



HAL
open science

Les processus d'attente dans les files d'attentes multi-serveurs saturées

Richard Loulou

► **To cite this version:**

Richard Loulou. Les processus d'attente dans les files d'attentes multi-serveurs saturées. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1978. tel-00287891

HAL Id: tel-00287891

<https://theses.hal.science/tel-00287891>

Submitted on 13 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée à

**Université Scientifique et Médicale de Grenoble
Institut National Polytechnique de Grenoble**

pour obtenir le grade de

**DOCTEUR ES SCIENCES
MATHEMATIQUES**

par

Richard LOULOU



**LES PROCESSUS D'ATTENTE DANS LES FILES D'ATTENT
MULTI - SERVEURS SATUREES**



Thèse soutenue le 9 mai 1978 devant la Commission d'Examen :

Président : B. VAN CUTSEM

Examineurs : D. DACUNHA CASTELLE

B. MAISONNEUVE

M. VERAN

M. SAKAROVITCH

UNIVERSITE SCIENTIFIQUE
ET MEDICALE DE GRENOBLE

Monsieur Gabriel CAU : Président

Monsieur Pierre JULLIEN : Vice Président

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

PROFESSEURS TITULAIRES

MM.	AMBLARD Pierre	Clinique de dermatologie
	ARNAUD Paul	Chimie
	ARVIEU Robert	I.S.N.
	AUBERT Guy	Physique
	AYANT Yves	Physique approfondie
Mme.	BARBIER Marie-Jeanne	Electrochimie
MM.	BARBIER Jean-Claude	Physique expérimentale
	BARBIER Reynold	Géologie appliquée
	BARJON Robert	Physique nucléaire
	BARNOU Fernand	Biosynthèse de la cellulose
	BARRA Jean-René	Statistiques
	BARRIE Joseph	Clinique chirurgicale
	BEAUDOING André	Clinique de pédiatrie et puériculture
	BELORIZKY Elie	Physique
	BERNARD Alain	Mathématiques pures
Mme.	BERTRANDIAS Françoise	Mathématiques pures
MM.	BERTRANDIAS Jean-Paul	Mathématiques pures
	BEZEZ Henri	Pathologie chirurgicale
	BLAMBERT Maurice	Mathématiques pures
	BOLLIET Louis	Informatique (IUT B)
	BONNET Jean-Louis	Clinique ophtalmologique
	BONNET-EYMARD Joseph	Clinique gastro-entérologique
Mme.	BONNIER Marie-Jeanne	Chimie générale
MM.	BOUCHERLE André	Chimie et toxicologie
	BOUCHEZ Robert	Physique nucléaire
	BOUSSARD Jean-Claude	Mathématiques appliquées
	BOUTET DE MONVEL Louis	Mathématiques pures
	BRAVARD Yves	Géographie
	CABANEL Guy	Clinique rhumatologique et hydrologique
	CALAS François	Anatomie
	CARLIER Georges	Biologie végétale
	CARRAZ Gilbert	Biologie animale et pharmacodynamie
	CAU Gabriel	Médecine légale et toxicologie
	CAUQUIS Georges	Chimie organique
	CHABAUTY Claude	Mathématiques pures
	CHARACHON Robert	Clinique oto-rhino-laryngologique
	CHATEAU Robert	Clinique de neurologie
	CHIBON Pierre	Biologie animale
	COEUR André	Pharmacie chimique et chimie analytique
	CONTAMTIN Robert	Clinique gynécologique
	COUDERC Pierre	Anatomie pathologique

Mme.	DEBELMAS Anne-Marie	Matière médicale
MM.	DEBELMAS Jacques	Géologie générale
	DEGRANGE Charles	Zoologie
	DELORMAS Pierre	Pneumophtisiologie
	DEPORTES Charles	Chimie minérale
	DESRE Pierre	Métallurgie
	DESSAUX Georges	Physiologie animale
	DODU Jacques	Mécanique appliquée (IUT I)
	DOLIQUE Jean-Michel	Physique des plasmas
	DREYFUS Bernard	Thermodynamique
	DUCROS Pierre	Cristallographie
	GAGNAIRE Didier	Chimie physique
	GALVANI Octave	Mathématiques pures
	GASTINEL Noël	Analyse numérique
	GAVEND Michel	Pharmacologie
	GEINDRE Michel	Electroradiologie
	GERBER Robert	Mathématiques pures
	GERMAIN Jean-Pierre	Mécanique
	GIRAUD Pierre	Géologie
	JANIN Bernard	Géographie
	KAHANE André	Physique générale
	KOSZUL Jean-Louis	Mathématiques pures
	KLEIN Joseph	Mathématiques pures
	KRAVICHENKO Julien	Mécanique
	KUNTZMANN Jean	Mathématiques appliquées
	LACAZE Albert	Thermodynamique
	LACHARME Jean	Biologie végétale
Mme.	LAJZEROWICZ Janine	Physique
MM.	LAJZEROWICZ Joseph	Physique
	LATREILLE René	Chirurgie générale
	LATURAZE Jean	Biochimie pharmaceutique
	LAURENT Pierre-Jean	Mathématiques Appliquées
	LEDRU Jean	Clinique médicale B
	LE ROY Philippe	Mécanique (IUT I)
	LLIBOUTRY Louis	Géophysique
	LOISEAUX Pierre	Sciences nucléaires
	LONGEQUEUE Jean-Pierre	Physique nucléaire
	LOUP Jean	Géographie
Melle	LUTZ Elisabeth	Mathématiques pures
MM.	MALINAS Yves	Clinique obstétricale
	MARTIN-NOEL Pierre	Clinique cardiologique
	MAZARE Yves	Clinique médicale A
	MICHEL Robert	Minéralogie et pétrographie
	MICOUD Max	Clinique maladies infectieuses
	MOURIQUAND Claude	Histologie
	MOUSSA André	Chimie nucléaire
	NOZIERES Philippe	Spectrométrie physique
	OZENDA Paul	Botanique
	PAYAN Jean-Jacques	Mathématiques pures
	PEBAY-PEYROULA Jean-Claude	Physique
	PERRET Jean	Semeiologie médicale (Neurologie)
	RASSAT André	Chimie systématique
	RENARD Michel	Thermodynamique
	REVOL Michel	Urologie
	RINALDI Renaud	Physique
	DE ROUGEMONT Jacques	Neuro-chirurgie
	SEIGNEURIN Raymond	Microbiologie et Hygiène
	SENGEL Philippe	Zoologie
	SIBILLE Robert	Construction mécanique (IUT I)

MM.	SOUTIF Michel	Physique générale
	TANCHE Maurice	Physiologie
	TRAYNARD Philippe	Chimie générale
	VAILLANT François	Zoologie
	VALENTIN Jacques	Physique nucléaire
	VAUQUOIS Bernard	Calcul électronique
Mme.	VERAIN Alice	Pharmacie galénique
MM.	VERAIN André	Physique
	VEYRET Paul	Géographie
	VIGNAIS Pierre	Biochimie médicale

PROFESSEURS ASSOCIES

MM.	CRABBE Pierre	CERMO
	DEMBICKI Eugéniuz	Mécanique
	JOHNSON Thomas	Mathématiques appliquées
	PENNEY Thomas	Physique

PROFESSEURS SANS CHAIRE

Melle	AGNIUS-DELORD Claudine	Physique pharmaceutique
	ALARY Josette	Chimie analytique
MM.	AMBROISE-THOMAS Pierre	Parasitologie
	ARMAND Gilbert	Géographie
	BENZAKEN Claude	Mathématiques appliquées
	BIAREZ Jean-Pierre	Mécanique
	BILLET Jean	Géographie
	BOUCHET Yves	Anatomie
	BRUGEL Lucien	Energétique (IUT I)
	BUISSON René	Physique (IUT I)
	BUTEL Jean	Orthopédie
	COHEN ADDAD Pierre	Spectrométrie physique
	COLOMB Maurice	Biochimie
	CONTE René	Physique (IUT I)
	DELOBEL Claude	M.I.A.G.
	DEPASSEL Roger	Mécanique des fluides
	FONTAINE Jean-Marc	Mathématiques pures
	GAUTRON René	Chimie
	GIDON Paul	Géologie et minéralogie
	GLENAT René	Chimie organique
	GROULADE Joseph	Biologie médicale
	HACQUES Gérard	Calcul numérique
	HOLLARD Daniel	Hématologie
	HUGONOT Robert	Hygiène et médecine préventive
	IDELMAN Simon	Physiologie animale
	JOLY Jean-René	Mathématiques pures
	JULLIEN Pierre	Mathématiques appliquées
Mme.	KAHANE Josette	Physique
MM.	KRAKOWIACK Sacha	Mathématiques appliquées
	KUHN Gérard	Physique (IUT I)
	LUU DUC Cuong	Chimie organique
	MAYNARD Roger	Physique du solide
Mme.	MINIER Colette	Physique (IUT I)
MM.	PELMONT Jean	Biochimie
	PERRIAUX Jean-Jacques	Géologie et minéralogie
	PFISTER Jean-Claude	Physique du solide
Melle	PIERY Yvette	Physiologie animale

MM.	RAYNAUD Hervé	M.I.A.G.
	REBECQ Jacques	Biologie (CUS)
	REYMOND Jean-Charles	Chirurgie générale
	RICHARD Lucien	Biologie végétale
Mme.	RINAUDO Marguerite	Chimie macromoléculaire
MM.	ROBERT André	Chimie papetière
	SARRAZIN Roger	Anatomie et chirurgie
	SARROT-REYNAULD Jean	Géologie
	SIROT Louis	Chirurgie générale
Mme.	SOUTIF Jeanne	Physique générale
MM.	STIEGLITZ Paul	Anesthésiologie
	VIALON Pierre	Géologie
	VAN CUTSEM Bernard	Mathématiques appliquées

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

MM.	ARMAND Yves	Chimie (IUT I)
	BACHELOT Yvan	Endocrinologie
	BARGE Michel	Neuro-chirurgie
	BEGUIN Claude	Chimie organique
Mme	BERIEL Hélène	Pharmacodynamie
MM.	BOST Michel	Pédiatrie
	BOUCHARLAT Jacques	Psychiatrie adultes
Mme.	BOUCHE Liane	Mathématiques (CUS)
MM.	BRODEAU François	Mathématiques (IUT B) (Personne étrangère habilitée à être directeur de thèse)
	CHAMBAZ Edmond	Biochimie médicale
	CHAMPETIER Jean	Anatomie et organogénèse
	CHARDON Michel	Géographie
	CHERADAME Hervé	Chimie papetière
	CHIAVERINA Jean	Biologie appliquée (EFP)
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculaire
	CORDONNIER Daniel	Néphrologie
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie
	CYROT Michel	Physique du solide
	DENIS Bernard	Cardiologie
	DOUCE Roland	Physiologie végétale
	DUSSAUD René	Mathématiques (CUS)
Mme.	ETERRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine légale
	FAURE Gilbert	Urologie
	GAUTIER Robert	Chirurgie générale
	GIDON Maurice	Géologie
	GROS Yves	Physique (IUT I)
	GUIGNIER Michel	Thérapeutique
	GUITTON Jacques	Chimie
	HICTER Pierre	Chimie
	JALBERT Pierre	Histologie
	JULIEN-LAVILLAVROY Claude	O.R.L.
	KOLODIE Lucien	Hématologie
	LE NOC Pierre	Bactériologie-virologie
	MACHE Régis	Physiologie végétale
	MAGNIN Robert	Hygiène et médecine préventive
	MALLION Jean-Michel	Médecine du travail
	MARECHAL Jean	Mécanique (IUT I)
	MARTIN-BOUYER Michel	Chimie (CUS)
	MICHOULIER Jean	Physique (IUT I)

MM.	NEGRE Robert	Mécanique (IUT I)
	NEMOZ Alain	Thermodynamique
	NOUGARET Marcel	Automatique (IUT I)
	PARAMELLE Bernard	Pneumologie
	PECCOUD François	Analyse (IUT B) (Personnalité étrangère habilité à être directeur de thèse)
	PEFFEN René	Métallurgie (IUT I)
	PERRIER Guy	Géophysique-Glaciologie
	PHELIP Xavier	Rhumatologie
	RACHAIL Michel	Médecine interne
	RACINET Claude	Gynécologie et obstétrique
	RAMBAUD André	Hygiène et hydrologie (Pharmacie)
	RAMBAUD Pierre	Pédiatrie
	RAPHAEL Bernard	Stomatologie
Mme.	RENAUDET Jacqueline	Bactériologie (Pharmacie)
MM.	ROBERT Jean-Bernard	Chimie physique
	Romier Guy	Mathématiques (IUT B) (Personnalité étrangère habilité à être directeur de thèse)
	SCHAERER René	Cancérologie
	SHOM Jean-Claude	Chimie générale
	STOEBNER Pierre	Anatomie pathologie
	VROUSOS Constantin	Radiologie

MAITRES DE CONFERENCES ASSOCIES

MM.	DEVINE Roderick	Spectro physique
	HODGES Christopher	Transition de phases

Fait à SAINT MARTIN D'HERES, NOVEMBRE 1976.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Monsieur Philippe TRAYNARD : Président

Monsieur Pierre-Jean LAURENT : Vice Président

PROFESSEURS TITULAIRES

MM.	BENOIT Jean	Radioélectricité
	BESSON Jean	Electrochimie
	BLOCH Daniel	Physique du solide
	BONNETAIN Lucien	Chimie minérale
	BONNIER Etienne	Electrochimie et électrometallurgie
	BOUDOURIS Georges	Radioélectricité
	BRISSENEAU Pierre	Physique du solide
	BUYLE-BODIN Maurice	Electronique
	COUMES André	Radioélectricité
	DURAND Francis	Métallurgie
	FELICI Noël	Electrostatique
	FOULARD Claude	Automatique
	LESPINARD Georges	Mécanique
	MOREAU René	Mécanique
	PARIAUD Jean-Charles	Chimie-Physique
	PAUTHENET René	Physique du solide
	PERRET René	Servomécanismes
	POLOUJADOFF Michel	Electrotechnique
	SILBER Robert	Mécanique des fluides

PROFESSEUR ASSOCIE

M.	ROUXEL Roland	Automatique
----	---------------	-------------

PROFESSEURS SANS CHAIRE

MM.	BLIMAN Samuel	Electronique
	BOUVARD Maurice	Génie mécanique
	COHEN Joseph	Electrotechnique
	LACOUME Jean-Louis	Géophysique
	LANCIA Roland	Electronique
	ROBERT François	Analyse Numérique
	VEILLON Gérard	Informatique fondamentale et appliquée
	ZADWORNÝ François	Electronique

MAITRES DE CONFERENCES

MM.	ANCEAU François	Mathématiques appliquées
	CHARTIER Germain	Electronique
	GUYOT Pierre	Chimie minérale
	IVANES Marcel	Electrotechnique
	JOUBERT Jean-Claude	Physique du solide
	MORET Roger	Electrotechnique nucléaire
	PIERRARD Jean-Marie	Mécanique
	SABONNADIÈRE Jean-Claude	Informatique fondamentale et appliquée
Mme.	SAUCIER Gabrièle	Informatique fondamentale et appliquée

MAITRE DE CONFERENCES ASSOCIE

M.	LANDAU Ioan	Automatique
----	-------------	-------------

CHERCHEURS DU C.N.R.S. (Directeur et Maîtres de Recherche)

MM.	FRUCHART Robert	Directeur de Recherche
	ANSARA Ibrahim	Maître de Recherche
	CARRE René	Maître de Recherche
	DRIOLE Jean	Maître de Recherche
	MATHIEU Jean-Claude	Maître de Recherche
	MUNIER Jacques	Maître de Recherche

REMERCIEMENTS

Je suis très sensible à l'honneur que me font Messieurs Van Cutsem et Dacunha Castelle, le premier en présidant le jury de cette thèse, le second en venant de Paris pour y participer.

Je remercie Messieurs Maisonneuve et Vêran d'avoir bien voulu faire partie de ce jury.

Je remercie Monsieur Sakarovitch et les membres de l'équipe de Recherche Opérationnelle de m'avoir chaleureusement accueilli dans cette équipe pendant l'année 1976-77 et d'avoir facilité la rédaction de ce travail.

TABLE DES MATIERES

<u>CHAPITRE LIMINAIRE</u> - INTRODUCTION ET RESUME	1
0.1. Introduction	1
0.2. Résumé de la thèse	2
<u>CHAPITRE I</u> : LA CONVERGENCE FAIBLE DES PROCESSUS STOCHASTIQUES	4
1.1. Introduction	4
1.2. Convergence Faible des éléments aléatoires sur un espace métrique	5
1.3. Caractérisation de la convergence faible	7
<u>CHAPITRE II</u> : LES FILES D'ATTENTE : VUE GENERALE	9
2.1. Généralités	9
2.2. Travaux antérieurs sur les systèmes saturés	10
<u>CHAPITRE III</u> : THEOREMES DE CONVERGENCE POUR LES SYSTEMES GI/G/1 SATURES	12
3.1. Introduction	12
3.2. Convergence des processus d'attente virtuelle	12
3.3. Convergence faible des processus d'attente réelle	16
3.4. Etude des suites de systèmes GI/G/1	17
3.5. Utilisation pratique des théorèmes de convergence	20
<u>CHAPITRE IV</u> : LES SYSTEMES A s SERVEURS GI/G/s	22
4.1. Introduction	22
4.2. Le système auxiliaire	23
4.3. Etude du système GI/G/s	24
4.4. Les suites de systèmes GI/G/s	26

<u>CHAPITRE V</u> : EXTENSIONS ET COMPLEMENTS	27
5.1. Introduction	27
5.2. Généralisations quasi-stationnaires du processus d'arrivée.	27
5.3. Un modèle non classique d'arrivées transitoires à s serveurs.	30
5.4. Quelques autres processus importants	34
 <u>CHAPITRE VI</u> : LA PERIODE D'ACTIVITE	 36
6.1. Introduction	36
6.2. Une borne supérieure pour E B dans le système GI/G/1	38
6.3. Une conjecture sur E B dans GI/G/s	40
 <u>CHAPITRE VII</u> : QUELQUES CONSIDERATIONS PRATIQUES	 41
7.1. Introduction	41
7.2. Bornes théoriques sur les vitesses de convergence dans les files d'attente	42
7.3. Considérations générales sur la simulation des systèmes saturés	43
7.4. Résultats empiriques pour les systèmes GI/G/s	44
7.5. Simulation du système transitoire de la section 5.3	46
7.6. Conclusion	46
 <u>BIBLIOGRAPHIE</u>	 47
 <u>ANNEXE I</u> (L4)	 49
 <u>ANNEXE II</u> (L5)	 66

0.1. Introduction

Cette thèse est le résultat de plusieurs travaux antérieurs dont certains ont été publiés, d'autres ont paru sous forme de rapports internes, et d'autres enfin sont actuellement soumis pour fins de publication. Dans la présente synthèse, nous ferons référence à six travaux, nommés par ordre chronologique comme suit:

- (L1) : "Weak convergence for Multi-Channel Queues in Heavy Traffic"
Rapport interne du centre de recherche opérationnelle,
Université de Californie, Berkeley. O.R.C. 71-31 (1971)
- (L2) : "Multi-Channel Queues in Heavy-Traffic" .Journal of Applied Probability ,vol.10,pp.769-777 (1973)
- (L3) : "On the extension of Congestion Theorems to Multi-Channel Systems". Mathematical Methods in Queueing Theory,pp.185-197
Springer-Verlag, New-York (1974)
- (L4)⁽¹⁾ : "An analytical Model for peak-hour transient multi-channel queues". Soumis à Naval Research Logistics Quarterly (1976)
- (L5)⁽⁰⁾ : "An explicit upper bound for the Mean Busy Period in a G1/G/1 Queue"
A paraître dans Journal of Applied Probability, Juin 1978.
De plus, les simulations rapportées au chapitre VII ont été effectuées expressément pour cette thèse.

