Random Allocations:
New and Extended Models and Techniques
with
Applications and Numerics

Raymond William Kennington

B.Sc.(Ma.Sc.)(Hons)	University of Adelaide	1978
Grad.Dip.Ed.	University of Adelaide	1990
B.Comp.Inf.Sc.(Hons)	University of South Australia	1996

Thesis submitted for the degree

of

Doctor of Philosophy

in

Applied Mathematics

at

The University of Adelaide

(Faculty of Engineering, Computer and Mathematical Sciences)

School of Mathematical Sciences

September 2006

Trademarks

Athlon^{TM} is a trademark of *AMD*.

Borland® is a registered trademark of *Borland Software Corporation*.

Borland C++® is a registered trademark of *Borland Software Corporation*.

Delphi^{TM} is a trademark of *Borland Software Corporation*.

Celeron^{TM}, i.e. *Intel*® *Celeron*^{TM}, is a trademark of *Intel Corporation*.

Intel® is a registered trademark of *Intel Corporation*.

Maple® is a registered trademark of *Waterloo Maple Software*.

MuPAD^{TM} is a trademark of *SciFace Software GmbH & Co. KG*.

Scientific Word® was a registered trademark of *TCI Software Research, Inc.*

Scientific WorkPlace® is a registered trademark of *MacKichan Software, Inc.*

SET® and *SET Enterprises*® are registered trademarks of *SET Enterprises Inc.*

Windows® is a registered trademark of *Microsoft Corporation*.

T^3^{TM} *was a trademark of TCI Software Research, Inc*

All other brand and product names are trademarks of their respective companies.

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

List of Figures

List of Figures

List of Tables

List of Tables

List of Tables

General Notation and Conventions

Symbol	Meaning
$\mathbb{Z}^+$	the set of non-negative integers.
$\sim$	*is asymptotically approximately equal to.*
$\simeq$	*is calculated to be approximately equal to.*
$\equiv$	(a) *is equal to* for all relevant values of the index or indices; e.g. $m_i \equiv m$ means $m_i = m \; \forall i$ and $A_i \cap A_j \equiv G$ means $A_i \cap A_j = G$ for all $i \neq j$.
	(b) *is defined as* for all relevant values of the index; e.g. $\beta_u \equiv \lvert B_u \rvert$ means $\beta_u \overset{def}{=} \lvert B_u \rvert \; \forall u$.
$\leftarrow$	*is assigned the value of*, usually for substitution of variables in formulae.
$\rightarrow$	*tends to* or *has the limit.*
$\forall$	*for all.*
$o\left(g\left(N\right)\right)$	Landau's little-o of N: $f\left(N\right) = o\left(g\left(N\right)\right)$ means for each constant $c > 0$ there exists a constant $k\left(c\right) > 0$ such that $0 \leq f\left(N\right) < cg\left(N\right)$ for all $N \geq k$.
$O\left(g\left(N\right)\right)$	Landau's big-O of N: $f\left(N\right) = O\left(g\left(N\right)\right)$ means there exists constants c and k such that $0 \leq f\left(N\right) \leq cg\left(N\right)$ for all $N \geq k$.

$[x]_\ell$	The ℓth rising factorial of x.
$(x)_\ell$	The ℓth falling factorial of x.
$\delta_{i,j}$	Kronecker's delta function.
0^0	will be taken as 1 for convenience of representation of formulae.
$\lvert B\rvert$	the number of elements in the set B.
#	The number of ways in which an event can occur; e.g. $\#(T = k)$.
$\binom{N}{n}$	(a) the binomial coefficient.
	(b) the number of ways of choosing n objects from N distinct objects without repetition.
	(c) for $n \geq 0$, defined to be $\frac{(N)_n}{n!}$.
$\binom{N}{n_1,n_2,\ldots,n_r}$	the multinomial coefficient.
$\left\{ {n \atop i} \right\}$	(a) the Stirling number of second kind.
	(b) the number of ways of partitioning a set of n elements into i non-empty subsets.
	(c) the coefficients in the factorial polynomial form of a power, with $k^n = \sum_{i=1}^{n} \left\{ {n \atop i} \right\} (k)_i$.
$\dot{\bigcup}_{i=1}^{\gamma}$	disjoint union.
C^+	a count of the number of addition-like operations in a formula, where C is any letter of an alphabet.

$C^{\times}$	a count of the number of multiplication-like operations in a formula, where C is any letter of an alphabet.
$\mathcal{C}_{\ell}^{s}$	same as $\left\{{s \atop \ell}\right\}$; see $\left\{{n \atop i}\right\}$.
$f(x)\big\vert_{[x^{\lambda}]}$	coefficient of x^k in the Maclaurin expansion of $f(x)$.
$\Vert(g,a,s,k)\Vert$	The norm $\Vert(g,a,s,k)\Vert$ is either equal to $g+a+s$ or $g+a+s+k$ for a valid state, (g,a,s,k), in a Markov Chain for a standard Ψ_1-process or Ψ_2-process, respectively, with $r=1$, $\sigma=\rho$ and $\alpha=m$.
iff	*if and only if.*
$\dot{r}$	the dot above a letter indicates the sum of indexed values of the variable; e.g. $\dot{r}=\sum_{i=1}^{\gamma} r_i$.
$\mathcal{S}_i^{(n)}$	(a) the Stirling numbers of first kind. (b) the number of ways to partition i objects into n non-empty parts and arrange the members of each part around a circle, where the order of the objects around the circle must be taken into account. (c) the coefficients in the polynomial expansion of a factorial polynomial, with $(k)_n=\sum_{i=1}^{m}\mathcal{S}_i^{(n)}k^i$.
Δ^m	the mth finite difference operator.
Δ^{-m}	the mth inverse finite difference operator.
χ^2	test statistic for the χ^2 distribution.

χ^2_ν	χ^2-distribution with ν degrees of freedom.
$\chi^2_{\nu,\alpha}$	upper αth percentile of the χ^2_ν distribution.

Glossary

The most-commonly-used symbols and their most-commonly-used usage are provided here.

Symbol	**Meaning**
Ψ-*distribution*	the probability distribution function for a Ψ-process.
Ψ-*number*	the number of ways in which the arrivals can occur in order to produce the specified waiting time in a Ψ-process.
Ψ-*probability*	the probability distribution function or a value thereof for a Ψ-process.
Ψ-*process*	(a) a *new* occupancy urn model in which we have a sequence of urns and throw the balls into them one by one until the appearance of a specified configuration occurs *after or at the same time as an initial specified configuration occurs.* (b) a Ψ-*process* is a *waiting-time* random process in which the waiting time is measured from the instant the process has first visited at least the σth element of G to the instant it has first visited at least ω elements of G and α_i elements of $A_i \backslash G$ for at least q elements $i \in \{1, \ldots, r\}$ but not β_u elements of B_u for at most w B-sets, B_u, $u \in \{1, \ldots, t\}$.

Ψ_1-process	a *without-replacement* Ψ-process.
Ψ_2-process	a *with-replacement* Ψ-process.
able to leave	in a *taboo* Ψ-process with r A-sets and t B-sets, a G-set G is *able to leave* when at least ω distinct elements of G and at least α_{i_j} distinct elements of at least q corresponding sets $A_{i_j} \backslash G$, $j \in \{1, \ldots, q\}$ have arrived, and at most w B-sets have at least β_u distinct elements of B_u with arrivals If not a *taboo* model, then omit all conditions that refer to B-sets in the above sentence.
alternative	In the *2-D Gap Problem*, an *alternative* at gap ℓ is said to occur for a collection of paths $\{A_{i_1}, \ldots, A_{i_s}\}$ if $\lvert\{A_{i_1}(\ell), \ldots, A_{i_s}(\ell)\}\rvert \geq 2$. When $\lvert\{A_{i_1}(\ell), \ldots, A_{i_s}(\ell)\}\rvert = \lambda \geq 2$ there is said to be λ alternatives at gap ℓ. When $\lvert\{A_{i_1}(\ell), \ldots, A_{i_s}(\ell)\}\rvert = 1$ there is said to be *no alternative* at gap ℓ.
A-set	a subset of $\mathcal{N}$; called a *required set*. In the new *waiting-time* model, measuring the wait ends when one or more A-sets are completed (or, in the more-general case, partially completed).
batch	a collection of one or more arrivals at an instant.
	in the game *SET*, the game will be called *batch* if it is not *linear*.
Bird-Watcher's Expectation	in the *Bird-Watcher's Problem*, this is the expected waiting time for the completion of pages one and two, measured from the sighting of the 5th unique bird on page 2, conditional on completing both pages.

