**Zeitschrift:** Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker = Bulletin / Association des Actuaires Suisses = Bulletin / Association of Swiss Actuaries

**Herausgeber:** Vereinigung Schweizerischer Versicherungsmathematiker

**Band:** 67 (1967)

**Artikel:** Introduction à la théorie des phénomènes d'attente

**Autor:** Dubois, Philippe

**DOI:** https://doi.org/10.5169/seals-966952

# Introduction à la théorie des phénomènes d'attente

Par *Philippe Dubois*, Zurich

## Résumé

L'auteur expose, à l'aide d'un exemple simple, les procédés d'analyse en application dans la théorie des phénomènes d'attente.

## 1. Remarques préliminaires et hypothèses de base

Les phénomènes d'attente – que l'on désigne communément par «files d'attente» ou «queues» – se rencontrent non seulement dans la vie courante, mais aussi dans maints secteurs de l'économie et de la technologie. D'une manière générale, ils reposent sur le modèle suivant:

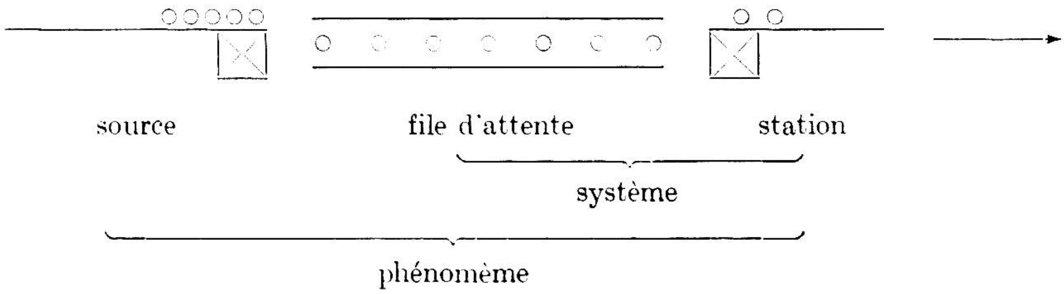
Des unités (personnes, objets) provenant d'une ou de plusieurs sources arrivent, à des intervalles de temps réguliers ou irréguliers, dans un système comprenant un centre d'attente et un centre de service. Le centre d'attente peut comporter une ou plusieurs voies d'accès; de même, le centre de service peut être constitué par une ou plusieurs stations. Chaque unité passe dans l'une des stations pour y bénéficier d'une certaine prestation de service ou pour y subir un certain traitement. La durée de service étant, en général, aléatoire, il se peut que les unités soient obligées d'attendre que l'une des stations soit disponible. Ces unités séjournent alors dans le centre d'attente et donnent lieu à une ou plusieurs files d'attente.

L'analyse de ce modèle montre que la nature d'un phénomène d'attente sera complètement caractérisée par l'indication des lois de distribution qui régissent les intervalles entre les arrivées des unités d'une part, et les durées aléatoires de service d'autre part.

Il s'ensuit que le traitement mathématique des phénomènes d'attente sera nécessairement fondé sur les méthodes du calcul des probabilités et sur la théorie des processus stochastiques.

Dans le cadre d'une introduction à la théorie des phénomènes d'attente, il ne saurait être question d'aborder l'étude de ces phénomènes dans toutes leurs généralités. Notre propos sera plutôt d'illustrer, à l'aide d'un exemple simple, les procédés d'analyse élaborés en vue du traitement mathématique de tels phénomènes. Plus précisément, nous examinerons le cas d'un système ouvert alimenté par des unités en nombre illimité provenant d'une seule source, le centre d'attente ne comprenant qu'une voie d'accès et le service étant assumé par une station unique. Les unités sont servies dans leur ordre d'arrivée après avoir éventuellement séjourné dans la file d'attente. Il est supposé que lorsque le centre de service est inoccupé le passage d'une unité du centre d'attente dans le centre de service s'effectue instantanément.

Le phénomène envisagé se présente schématiquement de la manière suivante:




En outre, il est supposé que le processus des arrivées et le processus de service sont du type poissonien. Ces hypothèses équivalent à supposer que les intervalles entre les arrivées des unités dans le système sont indépendants les uns des autres et sont distribués selon une loi exponentielle. De même, les durées de service sont indépendantes les unes des autres et sont également distribuées selon une loi exponentielle [5, vol. 1][1]).

