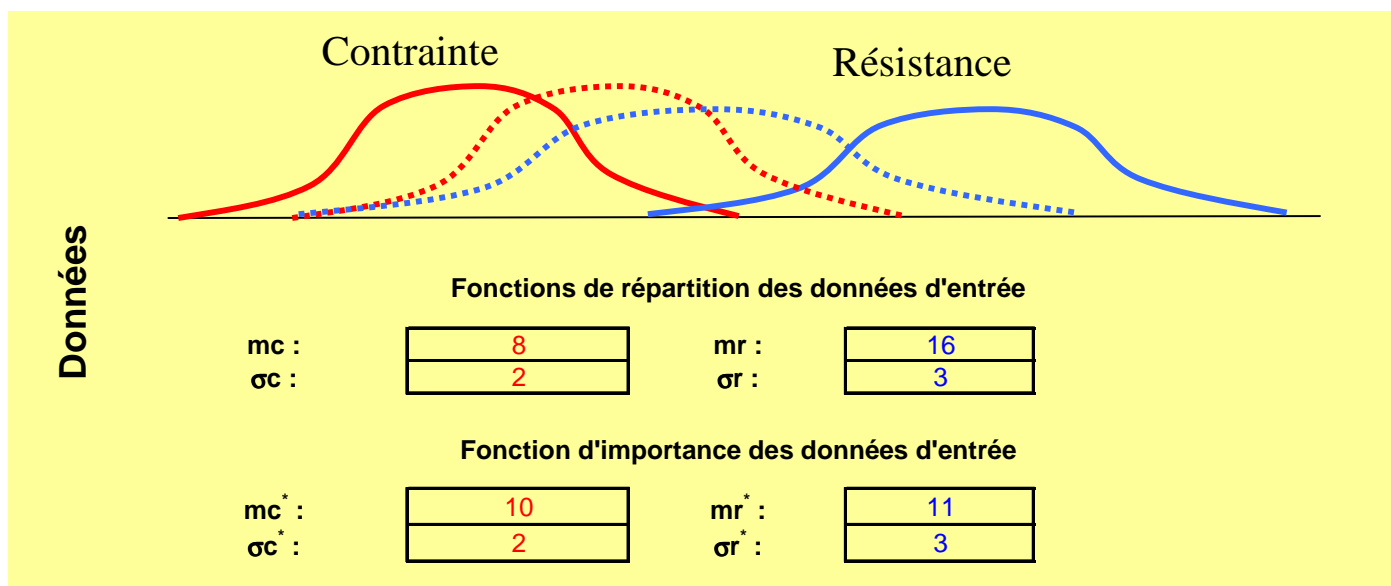


<b>Simulation</b>	<b>Monte Carlo classique</b>		<b>Résistance</b>	<b>Contrainte</b>	<b>Z=R-C</b>	<b>Proba (R&gt;C)</b>	
			12,89810558	8,350435467	4,54767011	1	
	<b>Monte Carlo par échantillonnage stratifié</b>	<b>Zone 1</b>	<b>Résistance</b>	<b>Contrainte</b>	<b>Z=R-C</b>	<b>Probabilité</b>	<b>Pondération</b>
		Zone 2	18,32781927	12,8648272	5,462992069	1	0,060713656
		Zone 3	12,00827291	7,087035496	4,921237413	1	0,848075062
Zone 4		11,71726915	11,0427977	0,674471451	1	0,006093573	
		11,95216833	8,089231867	3,862936468	1	0,085117709	
Nombre de simulations :		<b>1000</b>	Intervalle de confiance à :		<b>90%</b>		

<b>Résultats</b>	<b>Sans réduction de variance</b>	<b>moyenne</b>	<b>Ecart type</b>	<b>Borne inf.</b>	<b>Borne sup.</b>	<b>Delta</b>
		0,984	0,125537882	0,97747017	0,99052983	0,01305966
	<b>Avec réduction de variance</b>	0,986427107	0,026196898	0,98506448	0,987789734	0,002725254

### iii. METHODE DE MONTE CARLO PAR ECHANTILLONNAGE D'IMPORTANCE

Les fonctions d'importance de R et C sont centrées sur les zones de fortes probabilités de défaillance, c'est-à-dire lorsque la contrainte est importante et la résistance est faible (ouverture du fichier Excel sur double clic).