Bird-Watcher's Probability	in the *Bird-Watcher's Problem*, this is the probability of completing pages 1 and 2.
blockage covering	given a G-set G and A-sets $A_1, \ldots, A_r$ a collection of $t > 0$ blockage sets $\mathfrak{B} = \{B_1, \ldots, B_t\}$ is defined to be a *blockage covering* of the G-set G for A-sets $A_1, \ldots, A_r$ if for any blockage set B' of G for A-sets $A_1, \ldots, A_r$ there exists $B \in \mathfrak{B}$ s.t. $B' \supseteq B$.
blockage set	given a G-set G and A-sets $A_1, \ldots, A_r$, a set B is defined to be a *blockage set of G for A-sets* $A_1, \ldots, A_r$ if $B \subseteq \cup_{i=1}^{r} A_i \backslash G$ and contains at least one element from each set A_i.
blocking model	a Ψ-process in which success can be obtained only if at most w *taboo* sets have been completed.
blocking probabilities	the probability of a Ψ-process not able to complete due more than w *taboo* sets having been completed.
blocking sets	see *blockage set.*
Bonferroni's Inequalities	upper and lower bounds for the probability of a union of events that truncates the inclusion-exclusion formula to odd and even numbers of summation terms, respectively.
bounded partition	in the *2-D Gap Problem*, this is an $(L, \ell, \lambda, n, d, b)$*-partition* (q.v.).
B-set	a subset of $\mathcal{N} \backslash (G \cup \bigcup_{i=1}^{r} A_i)$; called a *taboo* set. In the new *waiting-time* model, if all elements of B have been visited prior to at least one A-set being completed, then measuring the wait ends and it is considered impossible for the G-set to leave.

classical occupancy problem	(Boltzmann-Maxwell statistics) r indistinguishable balls are distributed among n cells and all of the n^r possible distributions have equal probability.
classical occupancy urn model	see *classical occupancy problem.*
complete match	all elements in a d-tuple of elements of a G-set have an arrival.
conditional rising factorial moments	in a Ψ_2-process, these are the rising factorial moments for the waiting time of an arrival for the G-set given that both the G-set and at least one A-set completes; extensions to the more-general models are also included.
covered	a blockage set B is said to be *covered* by a blockage set B^* if $B^* \subset B$.
covering	a subset of A-sets that can be used in the *fundamental formulae* to produce an identical result for a model.
decomposition	the property of linear independence of a collection of functions is used to write complex expressions as a unique linear sum of those functions. In this thesis, this corresponds to writing the *fundamental formula* as a linear combination of distinct Ψ-numbers.
decomposition coefficient	the coefficient of a function when a formula is expressed in its *decomposition* form.
dynamic model	an *occupancy urn model* in which there is a sequence of urns into which balls are thrown one by one at random, and the sequence of configurations is considered.

empty match	no arrivals have occurred for a d-tuple of elements of a G-set.
first kind	associated with a *without-replacement* Ψ-process.
full covering	the complete collection of model-related A-sets associated with a particular G-set.
Fundamental Formula	the result of the *Fundamental Theorem.*
Fundamental Theorem	a theorem that provides a way to write the probability of at least t events occurring in Ψ-processes in terms of single events. It is both *fundamental* to the determination of distributions and moments for non-trivial cases and is also *fundamental* in the sense that it is a crucial piece of knowledge in the theory of the new *waiting-time* models.
gaps with alternatives	in the *2-D Gap Problem*, there are said to be ℓ *gaps with alternatives* if ℓ of the L gaps have at least 2 *alternatives* and the remaining $L-\ell$ gaps have *no alternative.*
G-set	a subset of $\mathcal{N}$. In the new *waiting-time* model, measuring the wait begins from the completion of (or partial completion of) a G-set.
limiting conditional rising factorial moments	the limit of the *conditional rising factorial moments* as $n \to \infty$.
linear	in the game *SET*, the game will be called *linear* if the cards are placed on the table one at a time.

$(L, \ell, \lambda, n, d, b)$-*partition*	in the *2-D Gap Problem*, given positive integers L and n, which are the numbers of gaps and lanes, respectively, and non-negative integers ℓ, λ, d and b with $\ell \leq L$ and $b\ell \leq \lambda \leq n\ell$, an $(L, \ell, \lambda, n, d, b)$-*partition* is a collection of L numbers of which $L - \ell$ of them are set to d, the other ℓ of them are bounded below by b, are bounded above by n and sum to λ.
minimal blockage covering	a *blockage covering*, $\mathfrak{B}$, is a *minimal blockage covering* if $\lvert\mathfrak{B}\rvert \leq \lvert\mathfrak{B}'\rvert$ for any other blockage covering, $\mathfrak{B}'$.
Minimal Blockage Covering Theorem	a theorem that provides a way to write the new waiting-time distribution for $T_B(A_1, \ldots, A_r)$ as a Ψ-distribution with parameters being the unions of *blockage sets* of a *minimal blockage covering* and the G-set G.
minimal covering	the minimal collection of A-sets that can be used in the *fundamental formulae* or the *minimal covering theorem for platoons* to produce the required probability for a model.
Minimal Covering Theorem for Platoons	provides a way to reduce the number of A-sets involved in the calculation of the platoon-size distribution by eliminating redundant A-sets.
MLE	maximum likelihood estimate.
(N, m, ρ)-*sequence*	in the standard Ψ_1-process, a sequence of m elements of $A \backslash G$, ρ of G and $N - m - \rho$ of S.
noset	in the game *SET*, a set of cards with no triad amongst them.

on Athlon	indicates that a computer program was run on a 100-mega-flop 1.2 GHz Athlon-based computer to obtain results, in particular, timing results.
on Celeron	indicates a computer program was run on a 466 MHz Celeron II-based computer to obtain results, in particular, timing results.
partial match	between 1 and $d-1$ elements of a d-tuple of elements of a G-set have an arrival.
platoon	when an arrival occurs, the group of cells of all G-sets that are considered complete and *able to leave* is called a *platoon.*
platoon size	the size of a *platoon* when an arrival occurs.
Principle of Inclusion and Exclusion for the Mini-Max	given r events $E_1, \ldots, E_r$ and a non-negative integral function f, $P(\min_{\{i_1,\ldots,i_t\}\subseteq\{1,\ldots,r\}} \max_{i\in\{i_1,\ldots,i_t\}} f(E_i) = k) = \sum_{s=t}^{r}(-1)^{s-1}\binom{s-1}{t-1} \sum_{i_1,\ldots,i_s} P\left(\max_{j\in\{1,\ldots,s\}} f\left(E_{i_j}\right) = k\right)$.
randomised varieties	a process with *varieties* in which there is no restriction on the number or type of varieties that can arrive simultaneously.
reduced formula	there are some alternative expressions for formulae that have been produced in order to speed up the calculations and also enable the determination of much simpler forms for the rising factorial moments; these are also more appealing due to their relative simplicity. These alternative formulae are referred to as either *simplified* or *reduced* formulae or expressions. Some of these expressions are far more complicated in immediate appearance, but are still referred to as *simplified*, because they allow finding closed-forms for the sum that produces the rising factorial moments and because there are orders-of-magnitude less calculations to be performed.

reduction numbers	in the *2-D Gap Problem*, consider ℓ specific gaps $\gamma_1, \ldots, \gamma_\ell$ that have corresponding particular alternatives with counts $\lambda_1, \ldots, \lambda_\ell$, with each $\lambda_\alpha \geq 2$, $\alpha \in \{1, \ldots, \ell\}$. For $j \in \{0, \ldots, \lambda_1 + \ldots + \lambda_\ell - \ell\}$ and $\alpha \in \{1, \ldots, \ell\}$, let $r_\alpha \in \{0, \ldots, \lambda_\alpha - 1\}$ satisfy $\sum_{\alpha=1}^{\ell} r_\alpha = j$. When j corresponds to the index in $\Lambda_j(n, \ell, \boldsymbol{\lambda}, s)$ (q.v), the r_α's are called *reduction numbers.*
required set	see *A-set.*
second kind	associated with *with-replacement* Ψ-processes.
simplified formulae	see *reduced formulae.*
simultaneous varieties	a process in which each arrival for each *variety* occurs simultaneously at each of the arrival-points and whose arrival streams are independent of each other.
standard Ψ-process	a Ψ-process in which there are no *taboo* sets.
static process	an *occupancy urn model* in which we have a sequence of urns and throw balls into them at random, and look at the final configuration.
taboo model	a modification of the new *waiting-time* model in which the completion of *taboo* sets prevents the G-set from ever leaving.
taboo set	see *B-set.*
transformation formula	a specific combinatorial identity that is used to convert sums of products of two binomials to an alternative form that is more efficient and also enables a much-simplified version of moments to be found.

varieties	when each of N urns has v distinct attributes or locations in which a ball may be placed, a process is said to be *with varieties*.
waiting-time process	an *occupancy urn model* in which we have a sequence of urns and throw balls into them at random, and wait until the appearance of a specified configuration; the new *waiting-time* process begins counting *after or at the same time as an initial specified configuration occurs*.
with multiplicities	this term refers to a model in which a G-set can have multiple partial completions (or matches), with specified counts determining the partial completion sets.
without multiplicities	this term refers to a model in which only an entire G-set is considered to be a match.

Symbols

Only the most-commonly-used symbols and their most-commonly-used usage are provided here.