Il résulte de ces hypothèses de base que

- la probabilité qu'une unité arrive dans le système dans l'intervalle de temps $dt$ est égal à $\lambda\, dt$, où $\lambda$ est le nombre moyen d'arrivées par unité de temps;

---

[1]) Les chiffres entre crochets se rapportent à la liste bibliographique à la fin du présent article.

- la probabilité que la fin de service d'une unité intervienne dans l'intervalle de temps $dt$ est égale à $\mu\, dt$, où $1/\mu$ représente le temps moyen de service;
- la probabilité que plusieurs arrivées ou plusieurs fins de service se produisent dans l'intervalle $dt$ est d'un infiniment petit en $dt$ d'ordre supérieur à 1; une telle probabilité peut être négligée.

En vue de l'analyse d'un tel phénomène, il importe que $\lambda < \mu$, car dans l'hypothèse contraire la durée moyenne de service $1/\mu$ serait supérieure à l'intervalle moyen de temps $1/\lambda$ entre les arrivées, et la file d'attente deviendrait alors de plus en plus longue avec le temps.

Les grandeurs intéressantes à déterminer dans le phénomène d'attente envisagé peuvent être rangées dans deux catégories:

1° Les grandeurs liées à l'état du système, par exemple le nombre d'unités dans le système à l'instant $t$. A ces grandeurs sont attachées diverses probabilités, par exemple la probabilité qu'il y ait $n$ unités dans le système à l'instant $t$.

2° Les grandeurs dont la définition ne fait pas intervenir l'état du système, par exemple le temps d'attente d'une unité dans la file. A ces grandeurs sont également attachées diverses probabilités, par exemple la probabilité d'un temps d'attente dans la file supérieur à une certaine valeur pour une unité arrivant à l'instant $t$.

A ces deux catégories de grandeurs correspondent deux méthodes d'analyse bien distinctes. La première dite «méthode différentielle» consiste à établir les équations d'état du phénomène envisagé sous la forme d'équations différentielles. La seconde dite «méthode intégrale» se ramène à la recherche d'équations intégrales permettant de définir certaines grandeurs caractéristiques du phénomène d'attente considéré.

## 2. Méthode différentielle

### *2.1. Matrice de transition et équations d'état*

En vertu des hypothèses énoncées au paragraphe 1 ci-dessus d'une part et des propriétés de la loi de Poisson d'autre part, il est important de remarquer que le phénomène d'attente étudié dans le présent article est du type markovien [3, 5]. En d'autres termes, l'état du système du phénomène à l'instant $t$ est complètement déterminé par la connaissance du nombre $n$ d'unités contenues dans ce système à l'instant considéré, et les prévisions que l'on peut être amené à formuler sur l'évolution future du phénomène ne peuvent être améliorées par la connaissance de données supplémentaires portant sur une période antérieure à l'instant $t$.

Compte tenu des remarques ci-dessus, il est facile d'établir la matrice de transition propre au phénomène d'attente étudié. A cet effet, il sera fait appel à la terminologie introduite dans la théorie des chaînes de Markov.

Soit $E_n$ l'état correspondant à $n$ unités dans le système à l'instant $t$. Les transitions pouvant se produire dans l'intervalle de temps $(t, t + dt)$ sont les suivantes:

$$E_0 \to E_0, \qquad E_0 \to E_1, \qquad E_1 \to E_0;$$
$$E_n \to E_n, \qquad E_n \to E_{n-1}, \qquad E_n \to E_{n+1}, \qquad n \geqq 1.$$

Soit $p_n(t)$, $n = 0, 1, 2, 3, \ldots$, la probabilité a priori de l'état $E_n$ à l'instant $t$ et

$$[p(t)] = [p_0(t), p_1(t), p_2(t), p_3(t), \ldots]$$

le vecteur d'état du système à l'instant $t$ avec

$$0 \leqq p_n(t) \leqq 1,$$
$$\sum_0^\infty {}_i\, p_i(t) = 1.$$

Compte tenu des hypothèses formulées au paragraphe 1, les probabilités de transition seront les suivantes:

| *Changement d'état* | | *Probabilités de transition* |
|---|---|---|
| $E_0 \to E_0$ | | $1 - \lambda\, dt$ |
| $E_0 \to E_1$ | | $\lambda\, dt$ |
| $E_1 \to E_0$ | | $(1 - \lambda\, dt)\,(\mu\, dt) \cong \mu\, dt$ |
| $E_n \to E_n$ , | $n \geqq 1$ | $(1 - \lambda\, dt)\,(1 - \mu\, dt) + (\lambda\, dt)\,(\mu\, dt) \cong 1 - \lambda\, dt - \mu\, dt$ |
| $E_n \to E_{n-1}$ , | $n \geqq 1$ | $(1 - \lambda\, dt)\,(\mu\, dt) \cong \mu\, dt$ |
| $E_n \to E_{n+1}$ , | $n \geqq 1$ | $(\lambda\, dt)\,(1 - \mu\, dt) \cong \lambda\, dt$ |