**Simulation**

Monte Carlo classique		Contrainte	Résistance	Z=R-C	Proba (R>C)
		6,413222712	19,45509068	-13,04186797	0

Monte Carlo par échantillonnage d'importance		Contrainte	Résistance	Z=R-C	Proba (R<C)
<i>ESTIMATEUR BIAISE</i>		12,7734598	9,939454895	-2,834004908	1
<i>PONDERATION</i>		Densité vrai	Densité biaisé	Pondération	
		0,00019975	0,009526929	<b>0,020966895</b>	

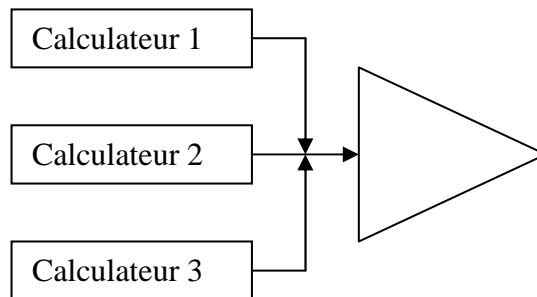
Nombre de simulations : **1000**      Intervalle de confiance à : **90%**

**Résultats**

		moyenne	Ecart type	Borne inf.	Borne sup.
Sans réduction de variance		0,986	0,117549214	0,979885699	0,992114301
Avec réduction de variance		0,985513389	0,024290208	0,984249938	0,986776839

## b. ESTIMATION DE FIABILITE D'UN SYSTEME A VOTE

### i. RESULTAT THEORIQUE



**Calculateur 1 :** MTTF = 1000 h

**Calculateur 2 :** MTTF = 1000 h

**Calculateur 3 :** MTTF = 1000 h

Le système réalise la comparaison des résultats en provenance de chacun des trois calculateurs. Il fonctionne tant que deux des trois calculateurs sont en état de marche. Il est donc équivalent à une redondance active 2 parmi 3.

La fiabilité d'un élément est donnée par la formule suivant :  $r = e^{-\lambda t}$