Symbol	**Meaning**
$\circledast$	in the Markov Chain for the static model, indicates when at least one of a G-set's corresponding r A-sets, A_1, A_2, …, A_r has completed.
$\bigoplus_{(\nu,\mu)} f(\nu,\mu)$	an operator that is an abbreviation for a specific double-sum of $f(\nu,\mu)$ over ν and μ with coefficients that are functions of the parameters N, m, ρ, σ, which have the usual meanings, and the parameters j, α and β, which are local to the problem of determining the conditional rising factorial moments. This is akin to the right-hand side of the *transformation formula* that we frequently use in Ψ_1-processes.
α	the number of states in A that are yet to be visited at the instant when the σth state of G is visited.
$\boldsymbol{\alpha}$	$(\alpha_1,\ \ldots,\ \alpha_r)$.
α_i	number of states in A_i that are yet to be visited at the instant when the σth state of G is visited..
β	number of elements in the B-set B.
$\boldsymbol{\beta}$	list of the numbers of elements in the t B-sets $B_1, \ldots, B_t$; $(\beta_1, \ldots, \beta_t)$.

β_ℓ	cumulative number of arrivals after the ℓth batch has arrived, with $\beta_0 = 0$.
β'_ℓ	number of arrivals yet to occur after the $(\ell - 1)$th batch has arrived.
β_u	number of elements in the B-set B_u.
Γ	the gamma function $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ for $x > 0$.
$\Gamma_g(\sigma; \pi)$	for $\pi \in \Pi$, the set of elements of G_g that are visited at or before the σth element of G_g.
γ	(a) number of G-sets when the sample space S is considered to be partitioned into not-necessarily disjoint G-sets.
	(b) in the game *SET*, γ is the number of possible triads.
$\gamma_1, \ldots, \gamma_\ell$	in the *2-D Gap Problem*, ℓ specific gaps that have a particular number of alternatives.
γ^*	MLE of the number of G-sets.
$\zeta(\ell, q, x)$	essentially is the generating function for a truncated sum of rising factorials.
$\zeta_1(\ell, p, q, x)$	$= \zeta(\ell, q, x) - \zeta(\ell, p - 1, x)$ for $p \leq q$ and zero otherwise; this enables writing the upper-sum form of $\zeta(\ell, q, x)$ as the difference of two expressions whose summations begin at zero.
$\zeta_\infty(\ell, x)$	$\lim_{q\to\infty} \zeta(\ell, q, x)$; used in determining the *limiting conditional rising factorial moments.*

$\Lambda(n, \ell, \boldsymbol{\lambda}, s)$	in the *2-D gap Problem*, the number of collections of s A-sets $\{A_{i_1}, \ldots, A_{i_s}\}$ such that $A_{j_1}(\gamma) = A_{j_2}(\gamma)$ $\forall j_1, j_2 \in \{i_1, \ldots, i_s\}$, $j_2 \neq j_1$, $\forall \gamma \in \{1, \ldots, L\} \setminus \{\gamma_1, \ldots, \gamma_\ell\}$, and $\Lambda(n, 0, \boldsymbol{\lambda}, s) \stackrel{def}{=} 1$. That is, it is the number of collections of s paths that are possible through the gaps that have no alternatives given gaps $\gamma_1, \ldots, \gamma_\ell$.have alternatives.
$\Lambda_j(n, \ell, \boldsymbol{\lambda}, s)$	in the *2-D Gap Problem*, consider ℓ specific gaps $\gamma_1, \ldots, \gamma_\ell$ that have corresponding particular alternatives with counts $\lambda_1, \ldots, \lambda_\ell$, with each $\lambda_\alpha \geq 2$, $\alpha \in \{1, \ldots, \ell\}$. For $j \in \{0, \ldots, \lambda_1 + \ldots + \lambda_\ell - \ell\}$, $\Lambda_j(n, \ell, \boldsymbol{\lambda}, s)$ is the number of collections of s A-sets $\{A_{i_1}, \ldots, A_{i_s}\}$ such that $A_{j_1}(\gamma) = A_{j_2}(\gamma)$ $\forall j_1, j_2 \in \{i_1, \ldots, i_s\}$, $j_2 \neq j_1$, $\forall \gamma \in \{1, \ldots, L\} \setminus \{\gamma_1, \ldots, \gamma_\ell\}$ *and* the s A-sets have j fewer alternatives than from the total, $\lambda_1 + \ldots + \lambda_\ell$.
λ	in the *2-D Gap Problem*, when $\|\{A_{i_1}(\ell), \ldots, A_{i_s}(\ell)\}\| = \lambda \geq 2$ there are λ alternatives at gap ℓ, and when $\lambda = 1$ there are no alternatives at gap ℓ.
$\boldsymbol{\lambda}$	abbreviation or vector notation for $\lambda_1, \ldots, \lambda_\ell$.
$\lambda(i)$	number of left-most bits of i that are positive in its binary representation.
$\lambda_1, \ldots, \lambda_\ell$	in the *2-D Gap Problem*, these are the corresponding numbers of particular alternatives for the ℓ specific gaps $\gamma_1, \ldots, \gamma_\ell$, with each $\lambda_\alpha \geq 2$
λ_i	see $\lambda_1, \ldots, \lambda_\ell$.
μ	(a) in the static and dynamic processes, μ represents either the number of G-sets that have at least one element visited but not all elements visited or those that need at least μ to be visited.

	(b) when measuring the expected duration of cakes on display in the *Cake Display Problem*, μ represents the minimal number of slices each cake on display must have.
	(c) in the game *SET*, μ is the maximal number of triads that can be formed with no triad amongst them.
	(d) in the *2-D Gap Problem*, the total number of vehicles in front of, and in the same lane as, the special vehicle labelled g.
μ'	in the *2-D Gap Problem*, the total number of vehicles behind and in the same lane as the special vehicle labelled g.
ν	(a) number of steps in a Markov Chain.
	(b) index of the sum in $v(r, n, N)$.
	(c) number of *varieties*..
Π	the set of permutations on the N elements of $\mathcal{N}$.
π	(a) ratio of the circumference to the diameter of a circle.
	(b) for $\pi \in \Pi$, represents an ordering of the visits to the elements of $\mathcal{N}$.
$\pi(a)$	the arrival position for $a \in \mathcal{N}$.
$\pi(\alpha)$	for $\pi \in \Pi$, $\pi(\alpha)$ is the position at which the visit to α occurs.
π^*	an element of Π that maximises the total wait, $\phi(\pi)$ over all $\pi \in \Pi$.
ρ	(a) number of elements in a single G-set; $\sigma \leq \omega \leq \rho$.

	(b) number of arrivals for a vehicle.
$\rho_1, \ldots, \rho_r$	numbers of elements in the G-sets $G_1, \ldots, G_r$.
ρ_i	number of elements in the ith G-set.
ρ^*	MLE of ρ.
σ	(a) an arrival position for an element in a single G-set; $\sigma \leq \omega \leq \rho$.
	(b) an arrival position for a member of a particular vehicle.
$\boldsymbol{\sigma}$	$(\sigma_1, \ldots, \sigma_\gamma)$.
$\sigma_1, \ldots, \sigma_r$	(a) arrival positions for members of the G-sets $G_1, \ldots, G_r$.
	(b) occupancy numbers of the G-sets $G_1, \ldots, G_r$.
σ_i	arrival position for a member of the ith G-set.
$\sigma\text{-}\max_{\alpha \in I} f(\alpha)$	the maximum of the first σ elements in the ordered list of elements in the set $\{f(\alpha) : \alpha \in I\}$.
τ	when measuring the expected duration of cakes on display in the *Cake Display Problem*, τ represents the minimal number of cakes on display.
$\phi(m)$	the *decomposition coefficient* of $P(T(m) = k)$ in the *decomposition formula*.
$\phi(\pi)$	abbreviation for $\phi(\boldsymbol{\sigma}; \pi)$ when $\sigma_g \equiv \rho_g$.
$\phi_{(N,\rho,\sigma)}(m)$	same as $\phi(m)$, but with an explicit specification of N, ρ and σ.