L'objectif de cette synthèse est donc de présenter sous une forme concise et logique, mais pas toujours détaillée, les différents résultats contenus et rigoureusement démontrés dans ces cinq travaux.

L'ensemble de ces travaux traite de l'étude de systèmes de service (files d'attente) généraux dans des conditions de surcharge (saturation) entraînant, en général, la non-stationnarité des processus étudiés. Les travaux (L1) à (L4) contiennent des théorèmes sur la convergence de ces processus, alors que (L5) consiste en l'obtention d'une borne supérieure.

Des phénomènes d'attente se manifestent très fréquemment dans le fonctionnement des systèmes hommes-machines de notre civilisation. Si l'étude de ces phénomènes est donc essentiellement motivée par des soucis d'ordre pratique, il n'en est pas moins vrai que les méthodes utilisées pour ce faire font souvent appel à des outils mathématiques relativement complexes, en particulier au calcul des probabilités. Cette thèse reflète cette double référence au pratique et au théorique: en effet, nous étudions des modèles assez généraux pour être applicables

(1) Cf. Annexes I et II

à de nombreuses situations, et, souvent, nos modèles sont dictés par des considérations d'ordre pratique (et non mathématique): tous les modèles du chapitre 5 le sont. D'autre part, il est vrai que l'essentiel de notre recherche est mathématique, puisqu'elle aboutit à des théorèmes ou à des bornes supérieures. Nous nous sommes donc efforcé de rendre évident le lien entre le résultat mathématique obtenu et l'utilisation pratique qu'il est possible d'en faire, lorsque ce lien est obscurci par le formalisme employé.

0.2. Résumé de la thèse

Le chapitre 1 est un bref résumé de la théorie de la convergence faible des processus stochastiques. Ce mode de convergence est utilisé systématiquement dans cette thèse en raison de sa plus grande généralité. Ce chapitre ne contient aucun résultat original de l'auteur.

Le chapitre 2 rappelle les définitions et les résultats essentiels relatifs aux modèles classiques de files d'attente, avec une référence toute particulière aux conditions de saturation. Il ne contient que des résultats connus.

Au chapitre 3 nous établissons plusieurs théorèmes de convergence faible pour des suites de processus construits sur des quantités jugées importantes dans une file d'attente à un seul serveur. Les limites obtenues sont explicites et sont toutes des processus gaussiens assez simples, et ceci pour des distributions générales des arrivées et des services (modèle GI/G/1). Les résultats de ce chapitre ont déjà été montrés par Iglehart et Whitt (1969), mais nos méthodes de démonstration sont différentes, plus simples, et, fait capital, elle peuvent être et seront généralisées aux systèmes multi-serveurs. Dans ce même chapitre, nous étudions aussi les suites de systèmes GI/G/1, ce qui ajoute à la flexibilité et la généralité de nos résultats; ce chapitre est basé sur (L1) et (L2).

Le chapitre 4 est une généralisation systématique des théorèmes

CHAPITRE I

LA CONVERGENCE FAIBLE DES PROCESSUS STOCHASTIQUES

1.1. Introduction

Pour justifier ce mode de convergence, prenons l'exemple du temps d'attente du $n^{\text{ième}}$ client, W_n , dans la file d'attente GI/G/1, et en supposant la saturation (par exemple $\rho > 1$). On pourrait, dans ce cas précis, montrer la convergence en distribution de la suite de variables aléatoires $(\frac{W_n - a_n}{b_n})$ vers une variable normale $N(0,1)$,

par un choix judicieux des constantes a et b . Au lieu de cela, nous allons plutôt construire une suite de processus stochastiques (A_n) définie comme suit :

$$A_n(t) = \frac{W_{[nt]} - a_n t}{b \sqrt{n}}, \quad 0 \leq t \leq 1$$

(où $[nt]$ est la partie entière de nt).

Cette construction qui semble à première vue gratuite et artificielle, crée une suite de processus (A_n) à paramètre dans $[0,1]$, dont on peut noter au passage que les réalisations sont presque sûrement des suites de fonctions réelles sur $[0,1]$, semi-continues à droite. L'espace de telles fonctions est appelé $D_{[0,1]}$ ou plus simplement D , et sera muni d'une distance d_0 qui le rendra métrique, complet et séparable. Ceci dit, pourquoi compliquer ainsi le problème et ne pas se contenter de la convergence des variables aléatoires $A_n(1)$? La réponse est double :

(a) Il s'avèrera aussi simple de s'attaquer à la convergence faible des processus A_n qu'à celle des simples variables aléatoires $A_n(1)$, et ce sans hypothèse supplémentaire dans la plupart des cas

(b) La convergence des processus A_n entraînera (grâce au théorème 1.1 ci-dessous) celle de toute application continue $f(A_n)$, ce qui permettra d'obtenir de nombreux corollaires sans effort supplémentaire.

1.2. Convergence Faible des éléments aléatoires sur un espace métrique

Définition 1.1

Soit S un espace métrique muni de la distance ρ , et soit \mathcal{B} la σ -algèbre engendrée par les ouverts de S . Soit (Ω, \mathcal{A}, P) un espace de probabilité. On appelle élément aléatoire de S toute application mesurable X :

$$X : (\Omega, \mathcal{A}, P) \longrightarrow (S, \mathcal{B})$$

Il est clair que X induit une mesure de probabilité P_X sur \mathcal{B} par :

$$\forall A \in \mathcal{B} : P_X A = P(X^{-1}A)$$

Remarque: si (S, ρ) est l'espace C des fonctions continues sur $[0, 1]$

muni de la distance ρ de la convergence uniforme :

$$\rho(X, Y) = \sup_{0 \leq t \leq 1} |X(t) - Y(t)|$$

alors tout élément aléatoire de C a ses trajectoires continues. De même, on peut considérer l'espace D des fonctions semi-continues à droite muni de la même distance ρ . Cependant, C est complet séparable alors que D ne l'est pas. Skorokhod a muni D d'une distance d_0 qui le rend complet séparable (voir Billingsley(1968), ainsi que pour tous les résultats de ce chapitre). Ces deux propriétés étant essentielles aux théorèmes qui suivront, nous supposons que l'espace D est toujours muni de la distance d_0 . Par ailleurs, il ne sera pas nécessaire de manipuler directement d_0 , qui a une expression compliquée car :

$$d_0(X, Y) \leq \alpha \cdot \rho(X, Y)$$

pour toute paire X, Y d'éléments de D . Chaque fois, par exemple, qu'il sera nécessaire de montrer que $d_0(X_n, Y_n) \xrightarrow{D} 0$, il sera suffisant de montrer la même propriété pour $\rho(X_n, Y_n)$.

Passons maintenant aux deux définitions équivalentes de la convergence faible, que nous noterons \xrightarrow{D}

Définition 1.2

La suite (X_n) d'éléments aléatoires de l'espace métrique (S, ρ) converge faiblement vers X si, pour toute application continue bornée

f de S dans R , la suite de variables aléatoires $f(X_n)$ converge en distribution vers $f(X)$.

Définition 1.3

La suite (X_n) converge faiblement vers X si, pour tout $A \in \mathcal{B}$ tel que $P_X(\partial A) = 0$, on a :

$$P_{X_n}(A) \rightarrow P_X(A)$$

(où ∂A représente la frontière de A dans S).

Remarques :

.Si la première définition est plus imagée que la seconde, elle n'est guère plus opérationnelle.

.On peut montrer que la projection d'un processus X_n de C en un point quelconque t de $[0, 1]$ est une fonction continue ; il en découle que la convergence de $X_n(t)$ est une conséquence de la convergence faible de X_n . De même, la même projection est une fonction continue de D dans R pour tout t qui n'est pas un point de discontinuité de X_n ; la convergence de $X_n(t)$ vers $X(t)$ sera donc assurée pour tous les t qui ne sont pas des discontinuités de X .

Théorème Fondamental (théorème 1.1)

Soit f une application mesurable de S dans S' , tous deux métriques. Soit $D_f \subset S$ l'ensemble des discontinuités de f .

Si

$$X_n \xrightarrow{D} X \quad \text{et} \quad P_X(D_f) = 0$$

Alors

$$f(X_n) \xrightarrow{D} f(X)$$

(En particulier, les hypothèses sont satisfaites si f est continue)

Les théorèmes suivants ont une grande importance pour nos applications.

Théorème 1.2

Si

$$X_n \xrightarrow{D} X \quad \text{et} \quad \int (X_n, Y_n) \xrightarrow{D} 0$$

Alors

Théorème 1.3

Si

$$X_n \xrightarrow{-D} X \quad \text{et } Y_n \xrightarrow{-D} a \text{ non aléatoire}$$

Alors

$$(X_n, Y_n) \xrightarrow{-D} (X, a)$$

(où (X_n, Y_n) est un élément de $S \times S$ muni de la topologie engendrée par $\mathcal{B}_X \times \mathcal{B}_Y$.)

1.3. Caractérisation de la convergence faible

Comment établit-on la convergence faible dans C ou dans D ?

Les définitions 1.2 et 1.3 étant peu utiles, il a fallu trouver des caractérisations plus opérationnelles. Notons d'abord que les trois théorèmes ci-dessus peuvent souvent servir à montrer la convergence de processus construits à partir d'autres dont la convergence est déjà assurée. Il existe en outre une théorie très complète des caractérisations de la convergence faible sur les espaces complets séparables, qui est incluse dans Billingsley(1968), et que nous n'utiliserons pas dans sa généralité. Nous nous contenterons de citer ici une condition suffisante qui sera utilisée au chapitre 5.

Théorème 1.4

Soit (X_n) une suite d'éléments aléatoires de D

Si

$$(a) \quad (X_n(t_1), \dots, X_n(t_k)) \xrightarrow{-D} (X(t_1), \dots, X(t_k))$$

pour tout $0 \leq t_1 \leq t_2 \leq \dots \leq t_k \leq 1$

Et si

(b) Il existe $\gamma > 0$ et $\alpha > 1$, et une fonction F continue monotone non décroissante sur $[0, 1]$, tels que:

$$E(|X_n(t_3) - X_n(t_2)|^\gamma \cdot |X_n(t_2) - X_n(t_1)|^\gamma) \leq (F(t_3) - F(t_1))^\alpha$$

pour tout $0 \leq t_1 \leq t_2 \leq t_3 \leq 1$

Alors

Pour finir, nous citons deux théorèmes de limite centrale dans le cadre de la convergence faible. Ils seront, associés aux théorèmes 1.1, 1.2, 1.3, à la base des résultats-clés de notre recherche.

Théorème de Donsker (th. 1.5)

Si (u_n) est une suite de variables aléatoires indépendantes et équidistribuées (i.e.d.) avec une moyenne nulle et une variance σ^2 , et si nous définissons

$$X_n(t) = \frac{\sum_{k=1}^{[nt]} u_k}{\sigma \sqrt{n}}, \quad 0 \leq t \leq 1$$

Alors,

$$X_n \xrightarrow{D} P \text{ sur } D$$

(ou P est le processus de Lévy-Wiener sur $[0,1]$, c'est à dire le processus Gaussien centré dont la variance au temps t est égale à t , et à accroissements indépendants et stationnaires)

Théorème 1.6 (Prohorov (1956))

Si (u_{nk}) , $k=1, \dots, K_n$ est une double suite de variables aléatoires centrées et indépendantes, et si l'on définit:

$$s_n^2 = \text{Var} \left(\sum_{k=1}^{K_n} u_{nk} \right)$$

$$X_n(t) = \frac{\sum_{k=1}^{[K_n t]} u_{nk}}{s_n}, \quad 0 \leq t \leq 1$$

Alors, sous la condition de Lindebergh (voir Loève (19)), on a:

$$X_n \xrightarrow{D} P$$

Remarque

Les deux derniers théorèmes établissent que le théorème de Limite Centrale classique est généralisable à la convergence faible des processus sans hypothèse supplémentaire.

CHAPITRE II

LES FILES D'ATTENTE : VUE GENERALE

2.1. Généralités

Pour nous, une file d'attente sera constituée d'un ou plusieurs serveurs placés en parallèle; des clients arrivent séquentiellement et sont servis dans l'ordre de leur arrivée par le premier serveur libre, après quoi ils quittent le système. Un système est donc parfaitement décrit par le nombre s de serveurs, la suite des temps entre les arrivées successives, et la suite des durées de service des clients :

u_n = temps écoulé entre les $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ arrivées

v_n = durée de service du $n^{\text{ième}}$ client

Il est utile de définir le processus de comptage, $A(x)$, égal au nombre d'arrivées dans $[0, x]$:

$$A(x) = 1 + \text{Max} \left(k ; \sum_{j=1}^k u_j \leq x \right)$$

Dans le modèle GI/G/s, il est supposé que les suites (u_n) et (v_n) sont i.e.d. et mutuellement indépendantes. On définit aussi :

$$E u_n = E u$$

$$E v_n = E v$$

$$\text{Var } u_n = \sigma_u^2$$

$$\text{Var } v_n = \sigma_v^2$$

$\rho = E v / E u \cdot s$, l'intensité de trafic

Les processus généralement étudiés sont :

$Q(x)$: le nombre de clients présents au temps x

$L(x)$: la charge du système au temps x ; c'est la somme des temps de service restant à accomplir pour tous les clients présents à l'instant x .

$W(x)$: le temps qu'un observateur arrivant à l'instant x devrait attendre avant de se trouver devant un serveur libre ; c'est l'attente virtuelle à l'instant x . Pour les systèmes à un seul serveur, $L(x) = W(x)$.

Dans le cas GI/G/s avec $\rho < 1$, tous ces processus ont une limite en distribution lorsque $x \rightarrow \infty$ (Kiefer et Wolfowitz (1957) ...)

limite n'existe pas lorsque $\rho \gg 1$, auquel cas les limites sont infinies en distribution .

D'autres variables aléatoires offrent un intérêt et sont en général induites par l'un des processus de base. Un exemple important est W_n , le temps d'attente du n^{ième} client (en excluant son service).

Il est clair que :

$$(2.1) \quad W_n = W\left(\sum_{j=1}^n u_j\right)$$

W_n a une limite en distribution dans les mêmes conditions ($\rho < 1$).

Dans le cas $\rho \gg 1$, la file est dite saturée et la question est de savoir si ces mêmes quantités ont une limite , une fois centrées et réduites. En répondant à cette question nous noterons deux points utiles :

(a) Les conditions d'existence de limites sont très générales (plus générales que GI/G/s)

(b) Les limites sont calculables et ne dépendent pas des distributions de (u_n) et (v_n) .

Ces deux constatations sont une puissante motivation à l'étude des systèmes saturés, et ceci d'autant plus qu'en pratique, les files d'attente éloignées de la saturation offrent en général peu d'intérêt, vu qu'elles ne créent pas trop de difficultés.

2.2. Travaux antérieurs sur les systèmes saturés

Kingman(1961,1962) fut le premier à s'intéresser à la saturation . Usant des transformées de Laplace-Stieljes (T.L.S.) , il montra que , pour une suite de systèmes GI/G/1, dont les intensités de trafic croissent vers un , la distribution d'équilibre de l'attente d'un client, convenablement réduite, converge vers une loi exponentielle. Borovkov (1964,1965,1966) et Prohorov (1963) montrèrent, par des méthodes semblables la convergence dans les cas $\rho \gg 1$. Whitt et Iglehart (1970) unifièrent la théorie pour les systèmes GI/G/1

Ils furent les premiers à utiliser la convergence faible pour leurs démonstrations. Nous indiquons ci-dessous le schéma de leur approche afin d'en montrer les similitudes et les différences avec la nôtre.

Pour GI/G/1 ils définissent :

$$X_n = v_n - u_n$$

$$S_n = \sum_{k=1}^n X_k, \quad S_0 = 0$$

Alors, Lindley (1952) a montré que :

$$(2.3) \quad W_n = S_n - \min_{0 \leq k \leq n} S_k$$

Se basant sur (2.3), et sur le th. de Donsker, ils montrent la convergence faible des processus γ_n lorsque $\rho > 1$:

$$\gamma_n(t) = \frac{W_{[nt]} - (Ev - Eu)nt}{\sigma \sqrt{n}}, \quad 0 \leq t \leq 1$$

$$(\text{ou } \sigma^2 = \sigma_u^2 + \sigma_v^2 = \text{Var } X_n)$$

Notons que γ_n est un élément aléatoire (e.a.) de D.

La limite est le processus de Lévy, P, pour $\rho > 1$, et, pour $\rho = 1$, la limite est ce même processus de Lévy muni d'une barrière infranchissable à l'origine, B, c'est à dire :

$$(2.4) \quad BoP(t) = P(t) - \inf_{0 \leq s \leq t} P(s), \quad 0 \leq t \leq 1$$

Ils montrent ensuite des théorèmes analogues pour les processus relatifs à Q(x) et généralisent au cas des suites de systèmes GI/G/1.