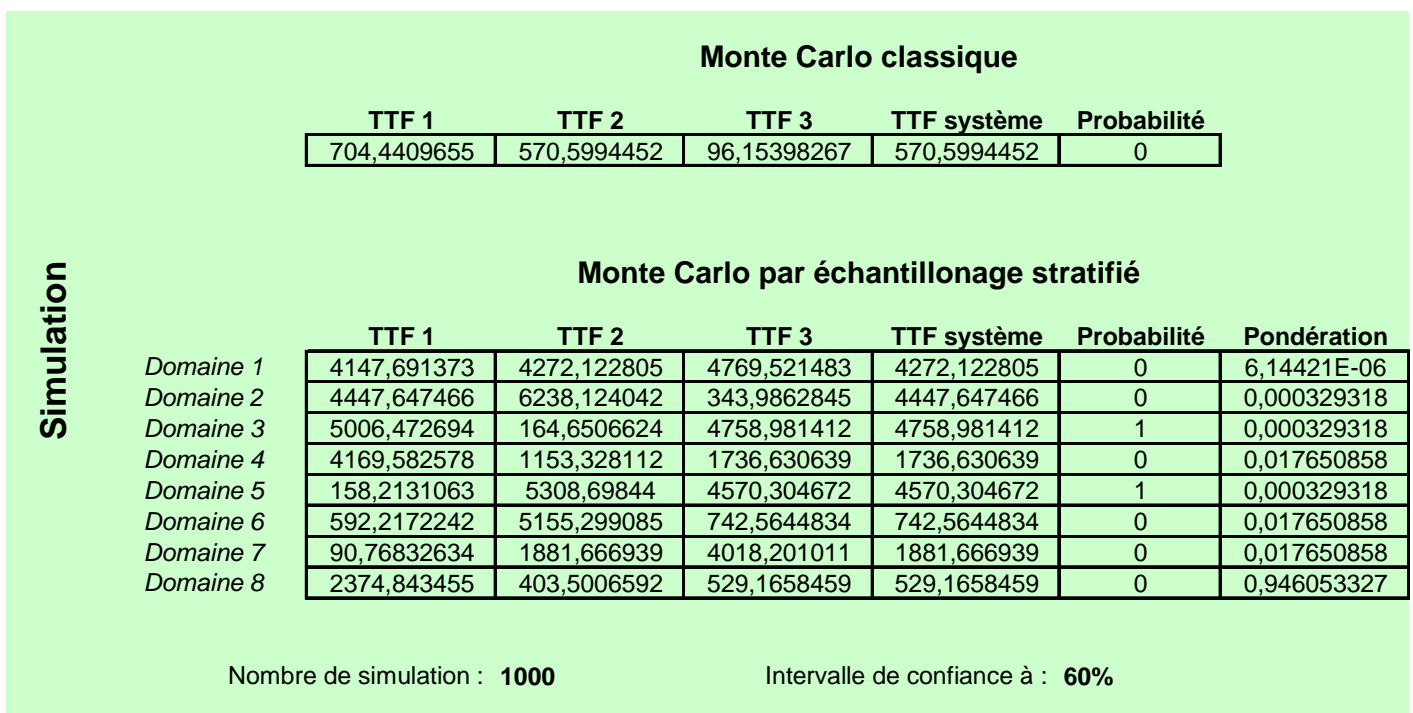
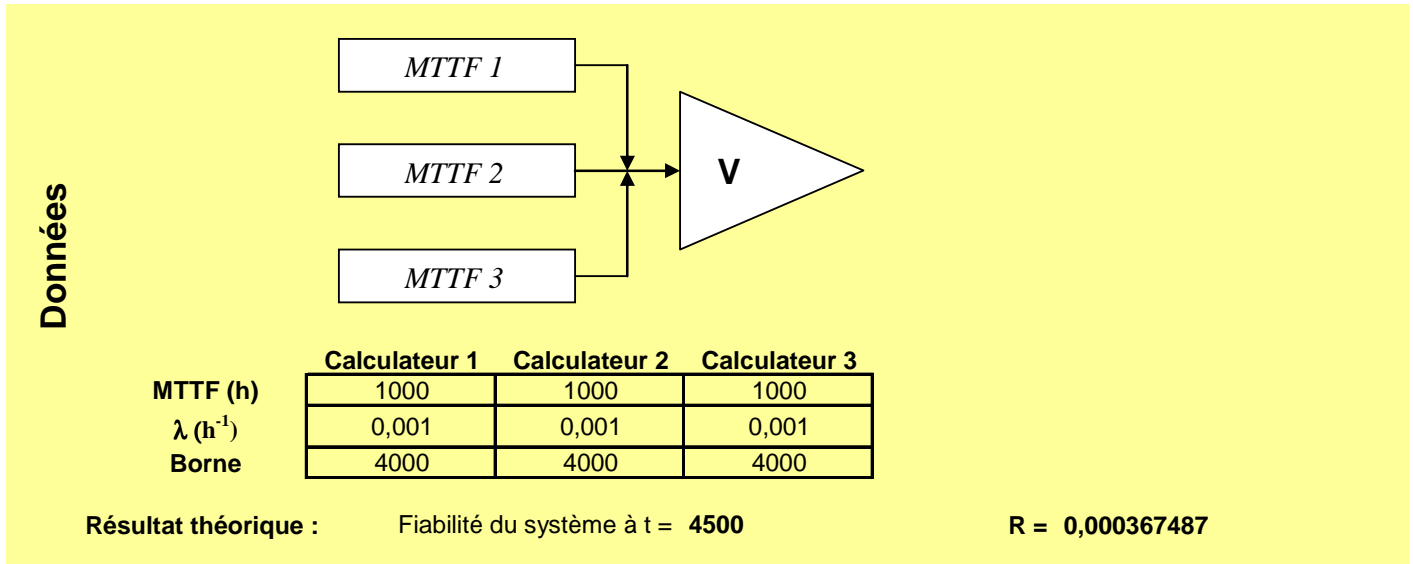
La fiabilité du système est donné par :

