Laboratoire 2 : moyenne des lots, variables de contrôle et temps moyen d'attente

Conditions particulières :

- Suivre les recommandations et respecter les contraintes énoncées dans les conditions générales distribuées.
- À rendre au plus tard le lundi 28 février 2005 à 17h.
- L'archive `Labo2.tgz` est disponible sur `http://ina.eivd.ch/collaborateurs/etr/msd`
- Les seuls (ceci implique qu'il n'y en a pas d'autres) fichiers à rendre sont `simulation.c` et `statistiques.c`.

Rappel de théorie des files d'attente :

Intéressons-nous à une file d'attente $A|S|K$, dotée d'un buffer de capacité infinie et de discipline FIFO, où

- les temps inter-arrivées sont i.i.d. selon une loi A,
- on dispose de K serveurs identiques et indépendants,
- les durées de service sont i.i.d. selon une loi S.

Lorsqu'une telle file d'attente est stable, c'est-à-dire lorsque son taux d'occupation

$$\rho = \frac{E(S)}{K\,E(A)}$$

est inférieur à 1.0, elle possède un régime stationnaire et il est possible de définir son temps moyen d'attente T_a. Dans le cas où les durées inter-arrivées sont distribuées selon une loi exponentielle et pour autant qu'il n'y ait qu'un seul serveur, *i.e.* pour une file $M|G|1$, la formule de Pollaczek-Khintchine pour le temps moyen d'attente

$$T_a = \frac{\rho\,E(S)}{1-\rho}\left(\frac{1+C_S^2}{2}\right)$$

nous fournit directement et de manière exacte[1] le résultat cherché. Rappelons que la notation C_S^2 désigne le carré du coefficient de variation du temps de service S

$$C_S^2 = \frac{\mathrm{Var}(S)}{(E(S))^2}.$$

Pour des modèles plus généraux (comme les files $G|G|K$ que nous allons considérer), nous ne connaissons pas de résultat exact et il faut donc nous contenter de formules approximatives et/ou utiliser la simulation.

Objectifs :

- Mettre en œuvre la méthode de la moyenne des lots pour l'estimation par simulation du temps moyen d'attente dans le buffer d'une file d'attente $G|G|K$ stable.
- Tester l'efficacité de la méthode des variables de contrôle dans ce contexte particulier. Choisir comme variable de contrôle (associée au temps d'attente d'une requête donnée) le « nombre d'arrivées durant les $d \cdot E(A)$ unités de temps précédant l'arrivée de la requête dans le système (la nouvelle requête elle-même n'est pas à comptabiliser) » (où $d \in \mathbb{N}^*$ est un paramètre fixé par l'expérimentateur).
- Tester par simulation, en fonction de C_A^2, C_S^2, K et ρ, la qualité de la formule approximative suivante (due à Krämer/Langenbach-Belz) pour le temps moyen d'attente dans le buffer d'une file $G|G|K$ stable :

$$T_a \approx \frac{W\,E(S)}{1-\rho}\left(\frac{C_A^2+C_S^2}{2K}\right)e^{-G},$$

  où W est une approximation de la probabilité de devoir attendre dans le buffer

$$W = \begin{cases} \rho^{\frac{K+1}{2}} & \text{pour } \rho \le 0.7 \\ \frac{\rho+\rho^K}{2} & \text{pour } \rho > 0.7 \end{cases}$$

  et G est un facteur correctif

$$G = \begin{cases} \frac{2(1-\rho)(1-C_A^2)^2}{3W(C_A^2+C_S^2)} & \text{pour } C_A^2 \in [0,1] \\ \frac{(1-\rho)(C_A^2-1)}{C_A^2+4C_S^2} & \text{pour } C_A^2 > 1. \end{cases}$$

[1] Notons qu'une telle expression aurait pu s'avérer précieuse s'il s'agissait de valider un modèle de simulation de files $M|G|1$.