Leur approche ne peut être généralisée aux systèmes multi-serveurs sauf dans le cas de Q(t) qui reçoit donc un traitement complet.

Nous avons conservé le mode de convergence faible "inauguré" par Iglehart et Whitt, car il est indubitablement le plus élégant et général. Cependant, nous commenterons notre étude par le processus W(x) pour ensuite passer à W_n , approche inverse de celle de nos prédécesseurs, et qui a l'avantage de permettre l'étude complète du cas GI/G/s. Cette même approche permet

CHAPITRE III

THEOREMES DE CONVERGENCE POUR LES SYSTEMES GI/G/1 SATURES

3.1. Introduction

Ce chapitre contient les résultats de base de nos travaux. Les démonstrations seront esquissées car elles illustrent très bien l'emploi des outils de la convergence faible. La section 3.2. présente le résultat-clé en deux théorèmes de convergence de processus construits sur $W(x)$. Dans la section 3.3, nous esquisserons le passage de $W(x)$ à W_n , et dans la section 3.4, nous donnerons les résultats correspondants pour les suites de systèmes. Enfin nous indiquerons un mode possible d'utilisation de ces théorèmes dans la section 3.5 qui clôt ce chapitre.

3.2. Convergence des processus d'attente virtuelle

Théorème 3.1

Si $\rho = 1$ dans le système GI/G/1, alors:

$$\delta_n \xrightarrow{D} \text{BoP}$$

où :

$$\delta_n(t) = \frac{W(nt)}{\sigma \sqrt{n/Eu}}, \quad 0 \leq t \leq 1$$

Esquisse de démonstration:

Cette dernière consiste à montrer successivement la convergence de plusieurs suites d'éléments de D , définis comme suit:

$$(3.1) \quad Y_n(t) = \frac{\sum_{i=1}^{[nt]} v_i - E v n t}{\sigma_v \sqrt{n}}$$

$$Z_n(t) = \frac{\sum_{i=1}^{[nt]} u_i - E u n t}{\sigma_u \sqrt{n}}$$

$$X_n(t) = \frac{\sum_{i=1}^{[nt]} (v_i - \rho \cdot u_i)}{\sqrt{n(\sigma_v^2 + \rho^2 \sigma_u^2)}} \quad , \quad 0 \leq t \leq 1$$

$$\eta_n(t) = \frac{\sum_{i=1}^{A(nt)} v_i - \rho^{nt}}{\mathcal{K} \sqrt{n}} \quad , \quad 0 \leq t \leq 1$$

$$\eta'_n(t) = \frac{\sum_{i=1}^{A(nt)} v_i - \rho \sum_{i=1}^{A(nt)} u_i}{\mathcal{K} \sqrt{n}} \quad , \quad 0 \leq t \leq 1$$

(3.1)

$$\zeta_n(t) = \frac{\sum_{i=1}^{A(nt)} v_i - nt}{\mathcal{K} \sqrt{n}} \quad , \quad 0 \leq t \leq 1$$

$$\delta_n(t) = \frac{W(nt) - (\rho - 1)nt}{\mathcal{K} \sqrt{n}} \quad , \quad 0 \leq t \leq 1$$

avec

$$\mathcal{K} = \left[(\sigma_v^2 + \rho^2 \sigma_u^2) / E u \right]^{1/2}$$

Y_n et Z_n convergent vers deux processus de Lévy-Wiener par le théorème de limite centrale de Donsker (th. 1.5.) et donc X_n converge aussi vers un processus P de Lévy - Wiener. D'autre part, η'_n est simplement la composition de X_n avec le processus ϕ_n :

$$\phi_n(t) = A(nt) \quad , \quad 0 \leq t \leq 1$$

on montre d'abord que $\phi_n \xrightarrow{D} \phi$, où ϕ est un processus non aléatoire

$\phi(t) = t/Eu$. Puis, à l'aide d'un théorème de Billingsley (1968) sur les compositions, il découle que $\eta_n' \xrightarrow{D} P$.

Pour montrer la convergence de η_n , on prouve que la distance $\rho(\eta_n, \eta_n')$ converge vers zéro en probabilité ; plus précisément :

$$\rho(\eta_n, \eta_n') = \frac{\rho}{K} \sup_{0 \leq t \leq 1} \left| \frac{\sum_{i=1}^{A(nt)} u_i - nt}{\sqrt{n}} \right| \leq \frac{\rho}{K} \max_{1 \leq i \leq A(n)} \dots$$

Et cette dernière quantité converge vers zéro en probabilité. Le théorème permet donc d'affirmer que $\eta_n \xrightarrow{D} P$. Finalement, (et dans le cas $\rho = 1$ se il existe une relation directe entre η_n et δ_n , qui découle directement de la relation suivante, établie par Reich (1958) pour GI/G/1 :

$$(3.2) \quad L(x) = W(x) = N(x) - \inf_{0 \leq y \leq x} N(y)$$

où

$$N(x) = \sum_{i=1}^{A(x)} v_i - x$$

(notons que $N(x)$ représenterait exactement la charge $L(x)$ si le serveur n'avait jamais été libre dans $[0, x]$; la relation (3.2) ajoute donc à $N(x)$ la quantité $\left\{ \begin{array}{l} -\inf_{0 \leq y \leq x} N(y) \end{array} \right\}$, qui représente le temps total d'inactivité.

dans $[0, x]$). (3.2) se transpose immédiatement, quand $\rho = 1$, en :

$$(3.3) \quad \delta_n = B \circ \eta_n$$

et le théorème 1.1 permet d'affirmer que $\delta_n \xrightarrow{D} B \circ P$, puisque Whitt (1968) a prouvé que B est une application continue dans D .

Remarque: Pour montrer la plus grande généralité introduite par la convergence faible nous pouvons déduire par exemple que :

$$\begin{array}{ccc} \text{Sup } \delta_n(t) & \xrightarrow{D} & \text{Sup } B \circ P(t) \\ 0 \leq t \leq 1 & & 0 \leq t \leq 1 \end{array}$$

puisque l'application de D dans R constituée par Sup_{0 ≤ t ≤ 1} est aussi continue.

Nous obtenons ainsi un théorème sur la convergence de $\frac{\text{Sup}_{0 \leq x \leq n} W(x)}{\sqrt{n} \cdot K}$ qu'il serait difficile d'obtenir directement.

Théorème 3.2 : si $\rho > 1$ dans GI/G/1, alors :

$$\delta_n \xrightarrow{D} P$$

(où δ_n et P sont définis au théorème 3.1)

Esquisse de la démonstration:

η_n converge toujours vers P comme précédemment, mais la relation (3.3) n'est plus vraie, à cause du terme constant de δ_n , qui n'est pas nul.

Par contre, examinons la distance $\rho(\delta_n, \eta_n)$. On montre que:

$$\rho(\delta_n, \eta_n) = \text{Sup}_{0 \leq t \leq 1} (-\eta_n(t) - (\rho-1) \sqrt{n} t) \xrightarrow{D} 0$$

Puisque cette quantité converge vers zéro en probabilité (par un argument classique), δ_n a donc la même limite que η_n par le théorème 1.2, ce qui achève la démonstration.

3.3 Convergence faible des processus d'attente réelle

Ici encore, la relation (2.1) permet un passage facile de $W(x)$ à W_n quand $\rho=1$, mais non dans le cas $\rho>1$, à cause des termes constants. Plus précisément, si nous définissons :

$$\gamma_n(t) = \frac{W_{[nt]} - (Ev - Eu) nt}{\sigma \sqrt{n}}, \quad 0 \leq t \leq 1$$

où :

$$\sigma^2 = \sigma_u^2 + \sigma_v^2$$

W_n est l'attente du n^{ième} client, à l'exclusion de son service

Alors, pour $\rho = 1$, et grâce à (2.1) :

$$\gamma_n(t) = (Eu)^{-1/2} \cdot \delta_n \circ \psi_n(t), \quad 0 \leq t \leq 1$$

où :

$$\psi_n(t) = \frac{\sum_{i=1}^{[nt]} u_i}{n}, \quad 0 \leq t \leq 1$$

Comme précédemment, nous montrons d'abord la convergence de ψ_n vers un processus non-aléatoire ψ défini par : $\psi(t) = Eu \cdot t$; il en résulte que γ_n converge vers $(Eu)^{-1/2} B \circ P \circ \psi$, c'est à dire $B \circ P$:

Théorème 3.3

Si $\rho=1$ dans GI/G/1, alors :

$$\gamma_n \xrightarrow{D} B \circ P$$

Dans le cas $\rho>1$, la relation (2.1) entraîne :

$$(3.4) \quad \delta_n \circ \psi_n(t) = \frac{W_{[nt]} - (\rho-1) \sum_{i=1}^{[nt]} u_i}{K \cdot \sqrt{n}}$$

Or, nous avons déjà indiqué à la section précédente que, pour $\rho>1$,

$$(3.5) \quad \delta_n(t) = \eta_n' + E_n$$

où \vdots

$$\sup_{0 \leq t \leq 1} E_n(t) \xrightarrow{D} 0 \quad \text{quand } n \rightarrow \infty$$

De plus, (3.4) peut s'écrire schématiquement :

$$a \gamma_n(t) = \delta_n \circ \Psi_n(t) + b Z_n(t)$$

Et, par (3.5) et (3.1) :

$$a \gamma_n(t) = \left[(cY_n + dZ_n) \circ \phi_n + E_n \right] \circ \Psi_n + bZ_n$$

Ce qui se simplifie en :

$$\gamma_n = \alpha Y_n + \beta Z_n + E_n \circ \Psi_n, \quad \alpha \text{ et } \beta = \text{constantes}$$

Le dernier terme étant négligeable, γ_n converge donc vers un processus de Lévy-Wiener . Nous énonçons :

Théorème 3.4

Si $\rho > 1$ dans GI/G/1 , alors :

$$\gamma_n \xrightarrow{D} P$$

3.4. Etude des suites de systèmes GI/G/1

L'idée centrale de cette section est d'étendre les résultats précédents aux suites de systèmes de la manière suivante : considérons une suite de files d'attente GI/G/1 , soit $(S^k, k=1,2,..)$; le $k^{\text{ième}}$ système étant défini par les suites i.e.d. (u_n^k) et (v_n^k) , avec $\rho^k = E v^k / E u^k$. Si l'on définit comme précédemment les suites (doubles) d'éléments de D , (δ_n^k) et (σ_n^k) , peut-on montrer la convergence de ces suites sous certaines conditions sur le comportement de ρ^k ? Il est clair que ce problème généra-

-lise ceux résolus aux sections précédentes (que l'on obtient en faisant $S^k = S$ quel que soit k). Afin de rendre ce problème plus abordable , nous nous contenterons d'étudier les suites diagonales $\{\delta_n^n\}$ et $\{\sigma_n^n\}$, ce qui conserve suffisamment de variété au problème pour obtenir des résultats en globant ceux des sections précédentes et les généralisant. Il s'avère, et en cela nous suivons Whitt (1968), que l'on peut distinguer cinq cas différents, suivant le comportement de la quantité $(\rho^{n-1}) \cdot \sqrt{n}$:

cas 1 : $\rho^n \uparrow 1$ et $(\rho^{n-1}) \cdot \sqrt{n} \rightarrow -\infty$

cas 2 : $\rho^n \uparrow 1$ et $(\rho^{n-1}) \cdot \sqrt{n} \rightarrow c$, $-\infty < c < 0$

cas 3 : $\rho^n \rightarrow 1$ et $(\rho^{n-1}) \cdot \sqrt{n} \rightarrow c$, $c=0$

cas 4 : $\rho^n \downarrow 1$ et $(\rho^{n-1}) \cdot \sqrt{n} \rightarrow c$, $0 < c < +\infty$

cas 5 : $\rho^n \rightarrow \rho > 1$ et $(\rho^{n-1}) \cdot \sqrt{n} \rightarrow +\infty$

Nous supposons en outre , dans tous les cas, que :

$$\begin{aligned} Eu^n &\rightarrow Eu > 0 & Ev^n &\rightarrow Ev & \rho^n = Ev^n/Eu^n &\rightarrow \rho \\ \sigma_u^n &\rightarrow \sigma_u & \sigma_v^n &\rightarrow \sigma_v & \sigma^2 = \sigma_u^2 + \sigma_v^2 &> 0 \end{aligned}$$

Théorème 3.5 Et : $\exists \delta > 0$ tel que : $E \left| v_1^n - u_1^n \right|^{\delta+2}$ unif. borné $\forall n$

Dans les cas 2,3,4 , on a :

$$\delta'_n \xrightarrow{D} B o (P + \alpha)$$

et

$$\sigma'_n \xrightarrow{D} B o (P + \beta)$$

où :

$$\delta'_n(t) = \frac{W^n(nt)}{\mathcal{K} \sqrt{n}} , \quad 0 \leq t \leq 1$$

$$\sigma'_n(t) = \frac{W^n[nt]}{\sigma \sqrt{n}} , \quad 0 \leq t \leq 1$$

$$\beta(t) = ct/\sigma \qquad \alpha(t) = ct/\mathcal{K} , \quad 0 \leq t \leq 1$$

$W^n(x)$ = charge du système S^n à l'instant x

W^n - ...

Remarque

Les méthodes utilisées pour montrer ce théorème sont en tous points semblables à celles des théorèmes 3.1 et 3.3, excepté qu'au lieu du théorème de limite centrale de Donsker, nous faisons appel au théorème de Prohorov (th. 1.6) relatif aux sommes diagonales. On observe cependant que les processus-limites contiennent des translations $\alpha(t)$ et $\beta(t)$ linéaires en t . Le processus $B_0(P; +\alpha)$ a des distributions marginales connues de façon explicite, bien que moins simples que celles des processus-limites des sections précédentes, (voir par exemple Lévy (1948)).

Théorème 3.6

Dans le cas 5, on a :

$$\delta_n^n \xrightarrow{D} P$$

et

$$\gamma_n^n \xrightarrow{D} P$$

où :

$$\delta_n^n(t) = \frac{W^n(nt) - (\rho^{n-1})}{\mathcal{K} \sqrt{n}}, \quad 0 \leq t \leq 1$$

$$\gamma_n^n(t) = \frac{W_{[nt]}^n - (E v^n - E u^n)nt}{\sigma \sqrt{n}}, \quad 0 \leq t \leq 1$$

Cas 1 : Limite Exponentielle

On peut montrer que dans ce cas, les éléments δ_n^n et γ_n^n ne peuvent pas converger faiblement. Par contre, Prohorov (1963) a montré la convergence en distribution des variables aléatoires $\gamma_n^n(1)$, convenablement réduites :

Théorème 3.7 (Prohorov (1963)) :

Dans le cas 1, on a :

$$P \left\{ \frac{W_n^n (E v^n - E u^n)}{\sigma \sqrt{n}} \leq x \right\} \longrightarrow \int_0^x 1 - e^{-2x} \quad , \quad x \geq 0$$

La démonstration de ce théorème donnée par Prohorov s'applique exclusivement au modèle GI/G/1 et n'est pas généralisable à GI/G/s, non plus qu'aux divers systèmes décrits au chapitre 5, alors que les cas 2,3,4,5 se verront généralisés sans peine.

Si, au lieu de choisir la suite diagonale (W_n^n) , nous étudions n'importe quelle suite double (W_n^k) , où n et k deviennent infinis, le même théorème est vrai. Il est légitime, en particulier, de faire tendre d'abord n vers l'infini et d'obtenir ainsi, après quelques manipulations, le résultat de Kingman (1961) :

Théorème 3.8 (Kingman (1961))

Soit (W_∞^k) la suite de variables aléatoires ayant comme distribution la distribution-limite du temps d'attente d'un client dans le système S^k (cette limite existe car $\rho^k < 1$). Quand $\rho^k \rightarrow 1$ par valeurs croissantes alors :

$$P \left\{ \frac{W_\infty^k (E v^k - E u^k)}{\sigma^2} \leq x \right\} \rightarrow \begin{cases} 1 - e^{-2x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

3.5 Utilisation pratique des théorèmes de convergence

(a) Notons tout d'abord qu'à toute convergence faible d'une suite d'éléments de D, (X_n) , correspond la convergence en distribution de la suite de variables aléatoires $(X_n(1))$. Par exemple, les Th. 3.3 et 3.4 ont pour corollaire :

Corollaire 3.1

Pour le système GI/G/1 :

Si $\rho = 1$, alors :

$$(3.6) \quad P \left\{ \frac{W_n}{\sigma \sqrt{n}} \leq x \right\} \rightarrow \begin{cases} \int_0^x \frac{\sqrt{2}}{\pi} e^{-y^2/2} dy & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

Si $\rho > 1$, alors :

$$P \left\{ \frac{W_n - (E v - E u)n}{\sigma \sqrt{n}} \leq x \right\} \rightarrow \begin{cases} 1 - e^{-2x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

Remarque

Le membre de droite de (3.6) représente la fonction de répartition de la variable aléatoire B ou $P(1) = P(1) - \inf_{0 \leq t \leq 1} P(t)$, alors que le

le membre de droite de (3.7) est la fonction de répartition normale centrée réduite .

Si nous étudions donc un système GI/G/1 avec $\rho = 1$ (resp. $\rho > 1$) , nous pouvons, après un temps suffisamment long, supposer que W_n possède la distribution - limite de (3.6) (resp. de (3.7) .

Notons de plus que les distributions de u_n et de v_n n'entrent pas en ligne de compte, si ce n'est qu'elles peuvent affecter la vitesse de convergence des suites concernées ; cette vitesse de convergence est d'une grande importance pratique , et sera étudiée au chapitre 7 .

(b) Une question peut se poser quant à l'utilité pratique des résultats concernant les suites de systèmes étudiées à la section 3.4 : en effet , en pratique, nous n'avons affaire qu'à des systèmes bien déterminés, et non à des suites de systèmes . Nous proposons ici deux idées dont la pertinence sera testée au chapitre 7 :

(i) Approximation d'un système stable ($\rho < 1$) mais quasi-saturé (ρ proche de l'unité) : dans ce cas, la distribution d'équilibre de l'attente d'un client est approximativement exponentielle, en vertu du Th. 3.8 , c'est à dire :

$$(3.8) \quad P \left\{ \frac{W_\infty (Eu - Ev)}{\sigma^2} \leq x \right\} \approx \begin{cases} 1 - e^{-2x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

Cependant, pour n fini, la distribution de W_n pourra être approximée soit par (3.8), soit par la distribution-limite du cas 2 décrit à la section précédente . Nous différencierons ces deux situations au chapitre 7 .

