

现代通信中的排队论



全国高技术重点图书·通信技术领域

现代通信中的排队论

陈鑫林 编著



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.com.cn>

全国高技术重点图书·通信技术领域

现代通信中的排队论

陈鑫林 编 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

排队论问题是从通信中提炼出来的,而通信的发展是排队论发展的重要推进力量和源泉。本书从经典排队论开始,针对通信的发展详细讨论了适应其发展的排队论方法。本书共分5章,从Markov排队模型讲起,讨论了Markov排队模型的推广、Neuts的矩阵几何理论、排队输入为MMPP(2)的排队系统的求解、分组数字图像业务流的建模,并展望了排队论在宽带数字网研究领域的发展。

本书的雏形是作者为邮电科学研究院研究生部“现代通信中的排队论”一课所编定的讲义,经过三年教学实践,作者不断修改,形成了本书循序渐进、说理清晰、计算严谨的风格。

本书可作为通信专业高年级本科生、研究生教材,也可供相关领域教师和科研人员参考使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,翻版必究。

图书在版编目(CIP)数据

现代通信中的排队论/陈鑫林编著. - 北京:电子工业出版社,1999.11

(全国高技术重点图书·通信技术领域)

ISBN 7-5053-5480-9

I. 现… II. 陈… III. 通信-排队论 IV. TN911.

中国版本图书馆 CIP 数据核字(1999)第 14536 号

丛 书 名: 全国高技术重点图书·通信技术领域

书 名: 现代通信中的排队论

编 著 者: 陈鑫林

策 划: 张 欣

责任编辑: 徐 堃

特约编辑: 荆显英

排版制作: 电子工业出版社计算机排版室

印 刷 者: 北京朝阳隆华印刷厂

装 订 者: 三河市新伟装订厂

出版发行: 电子工业出版社 URL: <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销: 各地新华书店

开 本: 787×1092 1/16 印张: 13 字数: 320 千字

版 次: 1999 年 11 月第 1 版 1999 年 11 月第 1 次印刷

书 号: ISBN 7-5053-5480-9
TN·1293

印 数: 2000 册 定价: 30.00 元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁盘或光盘有问题者,请向购买书店调换。

若书店售缺,请与本社发行部联系调换。电话 68279077

《全国高科技重点图书》出版指导委员会

主任: 朱丽兰

付主任: 刘 杲

卢鸣谷

总干事: 罗见龙

梁祥丰

委员: (依姓氏笔划为序)

王大中

王为珍

牛田佳

王守武

刘 仁

刘 杲

卢鸣谷

叶培大

朱丽兰

孙宝寅

师昌绪

任新民

杨牧之

杨嘉墀

陈芳允

陈能宽

罗见龙

周炳琨

欧阳莲

张兆祺

张钰珍

张效祥

赵忠贤

顾孝诚

徐修存

谈德颜

龚 刚

梁祥丰

《全国高科技重点图书·信息获取与处理领域》 编审委员会

主任委员: 陈芳允

委员: 汪成为

杨震明

袁宝宗

邓又强

《全国高科技重点图书编审委员会人员名单》 通信技术领域

主任委员: 叶培大

委员: 陈俊亮

李树岭

陈芳烈

姚 彦

徐大雄

程时昕

前 言

排队论的第一篇论文,即A.K.Erlang的著作《Solution of Some Problems in the Theory of Probability of Significance in Automatic Telephone Exchanges》,是针对电话交换机的使用情况进行分析的。因此排队论问题最初是从通信中提炼出来的。

在这以后,通信始终是排队论发展的重要推进力量和源泉。排队论经过大约半个世纪的发展,已经成为一个独立的数学分支(有时把它归在更大的类——“运筹学”中),特别是它在国防和经济发展中得到了直接的应用,所以吸引了大量学者的研究兴趣。关于独立到达和独立服务的排队论问题,在主要方面已基本趋于完善,大量的实际问题又给我们提出了突破两个独立性的新要求。在最近20年来,通信发展异常迅速,它与计算机的结合对排队论提出了许多新的课题,吸引了通信、计算机和应用数学三方面的学者共同的关心和努力。有关排队论的学术论文数以千计,它们主要发表在像IEEE Transactions on Commun.(on Information Theory)、IEEE Journal on SAC、Communication Statistics in Stochastic Models、Journal of Applied Probability、Advance Applied Probability、Operation Research、Bell Syst.Tech.Journal等著名的期刊和相关的专业性学术会议上。