Loi log-normale $\mathcal{L}(\mu, \sigma^2)$:

Dans le cadre de ce laboratoire, des lois log-normales sont utilisées pour la distribution des temps inter-arrivées A et de service S. Une loi log-normale $\mathcal{L}(\mu, \sigma^2)$ est donnée par deux grandeurs : $\mu \in \mathbb{R}$ (paramètre d'échelle) et $\sigma > 0$ (paramètre de forme). Notons au passage que sa fonction de densité est

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{si } x > 0, \\ 0 & \text{sinon.} \end{cases}$$

Comme son nom le laisse supposer, il existe un lien de nature logarithmique entre cette distribution et la loi normale :

$$X \sim \mathcal{L}(\mu, \sigma^2) \qquad \Longleftrightarrow \qquad \ln X \sim \mathcal{N}(\mu, \sigma^2)$$

Du coup, il suffit de prendre l'exponentielle de réalisations d'une $\mathcal{N}(\mu, \sigma^2)$ pour générer des réalisations d'une $\mathcal{L}(\mu, \sigma^2)$ (voir code source). L'espérance et la variance d'une variable aléatoire log-normale $X \sim \mathcal{L}(\mu_X, \sigma_X^2)$ sont

$$E(X) = e^{\mu_X + \frac{\sigma_X^2}{2}} \qquad \text{et} \qquad \mathrm{Var}(X) = e^{2\mu_X + \sigma_X^2}(e^{\sigma_X^2} - 1),$$

si bien que son coefficient de variation ne dépend que de son paramètre de forme σ_X

$$C_X^2 = \frac{\mathrm{Var}(X)}{(E(X))^2} = e^{\sigma_X^2} - 1.$$

Paramétrisation du modèle :

Dans le cadre de ce laboratoire, le modèle pourrait être entièrement spécifié par les paramètres $\mu_A, \sigma_A^2, \mu_S, \sigma_S^2$ associés aux lois $A \sim \mathcal{L}(\mu_A, \sigma_A^2)$ et $S \sim \mathcal{L}(\mu_S, \sigma_S^2)$, ainsi que par la connaissance du nombre de serveurs $K \in \mathbb{N}^*$. Cette paramétrisation n'étant absolument pas pratique pour étudier l'influence des grandeurs C_A^2, C_S^2 et ρ sur la qualité de la formule approximative à considérer, nous allons plutôt procéder comme suit :

- ▷ Les carrés des coefficients de variation $C_A^2, C_S^2 > 0$ sont choisis par l'expérimentateur. Utilisant l'identité $C_X^2 = e^{\sigma_X^2} - 1$ obtenue ci-dessus pour les lois log-normales, on obtient les paramètres

$$\sigma_A^2 = \ln(C_A^2 + 1) \qquad \text{et} \qquad \sigma_S^2 = \ln(C_S^2 + 1).$$

- ▷ Compte tenu du fait que, sans perte de généralité, il est possible de choisir l'unité de temps utilisée, nous fixons arbitrairement le temps moyen inter-arrivées à $e^3 \approx 20.08$. Ceci revient donc à poser

$$E(A) = e^{\mu_A + \frac{\sigma_A^2}{2}} = e^3 \qquad \Longleftrightarrow \qquad \mu_A = 3 - \frac{\sigma_A^2}{2}$$

- ▷ L'expérimentateur choisit également le nombre de serveurs K et le taux d'occupation ρ. Ces grandeurs étant reliées par l'expression

$$E(S) = K\rho E(A) = K\rho\, e^3 = e^{3 + \ln(K\rho)} = e^{\mu_S + \frac{\sigma_S^2}{2}},$$

on en déduit la valeur du dernier paramètre du modèle

$$\mu_S = 3 + \ln(K\rho) - \frac{\sigma_S^2}{2}.$$

En résumé, la syntaxe d'appel du programme est donc

$$\texttt{Labo2}\ C_A^2\ C_S^2\ K\ \rho$$

lorsque l'on ne souhaite pas utiliser la méthode des variables de contrôle et

$$\texttt{Labo2}\ C_A^2\ C_S^2\ K\ \rho\ d$$

si l'on veut appliquer cette méthode pour une certaine valeur du paramètre $d \in \mathbb{N}^*$ (*i.e.* avec une variable de contrôle mesurant le nombre d'arrivées sur une durée $d \cdot E(A) = d \cdot e^3$).