(ii) Approximation d'un système saturé ($\rho > 1$): dans ce cas encore les limites des cas 4 et 5 sont des approximations possibles de la variable aléatoire W_n , suivant les valeurs prises par n et ρ . Nous distinguerons

LES SYSTEMES A s SERVEURS GI/G/s

4.1. Introduction

Ce chapitre généralise tous les résultats sur GI/G/1 au cas multi-serveurs. La difficulté de cette généralisation est double :

(a) Les relations (2.3) et (3.2) qui permettaient de lier les quantités $W(x)$ et W_n d'une part aux quantités S_n et $N(x)$ d'autre part ne sont plus vraies pour des systèmes multi-serveurs.

(b) Dès qu'il y a plusieurs serveurs, la charge $L(x)$ diffère de l'attente virtuelle $W(x)$. Il faudra donc une étape supplémentaire pour passer de $L(x)$ à $W(x)$.

La première difficulté sera surmontée grâce à la définition d'un système auxiliaire permettant de ramener l'étude de GI/G/s à celle d'un certain système qui, aux fins d'étude de la charge $L(x)$, est indiscernable d'un système GI/G/1 : La seconde difficulté est moins considérable et cèdera grâce à une inégalité établie sans peine. Nous nous efforcerons donc, dans cette présentation, de ne pas répéter ce qui est commun à l'étude de GI/G/s et à celle de GI/G/1, mais au contraire de faire ressortir l'originalité des méthodes utilisées pour résoudre les deux différences importantes mentionnées ci-dessus.

Tout d'abord, mettons à jour les définitions essentielles de ce chapitre, qui sont légèrement différentes des définitions (3.1).

Le système GI/G/s est comme avant parfaitement défini par les deux de variables aléatoires i.é.d. :

u_n = temps écoulé entre les arrivées n et $(n+1)$

v_n = durée du service du $n^{\text{ième}}$ client

$$\rho = E v / s E u$$

$$\sigma_u^2 = \text{Var } u_n$$

$$\sigma_v^2 = \text{Var } v_n$$

$$\sigma^2 = \sigma_v^2 + s^2 \sigma_u^2$$

$$\mathcal{K} = \left[(\sigma_v^2 + \rho^2 s^2 \sigma_u^2) / Eu \right]^{1/2}$$

$$(4.1) \quad \eta_n(t) = \frac{\sum_{i=1}^{A(nt)} v_i - snt}{\mathcal{K} \sqrt{n}}$$

$$\xi_n(t) = \frac{L(nt) - s(\rho-1)nt}{\mathcal{K} \sqrt{n}}$$

$$0 \leq t \leq 1$$

$$\zeta_n(t) = \frac{s W(nt) - s(\rho-1)nt}{\mathcal{K} \sqrt{n}}$$

$$\delta_n(t) = \frac{s W_{[nt]} - s(\rho-1)Eu nt}{\sigma \sqrt{n}}$$

4.2. Le système auxiliaire

Considérons le système fictif S' défini par les mêmes suites (u_n) et (v_n) que le système GI/G/s ci-dessus, mais avec une discipline de service différente : en effet, nous supposons que, chaque fois que le système S' aura h serveurs libres, $h < s$, le taux de service des $s - h$ clients présents dans le système sera égal à $\frac{s}{s - h}$, au lieu d'être égal à s . Ceci revient à dire que, tant qu'il y a, ne serait-ce qu'un client dans S' , tous les serveurs sont occupés à servir les clients présents, de façon que le taux global de service reste égal à s . (De toute évidence, un tel système n'est peut-être pas réalisable en pratique, mais pour nous il n'est qu'un subterfuge utile). Notons que puisque la charge $L'(x)$ de ce système s'épuise à un taux exacte-

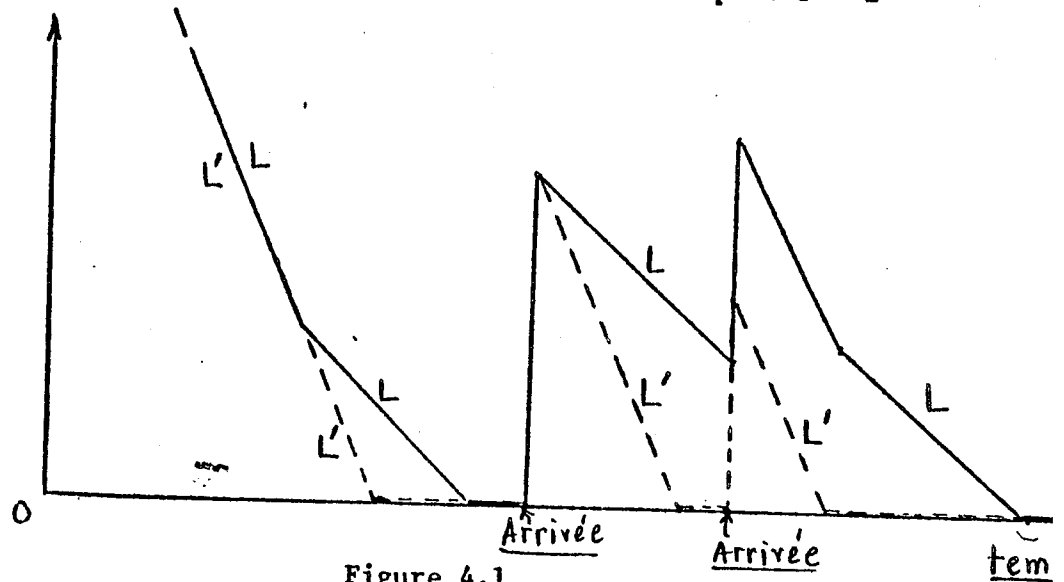
ment égal à s tant que S' n'est pas vide, $L'(x)$ se comporte exactement comme s fois la charge d'un système GI/G/1 ayant comme suites (u_n) et (v_n/s) . De ce fait important, il résulte que, comme pour GI/G/1

$$(4.2) \quad L'(x) = N'(x) - \inf_{0 \leq y \leq x} N'(y)$$

où :

$$N'(x) = \sum_{i=1}^{A(x)} v_i - sx = s \left(\sum_{i=1}^{A(x)} v_i/s - x \right)$$

La fig. 4.1 montre une réalisation de L et L' pour $s = 2$



En raison de (4.2), les résultats du chapitre 3 s'appliquent à ξ'_n défini par :

$$\xi'_n(t) = \frac{L(nt) - s(\rho - 1)nt}{K\sqrt{n}}, \quad 0 \leq t \leq 1$$

Il reste bien sûr à montrer que L et L' restent assez proches l'un de l'autre pour que ξ'_n et ξ_n aient la même limite. Ceci est accompli dans la section suivante .

4.3. Etude du système GI/G/s.

Lemme 4.1

$$P(\xi_n, \xi'_n) \xrightarrow{D} 0$$

Démonstration :

(a) lorsque s (ou plus) clients se trouvent dans le système GI/G/s, alors les deux systèmes travaillent au même taux, s, et donc la différence $L(x) - L(x')$ reste constante

(b) Dans tous les cas, nous avons aussi :

$$L(x) \geq L'(x) \geq 0$$

(c) Lorsque le système GI/G/s a moins de s clients, alors il est certainement vrai que $L(x)$ ne peut être plus grand que s fois le plus long service qui se soit manifesté depuis l'instant zéro, c'est-à-dire :

$$L(x) \leq s \max_{i \leq A(x)} v_i$$

(d) En joignant les conclusions a, b, et c, on obtient pour tout x :

$$0 \leq L(x) - L'(x) = |L(x) - L'(x)| \leq s \max_{0 \leq i \leq A(x)} v_i$$

ce qui entraîne :

$$\sup_{0 \leq t \leq 1} \left| \xi_n(t) - \xi'_n(t) \right| \leq s \cdot \frac{\max_{0 \leq i \leq A(nt)} v_i}{K \sqrt{n}}$$

Et cette dernière quantité converge vers zero en probabilité.

Il est donc désormais acquis que les théorèmes 3.1 et 3.2 peuvent être étendus à ξ_n , ce que nous énonçons :

Théorème 4.1

Pour un système GI/G/s

si $\rho = 1$ $\xi_n \xrightarrow{D} B \circ P$

si $\rho > 1$ $\xi_n \xrightarrow{D} P$

Il reste à faire le lien entre ξ_n et δ_n . Pour cela, dénotons par $L_j(x)$ la charge à l'instant x du serveur j, c'est-à-dire la part de $L(x)$ qui écherra au serveur j, $j = 1, 2, \dots, s$. Il apparaît que la différence de charge de deux serveurs quelconques à l'instant x ne peut pas excéder le plus long service qui se soit manifesté depuis

l'instant zéro, c'est-à-dire :

$$|L_1(x) - L_j(x)| \leq \max_{0 \leq k \leq A(x)} v_k$$

Et donc, un raisonnement tout à fait parallèle à celui du théorème

précédent permet d'affirmer que les processus ξ_n^j , $j=1, \dots, s$, ont la même limite, s'ils en ont une. Or, leur moyenne $1/s \cdot \sum_{j=1}^s \xi_n^j$ est égale à ξ_n qui, lui, a une limite; donc chaque ξ_n^j converge faiblement vers la limite de ξ_n . Mais puisque :

$$\delta_n(t) = \min_{1 \leq j \leq s} \xi_n^j(t)$$

on peut établir que δ_n a même limite que ξ_n , et, pour les mêmes conditions que celles du chapitre 3, γ_n se comporte de même :

Théorème 4.2

Pour un système GI/G/s

$$\text{si } \rho = 1 \quad \delta_n \xrightarrow{D} B \circ P \quad \text{et} \quad \gamma_n \xrightarrow{D} B \circ P$$

$$\text{si } \rho > 1 \quad \delta_n \xrightarrow{D} P \quad \text{et} \quad \gamma_n \xrightarrow{D} P$$

4.4. Les suites de systèmes GI/G/s

Les outils développés dans ce chapitre s'appliquent intégralement à l'étude des suites de systèmes GI/G/s. Les résultats sont absolument identiques à ceux de la section 3.4 dans les cas 2, 3, 4, et 5 décrits plus haut. Nous les omettons par souci de concision. Par contre il n'a pas été démontré de théorème semblable au Th.3.7 dans le cas 1 pour plusieurs serveurs. Le théorème de Kingman (Th.3.8) relatif à la convergence exponentielle des distributions stationnaires de l'attente a été, en revanche, généralisé par Köllerström (1973) au cas GI/G/s par une approche élaborée utilisant des méthodes de Transformées de Laplace-Stieljes. (Remarquons en passant que le procédé du système auxiliaire développé dans ce chapitre ne réussit pas à généraliser le th. de Kingman de façon simple, alors qu'il généralise aisément les autres théorèmes.).

CHAPITRE V

EXTENSIONS ET COMPLEMENTS

5.1. Introduction

Les chapitres 3^{ème} et 4 représentent l'ossature principale de l'étude des systèmes saturés. A partir des résultats de ces chapitres, diverses généralisations peuvent être obtenues à peu de frais. Ces généralisations consistent habituellement à admettre comme processus d'arrivées $A(x)$ autre chose que le processus de renouvellement du modèle GI/G/s. Tour à tour, nous examinerons quatre types de processus d'arrivées admissibles, qui ne seront pas tous stationnaires, mais qui, tous quatre, répondent à des soucis d'ordre pratique.

Et puis, nous indiquerons dans la section terminale de ce chapitre comment les théorèmes des chapitres précédents peuvent être utilisés pour montrer la convergence d'autres processus et variables aléatoires d'intérêt dans les files d'attente.

5.2. Généralisations quasi-stationnaires du processus d'arrivée

(a) $A(x)$ est la superposition de r processus de renouvellement indépendants.

En pratique, il peut être intéressant d'admettre que les clients proviennent de r sources indépendantes, chacune émettant ses clients par un processus de renouvellement $A^i(x)$, $i = 1, \dots, r$. Le processus d'arrivée est, dans ce cas, très complexe à analyser. Plus précisément, à chaque source i est attachée une suite de variables aléatoires (u_k^i) telle que :

$$E u_k^i = E u^i \qquad \text{Var } u_k^i = \sigma_u^2$$

De plus, nous admettrons, pour plus de généralité, que les clients de

la source i peuvent avoir des services (v_k^i) ayant une distribution propre à cette source, avec :

$$Ev_k^i = Ev^i \qquad \text{Var } v_k^i = \sigma_{v^i}^2$$

on définit:
$$\rho_i = Ev^i / Eu^i \qquad \rho = \sum_{i=1}^r \rho_i$$

$$(5.1) \quad K_i = \left[(\sigma_{v^i}^2 + \rho_i^2 \sigma_{u^i}^2) / Eu^i \right]^{1/2}, \quad K^2 = \sum_{i=1}^r K_i^2$$

on appellera ce modèle GI - r/G/s.

Il est remarquable que ces nouvelles complexités n'affectent pratiquement pas les démonstrations, et l'on peut établir:

Théorème 5.I

Dans le système GI - r/G/s

si $\rho = 1$, alors :

$$\delta_n \xrightarrow{D} B \circ P$$

si $\rho > 1$, alors :

$$\delta_n \xrightarrow{D} P$$

(où δ_n , B, P, sont définis comme précédemment, avec les valeurs de ρ et K décrites par (5.1)).

(b) Chaque $A(x)$ est "proche" d'un processus de renouvellement :

Par proche nous voulons dire que :

$$(5.2) \quad \sup_{0 \leq y \leq x} \left| \frac{A^1(y) - B^1(y)}{\sqrt{x}} \right| \xrightarrow{D} 0, \text{ quand } x \longrightarrow \infty$$

où B^1 est un processus de renouvellement.

Sous la condition (5.2), le théorème 5.I ci-dessus est encore vrai.

Cette généralisation a deux applications pratiques assez intéressantes que nous citons :

I° - $A(x)$ est un processus de rendez-vous : plus précisément, le $n^{\text{ième}}$ client est prévu à l'instant $n Eu$, mais il peut être en retard (ou en avance) du temps Z_n , (Z_n) étant une suite de variables aléatoires I.I.D., avec $E|Z| < \infty$. Alors, si l'on définit le processus de renouvellement décalé $B(y) = \lceil y/Eu \rceil$ on peut montrer que le

condition (5.2) est satisfaite. Donc, tous les résultats obtenus jusqu'à présent sont encore vrais si les processus d'arrivées sont des processus de rendez-vous.

2° - A(x) est constitué des départs de plusieurs autres queues stables :

supposons un réseau de r systèmes $GI/G/s_j$, $j = 1, 2, \dots, r$ placés en parallèle et qui sont stables ; les clients quittant ces systèmes constituent les arrivées du système à étudier, comme le décrit la figure 5.1 ci-dessous :

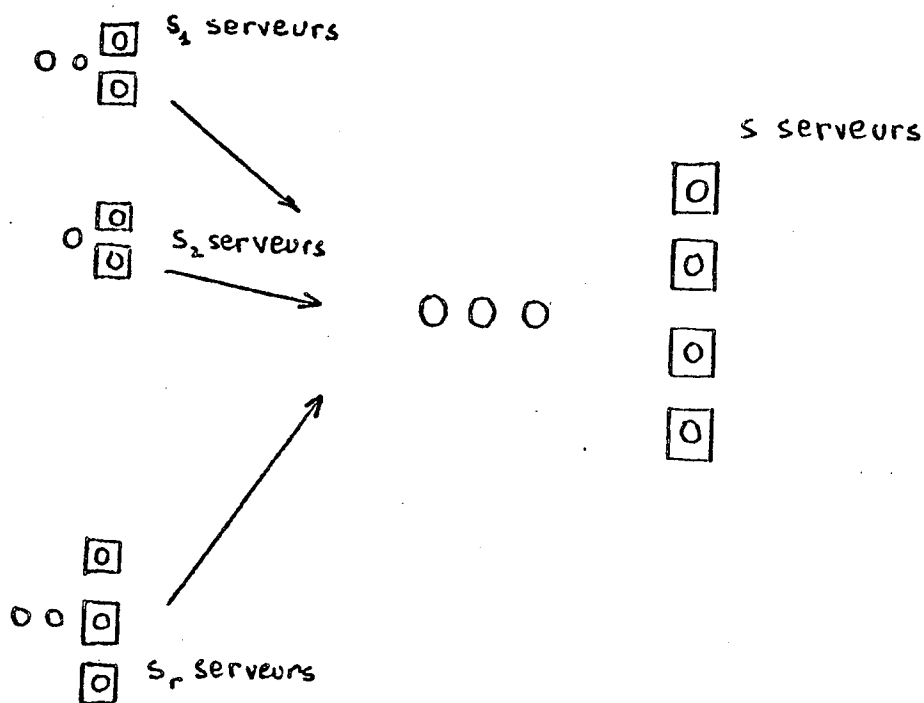


Figure 5.1

Si nous dénotons par $A^i(x)$ le nombre de clients servis par le $i^{\text{ème}}$ système dans l'intervalle $[0, x]$, et par $B^i(x)$ le nombre d'arrivées de ce même $i^{\text{ème}}$ système dans $[0, x]$, alors il est clair que la différence $Q^i(x) = B^i(x) - A^i(x)$ représente exactement le nombre de clients présents dans le $i^{\text{ème}}$ système à l'instant x . Mais puisque ce système est stable, il est aisé de montrer que l'on a :

$$\sup_{0 \leq y \leq x} \frac{Q^i(y)}{\sqrt{x}} \xrightarrow{D} 0 \quad \text{quand } x \rightarrow \infty$$

Nous sommes donc dans les conditions d'application (5.2) du théorème 5.2 généralisé.