通信的发展给排队论提出了许多新的研究课题,这些课题的解决又促进了通信的发展。在宽带综合业务数字网中,异步传输模式、统计复用和随机多址接入中都涉及许多排队论问题,而且是一些尚未完全解决的问题。在新的通信体制中,到达的业务流有其相关特性。同时,在以往的许多情况下,到达时间间隔和服务时间二者之中有一个是指数分布的,而指数分布在连续型概率分布中是唯一的具有无记忆性质的分布函数。无记忆性在数学上又有许多使问题简化的特性可以加以利用,在新的情况下不能利用这些条件,致使数学处理变得困难起来。因此,必须开创排队论的新理论和新方法。通信理论专家、数学家和计算机科学家们经过近20多年的努力,创立了许多适应通信发展的排队论方法。

近年来,笔者有幸得到国家自然科学基金的资助,研究ATM中业务流的数学模型及其排队分析方法。从所看到的国内出版的包括教科书在内的排队论著作中,还未见到涉及近10~20年发展起来的内容。鉴于通信理论和技术的飞速发展,需要大量各种层次的通信人才,我国已经和正在培养通信学科的研究生。笔者于3年前,在邮电科学研究院研究生部开设了“现代通信中的排队论”课程,并编写了讲义,它是本书的雏形。经过3年的教学实践,对讲义大改了两次就形成了本书的框架。笔者以为,当前需要有一本对研究生介绍排队论在通信中应用的书,以便他们深入研究。但是,要想在一本教科书中反映排队论的许多较新成果是有困难的,因而只能做到介绍比较成熟的典型方法,但这

并不排斥写一本“引进门”的书，为的是节省读者的时间，以免每一位要跨入本研究领域的人都要钻进浩繁的文献中去搜索。在本书的前身讲义中取材较多，后来经多次删简才成为日前的篇幅。如果读者想了解得更多，可以阅读本书最后的结束语。

本书虽然也从经典排队论开始讨论，但这仅是为了阅读上的方便，然而即使如此，在经典排队论的篇幅中仍然包含着大量的新内容。

排队论涉及的基础数学知识是概率论、Laplace 变换、母函数、特征值与特征向量以及非线性方程求解等。特别要建议的是，无论是做练习还是课题研究，最好手头有计算机软件，如 Mathematica、Matlab、MathCad 等。

以本书为教材，学时可安排在60课时左右。

由于笔者学识有限，错误难免，敬请读者指正。

陈鑫林

于电信科学技术研究院

目 录

第1章 Markov 排队模型及其应用	(1)
1.1 记号与分类	(1)
1.2 指数分布和普松(Poisson)过程	(2)
1.2.1 指数分布	(2)
1.2.2 Poisson 过程	(4)
附录: Laplace 变换和母函数	(6)
1.3 Markov 链	(8)
1.3.1 离散时间的 Markov 链(DTMC)	(8)
1.3.2 连续时间的 Markov 链(CTMC)	(13)
1.3.2.1 Q-矩阵	(14)
1.3.2.2 求 $\pi(t)$	(15)
1.3.2.3 连续时间的齐次 Markov 链	(16)
1.3.2.4 求平稳概率分布	(17)
1.3.3 生-灭过程	(19)
1.4 Poisson 到达指数服务的排队系统及其应用	(19)
1.4.1 M/M/1 占有(系统状态)分布的瞬时解	(19)
1.4.2 一般(有限状态)生-灭过程的瞬时解	(23)
1.4.2.1 特征值分析方法	(24)
1.4.2.2 随机化方法(又称 Jensen 方法)	(25)
1.4.2.3 应用举例(一种路由选择方法)	(26)
1.5 M/M/1 的稳态解	(28)
1.5.1 M/M/1 占有分布	(28)
1.5.2 嵌入过程	(29)
1.5.3 Little 定理	(31)
1.5.4 M/M/1 的稳态逗留时间与等待时间	(32)
1.5.5 M/M/1 排队系统的忙期分析	(33)
1.5.6 在稳定状态下指数系统的平衡方程法	(35)
1.5.7 应用: 话务工程中的计算方法	(36)
1.5.7.1 有限源	(36)
1.5.7.2 无限源	(38)
1.6 M/M/1 排队模型的初步推广	(42)
1.6.1 排队模型 $M/E_r/1$	(42)
1.6.2 排队模型 $E_r/M/1$	(44)