$\phi\left(\boldsymbol{\sigma};\pi\right)$	for $\pi\in\Pi$, the total wait by all elements of all G-sets G_g, $g\in\{1,\ldots,\gamma\}$.
$\phi_g\left(\pi\right)$	abbreviation for $\phi_g\left(\sigma;\pi\right)$ when $\sigma_g=\rho_g$.
$\phi_g\left(\sigma;\pi\right)$	for $\pi\in\Pi$, the wait by the σ'th element of G_g for the completion of at least one of the A-sets $A_{g\nu}$, $\nu=1,\ldots,r_g$.
$\phi^+\left(m\right)$	number of occurrences of $\Psi_1\left(N,m,k,\rho,\sigma\right)$ in $P\left(T=k\right)$.
$\phi^+_{(N,\rho,\sigma)}\left(m\right)$	same as $\phi^+\left(m\right)$, but with an explicit specification of N, ρ and σ.
ϕ^*	maximum total wait over all $\pi\in\Pi$.
$\varphi_1\left(\ldots\right)$	$\varphi_1\left(a,b,\alpha,\beta\right)$; a formula based on $v\left(r,n,N\right)$ that is used in the process of simplifying the probability distribution of a Ψ_2-process.
$\varphi_2\left(\ldots\right)$	$\varphi_2\left(k,j,a,b,\alpha,\beta\right)$; a formula based on $v\left(r,n,N\right)$ and $\varphi_1\left(a,b,\alpha,\beta\right)$ that is used in the process of simplifying the probability distribution of a Ψ_2-process.
$\varphi_3\left(\ldots\right)$	$\varphi_3\left(\ell,j,a,b,\alpha,\beta,e,f\right)$; a formula based on $\varphi_2\left(k,j,a,b-k,\alpha,\beta\right)$ that is used in the process of simplifying the *conditional rising factorial moments*; $\sum_{k=\max(e,1)}^{f}\left[k\right]_\ell\varphi_2\left(k,j,a,b-k,\alpha,\beta\right)$.
$\varphi_4\left(\ldots\right)$	$\varphi_4\left(\ell,j,a,b,\alpha,\beta,e,f\right)$; a formula based on $\varphi_2\left(k,j,a,b-k,\alpha,\beta\right)$ that is used in the process of simplifying the *conditional rising factorial moments*; $\sum_{k=\max(e,1)}^{f}\left[k\right]_\ell\varphi_2\left(k,j,a-k,b-k,\alpha,\beta\right)$.
$\varphi_5\left(\ldots\right)$	$\varphi_5\left(\ell,j,a,\alpha,\beta,e,f\right)$; a formula used in the determination of the *limiting conditional rising factorial moments*.

$\varphi_5'(\ldots)$ — $\varphi_5'(\ell, j, a, \alpha, \beta, e)$; a formula used in the determination of the *limiting conditional rising factorial moments.*

$\varphi_6(\ldots)$ — $\varphi_6(\ell, j, a, \alpha, \beta, e, f)$; a formula used in the determination of the *limiting conditional rising factorial moments.*

$\Psi_1(N, m, k)$ — Ψ-probabilities of first kind with $\sigma = \rho = 1$ and non-provided parameters having their default values.

Ψ_1' — the probabilities corresponding to ψ_1'.

$\Psi_1(\ldots)$ — $\Psi_1(N, n, m, \rho, \sigma, \omega, \boldsymbol{\alpha}, w, \boldsymbol{\beta}, \mathbf{n}, \nu, k)$; Ψ-probabilities of first kind.

$\Psi_2(\ldots)$ — $\Psi_2(N, n, m, \rho, \sigma, \omega, \boldsymbol{\alpha}, w, \boldsymbol{\beta}, \mathbf{n}, \nu, k)$; Ψ-probabilities of second kind.

ψ' — (a) the $\Psi-$numbers for a Ψ-process, usually with parameters indicating which model the values corresponds to, and with an index to specify which Ψ-process.

(b) the components of Ψ-numbers that do not include N in their expressions.

$\psi_1(\ldots)$ — $\psi_1(N, n, m, \rho, \sigma, \omega, \boldsymbol{\alpha}, w, \boldsymbol{\beta}, \mathbf{n}, \nu, k)$; Ψ-numbers of first kind.

$\psi_2(\ldots)$ — $\psi_2(N, n, m, \rho, \sigma, \omega, \boldsymbol{\alpha}, w, \boldsymbol{\beta}, \mathbf{n}, \nu, k)$; Ψ-numbers of second kind.

$\Omega(L, f, c, e)$ — either side in the *transformation formula.*

ω — for $\sigma \leq \omega \leq \rho$ and in a Ψ-process, it is not necessary to wait for more than the ω'th arrival of G; in general, there are also other conditions that determine when measuring the wait ceases.

A	an A-set.
$\mathbf{A}$	(a) abbreviation for the list $A_1, \ldots, A_r$.
	(b) represents the vector or list $(A_1, \ldots, A_r)$.
$A_1, \ldots, A_r$	r A-sets.
$\mathbf{A}_i$	abbreviation for $(A_{i1}, \ldots, A_{ir_i})$.
A_{ij}	the jth A-set of the ith G-set.
$A_i(\ell)$	in the *2-D Gap Problem*, the ℓth element of the path corresponding to A_i.
$\mathbf{A}^{(\gamma)}$	$(\mathbf{A}_1, \ldots, \mathbf{A}_\gamma)$.
a	in the game *SET*, a is the number of attributes.
B	(a) a *taboo* set.
	(b) number of batches when batch arrivals are permitted.
$\mathbf{B}$	(a) abbreviation for the list $B_1, \ldots, B_t$.
	(b) represents the vector or list $(B_1, \ldots, B_t)$.
$B_1, \ldots, B_t$	a collection of *taboo* sets.
$B_1', \ldots, B_t'$	a collection of sets of the form $G \cup B_u$, where B_u is a blocking set.

B_u	the uth blocking set.
C_ℓ	expected number of completions during the ℓth interval.
D	destination node in network problems.
d	size of a subset of a G-set that is considered to be a match.
d_i	size of a subset of the G-set G_i that is considered to be a match.
$d_{i1}, \ldots, d_{i\tau_i}$	the sequential sizes of the subsets of the G-set G_i that are considered matches.
E	(a) an expected value.
	(b) expectation of a distribution.
$E\left(k\right)$	expected number of completions of G-sets at the kth arrival when considering the clustering of completions.
$E_k^{(1)}$	expected platoon sizes as a result of the kth arrival for the uni–directional car parking model.
$E_k^{(2)}$	the expected platoon sizes as a result of the kth arrival for the bi-directional car parking model.
E_ℓ	abbreviation for $E_{\ell,r}$ when $r = 1$.
E_ℓ^*	limit of the ℓth *conditional rising factorial moment.*
$E_\ell\left(i, \sigma\right)$	ℓth rising factorial moment for the σth arrival for the G-set G_i.

$E_{\ell,r}$	ℓth rising factorial moment in the case of $r \geq 1$ A-sets.
$E_v\,(m)$	expectation when there are v varieties.
$\dot{E}_{1,K}$	in the game *SET*, the expected number of triads in K cards.
F_ℓ	(a) ratio of the ℓth rising factorial moments of Z_2 to Z_1.
	(b) expected number of completions of G-sets by the end of the ℓth interval.
f	in the *Interconnected Parallel Lines* problem, the number of elements (feet) in a connection-set.
G	a G-set.
$G_{(i,j)}$	the G-set corresponding to car $j \in \{1, \ldots, s_i\}$ of lane $i \in \{1, \ldots, t\}$ in a multi-lane car parking model or a similar two-dimensional grid of cells.
$G_1, \ldots, G_\gamma$	a collection of γ G-sets; typically $\dot{\cup}_{i=1}^{\gamma} G_i = \mathcal{N}$.
g	(a) a particular element of a G-set.
	(b) in the *2-D Gap Problem*, g is the special vehicle of interest.
	(c) in the *Interconnected Parallel Lines Problem*, the minimum number of elements required to be visited (shoed).
	(d) an index for a collection of G-sets as used in the determination of the maximum total wait.

(g, a, s, k) state of the Markov Chain for a Ψ-process, where g, a and s are the numbers of distinct elements of G, A and S that have been visited, respectively, and k is the waiting-time, which may also take on the values $k = -3$, -2, -1 and ∞.

$g(x)$ generating function of a Ψ-distribution.

H mean inter-arrival times between all pairs of consecutive arrivals.

h in the game *SET*, h is the number of triads each triad intersects with.

i (a) the ith A-set.

(b) the ith lane of vehicles.

j (a) the jth element of an A-set.

(b) position of a vehicle in a lane.

K (a) cumulative value of a random variable.

(b) in the game *SET*, the number of cards for which the number of matches is sought.

(c) the arrival at which properties of the cumulative number of completions are determined.

k (a) a value of a random variable.

(b) waiting time.

(c) number of arrivals.

L — in the *2-D Gap Problem*, the number of physical gaps within each lane in front of the special vehicle labelled g.

L' — in the *2-D Gap Problem*, the number of physical gaps within each lane behind the special vehicle labelled g.

$(L, \ell, \lambda, n, d, b)$ — in the *2-D Gap Problem*, these are the parameters of an $(L, \ell, \lambda, n, d, b)$-*partition*, where L and n are the numbers of gaps and lanes, respectively, ℓ, λ, d and b are non-negative integers with $\ell \leq L$ and $b\ell \leq \lambda \leq n\ell$.

ℓ — (a) a position in the arrival stream for the arrival of the last element of a G-set or an A-set.

(b) an index for the ℓth rising factorial moment.

m — (a) number of elements in an A-set other than those in G.

(b) in the *2-D Gap Problem*, the number of vehicles directly in front of the special vehicle labelled g and before the first gap.

$\mathbf{m}$ — abbreviation for $(m, m_1, m_2, \ldots, m_L)$; used in the *2-D Gap Problem*.

$m_1, \ldots, m_L$ — in the *2-D Gap Problem*, the number of vehicles directly in the front of the special vehicle labelled g that are in the regions between the L gaps.

$m'_1, \ldots, m'_{L'}$ — in the *2-D Gap Problem*, the number of vehicles directly behind the special vehicle labelled g that are in the regions between the L' gaps.