5.3. Un modèle non classique d'arrivées transitoires à s serveurs

Le modèle décrit dans cette section est vraiment très différent des divers modèles classiques de files d'attente généralement étudiées. Ces derniers s'attachent à l'étude des divers processus "après un temps infini", c'est-à-dire lorsque le processus d'arrivée est devenu stationnaire. En pratique, ces modèles s'appliquent à beaucoup de situations où le système, à toutes fins utiles, ne cesse pas d'opérer. Il est par contre d'autres situations où le phénomène d'attente est essentiellement transitoire (par exemple un poste de péage à l'heure de pointe). Dans ce cas, il serait assez vain d'espérer analyser ces situations "après un temps infini", surtout si la période d'activité du système est assez courte et si le système se sature durant cette période. Nous avons donc été amené à formuler et analyser un exemple d'un tel modèle, sans vouloir prétendre à une grande généralité.

Dans notre modèle, nous supposons que N clients exactement arrivent dans le système, et ce durant une période finie $[0, T]$. Nous supposons que l'instant d'arrivée d'un client quelconque,

est une variable aléatoire τ , distribuée suivant la fonction de répartition $F(\cdot)$ concentrée sur $[0, T]$, c'est-à-dire :

$$P(\tau \leq x) = F(x) \quad ; \quad F(0) = 0 \quad ; \quad F(T) = 1$$

Notons que ce modèle d'arrivées est très différent d'un processus de renouvellement, puisqu'ici chaque client a la même distribution pour l'instant de son arrivée que n'importe quel autre client. Les (τ_n) sont supposées indépendantes. Il est clair que, suivant le choix de $F(\cdot)$, nous pourrions donner aux arrivées la forme voulue (par exemple : heure de pointe, deux heures de pointe successives, etc...). Sans perte de généralité supposons $T = 1$ et appelons S_N le système ainsi décrit. Nous noterons :

$v_{N,n}$ = temps de service de la $n^{\text{ème}}$ arrivée dans le système S_N
 $A_N(t)$ = nombre d'arrivées dans $[0, t]$ dans le système S_N , $t \leq 1$
 $W_N(t)$ = temps d'attente virtuelle à l'instant t dans S_N
 s = nombre (fixe) de serveurs dans S_N

Il serait intéressant d'étudier le temps nécessaire à la résorption de l'embouteillage après l'instant $T = 1$, c'est-à-dire $W_N(1)$. Nous étudierons, plus généralement $W_N(t)$, mais uniquement pour des valeurs élevées de N , afin d'obtenir des résultats simples et utilisables, sous forme de limites. Pour cela, il est nécessaire de décrire le comportement de S_N quant $N \rightarrow \infty$. Nous supposerons simplement que :

$$\begin{aligned}
 (5.3) \quad & P(v_{N,n} \leq y) = P(v \leq y) = G(y), \quad \forall N, \quad n = 1, 2, \dots, N \\
 & E v_{N,n} = E v \\
 & \text{Var } v_{N,n} = \text{Var } v = \sigma^2 \\
 & (v_{N,n}) \text{ i.e.d. pour chaque } N
 \end{aligned}$$

Notons le fait important que la variable aléatoire $A^N(t)$ a une distribution binomiale, c'est-à-dire :

$$P(A_N(t) = k) = C_N^k \cdot F(t)^k \cdot (1 - F(t))^{N - k}$$

et que, donc, le taux moyen d'arrivées dans $[0, t]$ est :

$$\lambda(t) = E A_N(t) / t = N F(t) / t$$

L'intensité moyenne du trafic dans $[0, t]$ est :

$$\rho_N(t) = N E v F(t) / t$$

Théorème 5.2 : Si $F(\cdot)$ est continue sur $[0, 1]$ et $\liminf_{t \rightarrow 0^+} \frac{F(t)}{t} > 0$, alors :

$$\delta_N \xrightarrow{D} X, \text{ quant } N \rightarrow \infty$$

où :

$$\delta_N(t) = \frac{s W_N(t) - (N E v F(t) - s t)}{\sigma \sqrt{N}}, \quad 0 \leq t \leq 1$$

- X est un processus gaussien centré défini sur $[0, 1]$ et

caractérisé par sa fonction de covariance :

$$E X(t) \cdot X(s) = F(s) (E v^2 - (E v)^2 F(t)) / \sigma^2 \quad 0 \leq s \leq t \leq 1$$

Esquisse de la démonstration :

Contrairement aux démonstrations précédentes, nous ne pouvons pas nous servir ici d'un des théorèmes de limite centrale comme point de départ. Force nous est donc de montrer la convergence des distributions conjointes d'un certain processus η_N , puis de montrer que η_N satisfait à un critère supplémentaire, suffisant pour impliquer la convergence faible. (Nous nous servirons du critère exposé dans le théorème 1.4). Puis, nous montrons que : $\rho(\eta_N, \delta_N) \xrightarrow{D} 0$ pour les systèmes à serveur unique, et finalement, utilisant le système auxiliaire défini à la section 4.2. il nous est possible de généraliser le théorème pour s serveurs.

Définissons d'abord η_N comme aux chapitres 3 et 4 :

$$\eta_N(t) = \frac{\sum_{k=1}^{A_N(t)} v_{Nk} - NEv F(t)}{\sigma \sqrt{N}}, \quad 0 \leq t \leq 1$$

Alors, nous servant du fait que $A_N(t)$ a une distribution binomiale, il nous est possible, par des méthodes absolument classiques de Transformées de Laplace-Stieljes (TLS) d'écrire le développement limité de la TLS des distributions conjointes de η_N , ce qui nous permet de montrer que ces distributions conjointes convergent vers celles du processus X décrit plus haut.

Puis, il nous est possible de montrer que :

$$(5.4) \quad E (\eta_N(u) - \eta_N(v))^2 (\eta_N(v) - \eta_N(w))^2 \leq \frac{(Ev^2(F(u) - F(w)))}{\sigma^4}$$

ce qui satisfait au critère du théorème 1.4, avec $\alpha = \gamma = 2$.

(Notons aussi que la démonstration de (5.4) est simple puisque nous avons déjà calculé un développement limité des distributions conjointes de η_N .

Finalement, les deux dernières étapes de la démonstration suivent le schéma désormais classique ébauché aux chapitres précédents.

Remarques :

(i) l'hypothèse restrictive que N soit déterministe peut être relaxée, en lui attribuant une distribution au choix. Dans ce cas, le calcul de la TLS des distributions conjointes de η_N est plus complexe, mais nullement impossible. Nous avons effectué ces calculs lorsque N est soit poissonien, soit binomial, et le théorème 5.2 est toujours vrai, bien que la limite X soit différente.

(ii) Il peut sembler un peu cavalier d'exiger comme nous le faisons que $v_{N,n}$ reste constamment distribué comme $v, \forall N$, puisque ceci entraîne que $\rho_N \rightarrow \infty$, lorsque $N \rightarrow \infty$! Nous voulons faire deux remarques à ce sujet :

(a) Tout autre hypothèse sur $(v_{N,n})$ conduit à un modèle absurde : par exemple, si $v_{N,n}$ était distribuée comme $\frac{v}{N}$ (de façon à garder ρ_N constant $\forall N$), tous nos résultats sont encore valables intégralement, mais ils impliquent que $W_N(t)$ converge vers zéro. Un autre choix pourrait être que v_{Nn} soit distribuée comme v/N , et dans ce cas, la définition de δ_N est parfaitement raisonnable :

$$\delta_N(t) = \frac{sW_N(t) - (Ev\sqrt{N} F(t) - st)}{\sigma}, \quad 0 \leq t \leq 1$$

(notons cependant que, dans ce cas, $\rho_N(t) \rightarrow \infty$, aussi bien). Alors que conclure ? Tout d'abord que le choix des suites $(v_{N,n})$ est peu important pour les résultats mathématiques, et qu'il doit être dicté par les valeurs observées de $E v_n$ dans le système réel que nous voulons approximer (car n'oublions pas que la création d'une suite (S_N) de systèmes n'est qu'un stratagème pour approximer un système réel). En second lieu, il est tout à fait possible que N soit grand, sans que pour cela ρ_N soit très grand. Le chapitre 7 a d'ailleurs pour objet d'examiner par voie de simulation le champ pratique d'utilisation rationnelle de tous nos modèles.

En conclusion à cette section, notons seulement que le modèle étudié ici n'est qu'un premier pas vers l'étude de systèmes fortement non-stationnaires à caractère transitoire, champ qui jusqu'alors était resté du domaine de la simulation, mis à part quelques brillantes études de Newell.

5.4. Quelques autres processus importants

Nous n'avons pas parlé du nombre de clients à l'instant x , $Q(x)$, que Whitt et Iglehart (1969) ont étudié de façon très complète. Nous voulons simplement citer quelques résultats qui sont des conséquences directes des théorèmes ci-dessus :

(a) Le temps d'inactivité : soit $I(x)$ le temps total d'inactivité du serveur dans GI/G/1. Alors, (comme nous l'avons noté en commentant l'expression (3.2)), on peut écrire :

$$I(x) = - \inf_{0 \leq y \leq x} N(y)$$

ce qui, après normalisation, s'écrit :

$$\lambda_n(t) = \frac{I(nt)}{K_n} = - \inf_{0 \leq s \leq t} \zeta_n(s) \quad , \quad 0 \leq t \leq 1$$

et, puisque la fonction $f : D \rightarrow D$, définie par :

$f \circ \zeta(t) = - \inf_{0 \leq s \leq t} \zeta(s)$, est continue, on déduit que λ_n converge faiblement, ce que l'on cite :

Théorème 5.3

Si $\rho = 1$ dans GI/G/1, alors :

$$\lambda_n \xrightarrow{D} f \circ P$$

si $\rho > 1$ dans GI/G/1, alors :

$$\lambda_n \xrightarrow{D} 0$$

(b) Temps maximal d'attente, lorsque $\rho = 1$: Considérons la fonction $g : D \rightarrow D$, telle que: $g \circ x(t) = \sup_{0 \leq s \leq t} x(s)$. Elle est continue, et il découle du théorème 1.1 que :

$$g \circ \gamma_n \xrightarrow{D} g \circ B \circ P$$

En particulier : $g \circ \gamma_n(1) \xrightarrow{D} g \circ B \circ P(1)$

c'est-à-dire :

$$\frac{\max_{1 \leq k \leq n} W_k}{\sigma \sqrt{n}} \xrightarrow{D} \sup_{0 \leq t \leq 1} (P(t) - \inf_{0 \leq s \leq t} P(s))$$

(c) Moyenne des temps d'attente des n premiers clients :

La quantité $\bar{W}_n = \frac{1}{n} \sum_{i=1}^n W_i$ donne naissance à l'élément $\bar{\gamma}_n$ de D

$$\bar{\gamma}_n(t) = \frac{W_{[nt]} - (\rho-1) E u n t^2/2}{\sigma \sqrt{n}}, \quad 0 \leq t \leq 1$$

D'autre part, \bar{W}_n peut aussi s'écrire: $\bar{W}_{[nt]} = \int_0^t W_{[ns]} ds$, ce qui entraîne :

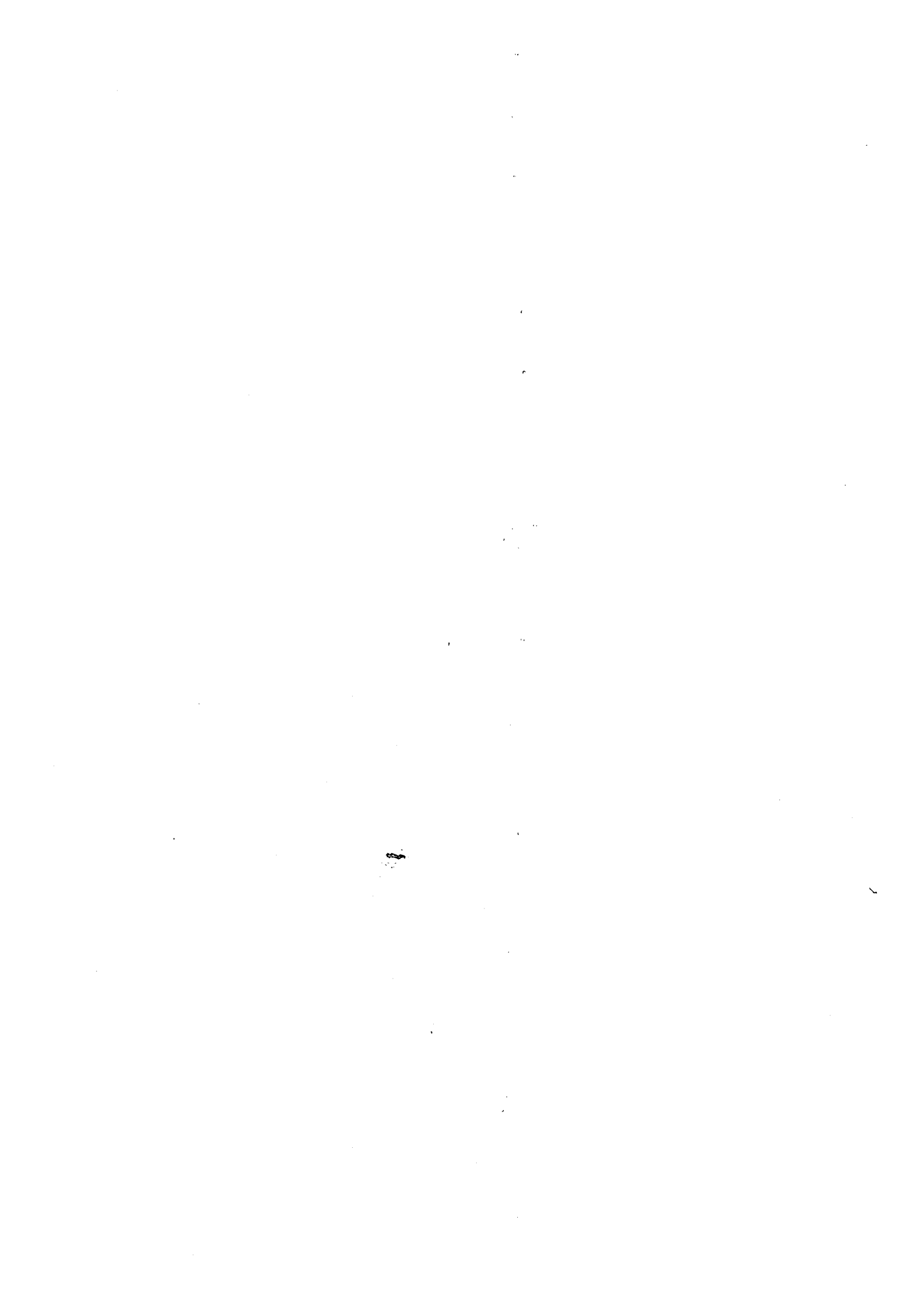
$$\bar{\gamma}_n(t) = \int_0^t \gamma_n(s) ds$$

La fonction $\Lambda : D \rightarrow D$ définie par : $\Lambda \circ x(t) = \int_0^t x(s) ds$ est continue et donc le théorème 1.1 permet de conclure :

Si $\rho = 1$, alors : $\bar{\gamma}_n \xrightarrow{D} \Lambda \circ B \circ P$

Si $\rho > 1$, alors : $\bar{\gamma}_n \xrightarrow{D} \Lambda \circ P$

Ce type de résultats montre bien la puissance de la convergence faible et augure bien de la convergence des diverses fonctions économiques construites sur les files d'attente.



CHAPITRE VI

LA PERIODE D'ACTIVITE

6.1. Introduction

Dans une file d'attente à un seul serveur, le serveur peut être soit occupé à servir (actif), soit inactif. Son état peut donc passer par des périodes successives d'activité et d'inactivité: $B_1, I_1, B_2, I_2, B_3, \text{etc...}$, illustrées par les figures 6.1, 6.2 :

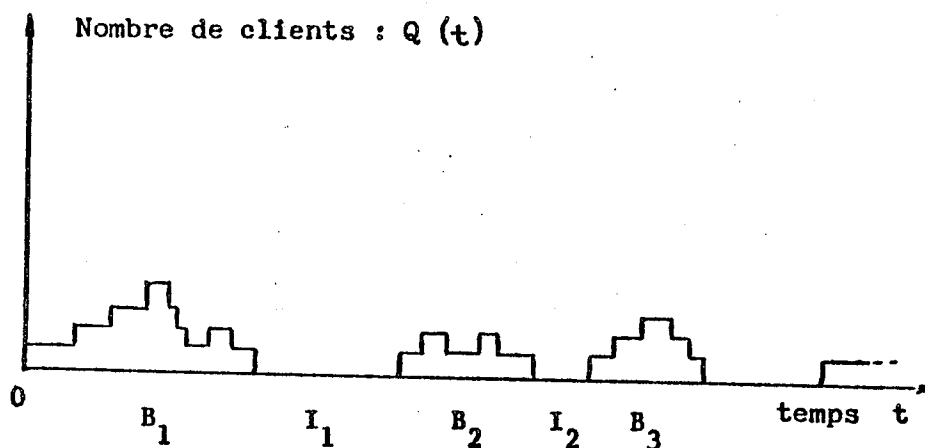


fig. 6.1

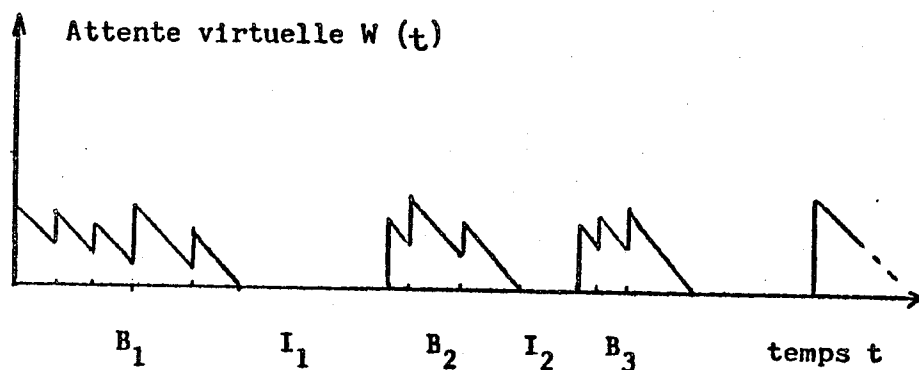


fig. 6.2

Pour un système GI/G/1 stable ($\rho < 1$), il existe une suite infinie de telles alternances qui constituent alors un processus de renouvellement alterné puisque au début de chaque nouvelle période d'activité, le système "repart à zéro", c'est-à-dire que ces instants sont des points de régénération pour le système (voir Smith (1955))

De plus, si $\rho < 1$, il est démontré que $E B < \infty$ et $E I < \infty$ avec :

$$(6.1) \quad E B / (E B + E I) = \rho$$

Ce processus de renouvellement alterné a une très grande importance en théorie des files d'attente, et a servi à établir de nombreux résultats (voir Prabhu (1965) et Jaiswal (1968) par exemple).