Dans ce qui précède, les infiniment petits en $dt$ d'ordre supérieur à 1 ont été négligés. C'est aussi la raison pour laquelle le changement d'état

$E_n \to E_{n-2}$ conduisant à une probabilité de l'ordre $(dt)^2$ n'a pas été envisagé dans le schéma ci-dessus.

Sur la base des probabilités de transition ainsi obtenues, la matrice carrée de transition peut s'écrire:

*Etat à l'instant $t + dt$*

| | | $n = 0$ | $n = 1$ | $n = 2$ | $n = 3$ | ... |
|---|---|---|---|---|---|---|
| *Etat à l'instant $t$* | $n = 0$ | $1 - \lambda\, dt$ | $\lambda\, dt$ | $0$ | $0$ | ... |
| | $n = 1$ | $\mu\, dt$ | $1-(\lambda+\mu)\, dt$ | $\lambda\, dt$ | $0$ | ... |
| | $n = 2$ | $0$ | $\mu\, dt$ | $1-(\lambda+\mu)\, dt$ | $\lambda\, dt$ | ... |
| | $n = 3$ | $0$ | $0$ | $\mu\, dt$ | $1-(\lambda+\mu)\, dt$ | ... |
| | . | . | . | . | . | ... |
| | . | . | . | . | . | ... |
| | . | . | . | . | . | ... |

Si l'on désigne par $[\tau]$ la matrice de transition, les équations d'état seront définies par la relation matricielle

$$[p(t+dt)] = [p(t)]\,[\tau],$$

ou sous une forme explicite:

$$p_0(t+dt) = (1-\lambda\, dt)\, p_0(t) + \mu\, dt\, p_1(t)$$

$$p_n(t+dt) = \lambda\, dt\, p_{n-1}(t) + (1-\lambda\, dt - \mu\, dt)\, p_n(t) + \mu\, dt\, p_{n+1}(t), \qquad \text{pour } n \geqq 1,$$

ou encore

$$\text{(A)} \qquad p_0'(t) = -\lambda p_0(t) + \mu p_1(t),$$

$$\text{(B)} \qquad p_n'(t) = \lambda p_{n-1}(t) - (\lambda+\mu)\, p_n(t) + \mu p_{n+1}(t), \qquad n \geqq 1.$$

Ces équations différentielles ne sont qu'un cas particulier des équations plus générales définissant les processus poissoniens dits de naissance et de mort dans lesquels les paramètres $\lambda$ et $\mu$ sont fonction de $n$. Elles peuvent être obtenues directement à partir des équations différentielles de Chapman-Kolmogorov, lesquelles jouent un rôle fondamental dans la théorie des chaînes de Markov.

La solution du système d'équations différentielles (A) et (B), compte tenu des conditions initiales

$$p_{n_0}(0) = 1 \qquad \text{et} \qquad p_n(0) = 0 \qquad \text{pour } n \neq n_0,$$

où $n_0$ désigne le nombre d'unités dans le système à l'instant 0, est assez compliquée et fait intervenir les fonctions de Bessel.

Dans le présent article, nous nous bornerons à indiquer la solution de ces équations dans le cas pratique important où le phénomène d'attente se trouve en régime permanent.

### 2.2. *Régime permanent*

La matrice de transition indiquée au chiffre 2.1. étant ergodique, c'est-à-dire n'étant ni réductible ni périodique, la probabilité $p_n(t)$ – selon un théorème classique [5, vol. 1] de la théorie des chaînes de Markov – tend, avec $t$ croissant, vers une limite indépendante de $t$ et du nombre initial $n_0$ d'unités dans le système. On a donc:

$$\lim_{t \to \infty} p_n(t) = p_n, \qquad n = 0, 1, 2, \ldots.$$

Le phénomène est alors dit se trouver en régime permanent.