1.6.3	批量到达的排队系统	(46)
1.7	单服务器的指数服务网路	(47)
1.7.1	任意互联(随机路由)	(47)
1.7.2	单服务器的封闭网路(随机路由)	(48)
第2章	Markov 排队模型的推广	(52)
2.1	M/G/1 排队系统	(52)
2.1.1	M/G/1 系统的逗留时间	(55)
2.1.2	M/G/1 系统的等待时间	(56)
2.1.3	M/G/1 排队系统的忙期	(56)
2.2	求 M/G/1 排队系统队长分布的两种方法	(57)
2.2.1	DFT 方法	(57)
2.2.2	递推算法	(61)
2.3	G/M/m 排队系统	(63)
2.3.1	G/M/1 排队系统的队长分布	(63)
2.3.2	G/M/m 排队系统	(67)
2.3.3	等待时间的分布	(71)
2.4	依赖于相位的到达率和服务率	(72)
2.5	更新过程的若干基本结果	(84)
2.5.1	更新过程	(84)
2.5.2	由更新过程导出 M/G/1 系统等待时间的平均值	(86)
2.5.3	忙期和交替更新过程	(87)
2.5.4	更新过程的矩量表示及其 Laplace 变换	(88)
第3章	Neuts 的矩阵几何理论	(92)
3.1	矩阵几何方法	(92)
3.1.1	简单情形	(92)
3.1.2	用特征值分析方法求率矩阵 R	(98)
3.1.3	相位型服务时间分布的 M/PH/1 排队模型	(100)
3.2	M/G/1 型和 G/M/1 型排队模型	(104)
3.2.1	G/M/1 型排队模型的求解	(104)
3.2.2	M/G/1 型排队模型的求解	(105)
3.2.3	在统计复用中的应用	(108)
第4章	排队输入为 MMPP(2) 的排队系统的求解	(115)
4.1	(GI+M)/M/1 模型的排队分析	(115)
4.1.1	分段马尔可夫过程(Piecewise Markov Processes)	(115)
4.1.2	嵌入 Markov 过程的平稳分布	(116)
4.1.3	两类顾客的等待时间分布	(120)
4.1.4	例及开关 Poisson 过程	(122)

4.1.4.1	GI 是参数为 ν 的指数分布	(122)
4.1.4.2	GI 是二阶参数为 ν 的 Erlang 分布	(122)
4.1.4.3	开关 Poisson 过程 (Interrupted Poisson Processes, IPP)	(124)
4.1.4.4	GI 是参数为 ω, γ, λ 的 IPP	(127)
4.2	MMPP(2) 的某些统计量	(129)
4.2.1	MMPP(2) 的相关特性	(130)
4.2.2	MMPP(2) 在 $(0, t)$ 的到达数的母函数	(133)
4.2.3	信包话音过程的统计特性	(142)
4.3	排队输入为 MMPP(2) 的排队系统的求解	(145)
4.3.1	MMPP(2)/M/1 的排队性能分析	(145)
4.3.2	MMPP(2)/G/1 的排队性能分析	(147)
第 5 章	分组数字图像业务流的建模举例	(161)
5.1	流体流动模型	(161)
5.1.1	模型	(161)
5.1.2	特征值和特征向量	(162)
5.1.3	求解	(167)
5.2	频域分析方法	(171)
5.2.1	输入相关函数	(172)
5.2.2	输入功率谱	(172)
5.2.3	输入周期链	(174)
5.2.4	周期链输入的排队分析	(175)
结束语		(178)
练习		(187)
参考文献		(193)