$m_1, \ldots, m_r$ — the numbers of elements in the A-sets $A_1, \ldots, A_r$ other than those in G.

m_{ij}	$\|A_{ij}\backslash G_i\|$ in a model that has A-sets for each G-set G_i.
m^*	in a Ψ-process with $\sigma = \rho$, the number of elements minus one in the smallest collection of elements from $(r-1)$-collections of the sets $A_i\backslash G$ with one element from each of the $(r-1)$-collections of the sets $A_i\backslash G$. In the *taboo* models, there must be at least one element from each of the B-sets too.
m^*_σ	the same as m^*, but for the case $\sigma \leq \rho$.
m^{**}	the batch equivalent of m^*.
N	(a) total number of elements in $\mathcal{N}$.
	(b) total number of vehicles, distinct coupons, etc.
$\mathcal{N}$	(a) the sample space $\{1, 2, \ldots, N\}$.
	(b) the set of all elements (or cells, cars, etc.).
$(N, m, \rho)_k$	number of (N, m, ρ)-sequences for which $T(m) = k$.
$(N, m, \rho, \sigma)_k$	number of (N, m, ρ)-sequences for which $T(m) = k$ when $\sigma \leq \rho$.
(N, m, ρ, σ, n)	represents the parameters in the *Bird-Watcher's Problem.*
(N, n, m, ρ)	number of distinguishable (N, m, ρ)-sequences when the number of arrivals, n, is different to N.
$(N, n, m, \rho)_k$	number of (N, n, m, ρ)-sequences for which $T = k$, where $k \in \{0, 1, \ldots n - \rho\} \cup \{-1, \infty\}$.

$N_{ap}(k)$	mean number of arrivals that are in partial matches at the kth arrival.
$N_{ar}(k)$	mean number of arrivals required to complete partial matches at the kth arrival.
N_b	number of distinct *blockage sets* of the G-set G for A-sets $A_1, \ldots, A_r$.
$N_c(k)$	mean number of complete matches at the kth arrival.
$N_e(k)$	mean number of empty matches at the kth arrival.
$N_{pa}(k, \mu)$	mean number of partial matches with at least μ arrivals at the kth arrival.
$N_{pr}(k, \mu)$	mean number of partial matches with at least μ arrivals still required at the kth arrival.
N_σ	maximum finite wait possible for the σth arrival of G.
N^*	MLE of N.
N^*_∞	an approximation of N^* for small values of m and a large population in a Ψ_1-process.
N^*_{-1}	MLE of N for the observation $T = -1$ in a Ψ_2-process' limiting distribution for T as $N \to \infty$ with $n = \alpha N$ for $\alpha > 0$.
N^*_k	MLE of N for the observation $T = k$ in a Ψ_2-process' limiting distribution for T as $N \to \infty$ with $n = \alpha N$ for $\alpha > 0$.
N^*_∞	MLE of N for the observation $T = \infty$ in a Ψ_2-process' limiting distribution for T as $N \to \infty$ with $n = \alpha N$ for $\alpha > 0$.

n	(a) number of observations.
	(b) number of lanes in the *2-D Gap Problem*.
	(c) number of intervals when considering the clustering of completions.
$\mathbf{n}$	abbreviation for $(n_1, n_2, \ldots, n_B)$.
$(n_1, \ldots, n_B)$	vector representation of batch sizes.
n_a	number of absorbing states in a Markov Chain.
n_b	batch size of batch $b \in \{1, \ldots, B\}$, where B is the number of batches.
n_s	total number of valid states in a Markov Chain.
n^*_{-1}	MLE of n for the observation $T = -1$ in a Ψ_2-process' limiting distribution for T as $N \to \infty$ with $n = \alpha N$ for $\alpha > 0$.
n^*_k	MLE of n for the observation $T = k$ in a Ψ_2-process' limiting distribution for T as $N \to \infty$ with $n = \alpha N$ for $\alpha > 0$.
n^*_∞	MLE of n for the observation $T = \infty$ in a Ψ_2-process' limiting distribution for T as $N \to \infty$ with $n = \alpha N$ for $\alpha > 0$.
O	origin node in network problems.
$P(T = k)$	the probability that the random variable T takes on the value k. Typically, the parameters of T are understood by context. Where a parameter is specified, it is included for emphasis or because this parameter is going to be given various values for comparison purposes; e.g. $P(T(m) = k)$.

$P(\boldsymbol{\alpha})$	the probability α_i states of $A_i \backslash G$ have not been visited for all $i \in \{1, \ldots, r\}$ at the instant the σth state of G arrives.
$P(\boldsymbol{\sigma}; k)$	the probability that at time k each G-set G_i, $i \in \{1, \ \ldots, \gamma\}$, has σ_i arrivals.
P_{ij}	$P_{ij}^{(\nu)}$ with $\nu = 1$.
$P_{ij}^{(\nu)}$	the probability of Markov Chain going from state i to state j in ν steps; states may be of the form (g, a, s, k).
$P_k(G)$	$P_k(G, \mathbf{A})$ when $r = 0$.
$P_k(G, A)$	$P_k(G, \mathbf{A})$ when $r = 1$.
$P_k(G, \mathbf{A})$	for $s \in \{1, \ \ldots, \ r\}$, the probability that the first of the r sets $G \cup A_s$ completes at the k'th arrival, and for $r = 0$, the probability that G completes upon the kth arrival.
$P_{k,n_1,n_2}(.,.)$	for a pair of G-sets G_1 and G_2 with corresponding A-sets A_1 and A_2, $P_{k,n_1,n_2}((G_1, \mathbf{A}_1), (G_2, \mathbf{A}_2))$ is the joint probability that, upon the kth arrival, $G_1 \cup A_{1s_1}$ completes for at least one of $A_{11}, \ \ldots, \ A_{1r_1}$ and $G_2 \cup A_{2s_2}$ completes for at least one of $A_{21}, \ \ldots, \ A_{2r_2}$, where $n_1, n_2 \in \{0, 1\}$ with a value of 1 indicating completion and 0 indicating non-completion. When $r_1 = r_2 = 1$, it is written as $P_{k,n_1,n_2}((G_1, A_1), (G_2, A_2))$.
$P_{k_1,k_2}(.,.)$	for a pair of G-sets G_1 and G_2 with corresponding collections of associated A-sets $A_{11}, \ \ldots, \ A_{1r_1}$ and $A_{21}, \ \ldots, \ A_{2r_2}$, $P_{k_1,k_2}((G_1, \mathbf{A}_1), (G_2, \mathbf{A}_2))$ is the joint probability that $G_1 \cup A_{1s_1}$ completes upon the k_1th arrival for at least one of $A_{11}, \ \ldots, \ A_{1r_1}$ and $G_2 \cup A_{2s_2}$ completes upon the k_2th arrival for at least one of $A_{21}, \ \ldots, \ A_{2r_2}$. When $r_1 = r_2 = 1$, it is written as $P_{k_1,k_2}((G_1, A_1), (G_2, A_2))$.

$P_L(n)$	in the *2-D Gap Problem*, the expected waiting-time as a percentage of the total population.
$P_m(\alpha)$	the probability α states of $A \backslash G$ have not been visited at the instant the σth state of G is visited.
$P_m(\alpha_i)$	the probability α states of $A_i \backslash G$ have not been visited at the instant the σth state of G is visited.
$P_{\sigma\tau}(k)$	the probability of waiting k for the τth element of G, measured from σth..
$R(A;B)$	$R(A_1,\ldots,A_r;B_1,\ldots,B_t)$ when $r=1$ and $t=1$.
$R(\mathbf{A};\mathbf{B})$	random variable for the event that the process visits all the states of at least one of the A-sets $A_1,\ldots,A_r$ but not all the elements of any the B-sets $B_1,\ldots,B_t$; the possible values of R are *true* and *false* and are represented by 1 and 0, respectively.
r	(a) number of A-sets.
	(b) in the game *SET*, r is the number of triads each card is a member of.
r_i	number of A-sets for the ith G-set.
S	$S=\mathcal{N}\backslash\left(G\cup\bigcup_{i=1}^{r}A_i\cup\bigcup_{u=1}^{t}B_u\right)$ is the set of non-specified elements of $\mathcal{N}$.
s	(a) in car parking models, the number of vehicles in a lane.
	(b) in the *fundamental formulae*, the index for the outer sum over the number of A-sets.
s_i	the number of cars (or cells) in the ith lane (or set).