$$R = \sum_{i=0}^1 C_3^i (1-r)^i r^{3-i}$$

$$R = r^3 + 3r^2 \cdot (1-r)$$

## ii. METHODE DE MONTE CARLO PAR ECHANTILLONNAGE STRATIFIE

On peut, à nouveau, se limiter à deux domaines sur chacun des espaces de départ (TTF : durée de bon fonctionnement de plus ou moins de 4000 heures par exemple), et donc à 8 domaines différents (ouverture du fichier Excel sur double clic).



**Résultats**

	moyenne	Ecart-type	Borne inf.	Borne sup.	Delta
Sans réduction de la variance	0	0	0	0	0
Avec réduction de la variance	0,000369586	0,000280287	0,000362127	0,000377046	1,49193E-05
Résultat théorique	0,000367487				

La méthode de l'échantillonnage stratifié permet ici d'estimer une probabilité que l'on ne pouvait pas atteindre avec la méthode de Monte-Carlo classique.



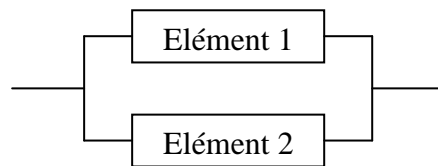
### 3. APPORT A L'ESTIMATION DE DISPONIBILITE

Le problème se complexifie dans la mesure où le système n'a plus un état absorbant (état de panne) mais des états discrets (certains pour lesquels le système est disponible et d'autres pas) qui doivent être simulés pendant toute la durée de la mission.

L'estimation est alors réalisée au moyen d'un outil de simulation des systèmes à états discrets. La technique des modèles de simulation récursive, proposée par l'outil SIMCAB, est utilisée pour traiter l'exemple ci-après.

#### a. REDONDANCE ACTIVE 1 PARI 2

##### i. RESULTAT THEORIQUE



$$\begin{array}{ll} \text{MTTF}_1 = 1000 \text{ h} & \text{MTTR}_1 = 500 \text{ h} \\ \text{MTTF}_2 = 1000 \text{ h} & \text{MTTR}_2 = 500 \text{ h} \end{array}$$

La disponibilité asymptotique d'un élément est donnée par la formule suivante :

$$A_1 = \frac{\text{MTTF}_1}{\text{MTTF}_1 + \text{MTTR}_1}$$

La disponibilité de deux éléments en redondance active est donnée par :  $A = (1 - (1 - A_1)) * (1 - A_2)$

La disponibilité asymptotique de ce système est donc  $A = \frac{8}{9} \approx 0.88889$

## ii. METHODE DE MONTE CARLO PAR ECHANTILLONNAGE STRATIFIE

Le système possède quatre paramètres d'entrée qui ont chacun une influence sur la disponibilité : MTTF1, MTTF2, MTTR1, MTTR2. Aussi apparaît-il judicieux de considérer deux domaines pour chacun d'entre eux, ce qui revient à étudier 16 (2<sup>4</sup>) strates disjointes.

Afin de simplifier cet exemple, l'échantillonnage est réalisé uniquement sur les variables MTTF1 et MTTR1, ce qui limite à 4 le nombre de strates.