Peut-on en dire plus sur les variables aléatoires B_1 et I_1 ? Dans le cas des arrivées poissonniennes, il existe une équation implicite ayant comme solution unique la transformée de Laplace - Stieljes de B_1 , $\tilde{B}(s)$; plus précisément, définissons :

$$\tilde{G}(s) = E \exp(-s v_1)$$

v_K = durée de service du $K^{\text{ième}}$ client

λ = intensité de processus de Poisson des arrivées

$$\rho = \lambda E v_1$$

alors (voir par exemple Prabhu (1965)) :

$$(6.2) \quad \tilde{B}(s) = \tilde{G}(s + \lambda - \lambda \tilde{B}(s))$$

De la relation (5.2) on peut déduire :

$$(6.3) \quad E B = -\frac{d\tilde{B}(0)}{ds} = \frac{1/\lambda}{1-\rho}$$

La relation (6.3) n'a pas été généralisée à GI/G/1. Nous montrerons cependant dans la section suivante l'inégalité :

$$(6.4) \quad E B \leq \alpha \frac{1}{1-\rho}$$

Des relations telles que (6.3), ou, à défaut, (6.4) sont très utiles par exemple dans les études de simulation ; en effet, nous savons maintenant que les simulations de files d'attente devraient couvrir un nombre entier N de cycles d'activité pour utiliser des méthodes efficaces d'estimation. Il est donc important de pouvoir prévoir (au moins approximativement) le temps moyen nécessaire pour couvrir les N cycles choisis. Ce temps moyen

est égal à $N E v / (1-\rho)$ si (6.3) est vrai et est borné inférieurement par $\frac{\alpha N E v}{1-\rho}$ si (6.4) est vrai. Dans tous les cas, il est important de savoir que, lorsque ρ tend vers 1, ce temps moyen ne croît pas plus rapidement que $1/(1-\rho)$, toutes choses égales par ailleurs.

Dans les systèmes à plusieurs serveurs, et en particulier le système GI/G/s, la situation est loin d'être aussi claire, car la notion même d'inactivité est plus difficile à préciser : en effet, certains serveurs peuvent être inactifs pendant que d'autres sont actifs. L'existence de périodes I_1, I_2, \dots d'inactivité totale des s serveurs serait quand même importante, car cela assurerait que le système GI/G/s est régénératif. En général cependant, un système GI/G/s, même stable, peut fort bien n'être jamais totalement vide. Whitt (1973) en donne un exemple simple et propose la relation suivante (6.5), qui est suffisante pour que GI/G/s ($\rho \leq 1$) possède une infinité de périodes d'inactivité totale, chaque cycle d'occupation ayant alors une espérance finie si $\rho < 1$, et infinie si $\rho = 1$.

$$(6.5) \quad \text{Prob } (u_1 > v_1) > 0$$

Il serait très utile, en supposant (6.5), de trouver des relations telles que (6.2) (6.3) ou (6.4) dans le cas à s serveurs. Malheureusement ceci est extrêmement ardu et nous présenterons à la section (6.3) une remarque à ce sujet.

6.2. Une borne supérieure pour E B dans le système GI/G/1

Théorème 6.1 soit un système GI/G/1 défini par 1 suites (u_n) et (v_n) :

Si $E |v - u|^3 < \infty$, alors :

$$E N \leq \frac{H}{E u (1-\rho)} \exp \left(\frac{K E |v-u|^3}{\sigma^3} + \frac{L (1-\rho)}{E u \cdot \sigma} \right)$$

où

$$\sigma^2 = \sigma_u^2 + \sigma_v^2$$

N = nombre de clients servis durant une période d'activité du serveur

H, K, L sont des constantes numériques qui peuvent être prises égales à 0,9 6,3 et 0,8 respectivement.

Remarque : si B et I sont les durées d'une période d'activité et d'une période d'inactivité; respectivement, alors il est clair que

$$EB = Ev EN$$

$$EI = Ev EN (1-\rho) / \rho$$

et donc les quantités E B et E I ont des bornes déduites de celle de E N. En particulier, lorsque $\rho \rightarrow 1^-$, E I reste finie.

Esquisse de démonstration : Posons d'abord : $X_n = v_n - u_n$.

Notre résultat est basé sur une relation établie par Sparre et Andersen (1953) et que nous citons sans démonstration sous forme de :

Lemme : soit une suite de variables aléatoires I. E. D, (X_n) , telles que $E X < 0$, et soient $S_n = \sum_{i=1}^n X_i$, les suites partielles correspondantes. Définissons N comme le premier index k

tel que: $S_k < 0$. Alors E N existe et :

$$\text{Log } E N = \sum_{n=1}^{\infty} (1/n) P(S_n \geq 0)$$

Pour utiliser ce lemme, il suffit d'obtenir une borne supérieure suffisamment serrée pour $P(S_n \geq 0)$. Ceci est accompli en exploitant le fait que $\left(\frac{S_n + a n}{b \sqrt{n}} \right)$ tend en distribution vers N(0, 1) pour les valeurs appropriées de a et b. De plus, il existe une borne supérieure sur la vitesse de convergence. En combinant ces deux faits, on

arrive à la borne voulue, après des manipulations techniques peu instructives.

6.3. Une conjecture sur E B dans GI/G/s :

Conjecture :

Pour un système GI/G/s tel que $\rho < 1$ et $P(u_n > v_n) = \alpha > 0$ nous conjecturons qu'il existe une borne supérieure pour E B, en tous points semblable à celle obtenue pour GI/G/1, aux constantes près, et où X_n serait définie par : $X_n = v_n - s \cdot u_n$

Remarque

Lorsque, dans GI/G/s, l'un des serveurs est libre, alors il n'y a personne dans la file d'attente, et donc, la charge totale du système est constituée, au plus, de (s - 1) services. Il est donc naturel de conjecturer que la probabilité que cette charge s'épuise avant l'arrivée du client suivant est bornée inférieurement par une constante fixe, $\beta > 0$. Si cela était prouvé, alors on pourrait écrire :

$$EB \leq \frac{EB^1}{\beta}$$

où : B^1 = période d'activité du 1^{er} serveur

B = période d'activité de GI/G/s

Il resterait alors à établir une borne supérieure pour EB^1 (ce qui ne découle pas directement du théorème 6.1, puisque le 1^{er} serveur ne constitue pas un système GI/G/1 ; en effet ses arrivées ont des écarts équidistribués, mais non indépendants).

CHAPITRE VII

QUELQUES CONSIDERATIONS PRATIQUES

7.1. Introduction

Ce chapitre est assez différent des autres parties de cette thèse, à la fois dans son objet et dans sa méthodologie. Dans son objet, car il s'agira ici, non plus d'obtenir des résultats théoriques, mais des règles pratiques d'application de ces résultats ; ainsi, nous essaierons de nous rapprocher d'une réponse utilisable à la question des vitesse de convergence des théorèmes énoncés aux chapitres 3,4, et 5. En effet, s'il est essentiel de connaître la distribution-limite d'une suite de variables aléatoires (X_n) il est aussi très important de savoir pour quelles valeurs de n cette distribution pourra servir d'approximation valable à celle de X_n , car, en pratique un système est étudié pour des valeurs finies et non infinies de n . De plus il nous faudra clairement établir comment les théorèmes de la section 3.4 relatifs aux suites de systèmes sont utilisables.

Quant à l'approche à utiliser pour répondre à ces questions, il serait bien sûr souhaitable d'obtenir des bornes théoriques générales sur les vitesses de convergence. Une telle borne existe, par exemple, pour le Théorème de Limite Centrale des sommes partielles de variables aléatoires i.e.d. (X_n) avec $E|X_1|^3$. On peut en effet écrire (Esséen (1944)) :

$$(7.1) \quad \left| P\left(\frac{\sum_{i=1}^n X_i - n EX_1}{\sqrt{n \text{Var } X_1}} \leq x \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \right| \leq \frac{2,05 E|X_1|^3}{\sqrt{n} \cdot (\text{Var } X_1)}$$

Notons que (7.1) est vrai pour toutes les distributions possibles de X_n .

De tels résultats seraient malheureusement très difficiles à généraliser aux théorèmes de convergence sur les files d'attente, et une telle entreprise constituerait un important programme de recherche en soi. Il existe toutefois des résultats partiels sur ce sujet, que nous énoncerons à la section suivante.

Pour notre part, nous avons choisi de compléter cette thèse par une série de simulations qui tendent à établir des bornes empiriques sur les vitesses de convergence .

Nous examinerons les résultats théoriques disponibles à la section 7. puis nous définirons à la section 7.3 les conditions dans lesquelles nos simulations furent effectuées; enfin les sections 7.4 et 7.5 exposeront les résultats numériques, d'abord pour les systèmes GI/G/s , puis pour le système transitoire de la section 5.3 .

7.2. Bornes théoriques sur les vitesses de convergence dans les files d'a

En dépit d'un excellent travail d'unification accompli par Kennedy (1 la théorie est loin d'être achevée . Les meilleures bornes relatives aux processus décrits dans cette thèse sont celles de Kennedy et de Dudley(197 De façon générale, à chaque théorème de convergence faible de cette thèse correspond un taux de convergence d'ordre au moins égal à : $\log n / n^{1/4}$ Plus précisément, si nous prenons l'exemple typique de $\tau_n^n(1)$, défini au chapitre 4 pour les systèmes GI/G/s , nous pouvons affirmer que , si $E|u_1 - v_1|^5 < \infty$, alors, et dans les cas 2,3,4,5 seulement :

$$(7.2) \quad \left| P(\tau_n^n(1) \leq x) - P(\tau(1) \leq x) \right| \leq \frac{C_\sigma \log n}{n^{1/4}}$$

où $\tau_n^n(1)$ a pour limite $\tau(1)$ et C_σ dépend du cas étudié et des distributions de u_1 et v_1 . Remarquons que la borne (7.2) est moins bonne que (7.1) qui fait intervenir $n^{-1/2}$, et , de plus, (7.2) contient le coefficient inconnu C_σ . Dans un seul cas précis, celui de $\tau_n(1)$ pour un système GI/G/1 avec $\rho = 1$, Nagaev (1970) a réussi à établir une borne en $n^{-1/2}$, et ce à la seule condition : $E|v-u|^3 < \infty$.

$$\left| P\left(\frac{W_n}{\sigma_n^{1/2}} \leq x\right) - \sqrt{\frac{2}{\pi}} \int_0^x e^{-y^2/2} dy \right| \leq \frac{C}{n^{1/2}}$$

Ce résultat nous fait naturellement conjecturer que toutes les bornes

devraient être du même ordre $n^{-1/2}$, mais ceci reste à prouver. De plus, les coefficients C ne sont pas explicites.

Nous ne nous attarderons donc pas plus longuement sur ce sujet, dont un résumé figure dans Whitt (1973), pour nous tourner vers des méthodes pratiques.

7.3. Considérations générales sur la simulation des systèmes saturés

Dans toutes nos simulations, nous nous sommes contenté d'observer le temps d'attente du $n^{\text{ième}}$ client, W_n , afin de ne pas multiplier les calculs.

Pour simuler un système GI/G/s, il suffit de connaître les distributions de u_n et de v_n , ainsi que le nombre s de serveurs. Pour s nous avons choisi les valeurs 1, 3, et 10, afin de noter d'éventuelles différences dues à ces différentes distributions de u_n et v_n , nous avons choisi un éventail assez large de lois usuelles et nous les avons combinées deux à deux. En voici la liste :

- Loi à densité triangulaire sur $[a, b]$, et dont le mode est égal à m
- Loi Gamma (avec comme cas particulier la loi exponentielle).
- Loi à densité polynomiale : $f(x) = a x^b$ (avec comme cas particulier la loi uniforme).
- Lois discrètes diverses, dont les 3 premiers moments peuvent être contrôlés en faisant varier les paramètres.

L'ensemble de ces lois couvre un large éventail de formes diverses et leurs moments peuvent être contrôlés de façon à pouvoir dégager des "lois" les plus générales possibles étant donné l'envergure limitée de cette étude.

Pour une valeur de s donnée et des lois bien déterminées pour u_n et v_n , nous avons effectué 100 simulations indépendantes du même système. Pour chaque simulation, nous avons observé les valeurs de W_n , $n=1, 2, \dots, N$. Pour chaque valeur de n , les 100 observations nous ont permis de tester l'hypothèse statistique H_0 suivante :

H_0 : " Les observations de $\mathcal{Y}(n)$ proviennent d'une population dont la loi probabiliste est celle de la variable aléatoire $\mathcal{Y}(1)$."

[où $\mathcal{Y}(1)$ sera la limite de l'un des cas 1 à 5, judicieusement choisi.]

Pour tester H_0 , nous avons eu recours au test non paramétrique de Kolmogorov-Smirnov à un niveau de confiance de 0,9 .

Pour chaque système ainsi simulé, nous avons noté la valeur n_0 minimum pour laquelle H_0 n'est pas rejetée . Puis nous avons essayé de découvrir comment n_0 variait en fonction des 3 premiers moments de u_n et de v_n , et ceci indépendamment des lois choisies, si possible .

7.4. Résultats empiriques pour les systèmes GI/G/s

1°) $\rho > 1$:

Lorsque $\rho > 1$, deux approximations peuvent être a priori considérées pour $\sigma_n(1)$, à savoir celles correspondant aux cas 4 ou 5 décrits à la section 3.5. Notons que si n est très grand, alors $(\rho - 1)\sqrt{n}$ est grand et le cas 5 s'impose, alors que pour des valeurs plus modérées de n , il peut être plus judicieux de choisir le cas 4 . Les résultats empiriques confirment ces hypothèses et permettent de découvrir les seuils de validité de chacune des approximations . Voici les résultats de nos expériences :

(a) Pour toutes les valeurs de ρ entre 1,01 et 2, et pour toutes les lois envisagées, on peut écrire

$$(7.3) \quad n_0 \approx 26 \frac{E^3(v_1 + su_1)}{\sigma^3}$$

où :

$$\sigma^2 = \sigma_u^2 + \sigma_v^2 / s^2$$

(b) La meilleure approximation est obtenue en utilisant la limite du cas 5 dès que;

$$(7.4) \quad (\rho - 1)\sqrt{n} \geq 3,1 \sigma$$

et celle du cas 4 dès que:

$$(7.5) \quad (\rho - 1)\sqrt{n} \leq 2,7 \sigma$$

Les deux approximations sont à peu près équivalentes dans l'intervalle (2).

Remarque:

Le fait que les relations (7.3) et (7.4) soient valables pour toutes les lois envisagées n'était pas prévisible a priori . En fait les coefficients

obtenus pour chaque loi étaient légèrement différents , et nous avons utilisé dans nos deux formules empiriques des valeurs de ces coefficients valides pour toutes les lois envisagées .

D'autre part, rappelons ici la fonction de répartition de la limite dans le cas 4 (Lévy (1948)) :

$$(7.6) \quad P (B_0(P+\beta) (1) \leq x) = \Phi \left\{ \frac{x - ct/\sigma}{\sqrt{t}} \right\} - e^{2cx/\sigma} \cdot \Phi \left\{ \frac{-x - ct/\sigma}{\sqrt{t}} \right\}$$

où:

$$\bullet \beta(t) = ct/\sigma$$

• c est approximé par $(\rho - 1) \cdot \sqrt{n}$

• Φ est la fonction de répartition de la loi normale centrée rédu

2°) $\rho = 1$:

Ce cas n'admet qu'une approximation : B o P(1) , et nous avons vérifié que pour toutes les lois étudiées, cette limite est acceptée lorsque $n \gg n_0$ avec :

$$(7.7) \quad n_0 = \frac{17E^3(v+su)}{\sigma^3}$$

3°) $\rho < 1$:

Cette situation est presque symétrique à celle $\rho > 1$, c'est à dire que l'approximation pourra être soit celle du cas 1 soit celle du cas 2 . Ici cependant , si ρ est trop éloigné de la valeur unité, aucune de ces deux approximations n'est applicable. En plus donc de relations du type de (7.3) et (7.4) , nous avons obtenu une valeur minimale de ρ , ρ_0 . Il s'avère malheureusement que cette valeur minimale est très sensible aux lois de u et v . Voici les résultats numériques :

(a) Pour tous les systèmes étudiés, la distribution de $\sigma_n(1)$ accepte l'une des approximations des cas 1 ou 2 si :

$$(7.8) \quad \rho \gg \rho_0 = 0,93$$

et

(b) Le cas 1 doit être choisi si:

$$(7.10) \quad (1 - \rho) \sqrt{n} \geq 4.4 \sigma$$

et le cas 2 sinon (Rappelons que le cas 1 conduit à la loi exponentielle et le cas 2 à celle définie par (7.6) où c est remplacé par $-c$.)

Remarque

Suivant les lois adoptées pour u et v , la valeur critique ρ_0 varie de 0,67 à 0,93, cette dernière valeur étant donc valide pour toutes les lois considérées . Ce résultat est donc assez décevant en raison de la valeur élevée de ρ_0 .

7.5. Simulation du système transitoire de la section 5.3

Rappelons que ce système est défini par s , la distribution de v_n , et la fonction de répartition $F(t)$, $0 \leq t \leq 1$. Pour cette dernière , nous avons choisi tour à tour une fonction à densité uniforme, triangulaire, et polynomiale. Dans la section 5.3, nous avons vu que $W_N(t)$ avait une distribution limite normale pour tout t ; pour tous les systèmes simulés, nous avons vérifié que cette approximation est acceptée pour toutes les valeurs de n supérieures à 50 , et pour toutes les valeurs de t telles que $\rho_N(t) \geq 1$.

7.6. Conclusion

La forme même des résultats de ce chapitre montre que cette étude empirique ne peut en aucun cas remplacer une étude théorique des vitesses de convergence, étude qui actuellement s'annonce lente et ardue. Cependant, nous pouvons d'ores et déjà conclure (avec les distributions "usuelles" pour u et v) que les approximations décrites dans ce chapitre sont un outil à la fois simple et d'application assez générale aux systèmes saturés ou proches de la saturation .