Initialisation, régime transitoire, mesures et critère d'arrêt :

- ▷ Au début de la simulation, le système est vide (état initial).
- ▷ On ne tient pas compte des temps d'attente des 10 000 premières requêtes (*a priori* suffisant pour l'élimination du régime transitoire dans le cadre de ce laboratoire).
- ▷ La première tentative de recherche d'un taille de lots adéquate est effectuée sur les 1 000 observations suivantes.

- ▷ Le seuil de couverture des intervalles de confiance à construire est, comme d'habitude, fixé à 99%.
- ▷ Dans le cadre de ce laboratoire, le critère d'arrêt à utiliser est une mesure relative : on souhaite obtenir une précision à $\pm 5\%$, c'est-à-dire une demi-largeur d'intervalle de confiance inférieure ou égale au vingtième du temps moyen d'attente mesuré.
- ▷ Appliquer la formule en page 132 du cours (celle où l'on précise qu'il faut $(l/L)^2$ fois plus de réalisations) pour le nombre d'observations supplémentaires à générer (de plus, si aucune taille de lot adéquate n'est trouvée, alors doubler la taille de l'échantillon).
- ▷ L'éventuelle variable de contrôle associée à une requête est mesurée lors de son arrivée dans le système.
- ▷ Le temps d'attente d'une requête est mesuré au moment de sa prise en charge par le serveur qui va s'en occuper, juste avant que le service en question ne commence (et non à sa sortie du système).

Travaux à effectuer :

1. Prendre connaissance du code source contenu dans l'archive qui vous est généreusement fournie sur la page web du cours. L'architecture globale du programme, ainsi que la plupart des modules fastidieux (spécification des paramètres, gestion de l'échéancier, du buffer d'événements et de la liste pour les variables de contrôle, boucle principale, générateurs de variables aléatoires, gestion des événements,...) sont déjà codés. Cette archive comprend 16 fichiers (voir les en-têtes et commentaires pour plus de détails) :
   - ▷ `MT19937.c` et `MT19937.h` : le générateur pseudo-aléatoire,
   - ▷ `generateurs.c` et `generateurs.h` : la bibliothèque de générateurs de variables aléatoires,
   - ▷ `echeancier.c` et `echeancier.h` : implémentation de l'échéancier et spécification des événements de la simulation,
   - ▷ `buffer.c` et `buffer.h` : permet de stocker dans un buffer (FIFO) des événements (forcément de type arrivée) correspondant à des requêtes devant être mises en attente le temps qu'un serveur se libère,
   - ▷ `liste.c` et `liste.h` : permet de stocker dans une liste (FIFO) des dates d'arrivée afin de pouvoir facilement obtenir la variable de contrôle associée à une requête,
   - ▷ `simulation.c` et `simulation.h` : gestion des événements de la simulation et génération de la réalisation suivante (il n'y manque que la gestion des variables de contrôle),
   - ▷ `statistiques.c` et `statistiques.h` : analyse statistique des résultats par la méthode de la moyenne des lots (à réaliser),
   - ▷ `Labo2.c` : programme principal permettant de lancer la construction d'un intervalle de confiance pour le temps moyen d'attente dans une file $G|G|K$ et de comparer ce résultat à la formule approximative de Krämer/Langenbach-Belz,
   - ▷ `Makefile` : pour compiler en paix avec la commande `make`.
2. Coder le module `statistiques.c` compte tenu des spécifications ci-dessus (et des en-têtes des 4 fonctions définies dans `statistiques.h`). Une méthode de moyenne des lots devant être utilisée, il n'est pas possible de faire appel à des estimateurs itératifs : les observations doivent être stockées (en mémoire ou sur fichier). Le passage par fichier étant à éviter (pour des questions de performance), il convient d'utiliser l'allocation dynamique de mémoire pour stocker les observations obtenues par simulation.
3. Dans le module `simulation.c`, la fonction `simulation_prochaine_observation_avec_var_controle()` reste à coder. Pour ce faire, il suffit de recopier la fonction `simulation_prochaine_observation()` et de lui ajouter les éléments nécessaires à la prise en considération de la variable de contrôle définie plus haut. Compte tenu du fait que la variable de contrôle et le temps d'attente d'une requête ne sont pas mesurés simultanément (la variable de contrôle est mesurée au moment de l'arrivée de la requête, alors que le temps d'attente n'est connu qu'au moment de sa prise en charge par un serveur, c'est-à-dire après un éventuel passage par le buffer), la variable de contrôle associée à une requête doit pouvoir être stockée quelque part entre-temps. Pour une requête représentée par un événement `toto` (forcément de type arrivée), cette grandeur est à stocker dans le champ `toto.nombre_requetes`.