		Données		Simulation				Résultats						
				<b>DeltaT</b> 163,4086		<b>Tmission</b> 8000		<b>Taux de confiance : 60%</b>						
				<b>TO</b> 1		<b>Ti</b> 2000		<b>Tj</b> 163,4086		<b>Nombre de simulations : 30</b>				
<b>Monte Carlo classique</b>	<b>MTTF</b>	1000	<b>MTTR</b>	500	<b>TTF</b>	488,589658	<b>TTR</b>	325,181						
		1		1		727,434		564,026						
				<b>Disponibilité système :</b> 1				<b>Disponibilité mission :</b> 1						
								<b>Moyenne</b>		<b>Ecart-type</b>	<b>Min</b>	<b>Max</b>		
								Dispo		0,891172	0,057871	0,88228	0,900065	
<b>Monte Carlo par échantillonnage stratifié</b>	<b>Domaine 1</b>	1000	500	1	0	163,408596	1	36,9197	<i>Probabilité d'avoir TTF &lt; 100 :</i> 0,095163					
	<i>TTF1 &lt; 100</i>				1	1120,7009	1	957,292	<i>Probabilité d'avoir TTR &lt; 60 :</i> 0,698806					
	<i>TTR1 &lt; 600</i>				1	1120,7	1	957,292						
					<b>Disponibilité système 1 :</b> 1				<b>Moyenne</b>		<b>Ecart-type</b>	<b>Min</b>	<b>Max</b>	
									Dispo 1		0,726026	0,07196	0,714969	0,737083
	<b>Domaine 2</b>	1000	500	1	0	438,632202	0	275,224	<i>Probabilité d'avoir TTF &lt; 100 :</i> 0,095163					
	<i>TTF1 &lt; 100</i>				1	269,057052	0	105,648	<i>Probabilité d'avoir TTR &gt; 60 :</i> 0,301194					
	<i>TTR1 &gt; 600</i>				1	269,057	0	105,648						
					<b>Disponibilité système 2 :</b> 0				<b>Moyenne</b>		<b>Ecart-type</b>	<b>Min</b>	<b>Max</b>	
									Dispo 2		0,679354	0,077385	0,667463	0,691245
	<b>Domaine 3</b>	1000	500	1	0	1096,91588	0	933,507	<i>Probabilité d'avoir TTF &gt; 100 :</i> 0,904837					
	<i>TTF1 &gt; 100</i>				1	389,838646	0	226,43	<i>Probabilité d'avoir TTR &gt; 60 :</i> 0,301194					
<i>TTR1 &gt; 600</i>				1	389,839	0	226,43							
				<b>Disponibilité système 3 :</b> 0				<b>Moyenne</b>		<b>Ecart-type</b>	<b>Min</b>	<b>Max</b>		
								Dispo 3		0,840947	0,055593	0,832405	0,849489	
<b>Domaine 4</b>	1000	500	1	1	458,181817	1	294,773	<i>Probabilité d'avoir TTF &gt; 100 :</i> 0,904837						
<i>TTF1 &gt; 100</i>				1	235,082183	1	71,6736	<i>Probabilité d'avoir TTR &lt; 60 :</i> 0,698806						
<i>TTR1 &lt; 600</i>				1	235,082	1	71,6736							
				<b>Disponibilité système 4 :</b> 1				<b>Moyenne</b>		<b>Ecart-type</b>	<b>Min</b>	<b>Max</b>		
								Dispo 4		0,934544	0,037862	0,928727	0,940362	
								<b>Disponibilité mission pondérée :</b> 0,69880579						
								<b>Sans Réduction de variance</b>		0,891172	0,057871	0,88228	0,900065	
								<b>Avec réduction de variance</b>		0,887855	0,029153	0,883376	0,892335	
								<b>Valeur théorique</b>		0,888889				

## **b. GENERALISATION**

Les techniques de réduction de variance peuvent être appliquées à des systèmes de complexité quelconque dans la mesure où les paramètres influents considérés sont en nombre limité pour éviter l'explosion combinatoire ( $2^n$  strates pour  $n$  paramètres ayant chacun 2 domaines).

L'une des difficultés à prendre en compte lors de la modélisation est de s'assurer que le calcul de disponibilité dans chacune des strates ne perturbe pas le reste de la simulation. Aussi, est-il nécessaire de ne considérer **qu'un seul tirage** des temps de défaillance ou de réparation sans tenir compte du caractère markovien du système (ces durées sont décrémentées par la suite à l'occurrence de chacun des nouveaux événements).

De même dans le cas de systèmes comprenant des éléments ayant les mêmes caractéristiques, tel que celui ici traité, il est nécessaire de considérer **chaque variable de façon distincte** lors de la stratification.