Pour $t \to \infty$, les dérivées dans les équations d'état (A) et (B) tendent alors vers zéro. Il vient donc:

$$\text{(A')} \qquad \lambda p_0 = \mu p_1,$$

$$\text{(B')} \qquad (\lambda + \mu)\, p_n = \lambda p_{n-1} + \mu p_{n+1}, \qquad n \geqq 1.$$

On tire immédiatement de ces équations:

$$\begin{aligned} p_1 &= (\lambda/\mu)\, p_0 \\ p_2 &= (\lambda/\mu)^2\, p_0 \\ &\vdots \\ p_n &= (\lambda/\mu)^n\, p_0 \\ &\vdots \end{aligned}$$

D'autre part, on a

$$\sum_0^\infty p_i = 1,$$

d'où

$$1 = p_0[1 + (\lambda/\mu) + (\lambda/\mu)^2 + \ldots + (\lambda/\mu)^n + \ldots]$$

ou ($\lambda/\mu < 1$ par hypothèse)

$$1 = p_0 \frac{1}{1 - \lambda/\mu},$$

soit, en posant $\psi = \lambda/\mu$

$$p_0 = 1 - \psi.$$

Il vient alors

$$\text{(C)} \qquad p_n = (1 - \psi)\,\psi^n, \qquad n = 0, 1, 2, 3, \ldots.$$

La quantité $\psi$, appelée «coefficient d'utilisation» dans la théorie des phénomènes d'attente, est indépendante de l'unité de temps et joue un rôle important dans la pratique. Pour $\psi > 1$, le phénomène n'admet pas de régime permanent, la file d'attente devenant de plus en plus longue.

### *2.3. Grandeurs caractéristiques en régime permanent*

La probabilité que la variable aléatoire $N$ représentant le nombre d'unités dans le système soit inférieure ou égale à $n$ s'obtient facilement:

$$Pr\{N \leqq n\} = \sum_0^n p_i = \sum_0^n (1 - \psi)\,\psi^i,$$

$$Pr\{N \leqq n\} = (1 - \psi)\left[\frac{1 - \psi^{n+1}}{1 - \psi}\right] = 1 - \psi^{n+1}.$$

Il en résulte:

$$Pr\{N > n\} = \psi^{n+1}.$$

En particulier, la probabilité qu'une unité au moins se trouve dans le système est égale à

$$Pr\{N > 0\} = \psi.$$

Le nombre moyen $\bar{n}$ d'unités dans le système est égal à

$$\bar{n} = E(N) = \sum_0^\infty i\,p_i = (1 - \psi)\sum_0^\infty i\,\psi^i = \frac{\psi}{1 - \psi}.$$

Si $n$ désigne le nombre d'unités dans le système, le nombre d'unités $\nu$ dans la file d'attente sera évidemment :

$$\begin{aligned} \nu &= 0 \qquad \text{pour } n = 0, \\ \nu &= n-1 \quad \text{pour } n > 0. \end{aligned}$$

Le nombre moyen $\bar{\nu}$ d'unités dans la file d'attente sera donc égal à :

$$\bar{\nu} = \sum_{1}^{\infty} (i-1)\, p_i = \bar{n} - 1 + p_0 = \frac{\psi^2}{1-\psi}.$$

Comme autres grandeurs moyennes importantes, il convient de signaler le temps moyen d'attente d'une unité dans la file d'attente et le temps moyen d'attente d'une unité dans le système (temps de service compris). Ces grandeurs seront déterminées dans le paragraphe suivant.

## 3. Méthode intégrale

### *3.1. Equation intégrale du temps d'attente*

En vue de la construction de l'équation intégrale du temps d'attente relative au phénomène défini au paragraphe 1, il est judicieux d'individualiser les unités qui passent dans le système comprenant le centre d'attente et le centre de service. A cet effet, les unités seront numérotées dans leur ordre d'arrivée

$$1, 2, \ldots, n, \ldots$$

à partir de l'instant initial $t = 0$, et les définitions suivantes seront introduites :

Soit

| | |
|---|---|
| $t_n$ | l'instant d'arrivée de l'unité $n$ ; |
| $\tau_n = t_n - t_{n-1}$ | l'intervalle de temps entre les arrivées des unités $n-1$ et $n$ ; |
| $w_n$ | le temps d'attente de l'unité $n$ dans la file d'attente ; |
| $v_n$ | la durée de service de l'unité $n$ ; |
| $u_n = w_n + v_n$ | le temps d'attente de l'unité $n$ dans le système. |

Il est supposé que les unités qui entrent dans le système dans leur ordre d'arrivée en sortent également dans le même ordre.