T	random variable for the new waiting-time process.
$T(A; B)$	$T(A_1, \ldots, A_r; B_1, \ldots, B_t)$ when $r = 1$ and $t = 1$.
$T(A; \mathbf{B})$	$T(A_1, \ldots, A_r; B_1, \ldots, B_t)$ when $r = 1$.
$T(A_1, \ldots, A_r)$	same as T, specifically for r A-sets.
$T(\mathbf{A}; \mathbf{B})$	$T(A_1, \ldots, A_r; B_1, \ldots, B_t)$; same as T, specifically for r A-sets and t B-sets.
$T(m)$	same as T, specifically for a single A-set with m elements.
$T(\mathbf{m})$	same as T, specifically for the *2-D Gap Problem*.
$T(m_1, \ldots, m_r)$	same as T, specifically when $A_i \cap A_j \equiv G$.
$T_b(A_1, \ldots, A_r)$	random variable for the Ψ-process that measures the waiting time, possibly zero, from the completion time of the G-set G to the time when *at least one element* has been visited from *each* of the r sets $A_i \backslash G$.
T_q	$T_q(A_1, \ldots, A_r)$.
$T_q(A_1, \ldots, A_r)$	random variable for the new non-*taboo* waiting-time process when at least q of the A-sets are required instead of just one.
t	(a) number of lanes in car parking models.
	(b) number of B-sets in *taboo* and *blocking* models.
	(c) number of arrivals that constitute the interval length when considering the clustering of completions.

u	an index for a particular B-set in a *taboo* process or blocking model.
$u(r,n)$	the probability that N given cells are occupied in the classical occupancy problem (Boltzmann-Maxwell statistics).
V_k	variance of the platoon size at the kth arrival.
$\dot{V}_K$	in the game *SET*, the variance of the number of triads in K cards.
$V_{ap}(k)$	variance for the number of arrivals that are in partial matches at the kth arrival.
$V_{ar}(k)$	variance for the number of arrivals required to complete partial matches at the kth arrival.
$V_c(k)$	variance for the number of complete matches at the kth arrival.
$V_e(k)$	variance for the number of empty matches at the kth arrival.
$V_{pa}(k,\mu)$	variance for the number of partial matches with at least μ arrivals at the kth arrival.
$V_{pr}(k,\mu)$	variance for the number of partial matches with at least μ arrivals still required at the kth arrival.
v	in models with *varieties*, v is the number of *varieties*.
$v(r,n,N)$	number of ways of leaving each of N given cells occupied in the classical occupancy problem (Boltzmann-Maxwell statistics); $v(r,n,N) \stackrel{def}{=} 0$ if $N > n$, $r < N$ or $n < 0$.
W	expected total wait for all states.

W_i	expected total wait for the G-set G_i.
$W_{\max}^{(1)}$	maximum total wait with t parallel lanes, each containing s vehicles, with uni-directional exiting.
$W_{\max}^{(2)}$	maximum total wait with t parallel lanes, each containing s vehicles, with bi-directional exiting.
w	maximum number of taboo sets that may be completed.
Y_i	indicator function that has $Y_i = j$ if j d_i-tuples form complete matches.
Y_i'	indicator function that has $Y_i' = d_i Y_i$ if $j \in \{0, \ldots, r_i\}$ d_i-tuples form complete matches.
Y_i'	number of arrivals for complete d_i-tuples.
$Y_k(G, \mathbf{A})$	for $r \geq 1$, the indicator function for whether or not $G \cup A_s$ completes at the kth arrival for at least one A-set A_s.
$Y_k(G)$	same as $Y_k(G, \mathbf{A})$, but for $r = 0$.
$Y_k\left(\mathbf{G}, \mathbf{A}^{(\gamma)}\right)$	number of completions at time k of sets $G_i \cup A_{ij}$ for at least one A-set A_{ij}, $j \in \{1, \ldots, r_i\}$.
Z_1	in the uni-directional car parking model, the number of further arrivals for whom the driver of a randomly selected vehicle has to wait.
Z_2	in the bi-directional car parking model, the number of further arrivals for whom the driver of a randomly selected vehicle has to wait.

Preface

It was the author's original intention to generalise Hauer and Templeton's uni-directional exiting model of cars parked in lanes in order to determine the effects of allowing the more-realistic capability of allowing cars to reverse in addition to travelling forward. Also, their article was in a journal on transportation and it seemed like a good idea to make the model and the more-realistic solution apparent in mainstream probability circles.

The author discovered Hauer and Templeton's article in the *Transportation Science* journal whilst investigating models of traffic light queueing systems. A simple generalisation was considered immediately, but the full potential only developed later. The original derivation of the distribution and the rising factorial moments for the bi-directional model was an ad-hoc method based on Hauer and Templeton's work. It took over 400 hours during a hot summer month to derive the really simple form of the mean; suitable computer algebra tools were not available at the time, so all work was done with pencil and paper. The right-hand side of the initial equation took about 20 pages of 20 lines per page with about 10 terms per line, making about 4,000 terms per right-hand side. The terms were of the form of a fraction with numerator and denominator containing factorial terms and combinatorial terms. In order to simplify the expression, pairs of terms that could factorise nicely or had similar denominators were combined together. Not all terms were paired each time a new, simplified right-hand-side was produced. There were about 1,000 such pages in all.

This was painstaking work, but the result was so neat that it gave a great sense of joy and the knowledge that it was probably correct. Today it is quite simply produced as Corollary 11.52, because the form of the solution led to the idea of a simple generalisation whose basis is the new principle of inclusion and exclusion for the mini-max. Had it not reduced to such a simple and appealing form, it would have almost certainly ended any interest in this area of research. This is one of life's interesting coincidences.

The generalisation led to the idea of formalising the determination of the waiting-time for the completion of one set measured from the completion time of another set, where the former set includes the latter set; the sets could be made disjoint and the results would be the same, but applications are naturally modelled by the way chosen. This new waiting-time process is referred to as a psi-process, which is written as Ψ*-process*.

What happened was a gradual process of extending the models under investigation into more-general and/or more-complex models. Then the corresponding *with-replacement* models seemed a natural follow-on, but had there been foresight of the difficulties involved, these might not have been investigated. However, having produced the results, the author has no regrets for doing so.

After further thought it became apparent that there was an opportunity to investigate the *static* and *dynamic* aspects of the random processes in a more-general fashion than had been done before.

This led to a more-formal organisation of the thesis into descriptions of applications, probability theory and application of the theory. The abstraction process of disassociating models from applications has led to a work on the foundations of probability theory.

It became clear that some real numerical calculations were required, and this led to the need to analyse the numerical properties of the expressions for the probability distributions and their moments.

There was so much to do that has not been done before in this area, that limitations have had to be made on what theory, applications and examples have been provided in this work. The decisions were generally based on the degree of extension of existing knowledge, difficulty in determining the results, utility provided by the results, variety of the results, usefulness of the techniques used to develop the results, neatness of the formulae, applicability to existing and new applications, and completeness of the questions answered within the domain of random allocation theory.

In the search for more applications, an awareness of the mathematics done on *Bernoulli's Lot Problem* and *Bernoulli's Marriage Problem* and an awareness of rules of the game *SET* led to further generalisations. This led to a question the author had wondered many years earlier about the spoilage of cakes on display in a shop and yet further generalisations of existing models.

As this continued, the number of perceived gaps in the thesis became fewer and the content became an integrated work on several aspects of essentially one type of problem. This has also produced a large volume of information that is aesthetically pleasing to the author.

The material is in the form of notation, definition, formularisation, lemma, theorem, corollary, remark, scholium and example. A formularisation is a special theorem-like tag that is used to specify the waiting time in terms of arrival positions. This is not used except when applying the *principle of inclusion and exclusion for* Ψ*-processes*. It is not used in any combinatorial arguments. Scholia are used in place of remarks when the remark is used to amplify a proof or course of reasoning, or as a commentary on another author's work.

Examples are dotted throughout the thesis, and there are many tables and graphs to illustrate the theory. This is prefixed by a description of all of the applications and suffixed by an analysis of the major applications with emphasis on particularly pertinent aspects of the theory.

This work is therefore one of a fundamentally pure nature with examples, applications and numerical analysis provided to supplement the theoretical material with elucidations, the power of the results, surprising consequences, comparisons between and within models, demonstrations of deriving known results as special cases of the general theory — sometimes with extraordinary ease and sometimes demonstrating the level of generalisation that has been achieved — illustrations of the theory, difficulty of implementation, speed of convergence and sometimes the elegance of the results.

One of the consequences of such organisation is that it showed the first format of the thesis to include some applications as generalisations of the theory when they could, in fact, be formulated as applications, albeit complex ones. It has also enabled the development of a unified set of notations, definitions and approaches for related models, including the static, waiting-time and dynamic models. This provides a pleasing experience when modelling different aspects of the same physical process.

Sometimes one resorts to utilising other people's results. This can be done blindly or by first investigating the value of those results in both the general and current context. In this case, such an investigation has led to a correct, but perhaps dislikable, conclusion about the practical use of some commonly-used inequalities.

It is common today to employ calculus to solve many problems, even for those of a discrete nature. This thesis employs the calculus of finite differences wherever possible, although this is no longer taught in mathematics classes at many universities. Of particular importance is the determination of differences by parts. This enabled the first method of simplification of a distribution, and it was the form of this alternative form of the distribution that provoked its current combinatorial interpretation. The author studied the complete work of Jordan's 1947 book on the subject[1]; this was after having completed reading and answering every question in Feller's book on discrete probability and its applications[2].

Less analysis has been included for *with-replacement* processes than for *without-replacement* processes, but one must stop somewhere. However, an adequate amount has been done for the

[1] C. Jordan, *Calculus of Finite Differences*, *2nd edition*, Chelsea Publ. Co., New York (1947).