BIBLIOGRAPHIE

- BILLINGSLEY ,P. "Weak convergence of probability measures" , John Wiley and Sons , New-York , 1968 .
- BOROVKOV , A. "Some limit theorems in the theory of mass service, I " Theory Proba. Appl. vol 9 pp.550-565 (1964)
- BOROVKOV , A. "Some limit theorems in the theory of mass service, I " Theory Proba. Appl. vol. 10 pp.375-400 (1965)
- BOROVKOV , A. "Asymptotic Analysis of service systems" Theory Proba. Appl vol. 11 pp.596-602 (1966)
- DONSKER , M. "An invariance principle for certain probability limit theorems" Mem. Amer. Math. Soc. vol. 6 (1951)
- DUDLEY , R.M. "Speeds of metric probability convergence" Z. Wahrscheinlichkeitstheorie verw. Geb. vol. 22 pp.323-332 (1972)
- ESSEEN , G. "Fourier analysis of distribution functions , a mathematical study of the Laplace-Gaussian law" Acta Math. (1944)
- IGEHAERT , D. ET WHITT , W. " Multiple channel queues in heavy traffic I et II " Adv. Appl. Proba. vol. 2 pp. 150-177 et 355-369 (1970)
- JAISWAL, N.K. "Priority Queues" Academic Press New-York (1968)
- KENNEDY , D.P. " Rates of convergence for queues in heavy traffic I and II" .Adv. Applied Proba. vol. 4 pp. 357-391 (1972)
- KIEFER ,J. et WOLFOWITZ , J. "On the theory of queues with many servers" Trans.Amer.Math.Soc. vol. 78 pp.1-18 (1955)
- KINGMAN , J. "The single server queue in heavy traffic " Proc.Camb.Phil. Soc. vol. 57 pp.902-904 (1961)
- KINGMAN , J. "On queues in heavy traffic" J.Roy.Stat.Soc. Ser. B vol.24 pp.383-392 (1962)
- KÖLLERSTRÖM , J. " Heavy traffic theory for queues with several servers " Working Paper , Univer. Oxford (1973)

- LEVY , P. " Processus stochastiques et mouvement brownien "
Gauthier-Villard Paris (1948)
- LINDLEY , D. " Theory of queues with a single server " Proc.Camb.Phil.Soc.
vol. 48 pp.277-289 (1952)
- LOÈVE , M. " Probability Theory " Van Nostrand (19)
- NAGAEV , S.V. "On the speed of convergence in a boundary problem I et II "
Th.Proba.Appl. vol. 15 pp.179-199 et 419-441 (1970)
- PRABHU , N.U. " Queues and Inventories " John Wiley New-York (1965)
- PROHOROV , Y. " Transient phenomena in the theory of mass service "
LitovskMath.Sb. vol. 3 pp.199-205 (1963)
- REICH , E. "On the integro-differential equation of Takács I " Annals of
Math.Stat. vol.29 pp.563-570 (1958)
- SMITH, W.L. " Regenerative stochastic processes " Proc.Roy.Soc. A
vol.232 pp.6-31 (1955)
- SPARRE ANDERSEN " On the fluctuations of sums of random variables "
Math.Scand. vol. 1 pp. 263-285 et vol.2 pp.195-223 (1954)
- WHITT ; W. " Weak convergence theorems for queues in heavy traffic "
Ph.D. Thesis, Dept. of O. R. Stanford Univ. (1968)
- WHITT , W. " Embedded renewal processes in the GI/G/s queue " J.Appl.Prob.
vol.9 pp.650-658
- WHITT , W. " Heavy traffic limit theorems for queues : a survey "
Mathematical Methods in Queueing Theory Springer-Verlag N-Y (1973)

ANNEXE I (L4)

AN ANALYTICAL MODEL FOR PEAK-HOUR
TRANSIENT MULTICHANNEL SERVICE SYSTEMS

by

RICHARD J. LOULOU

FACULTY OF MANAGEMENT
MCGILL UNIVERSITY

AN ANALYTICAL MODEL FOR PEAK-HOUR
TRANSIENT MULTICHANNEL SERVICE SYSTEMS

Richard J. Loulou
Faculty of Management
McGill University

Abstract

This paper studies the transient behavior of a multi-server queueing system under heavy traffic conditions. The system is examined during the fixed time interval $[0, T]$ during which it is assumed that a large but known number N of customers arrive. To each value of N corresponds such a system, say S_N . Of particular interest is the limiting behavior of S_N as $N \rightarrow \infty$. Weak convergence theorems are obtained for stochastic elements of the space $D_{[0,1]}$ constructed upon the workload process and the virtual wait process. The limiting processes are gaussian processes on $[0,1]$ which are explicitly described by their covariance functions.

INTRODUCTION

This paper studies the transient behavior of a multi-server queueing system under heavy traffic conditions. The system is examined during the fixed time interval $[0, T]$ during which it is assumed that a large but known number N of customers arrive. To each value of N corresponds such a system, say S_N . Of particular interest is the limiting behavior of S_N as $N \rightarrow \infty$. The mathematical assumptions made about S_N are stated now:

A. Arrivals

The N customers' arrival epochs are I.I.D. random variables distributed as τ and

$$F(t) = P(\tau \leq t) \quad , \quad 0 \leq t \leq 1$$
$$F(0) = 0 \quad , \quad F(T) = 1$$

From now on we assume $T=1$, without loss of generality.

We will assume F to be continuous on $[0, 1]$.

By choosing a suitable function F , the pattern of arrivals during $[0, 1]$ can be adapted to various typical patterns (peak-hours, double-peaks, stationary, etc.).

The number of arrivals by t , $A_N(t)$, is clearly a binomial random variable with parameters $\{F(t), N\}$, i.e.

$$P\{A_N(t) = k\} = \binom{N}{k} (F(t))^k (1 - F(t))^{N-k} \quad , \quad 0 \leq t \leq 1$$

Hence the average arrival rate during $(0, t)$ is

$$\lambda_N(t) = EA_N(t)/t = NF(t)/t$$

B. Service Times: m identical servers process the customers in order of arrival. The k^{th} customer requires a service time v_{Nk} and the sequence $\{v_{Nk}, k = 1, 2, \dots, N\}$ is I.I.D. for each N .

We make the supplementary assumption that for every N and k , v_{Nk} is distributed as v , a random variable with $0 < E v < \infty$ and $0 < \text{var } v = \sigma^2 < \infty$. Let further $\tilde{v}(s)$ denote the Laplace-Stieljes transform (LST) of v_{Nk} , i.e.

$$\tilde{v}(s) = E \exp(-s v_{Nk}) = E \exp(-s v)$$

Remark: the above assumption implies that, as $N \rightarrow \infty$, the average traffic intensity becomes larger and larger. Indeed, if m is the number of servers, the average traffic intensity in $[0, T]$ is:

$$(1) \quad (1) \quad \rho_N(t) = \lambda_N(t) / m \cdot \mu = N E v F(t) / m t$$

This superficially disturbing fact is discussed in the conclusion

Let us now define the following quantities:

Gross Input: $g_N(t) = \sum_{k=1}^{A_N(t)} v_{Nk}$ = amount of time required to serve all arrivals by t

Net Input: $N_N(t) = g_N(t) - m t$

Work-load: $L_N(t)$ = amount of work facing the servers at time t
 [Note that if the m servers have never been idle during $(0, t)$, then $L_N(t) = N_N(t)$]

Virtual Wait: $W_N(t)$ = time a hypothetical customer arriving at time t would have to wait before undergoing service.

[Note that for $m=1$ $W_N(t) = L_N(t)$]

In the next section we shall use the weak convergence theory to prove our main results. To do so we need to modify the above quantities by renormalizing them into

$$(2) \quad \eta_N(t) = \frac{g_N(t) - NEv F(t)}{\sigma \sqrt{N}} = \frac{N_N(t) - [NEv F(t) - mt]}{\sigma \sqrt{N}}$$

$$(3) \quad \xi_N(t) = \frac{L_N(t) - [NEv F(t) - mt]}{\sigma \sqrt{N}}, \quad 0 \leq t \leq 1$$

$$(4) \quad \delta_N(t) = \frac{W_N(t) - [NEv F(t) - mt]}{\sigma \sqrt{N}}, \quad 0 \leq t \leq 1$$

These three random processes are elements of $D_{[0, 1]}$. We shall now use the weak convergence theory on $D_{[0, 1]}$ to prove the convergence theorems.

Since this approach has already been followed by Whitt,[1968], Igle and Whitt [1970], and Loulou [1973], we refer the reader to these papers for details and justifications. Most of the theoretical results used here are found in Billingsley [1968] referred to as [B] throughout. The weak convergence will be noted \xrightarrow{D} .

Convergence theorems when m=1

For one-channel systems $\xi_N = \delta_N$ and furthermore the following relation exists between $L_N(t)$ and $N_N(t)$ (Reich [1958])

$$(5) \quad L_N(t) = W_N(t) = N_N(t) - \inf_{0 \leq s \leq t} N_N(s)$$

To understand (5), note that $\inf_{0 \leq s \leq t} N_N(s)$ represents the idle time of the server during $[0, t]$. Using (5) we shall link the convergence of ξ_N to that of η_N . We study the latter now in our main result:

Theorem 1: If F is continuous on $[0, 1]$, then:

$$\eta_N \xrightarrow{D} X \text{ in } D_{[0, 1]}, \text{ as } N \rightarrow \infty$$

where η_N is defined by (2) and X is a Gaussian process on $[0, 1]$ with covariance function

$$E X(t)X(s) = F(s) [Ev^2 - (Ev)^2 F(t)]/\sigma^2 \text{ for } 0 \leq s \leq t \leq 1$$

Proof: We first show that the finite dimensional distributions of η_N converge to those of X . We then proceed to show that the sequence $\{\eta_N\}$ is tight in $D_{[0, 1]}$. Weak convergence of η_N to X will follow from a basic theorem of weak convergence in $D_{[0, 1]}$ ([B] theorem 15.1).

The symbol Σ stands for $\Sigma_{k=1}^n$ throughout this proof.

Let then: $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t_{n+1} = 1$

and set $\Delta_k = F(t_k) - F(t_{k-1})$, $k = 1, 2, \dots, n+1$

We first consider the vector

$$\{A_N(t_1), A_N(t_2) - A_N(t_1), \dots, A_N(1) - A_N(t_n)\}$$

It has a multinomial distribution with parameters

$\{\Delta_1, \Delta_2, \dots, \Delta_{n+1}; N\}$ since its k^{th} coordinate

represents the number of customers who arrived during

(t_{k-1}, t_k) . Therefore its n -fold Z transform is:

$$(6) \quad G_N(Z_1, \dots, Z_n) = [\Delta_1 Z_1 + \dots + \Delta_n Z_n + \Delta_{n+1}]^N$$

We now turn to the vector of increments of g_N :

$$\{g_N(t_1), g_N(t_2) - g_N(t_1), \dots, g_N(1) - g_N(t_n)\}$$

Since $\{v_{Nk}\}$ is an I.I.D. sequence, the n -fold L.S.T. of this

vector, $\tilde{g}_N(s_1, \dots, s_n)$ is obtained from $G_N(Z_1, \dots, Z_n)$

upon replacing Z_k by $\tilde{v}(s_k)$.

$$(7) \quad \tilde{g}_N(s_1, \dots, s_n) = [\tilde{v}(s_1)\Delta_1 + \dots + \tilde{v}(s_n)\Delta_n + \Delta_{n+1}]^N$$

Finally the vector of increments of η_N

$$\{\eta_N(t_1), \eta_N(t_2) - \eta_N(t_1), \dots, \eta_N(1) - \eta_N(t_n)\}$$

has L.S.T.

$$\tilde{f}_N(s_1, \dots, s_n) = \tilde{g}_N(s_1/\sigma\sqrt{N}, \dots, s_n/\sigma\sqrt{N}) \cdot \exp\left(\frac{Ev\sqrt{N}}{\sigma} \sum s_k \Delta_k\right)$$

(to see this, note that $E \exp(-s(x-a)/b) = e^{sa/b} E \exp(-sx/b)$)

we thus obtain

$$(8) \quad \log \tilde{f}_N(s_1, \dots, s_n) = \frac{Ev\sqrt{N}}{\sigma} \sum s_k \Delta_k + N \ln(\Delta_{n+1} + \sum \Delta_k \tilde{v}(s_k/\sigma\sqrt{N}))$$

The McLaurin development of \tilde{v} is

$$\tilde{v}(s_k/\sigma\sqrt{N}) = 1 - Evs_k/\sigma\sqrt{N} + Ev^2 s_k^2 / 2\sigma^2 N + o_k(1/N)$$

hence, the second term of the RHS of (8) becomes

$$\begin{aligned} N \log \{ \sum \Delta_k - (Ev/\sigma\sqrt{N}) \sum s_k \Delta_k + (Ev^2/2\sigma^2 N) \cdot \sum s_k^2 \Delta_k + o(1/N) \} \\ = N \{ (-Ev/\sigma\sqrt{N}) \sum s_k \Delta_k + (Ev^2/2\sigma^2 N) \cdot \sum s_k^2 \Delta_k \\ - [(Ev)^2/2\sigma^2 N] \sum s_k^2 \Delta_k^2 + (1/\sigma^2 N) \sum_{k < l} s_k s_l \Delta_k \Delta_l \} + o(1) \end{aligned}$$

Hence

$$\begin{aligned} \log \tilde{f}_N(s_1 \dots s_n) = (1/2\sigma^2) \sum s_k^2 \Delta_k (Ev^2 - \Delta_k (Ev)^2) \\ - [(Ev)^2/\sigma^2] \sum_{k < l} s_k s_l \Delta_k \Delta_l + o(1) \end{aligned}$$

and, as $N \rightarrow \infty$, $o(1) \rightarrow 0$, and

$$(9) \quad \text{Log } \tilde{f}_N(s_1 \dots s_n) \rightarrow \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n C_{kl} s_k s_l$$

which shows that the finite dimensional distributions of η_N

converge to those of a Gaussian process X with zero mean,

and whose covariance function can be found from (9)

namely,

$$EX(s) \cdot X(t) = (1/\sigma^2) F(s) [Ev^2 - (Ev)^2 F(t)] \quad \text{for } s \leq t$$

This completes the first part of the proof.

Tightness of $\{\eta_N\}$ is established using the Product-Moment Criterion ([B] theorem 15.6) which states that $\{\eta_N\}$ is tight in $D_{[0,1]}$ if there exists a non-decreasing continuous function $H(t)$ on $[0, 1]$, and real numbers $\gamma > 1$, $\alpha \geq 0$, such that:

$$(10) \quad E[\eta_N(u) - \eta_N(v)]^\alpha [\eta_N(v) - \eta_N(w)]^\alpha \leq [H(u) - H(w)]^\alpha$$

for every $0 \leq w \leq v \leq u \leq 1$

for $\alpha = 2$, the LHS of (10) is equal to $\left. \frac{\partial^4 \tilde{f}_N(s_1, s_2)}{\partial s_1^2 \partial s_2^2} \right|_{s_1=s_2=0}$

if the latter exists.

where: $\tilde{f}_N(s_1, s_2)$ is the two fold L.S.T.

The result of the derivations is equal to

$$\frac{(Ev^2)^2}{\sigma^4} \Delta_1 \Delta_2 \left(1 - \frac{1}{N}\right) + (\Delta_1 \Delta_2^2 + \Delta_2 \Delta_1^2) \frac{(Ev^2)^3}{N^2 \sigma^2} +$$

$$\frac{\Delta_1^2 \Delta_2^2}{\sigma^4} \left(\frac{(Ev^2)^2}{N^2 \sigma^4} - \frac{4(Ev)^4}{N} \right)$$

which is, for N large enough, bounded above by $\frac{(Ev^2)^2}{\sigma^4} \Delta_1 \Delta_2$,

which in turn is smaller than $\frac{(Ev^2)^2}{\sigma^4} (\Delta_1 + \Delta_2)^2 = \left[\frac{Ev^2}{\sigma^2} (F(u) - F(w)) \right]^2$

hence, the function $H(x) = \frac{Ev^2}{\sigma^2} F(x)$ satisfies the hypothesis

of the product-moment criterion with $\gamma = 2$, and thus $\{\eta_N\}$

is a tight sequence, which completes the proof of theorem 1.

We now turn the load processes ξ_N .

Theorem 2: If $\lim_{t \rightarrow 0^+} \text{Inf} \frac{F(t)}{t} > 0$ (possibly ∞)

Then: $\xi_N \xrightarrow{D} X$

Where ξ_N is defined by (3), and X is the limit of η_N described in theorem 1.

Proof: It is sufficient to show that $d_0(\xi_N, \eta_N) \xrightarrow{D} 0$ as $N \rightarrow \infty$, where d_0 is the Skorohod distance on $D[0, 1]$ ([B] pp. 111-116)

It is further sufficient that $\rho(\xi_N, \eta_N) \xrightarrow{D} 0$ since

a.s
 $\rho \leq d_0$ ([B] Chapter 3), where ρ is the distance defined by

$$\rho(Y, Z) = \text{Sup}_{0 \leq t \leq 1} |Y(t) - Z(t)|$$

From (5) we have that:

$$\rho(\eta_N, \xi_N) = \text{Sup}_{0 \leq t \leq 1} \frac{- \text{Inf}_{0 \leq s \leq t} N_N(s)}{\sigma \sqrt{N}}$$

$$(11) \quad \rho(\eta_N, \xi_N) = - \text{Inf}_{0 \leq t \leq 1} \frac{N_N(t)}{\sigma \sqrt{N}}$$

And from (2)

$$\rho(\eta_N, \xi_N) = \text{Inf}_{0 \leq t \leq 1} (\eta_N(t) + \frac{(\text{NEv} F(t) - t)}{\sigma \sqrt{N}})$$

but since $\lim_{t \rightarrow 0^+} \text{Inf} \frac{F(t)}{t} > 0$, (possibly ∞) then the ratio

$\frac{F(t)}{t}$ is strictly positive on $[0, 1]$, and thus it has a

strictly positive lower bound α on $[0, 1]$.