On voit facilement que:

$$w_n = 0 \qquad \text{si } u_{n-1} - \tau_n \leqq 0,$$

$$w_n = u_{n-1} - \tau_n \qquad \text{si } u_{n-1} - \tau_n \geqq 0.$$

Soit

$$F_n(x) = Pr\{w_n \leqq x\}$$

respectivement

$$G_n(x) = Pr\{u_n \leqq x\}$$

la fonction de répartition du temps d'attente de l'unité $n$ dans la file, respectivement la fonction de répartition du temps d'attente de l'unité $n$ dans le système.

Soit encore

$$a(x)\,dx = Pr\{x \leqq \tau_n < x + dx\} = \lambda e^{-\lambda x}\,dx$$

respectivement

$$A(x) = Pr\{\tau_n > x\} = e^{-\lambda x}$$

la densité de probabilité respectivement la fonction de répartition complémentaire relative à l'intervalle $\tau_n$ entre les arrivées des unités $n-1$ et $n$ dans le système. Dans ces deux dernières relations, l'indice $n$ peut être supprimé, puisque par hypothèse (voir paragraphe 1 du présent article) les intervalles entre les arrivées des unités dans le système sont indépendants les uns des autres et sont distribués selon une même loi exponentielle. D'autre part, les variables $u_{n-1}$ et $\tau_n$ sont aussi indépendantes l'une de l'autre. En effet, $u_{n-1}$ représentant le temps d'attente de l'unité $n-1$ dans le système ne dépend que des intervalles de temps $\tau_{n-1}, \tau_{n-2}, \ldots$, desquels la variable $\tau_n$ est indépendante. Il vient donc:

$$Pr\{u_{n-1} - \tau_n \leqq x\} = \int_{\xi=0}^{\infty} Pr\{\xi \leqq \tau_n < \xi + d\xi\}\, Pr\{u_{n-1} \leqq x + \xi\}$$

ou, pour $x \geqq 0$

$$\text{(D)} \qquad F_n(x) = \int_0^{\infty} a(\xi)\, G_{n-1}(x+\xi)\, d\xi.$$

Soit

$$s(x)\,dx = Pr\{x \leqq v_n < x + dx\} = \mu e^{-\mu x}\,dx$$

respectivement

$$S(x) = Pr\{v_n > x\} = e^{-\mu x}$$

la densité de probabilité respectivement la fonction de répartition complémentaire relative à la durée de service $v_n$ de l'unité $n$, cette durée étant également indépendante de $n$ par hypothèse. On a alors:

$$
\text{(E)} \qquad \begin{aligned} G_n(x) &= Pr\{u_n \leqq x\} \\ &= \int_0^x Pr\{\xi \leqq v_n < \xi + d\xi\}\, Pr\{w_n \leqq x - \xi\} \\ &= \int_0^x s(\xi)\, F_n(x-\xi)\, d\xi . \end{aligned}
$$

Les équations (D) et (E) permettent de déterminer de proche en proche les fonctions $F_n(x)$ et $G_n(x)$, compte tenu des conditions initiales

$$
\text{(F)} \qquad \begin{aligned} F_1(x) &= 1 \\ G_1(x) &= 1 - e^{-\mu x}, \end{aligned}
$$

lesquelles expriment (en supposant que le nombre initial d'unités dans le système est nul) que l'unité $n = 1$ a un temps d'attente nul dans la file, et par conséquent un temps d'attente dans le système égal à la durée de service.

### *3.2. Régime permanent*

Les fonctions $F$ et $G$ étant indépendantes de $n$ en régime permanent, les équations (D) et (E) deviennent alors:

$$
\text{(D')} \qquad F(x) = \int_0^\infty a(\xi)\, G(x+\xi)\, d\xi = \int_0^\infty \lambda e^{-\lambda\xi}\, G(x+\xi)\, d\xi ,
$$

$$
\text{(E')} \qquad G(x) = \int_0^x s(\xi)\, F(x-\xi)\, d\xi = \int_0^x \mu e^{-\mu\xi}\, F(x-\xi)\, d\xi .
$$