[2] W. Feller, *An Introduction to Probability Theory and its Applications*, *Vol. I*, *3rd edition.* Wiley and Sons, New York (1968).

new *with-replacement* Ψ-process to illustrate the notation, theory and techniques required, and the similarities and differences between the two processes.

The scheme used for numbering sections, formulae and theorem-like statements is as follows. Each chapter has its own numbering of these. All theorem-like statements use the same numbering sequence. All references use the full reference to chapter, section, sub-section and subsub-section, etc.

There are a liberal number of references to various divisions, tables, figures and equations. This enables the reader of the PDF version of the thesis to easily review earlier material and view related later material earlier through a series of hyper-jumps.

Some symbols are included in the preliminary section on *General Notation and Conventions* instead of the section on *Symbols* because they are in common use. The *Symbols* section is reserved for symbols specific to this thesis or are in less-common use. A *Glossary* of terms appears in the prefatory matter; this appears before the list of symbols specific to this thesis because the latter makes more use of the former than vice versa.

Typing of this thesis was begun using T^3. Due to a glitch in the computer, over four months of secretarial typing, which was mostly technical typing, were lost as there was no warning as to any problem whilst working on a material that was still fine. The backup disks had also been overwritten with garbage. Most of what was lost does not appear in the thesis now. The lost text associated with the current topic was entered again. Then LaTeX appeared and soon after came Scientific Word, which provided a visual interface to LaTeX. The text was entered yet again. The product developed into Scientific WorkPlace, and version 5.5 of this software was used to typeset the final version of this thesis. The packages used are *eurosym*, *amssymb*, *sw20au* (a modified version of the *mitthesis* style that changes the front matter, now included only because of the matching .CST file that Scientific WorkPlace uses), *amsmath*, *geometry*, *float*, *fancyhdr*, *trees* (modified to avoid a naming conflict, to allow single branches and to display the branches from node zero further apart), *curves*, *alltt*, *supertabular*, *minitoc*, *setspace* and *hyperref*.

The electronic version of this thesis is provided as a PDF file. It has had bookmarks created for the table of contents in order to provide an additional navigation method. These bookmarks are in a different font, namely unicode, because the standard LaTeX fonts used for mathematics and other symbols are not included in the format's specification. With a few of these, it was necessary to use a standard-size font instead of a subscripted font, and in a few others, it was considered reasonable to use words instead of symbols. None of this changes the textual content of the thesis.

One might well ask why it took so long to complete this thesis, and the answer has to do with a combination of several things: the need to earn an income, a car accident (in which the other driver drove through a stop sign in top gear without slowing down), health (whiplash injury, memory loss and other brain damage caused by the car accident (e.g. inability to concentrate, think and calculate, and sometimes forgetting my own name), multiple sclerosis, chronic fatigue, chronic osteo-arthritic pain in one hip), fitness, marriage and a child, working as a computer science lecturer and being required to produce in that field, and later running my own computer software development business with several staff. However, the total effort expended on the material in this thesis is equivalent to 4 years and 8 months at 40 hours per week, or 3 years and 2 months at 60 hours per week.

When researching mathematics, one often travels down a path that leads nowhere, and must choose whether to continue or follow another path from some point already travelled or to abandon the endeavour altogether. Sometimes we are lucky and sometimes the hard work pays off. In this case, there were many who said problems in this thesis could not be solved or that the author would not be the one to solve them. In this case then, there was luck in finding the article by Hauer and Templeton, but there was no luck recognising that the author knew in an instant that he had the ability to generalise the result and make something of it. One must always be preparing oneself for good luck to be taken advantage of, and always be prepared to go down a path not yet trodden. After that, sheer determination, will-power and effort are the keys.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution, and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

A few small parts of this thesis have appeared in the following refereed papers in which my own work was that of a full pro-rata contributor:

- W. Henderson, R.W. Kennington and C.E.M. Pearce, *A Second Look at a Problem of Queueing in Lanes*, Trans. Sci. **1** (1984), 85-93.

- W. Henderson, R.W. Kennington and C.E.M. Pearce, *Stochastic Processes and Combinatoric Identities.* Combinatorial Mathematics X, Proceedings, Adelaide 1982. Springer-Verlag (1983).

These papers are included at the end of this thesis.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Raymond Kennington

Acknowledgements

There are many people and events that have affected my life in a way to enable me to complete this work. I mention the most important here, including some from my childhood.

I thank Professor Charles Pearce not only for his inspirational and lively discussions, but also for the encouragement and emotional support he provided me during some trying times, and also for taking over as my supervisor when my previous supervisor became critically ill. Charles provided me with the opportunity to do a postgraduate degree. Charles made mathematics exciting and a growing, living thing. He also introduced me (and the rest of his 3rd-year class) to the significance of the *squiggle* symbol.

I thank my supervisor Associate Professor Bill Henderson for our discussions and his continuance with me during the years that this thesis has spanned. Bill provided me a framework within which I could enhance my analytical skills and research techniques. As I write up this thesis, many of Bill's comments, suggestions and ways of thinking about the presentation of the material pop into consciousness at appropriate times and I recall them with the sound of his voice as if he were standing right beside me.

There are several reasons for my early interest and ability in mathematics. Foremost amongst these is the way my father, Noel, played mathematical games with me for fun; this began at about age 6. Teaching me Trachtenberg methods from the outset played an important role in developing a sound foundation for my development as a mathematician. I thank dad also for proof-reading the non-mathematical aspects of the first draft of this thesis, especially as he has reached his 81'st year.

Literally hundreds of games of many different kinds were played by everyone in the family. I consider these games as extremely important in the formation of my mathematical ability. Of particular note is the game of cribbage, which my father played with me often, for I attach to it a significant role in developing my ability in combinatorial analysis. My mother, Barbara, was particularly adept at board games and played, amongst others, chess, German draughts, Halma and Mühle with me, especially as a very young child. I recall playing hundreds of different card games with my younger brother, Duncan, using up to 40 standard packs of cards in a single game, with one of these games, *memory*, utilising every room in the house, and others of much greater complexity too.

My elder brother, Alan, introduced me to Martin Gardner's puzzles and diversions and to Scientific American; Martin Gardner makes the thought processes associated with mathematical discovery and insight a thrilling adventure, and every school library should have them. I thank Alan for providing me with an intellectual role model from early childhood, for explaining mathematics to me whenever I asked, and for never letting me take any written notes away with me after our discussions. He also lent me Lancelot Hogben's *Mathematics for the Million*, which includes a discussion of the algebra of choice and chance.

In memory of my mother, who did not live to see me finish this thesis, I thank her for her dreams for all her children, for her determination that I succeed at university in my mathematical endeavours, and for the hard life she endured so that her dreams could come true through her children.

The awarding of an Australian Commonwealth Scholarship to me is acknowledged, for without it I would not have been able to study at the university after high school. The awarding of a University of Adelaide Postgraduate Research Scholarship to me for 10 months and 13 days is acknowledged.

I give thanks to the teachers at Salisbury East High School, a school whose motto was during my time *Tenax ad Asperum*, meaning *fight to the bitter end.* We took this to mean *persist with your aspirations until the goal is achieved no matter what the difficulties.* Of those teachers, one in particular needs to be singled out, namely Mr. J.M. Dayman, who was determined to teach his students mathematics in a highly professional manner and was able to teach probability and statistics in a competent and exciting way. Mr. Duffield, my grade seven teacher, deserves an acknowledgement for allowing me to teach him and some of my class-mates some mathematics in class, including how to find cube roots of exact 9-digit cubes without hesitation, and the process of finding square and cube roots like long divisions.

I thank all of my lecturers and tutors in the mathematics departments at the University of Adelaide for their excellence in mathematics and the teaching of it. One more of particular note is Dr. Bill Venables, my foremost mathematical statistics lecturer, who is not only a superb teacher but also introduced me to the philosophy of probability. Another, Dr. Franz Salzborn, who taught optimisation theory, is a mathematician I admire greatly and thank for the long discussions we had.

The staff, in particular John MacKichan and George Pearson, at MacKichan Software Inc, home of Scientific WorkPlace, deserve a mention for all the questions they have answered about their product, the LaTeX language, and various packages. Also, the authors of the *minitoc* package, Jean-Pierre Drucbert, and the *hyperref* package, Heiko Oberdiek, have both been very helpful. Without all their help, the typeset appearance would have been quite different and it would have taken me far longer to produce the final typeset version of this thesis.

Finally, I would like to thank numerous friends and others throughout the years of my post-graduate study who have encouraged me to finish this work.

iam confecet

alea iacta est

quod feci, feci - et

quidquid id est, id est

In memory of

Bill Henderson (my supervisor)

Barbara Kennington (née Mehling; my mother)

Penny Barlow (my friend)

Abstract

This thesis provides a general methodology for classifying and describing many combinatoric problems, systematising and finding theoretical expressions for quantities of interest, and investigating their feasible numerical evaluation. Unifying notation and definitions are provided.