   Cette intégration des variables de contrôle dans la gestion des événements doit bien évidemment être expliquée dans votre rapport. On y trouvera également une discussion de la variable de contrôle elle-même (en quoi est-ce un choix *a priori* raisonnable ou non ?).

   Au niveau statistique, rappelons que la méthode des variables de contrôle nécessite des estimateurs de la variance (de la variable de contrôle) et de la covariance. Contrairement à l'exemple de la page 115 du cours, nous utiliserons des estimateurs empiriques itératifs (il est inutile de stocker tout l'historique à ce niveau-là). Nous référant à la page 111 du cours, lorsqu'il s'agira de retourner la réalisation $z = x + c(y - m_Y)$, soulignons le fait que la vraie espérance m_Y de la variable de contrôle Y est à utiliser (et non l'estimateur empirique itératif correspondant qui est forcément moins précis).

4. Tester l'efficacité de la méthode des variables de contrôle en fonction de

$$C_A^2 \in \{0.05, 0.5, 1.0, 2.0\} \qquad \text{et} \qquad \rho \in \{0.5, 0.75, 0.9\},$$

tout en fixant

$$C_S^2 = 1 \qquad \text{et} \qquad K = 2.$$

Plus précisément, pour les 12 combinaisons de jeux de paramètres ci-dessus, comparer le nombre de requêtes nécessaires à l'obtention d'un intervalle de confiance adéquat

- ▷ sans utiliser la méthode des variables de contrôle,
- ▷ en utilisant la méthode des variables de contrôle avec la meilleure valeur possible de $d \in \{1, 2, \ldots, 100\}$.

Discuter le gain observé, ainsi que les valeurs de d obtenues en fonction de C_A^2 et ρ.

5. Tester la qualité de la formule approximative de Krämer/Langenbach-Belz par simulation, pour des lois A et S log-normales et pour les combinaisons des paramètres suivants :

$$C_A^2 \in \{0.05, 0.5, 1.0, 2.0\}, \qquad C_S^2 \in \{0.05, 0.5, 1.0, 2.0\},$$

$$K \in \{1, 2, 4, 10\} \qquad \text{et} \qquad \rho \in \{0.25, 0.5, 0.75, 0.9\}.$$

Discuter l'ordre de grandeur des erreurs obtenues (avec la formule approximative de Krämer/Langenbach-Belz) et déterminer pour quelles plages de valeurs des paramètres la formule est raisonnablement correcte ou au contraire mauvaise. En d'autres termes, comment conseilleriez-vous un utilisateur de la formule ne disposant pas de votre outil de simulation ?

Remarque :

Pour ceux qui n'y seraient pas habitués, signalons qu'il est parfois pratique (globalement l'exécution des simulations à effectuer pour ce laboratoire peut prendre plusieurs heures) d'utiliser des scripts pour exécuter des séries de simulations. Pour fixer les idées, un exemple de script (mettre tout ce qui suit dans un `fichier`, le rendre exécutable avec `chmod u+x fichier` et l'exécuter avec `./fichier`) pourrait être

```
#!/bin/sh

for ca2 in 0.05 0.5 1.0 2.0
  do
  for rho in 0.5 0.75 0.9
    do
    Labo2 $ca2 1.0 2 $rho >> Resultats
  done
done
```

Dans cet exemple, la sortie est redirigée vers le fichier `Resultats`.