En posant

$$x + \xi = t,$$

il vient pour (D')

$$F(x) = \int_x^\infty \lambda e^{-\lambda(t-x)}\, G(t)\, dt$$

ou encore

$$F(x) = \lambda e^{\lambda x}\left[\int_0^\infty e^{-\lambda t}\, G(t)\, dt - \int_0^x e^{-\lambda t}\, G(t)\, dt\right].$$

En particulier, pour $x = 0$, on a:

$$F(0) = \lambda \int_0^\infty e^{-\lambda t}\, G(t)\, dt.$$

Il vient alors:

$$F(x) = \lambda\, e^{\lambda x} \left[ \frac{F(0)}{\lambda} - \int_0^x e^{-\lambda t}\, G(t)\, dt \right].$$

En appliquant la méthode de transformation de Laplace à l'équation ci-dessus et à l'équation (E'), on voit aisément [3] qu'elles peuvent se mettre sous la forme explicite:

$$\text{(D'')} \qquad F(x) = F(0) \left[ \frac{\mu}{\mu - \lambda} - \frac{\lambda}{\mu - \lambda}\, e^{-(\mu-\lambda)x} \right],$$

$$\text{(E'')} \qquad G(x) = 1 - e^{-(\mu-\lambda)x}.$$

En posant $\psi = \lambda/\mu$, il vient pour (D")

$$F(x) = F(0) \left[ \frac{1}{1-\psi} - \frac{\psi}{1-\psi}\, e^{-(\mu-\lambda)x} \right].$$

Comme $\lim_{x\to\infty} F(x) = 1$ et $(\mu - \lambda) > 0$, on en déduit:

$$F(0) = 1 - \psi = p_0.$$

Il vient finalement pour (D'')

$$\text{(D''')} \qquad F(x) = 1 - \psi\, e^{-(\mu-\lambda)x}.$$

Il résulte de l'équation (D''') que la probabilité que la durée d'une attente dans la file soit supérieure à $x$ est égale à

$$\psi\, e^{-(\mu-\lambda)x}.$$

En régime permanent, le temps moyen d'attente d'une unité dans la file est égal à

$$\bar{w} = \int_0^\infty x\, dF(x) = \frac{1}{\mu}\, \frac{\psi}{1-\psi}.$$

De l'équation (E'') on déduit qu'en régime permanent le temps moyen d'attente d'une unité dans le système est égal à :

$$\bar{u} = \int_0^\infty x\, dG(x) = \frac{1}{\mu} \frac{1}{1-\psi}.$$

Le temps moyen de service d'une unité est égal à

$$\bar{v} = \bar{u} - \bar{w} = \frac{1}{\mu}.$$

Dans le cas où le phénomène d'attente envisagé se trouve en régime permanent, ce résultat était facilement prévisible.

## 4. Remarques finales

Le but du présent article était d'illustrer, à l'aide d'un exemple particulièrement simple, les méthodes utilisées dans la théorie des phénomènes d'attente. Ces méthodes ont été appliquées avec succès à des phénomènes plus généraux et ont permis d'aboutir à des résultats valables dans la pratique. Le lecteur s'intéressant à de tels problèmes pourra se reporter aux ouvrages figurant dans la liste bibliographique indiquée ci-après, notamment à l'ouvrage [3]. Dans cet ordre d'idées, il n'est pas inutile de mentionner les recherches entreprises ces dernières années par certains auteurs [1, 2, 5] en vue d'arriver à un traitement mathématique unifié de la théorie des phénomènes d'attente dans le cadre de la théorie moderne des processus stochastiques.

## Bibliographie

[1] *D.G.Kendall:* Stochastic processes occuring in the theory of queues and their analysis by the method of the inbedded Markov chain, Ann. Math. Stat. 24, 1953.

[2] *Barucha-Reid:* Elements of the theory of stochastic processes and their applications, McGraw-Hill, 1960.

[3] *Kaufmann et Cruon:* Théorie des phénomènes d'attente, Dunod, 1961.

[4] *Cox and Smith:* Queues, Methuen's Monographs, 1961.

[5] *Feller:* An introduction to probability theory and its applications Vol. 1 et 2, Wiley 1957/1966.

## Zusammenfassung

Mit Hilfe eines einfachen Modells legt der Verfasser die analytischen Methoden dar, welche in der Theorie der Warteschlangen zur Anwendung gelangen.

## Summary

Resting on a simple example, the author expounds the methods of analysis used in the theory of queues.

## Riassunto

L'autore expone, basandosi su un semplice esempio, i metodi di analisi usati nella teoria dei fenomeni di attesa.