Hence, we have

$$\text{NEv} F(t) - t \geq (\text{NEv} \alpha - 1)t, \quad 0 \leq t \leq 1$$

and for N large enough

$$\text{NEv} F(t) - t \geq \frac{N}{2} \text{Ev} \alpha t$$

so that:

$$\rho(\eta_N, \xi_N) \leq - \text{Inf}_{0 \leq t \leq 1} (\eta_N(t) + \frac{\text{Ev} \alpha t}{2\sigma} \sqrt{N})$$

This last quantity converges weakly to zero as $N \rightarrow \infty$ (as was proved in Loulou [1971 b] lemma 3.3). Hence $\xi_N \xrightarrow{D} X$.

m-Server systems

The model of section 1 is now studied for $m > 1$. Clearly η_N has same definition as for $m=1$ and hence $\eta_N \xrightarrow{D} X$ by theorem 1. ξ_N however has a different expression, since (5) and hence (11) do not hold for multiserver systems. We shall therefore follow the approach used by Loulou [1971,1973], and define a sequence of auxiliary systems $\{S_N^*\}$. In system $\{S_N^*\}$ the service rate is either m/Ev or 0 , according to whether customers are present in system or not. This is equivalent to assuming that whenever the number present in system is k , $0 < k < m$, the k busy servers work then at a faster rate $m/(kEv)$ each, so that overall rate is m/Ev . Apart from this, S_N^* is identical to S_N . It is easy to show that if $L_N(t)$ is the workload at t for S_N and $L_N^*(t)$ for S_N^* , then:

$$\left| L_N(t) - L_N^*(t) \right| \leq m \text{ Max}_{1 \leq k \leq A_N(t)} v_{Nk}$$

(The proof is detailed in Loulou [1971 b], section 4). Hence, if ξ_N^* denotes the workload sequence for S_N^* , then

$$\begin{aligned} \rho(\xi_N, \xi_N^*) &= \frac{1}{\sigma\sqrt{N}} \text{Sup}_{0 \leq t \leq 1} \left| L_N(t) - L_N^*(t) \right| \\ &\leq \frac{m}{\sigma\sqrt{N}} \text{Max}_{1 \leq k \leq A(1)} v_{Nk} \end{aligned}$$

$$= \frac{m}{\sigma\sqrt{N}} \text{Max}_{1 \leq k \leq N} v_{Nk}$$

which is distributed as

$$\frac{m}{\sigma\sqrt{N}} \text{Max}_{1 \leq k \leq N} v_k$$

when $\{v_k\}_{k=1, N}$ is I.I.D. as v .

The last expression goes to zero in probability by a standard result of probability theory (see e.g. Breiman [1968]).

Hence $\rho(\xi_N, \xi_N^*) \xrightarrow{D} 0$ and ξ_N, ξ_N^* have same weak limit, if any.

But, from a weak load point of view, system S_N^* is equivalent to a one-server system. In particular (5) and (11) hold for S_N^* , and hence theorem 2 applies to ξ_N^* and thus to ξ_N . We state:

Theorem 3 If $\lim_{t \rightarrow a^+} \text{Inf} \frac{F(t)}{t} > 0$ (possibly ∞)

Then $\xi_N \xrightarrow{D} X$ in $D_{[0,1]}$

where ξ_N is defined by (3)

X is defined in theorem 1

For the sake of completeness we will state the equivalent of theorem 3 for the virtual wait processes.

Corollary 1: If we define

$$\delta_N(t) = \frac{m.W_N(t) - (NEvF(t) - mt)}{\sigma\sqrt{N}}$$

where $W_N(t)$ is the virtual wait at time t then, under the assumptions of theorem 3:

$$\delta_N \xrightarrow{D} X \quad \text{in } D_{[01]}$$

The complete proof of this corollary uses theorem 3, and will not be given here (see Loulou [1971 b], section 4). We conclude by saying that the same convergence theorems hold both for $m > 1$ and for $m=1$.

Conclusions

- a. Cost Functions: In many problems (e.g. design of facilities, cost-benefit analyses) it is desirable to know the distribution of some cost function which will act as an objective function for optimization purposes. As an example we consider the (time-wise) average virtual wait

$$\bar{W}_N = \int_0^1 W_N(t) dt$$

Then, the element of $D_{[01]}$ to consider is

$$\bar{\xi}_N = \int_0^1 \xi_N(t) dt$$

Since the functional \int_0^1 is a continuous mapping: $D_{[01]} \rightarrow$ Borel line, then we have that

$$\bar{\xi}_N \xrightarrow{D} \int_0^1 X(t) dt$$

by the continuous mapping theorem ([B] theorem 5.1).

Other cost functions may be devised.

b. Discussion of Assumptions

- (i) The assumption that the service times in System S_N , v_{Nk} are I.I.D. for all k and N seems (superficially) a bad one, since it theoretically implies that $\rho_N(t) \rightarrow \infty$ as $N \rightarrow \infty$. This is not quite the case however since we never need, in any of our proofs, that $\rho_N(t)$ be large. Indeed, we need

only that N be large (proofs of theorems 1, 2, 3), that $\frac{\sqrt{N} Ev t}{m\sigma}$ be large (proof of theorem 2), and that $\rho_N(t)$ be larger than 1.

The following plausible example shows that these two quantities may be large without making $\rho_N(t)$ large. For simplicity we take $F(t)=t$, $0 \leq t \leq 1$, and $m=1$

If the total time period is one hour, it is quite reasonable to choose service times such that $Ev = \sigma = 10^{-4}$ hour. With these figures, $\rho_N(t) = 10^{-4} N$, and $N = 1.5 \cdot 10^4$ makes $\rho_N = 1.5$. Thus, N is large, the quantity $\frac{\sqrt{N}Ev}{m\sigma}$ is large too (120), but ρ_N is only 1.5.

This example indicates that the conditions of convergence for theorems 1, 2 and 3 are likely to be reached for such values of N that will keep $\rho_N(t)$ reasonably small, although larger than one. Thus, in this respect, the present results are similar to previous weak convergence results by Whitt [1968], Loulou [1973], and others.

- (ii) It should also be noticed that if a different assumption is made regarding the distribution of v_{Nk} , our results still hold after adjusting the parameters Ev and σ . For example if v_{Nk} is distributed as $\frac{v}{N}$, ξ_N should be redefined as

$$\xi_N(t) = \frac{N(t) - [Ev F(t) - mt]}{\sigma/\sqrt{N}}, \quad 0 \leq t \leq 1$$

i.e. σ is replaced by σ/N and Ev by Ev/N . All theorems hold for the new ξ_N . δ_N is modified in a similar fashion.

(Note that this alternate assumption implies that L_N converges to zero in probability!) Another possible choice would be that v_{Nk} be distributed as $\frac{v}{\sqrt{N}}$ which gives

$$\xi_N(t) = \frac{L_N(t) - [Ev \sqrt{N} F(t) - mt]}{\sigma}, \quad 0 \leq t \leq 1$$

and the same theorems apply.

We believe that our assumption is the most natural to make, but as just shown, it is quite easy to depart from it for some applications.

- (iii) The assumption that N is not random would not be a bad one for the many situations where the standard deviation of the number of arrivals is small compared to its mean. The author is currently trying to relax this assumption by allowing N to be random with some given probability distribution.

BIBLIOGRAPHY

- [B] Billingsley, P. [1968]: Convergence of Probability Measure.
Wiley, New York.
- Iglehart, D. and Whitt, W. [1970]: Multiple Channel Queues in
Heavy Traffic (I and II). Advances in Applied Probability,
2, no. 1.
- Loulou, R. [1971a]: Weak Convergence for Multi-Channel Queues
in Heavy Traffic (Ph.D. Thesis). ORC Report. Operations
Research Center, University of California, Berkeley,
ORC 71-31.
- Loulou, R. [1971b]: General Classes of Multi-Channel Queues
in Heavy Traffic. Faculty of Management, McGill
University, November 1971.
- Loulou, R. [1973]: Multi-Channel Queues in Heavy Traffic.
Journal of Applied Probability. To appear.
- Reich, E. [1958]: On the Integrodifferential Equation of
Takacs, I. Ann. Math. Stat. 29
- Whitt, W. [1968]: Weak Convergence Theorems for Queues in
Heavy Traffic (Ph.D. Thesis). Technical Report No. 2.
Department of Operations Research, Stanford University.

ANNEXE II (L5)

An Explicit Upper Bound for

the Mean Busy Period in a GI/G/1 Queue

by

Richard Loulou*

McGill University

*Now visiting at Université de Grenoble

March 1977

Abstract

* * * * *

This short paper establishes an explicit upper bound for EB, the expected duration of a busy cycle in a stable GI/G/1 queueing system. The upper bound obtained involves the first three moments of the service and inter-arrival distributions. The bound extends easily to bounds for the expected durations of the busy period and the idle period. It is especially useful when the traffic intensity ρ approaches unity from below, in which case the mean busy cycle becomes infinite, but no faster than a constant times $1/1-\rho$. All bounds are readily computed and involve only numerical constants and the three moments mentioned above.

An explicit upper bound for the mean
busy period in a GI/G/1 Queue

* * * * *

Richard Loulou
McGill University

* * * * *

* * * * *

1. Introduction

This short paper establishes an explicit upper bound for EB, the expected duration of a busy cycle in a stable GI/G/1 queueing system. The upper bound obtained involves the first three moments of the service and inter arrival distributions. The bound extends easily to bounds for the expected durations of the busy period and the idle period. It is especially useful when the traffic intensity ρ approaches unity from below, in which case the mean busy cycle becomes infinite, but no faster than a constant times $1/1-\rho$. All bounds are readily computed and involve only numerical constants and the three moments mentioned above.

More precisely, consider a GI/G/1 system specified by two independent sequences of i.i.d. random variables (U_n) and (V_n) , $n=1, 2, \dots$, where U_n is the time elapsed between the n^{th} and $(n+1)^{\text{st}}$ arrivals, and V_n is the duration of the n^{th} service. Let $X_n = U_n - V_n$ and $\alpha = -EX_1$. The traffic intensity ρ is defined by $\rho = EV_1/EU_1 = 1 - \alpha/EU_1$.

Let us define the following quantities:

$$\sigma^2 = \text{Variance of } X_1$$

$$S_n = \sum_{k=1}^n X_k \quad ES_n = -n\alpha$$

N = the (random) number of customers served during the first busy cycle

B = the duration of the first busy cycle

T = the duration of the first busy period

I = the duration of the first idle period

Assuming that the first customer arrives at time zero, we have:

$$(1) \quad B = \sum_{k=1}^N U_k \quad , \quad EB = EU \cdot EN$$

$$(2) \quad T = \sum_{k=1}^N V_k \quad , \quad ET = EV \cdot EN$$

$$(3) \quad I = B - T \quad \quad EI = \alpha \cdot EN$$

In section 2, we derive an upper bound for EN , which will immediately provide upper bounds for EB , ET , and EI , by expressions (1), (2), (3) above.

2. The Upper Bound for EN

Theorem 1: If $E|X_1|^3 < \infty$ and $\rho < 1$, then $EN \leq H \frac{\sigma}{\alpha} \exp\left(\frac{K E|X_1|^3}{\sigma^3} + \frac{L\alpha}{\sigma}\right)$

where H and K are purely numerical constants which can be taken as

$$H = 0.9$$

$$K = 6.30$$

$$L = 0.8$$

Proof: We start from a well-known result for random walks, first obtained by Sparre-Andersen (1953) and stated in Feller (1966, p. 396). viz.

$$\text{Log EN} = \sum_{n=1}^{\infty} \frac{1}{n} \text{Prob}[S_n > 0]$$

(In fact, this formula was established for the expectation of the first strict descending ladder index N of a random walk with partial sums S_n , with $ES_n = -n\alpha < 0$, so that it applies directly in our case only if we agree that a busy period is considered completed if it is immediately followed by a strictly positive idle period. This convention deals with the case when $P[S_n = 0]$ is not necessarily zero.

Now, using the fact that the (X_n) are IID with finite mean and variance, we shall decompose the quantity $P[S_n > 0]$ in two terms, and find upper bounds for each. Calling Z a random variable with the standard normal distribution:

$$\begin{aligned} P[S_n > 0] &= P\left[\frac{S_n + \alpha_n}{\sigma\sqrt{n}} \geq \frac{\alpha\sqrt{n}}{\sigma}\right] \\ &\leq P\left[Z \geq \frac{\alpha\sqrt{n}}{\sigma}\right] + \left| P\left[Z \geq \frac{\alpha\sqrt{n}}{\sigma}\right] - P\left[\frac{S_n + \alpha_n}{\sigma\sqrt{n}} \geq \frac{\alpha\sqrt{n}}{\sigma}\right] \right| \end{aligned}$$

Berry (1941) and Esséen (1944) have shown that the term within the modulus signs is bounded above by¹:

$$\frac{2.05 E |X_1|^3}{\sigma^3 \cdot \sqrt{n}}$$

thus:

$$(4) \quad \text{Log EN} \leq \frac{2.05 E |X_1|^3}{\sigma^3} \sum_{n=1}^{\infty} n^{-3/2} + \sum_{n=1}^{\infty} \frac{1}{n} P\left[Z \geq \frac{\alpha\sqrt{n}}{\sigma}\right]$$

1. The constant 2.05 is an improvement due to Wallace, as mentioned in Feller (1966, p.516).

The first series of the RHS converges (see Dwight (1958))

whereas the second may be written

$$\sum_1^{\infty} a(n) = \sum_1^{\infty} \left(\frac{1}{n} \int_{\frac{\alpha\sqrt{n}}{\sigma}}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2) dt \right)$$

Since the general term $a(n)$ is positive and non increasing in n , we may use the corresponding integral as an upper bound for the series, viz.

$$\begin{aligned} \sum_1^{\infty} a(n) &\leq a(1) + \int_1^{\infty} a(x) dx \\ (5) \quad &\leq 1/2 + \int_1^{\infty} a(x) dx \\ &= 1/2 + \int_1^{\infty} \frac{dx}{x} \int_{\frac{\alpha\sqrt{x}}{\sigma}}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2) dt \\ (6) \quad &= 1/2 + (2\pi)^{-1/2} \int_{\alpha/\sigma}^{\infty} e^{-t^2/2} dt \int_1^{\sigma^2 t^2/\alpha^2} \frac{dx}{x} \\ (7) \quad &= 1/2 + (2\pi)^{-1/2} \int_{\alpha/\sigma}^{\infty} e^{-t^2/2} \log\left(\frac{\sigma^2 t^2}{\alpha^2}\right) dt \\ (8) \quad &= 1/2 + \frac{\alpha}{2\sigma\sqrt{2\pi}} \int_1^{\infty} \frac{\exp(-V\alpha^2/2\sigma^2) \cdot \log V}{\sqrt{V}} \cdot dV \end{aligned}$$

(5) holds since $a(1) \leq \int_0^{\infty} (2\pi)^{-1/2} e^{-t^2/2} dt = 1/2$.

(6) is obtained by inverting the order of integration, which is justified by Fubini Theorem since the integrand is positive. Finally, (8) is derived from (7) by a change of variable $t = \sqrt{V} \cdot \alpha/\sigma$. The integral in

(8) may be calculated as follows, setting $a = \alpha^2 / 2\sigma^2$

$$(9) \quad \int_1^{\infty} e^{-aV} V^{-1/2} \log V \, dV = \int_0^{\infty} e^{-aV} V^{-1/2} \log V \, dV + \int_1^0 e^{-aV} V^{-1/2} \log V \, dV$$

The first definite integral in the RHS of (9) is known (see Dwight 1961, p. 248) and equal to $-(\pi/a)^{1/2} (\log a + 2 \log 2 + c)$, where c is Euler's constant, $c \leq 0.5772158$. The second integral of the RHS of (9) is bounded above by (see Dwight 1961 p. 239).

$$(10) \quad \int_1^0 \frac{\log V}{\sqrt{V}} = 4$$

Putting together (4), (8), (9), (10) we obtain

$$\text{Log EN} \leq \frac{2.05 E |X_1|^3}{\sigma^3} \Sigma n^{-3/2} + 1/2 - 1/2 \left(\log \left(\frac{\alpha^2}{\sigma^2} \right) + 2 \log 2 + c \right) + 1$$

i.e.

$$\text{EN} \leq \frac{H\sigma}{\alpha} \exp \left(\frac{K E |X_1|^3}{\sigma^3} + \frac{L\alpha}{\sigma} \right)$$

which completes the proof of theorem 1.

3. Conclusions and Comments

(a) As noted above, theorem 1 and expressions (1), (2), (3) provide similar bounds for EB, ET, EI. For this last quantity, it is nice to note that EI remains finite if ρ approaches unity but $E |X_1|^3$ remains bounded.

(b) It should be noted that the bound on EN found in this paper is in itself a poor bound. Indeed, in the M/M/1 case, where EN is known and equal to $1/1-\rho$, our bound is approximately equal to $\frac{\exp(14)}{1-\rho}$.

Thus, the merit of our bound ^{resides} solely in that it shows that EN varies as $\frac{1}{1-\rho}$, up to a multiplicative constant. Thus, when $\rho \rightarrow 1^-$ and if $E|X|^3$ remains uniformly bounded, EN is of the order of magnitude of $1/1-\rho$. This result may prove very useful in the study of queues with high traffic intensity.

(c) We conjecture that similar bounds exist for GI/G/S queues under certain additional conditions. We know (see Whitt (1972)), that a general stable GI/G/S system need not have finite mean busy cycle, and a counter-example is given in Whitt (1972). However, EB is finite if $P(V_1 < U_1) > 0$. Under this same condition, we conjecture the existence of bounds for GI/G/s, similar to that of this paper.

BIBLIOGRAPHY

1. Berry, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variats - Trans. Amer. Math. Soc., vol. 49.
2. Dwight, H.B. (1958). Mathematical Tables of Elementary and Some Higher Mathematical Functions, Dover Publications, New York.
3. Dwight, H.B. (1961). Tables of Integrals and Other Mathematical Data, The MacMillan Company, New York.
4. Esséen, G. (1944). Fourier Analysis of distribution functions: a study of the Laplace - Gaussian Law - Acta Math., 1944.
5. Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley and Sons, New York.
6. Sparre-Andersen, E. (1953). On the fluctuations of sums of random variables - Math. Scand. Vol. 1, 263-285 and Vol. 2, 195-223 (1954).
7. Whitt, W. (1972). Embedded Renewal Processes in the GI/G/3 queue. Journal Appl. Prob., 9, 650-658.