Our knowledge of random allocations is also extended. This is achieved by investigating new processes, generalising known processes, and by providing a formal structure and innovative techniques for analysing them.

The random allocation models described in this thesis can be classified as *occupancy urn models*, in which we have a sequence of urns and throw balls into them, and investigate static, waiting-time and dynamic processes. Various structures are placed on the relationship(s) between cells, balls, and the selection of items being distributed, including *varieties*, *batch arrivals*, *taboo sets* and *blocking sets*.

Static, *waiting-time* and *dynamic* processes are investigated. Both *without-replacement* and *with-replacement* sampling types are considered. Emphasis is placed on the distributions of waiting-times for one or more events to occur measured from the time a particular event occurs; this begins as an abstraction and generalisation of a model of departures of cars parked in lanes. One of several additional determinations is the *platoon size* distribution.

Models are analysed using combinatorial analysis and Markov Chains. Global attributes are measured, including *maximum waits*, *maximum room required*, *moments* and *the clustering of completions*. Various conversion formulae have been devised to reduce calculation times by several orders of magnitude.

New and extended applications include *Queueing in Lanes*, *Cake Displays*, *Coupon Collector's Problem*, *Sock-Sorting*, *Matching Dependent Sets* (including *Genetic Code Attribute Matching* and the game *SET*), the *Zig-Zag Problem*, *Testing for Randomness* (including the *Cake Display Test*, which is a *without-replacement* test similar to the standard *Empty Cell* test), *Waiting for Luggage at an Airport*, *Breakdowns in a Network*, *Learning Theory* and *Estimating the Number of Skeletons at an Archaeological Dig*.

Fundamental, *reduction* and *covering* theorems provide ways to reduce the number of calculations required.

New combinatorial identities are discovered and a well-known one is proved in a combinatorial way for the first time.

Some known results are derived from simple cases of the general models.

Keywords: *random allocations*, *occupancy urn models*, *static models*, *dynamic models*, *waiting-time models*, *without-replacement*, *with-replacement*, *Psi-processes*, *Psi-numbers*, *Psi-probabilities, probability theory, P. of I. E. for the mini-max*, *Bonferroni's Inequalities*, *decomposition theorem*, *minimal coverings*, *taboo sets*, *blocking*, *platoon size*, *combinatorial identities*, *Markov Chains*, *incomplete arrival streams*, *batch arrivals*, *varieties*, *asymptotics*, *asymptotic distributions*, *approximations*, *moments, numerics*, *queueing in lanes*, *sock sorting*, the game *SET*, *coupon-collector's problem, cake display problem*, *the zig-zag problem.*

Summary

Chapter 1 provides a global introduction and motivation. It begins by placing the work into its elementary context of occupancy urn models. Then it provides a historic view and a brief survey of problems, analytical techniques, computational issues, and applications.

Chapter 2 provides a description and discussion of numerous applications of the theory, most of which have been used in the thesis to illustrate theory, examine numerical issues, and provide challenges that required further theoretical development.

Chapter 3 provides the formal structure and notation for the random processes. This includes the common elements, the models and the random processes.

Chapter 4 discusses some computational aspects of determining numerical results for these applications. This includes a discussion of precise and asymptotic forms, the problems with Bonferroni's inequalities, the number of calculations involved, the size of the numbers involved, the digits of accuracy required, the processing time required, and the need to convert combinatorial sums to simpler forms.

Chapter 5 lists formulae that have been used many times, some of which have been generalised, and introduces the new powerful principle of inclusion and exclusion for the minimum of the maximum of a collection of events.

Chapter 6 investigates several aspects of the basic *without-replacement* processes for both the *waiting-time* and *static* processes. The original distribution is converted to another form that is computationally simpler and enables the moments to be determined with a greatly reduced number of terms. It provides the Ψ-numbers of first kind and uses them to provide a much-simpler formula for calculation of probabilities by writing the distribution as a linear combination of Ψ_1-numbers, called the *decomposition formula*. It shows how to remove redundant sets prior to determining the distribution formula and demonstrates its significance with an application to *The 2-D Zig-Zag Problem*; this leads to the notion of a *minimal covering*. Also investigated are cumulative

distributions, approximations, the number still required upon arrival, the waiting time from the τth arrival till the σth arrival, estimating the number in the population from various observations of the process, and a Markov Chain for each of the *waiting-time* and *static* processes.

Chapter 7 investigates several aspects of the basic *with-replacement waiting-time* processes. The original distribution is converted to another form that is computationally simpler and enables the moments to be determined with a greatly reduced number of terms, albeit the formula looks quite complicated. It provides the Ψ-numbers of second kind and uses them to provide a much-simpler formula for calculation of probabilities by writing the distribution as a linear combination of Ψ_2-numbers, called the *decomposition formula*. It shows how to remove redundant sets prior to determining the distribution formula; this leads to the notion of a *minimal covering*. Also investigated are asymptotic approximations, estimating the number in the population, estimating the number of trials required to achieve specified observations with an application to *The Bird-Watcher's Problem*, and a Markov Chain for the *waiting-time* process.

Chapter 8 provides some new combinatorial identities, some generalisations of known identities, and a new combinatorial proof of well-known identity.

Chapter 9 extends the basic *without-replacement* model to waiting for a minimum number of completions, taboo sets that must not be completed before success is achieved, measuring the wait until the event can no longer occur due to one or more sets being completed that *block* completion of the desired sets, incomplete arrival streams, requiring only a partial completion of desired sets, batch arrivals, and considers models in which each of the N cells has v distinct attributes called *varieties*, whose arrivals may be either simultaneous and independent of each other or randomised.

Chapter 10 extends the basic *with-replacement* model to waiting for a minimum number of completions, taboo sets that must not be completed before success is achieved, and measuring the wait until the event can no longer occur due to one or more sets being completed that *block* completion of the desired sets.

Chapter 11 provides global properties for the *without-replacement* process. This includes demonstrating the complexity of determining the maximum possible wait in the general case, deriving the maximum possible wait for the uni-directional and bi-directional car parking models, determining the moments and a simplification of them for the basic process, for a minimum number of completions and for batch arrivals, determining the total expected wait for all arrivals, moments for the number still required upon arrival, and the expected platoon size when a completion occurs; this is useful in analysing, for example, the congestion at the exits of a parking lot.

This chapter also investigates the *dynamic* model and specialises the results to the classic *Bernoulli's Marriage Problem*, sock-matching with multi-legged beings, and the new *Cake Display Problem*; this includes a method of determining the expected duration of the maximum number of cakes on display. Then it investigates the rate at which completions occur during the entire process using means, variances and covariances, compares the expected waiting times of the original *Hauer-Templeton* uni-directional exit model with the new bi-directional exit model, and a generating function.

Chapter 12 provides global properties for the *with-replacement* process. This includes determining the conditional moments, conditioned on being able to complete the set, and a simplification of them for the basic process, and compares the means of this *with-replacement* process with the *without-replacement* process. This chapter also provides limiting conditional moments as the number of arrivals increases indefinitely.

Chapter 13 applies the theory to various *without-replacement* applications. Several of these applications seem like generalisations of the main model, but are modelled in such a way as to be able to apply the general theory.

For *Queueing in Lanes*, this includes the total waits for both the uni- and bi-directional models, waiting times, comparison of delays for a parking lot, platoon departure size, parallel arrivals and an investigation into the effect a parking attendant might have on the waiting times.

For *The 2d-Gap Problem*, which is a variation of the Hauer-Templeton *Parking Lot Problem* in which there are regular spaces between cars in lanes that allow cars to swap lanes at the gaps, applying the *decomposition formula* is a very complex proposition. In the *zig-zagging* models, the minimal covering theorem is aptly applied, and the notion of an asymptotic result based on increasing the number of completion-sets (or paths in this case) to be included in the calculation is summarily dismissed.

The game *SET* is investigated in the cases of single placement and batch placement of cards, and both waiting for a particular set and any set are investigated. Known results for the means of the number of sets in K cards and at the kth card are determined from the more-general formulae. Also, the variances are provided for the first time.

The *Cake Display Problem* is used to illustrate many aspects of the theory. Of particular note is the application of the general theory to apply to the case of allowing more that one physical cake for each type of cake, but with only one of each type of cake appearing on the display counter at a time.

Other aspects of the theory are also applied and discussed within the applications listed above. There are also other applications.

Chapter 14 applies the theory to various *with-replacement* applications, including *The Bird-Watcher's Problem* and *The Bombing Raid*, the latter of which illustrates the use of the decomposition coefficients and determines blocking probabilities.

Chapter 15 numerically analyses some formulae that were claimed as computationally superior to other forms, and also investigates the numerical accuracy problems involved in calculating probabilities for *with-replacement* processes.

Chapter 16 applies some of the distributions to testing the randomness of data. This includes a permutation test and a Chi-Square test based on the basic *without-replacement* process, three tests based on the *Cake Display Problem*, and a test based on the *with-replacement Bird-Watcher's Problem* that is applied to testing the randomness of the decimal digits of π.

Chapter 17 lists the major accomplishments, suggests some ideas for further research, and provides a final comment.