第1章 Markov 排队模型及其应用

本章内容提要: 首先引入排队论所必需的指数分布和Poisson过程的知识,特别是它们的性质,然后介绍Markov链和半Markov过程,并在此基础上开始进入排队论的讨论。先是最简单的M/M/1排队模型,讨论求瞬态精确解和近似解的方法。然后讨论一般的生-灭过程排队模型求稳态解的经典方法,这就基本上概括了话务工程中所用到的那些主要的排队模型。再进一步,我们对M/M/1排队模型初步地推广,讨论M/E_r/1和E_r/M/1这两种模型以及批量到达的排队模型。最后引进求解封闭网络的最新结果——J.Gordon方法。

在排队论的开创和发展的历史中,Markov排队模型始终起着重要的作用。排队论起源于A.K.Erlang的著名论文^[1],该论文的主题是电话交换机的使用情况分析。因此可以说排队论问题最初是从通信中提炼出来的。虽然在第二次世界大战之后排队论得到了迅速的发展,并在通信、交通、生产线、存储问题、救护、公共事业、计算机等方面都得到了应用,但通信始终是排队论发展的重要推进力量和源泉。人们把相继到达“顾客”的到达时间间隔和服务时间都相互独立的排队论内容称为经典(或古典)排队论。经典排队论的主要内容已可见于众多的教科书。在最近20年来,通信发展异常迅速,它与计算机的结合对排队论提出了许多新的课题,这些新课题的解决仍受到经典排队论方法的影响。因此在介绍排队论的许多最新成果之前,用一定的篇幅介绍经典内容是十分必要的。本章除了这一目的之外还将介绍经典排队论在通信中的应用,而有些方面也是最新的成果。

1.1 记号与分类

除非特别声明,我们在本书中将使用以下记号:

$\tilde{r}_n$ 表示第 n 个顾客的到达时刻;

$\tilde{x}_n$ 表示第 n 个顾客的服务时间;

$\tilde{w}_n$ 表示第 n 个顾客在系统中的等待时间;

$\tilde{s}_n$ 表示第 n 个顾客在系统中的逗留时间(包括等待与服务), $\tilde{s}_n = \tilde{w}_n + \tilde{x}_n$;

$\tilde{t}_{n+1}$ 表示第 $n+1$ 个顾客与第 n 个顾客到达系统的间隔时间;

第 $n+1$ 个顾客的等待时间=(第 n 个顾客的等待时间+第 n 个顾客的服务时间)-第 $n+1$ 个顾客与第 n 个顾客到达系统的间隔时间。

如果出现负数,说明到达间隔较长,在第 $n+1$ 个顾客到达时第 n 个顾客已服务完毕,他就不用等待,以致等待时间为0。用符号表示是

$$\tilde{w}_{n+1} = (\tilde{r}_n + \tilde{w}_n + \tilde{x}_n - \tilde{r}_{n+1})^+ = (\tilde{w}_n + \tilde{x}_n - \tilde{t}_{n+1})^+$$

这里符号 $(x)^+$ 的意义是当 $x \leq 0$ 时 $(x)^+ = 0$,当 $x > 0$ 时 $(x)^+ = x$ 。

在时刻 t 的未完成工作是:取某个观察时刻 t ,若在时刻 t 之后无新的到达,则从 t 开始直到系统中所有的顾客都完成服务而离开系统为止的这段时间。记未完成的工作为 $U(t^-)$ (因为在时刻 t 可能有顾客到达)。有时称未完成工作为虚等待时间。

排队系统通常按以下情况分类:

1. 到达间隔的分布形式: $F_i(t) = P[\tilde{t} < t]$;
2. 服务时间的分布形式: $F_s(x) = P[\tilde{x} < x]$ (有时也用 $H(\cdot)$ 表示服务时间的分布函数);
3. 在一个批量到达中的顾客数;
4. 服务器(台)的个数;
5. 服务的规则(在本书中, 我们只讨论先到先服务的排队规则, 简记FCFS);
6. 允许等待的顾客数, 即缓冲器的容量;
7. 在群体中的顾客数, 即顾客源是有限的还是无限的(通常只有在有限的顾客源时才做特别的交待)。

在排队论中, 人们感兴趣的量有:

等待时间的分布 $F_w(x) = P[\tilde{w} < x]$, 平均等待时间 $\bar{W}$;

系统时间(也称逗留时间)的分布 $F_s(x) = P[\tilde{s} < x]$, 平均系统时间 $\bar{S}$;

在系统中顾客数(也称系统占有数)的分布 $\{p_n, n = 0, 1, 2, \dots\}$, 其中 $p_n = P[\tilde{n} = n]$, $n = 0, 1, 2, \dots$, 系统的平均占有数(也称平均队长) $\bar{N}$;

服务器忙着(或空闲)的概率;

忙期长度的分布;

在忙期被服务顾客数的分布以及它的均值。忙期分析在优先排队中是一个有用的工具。

对于排队模型, D.G.Kendall 建议采用记号 $A/B/c/K$ 。其中 A 代表到达时间间隔的分布, B 代表服务时间的分布, c 代表服务器个数, K 代表等待位置的数目。特别地, 人们常用 M 表示指数分布, GI 表示一般的独立分布。

在以下的行文中, 我们用大写的PDF表示概率分布函数, 用小写的pdf表示概率分布密度函数, 对此我们将不再特别声明。

1.2 指数分布和普松(Poisson)过程

指数分布和普松过程在排队论中有着特殊的地位。这是因为, 一方面指数分布在连续型概率分布中是唯一的具有无记忆性质的分布, 普松分布又和指数分布有着紧密的关系。另一方面, 实验证明普松分布是电话呼叫数概率分布的一种比较好的近似, 而指数分布又是电话通话时间概率分布的一种好的近似, 它们在排队论的发展历史中起了很大的作用, 并继续起着重要的作用。

1.2.1 指数分布

定义 一个随机变量 $\tilde{x}$ 当且仅当对任意的 $\alpha, \beta \geq 0$ 满足条件

$$P[\tilde{x} > \alpha + \beta | \tilde{x} > \beta] = P[\tilde{x} > \alpha] \quad (1.1)$$

时, 则称 $\tilde{x}$ 的分布是无记忆的。无记忆性的直观理解是: 一个物体的使用寿命是指被使用的时间, 它是一个随机变量。如果该物体不论被使用了多久, 其剩余寿命的分布与总寿命的分布完全相同, 那么这种寿命分布是无记忆的, 体现了“永远年轻”。

按条件概率定义, 我们有

$$P[\tilde{x} > \alpha + \beta | \tilde{x} > \beta] = \frac{P[\tilde{x} > \alpha + \beta, \tilde{x} > \beta]}{P[\tilde{x} > \beta]} = \frac{P[\tilde{x} > \alpha + \beta]}{P[\tilde{x} > \beta]}$$

如果随机变量 $\tilde{x}$ 是无记忆的, 那么由(1.1)及上式, 可得

$$P[\tilde{x} > \alpha + \beta] = P[\tilde{x} > \alpha] \cdot P[\tilde{x} > \beta] \quad (1.2)$$

反之, 如果随机变量的分布满足(1.2), 则该分布是无记忆的。

因为指数分布的余分布函数为 $G(t) = P[T > t] = 1 - F(t) = e^{-\lambda t}$, 而 $G(t_1 + t_2) = e^{-\lambda(t_1 + t_2)} = e^{-\lambda t_1} e^{-\lambda t_2} = G(t_1)G(t_2)$, 故指数分布是无记忆的。当服务时间是指数分布时, 则不论顾客占用服务台多久, 其剩余的服务时间仍为指数分布的随机变量。在连续型随机变量中指数分布是唯一的具有无记忆分布的随机变量。在离散随机变量中, 几何分布是唯一的具有无记忆性质的随机变量。

指数分布有以下性质:

设 $\tilde{x}$ 和 $\tilde{y}$ 分别是参数为 α 和 β 的指数分布随机变量, 且相互独立, 则

1. $\tilde{z} = \min(\tilde{x}, \tilde{y})$ 的分布也是指数分布的, 且参数为 $\alpha + \beta$;

2. $P[\tilde{x} < \tilde{y}] = \frac{\alpha}{(\alpha + \beta)}$;

3. $F_{\tilde{z}/\tilde{x} < \tilde{y}}(z) = F_{\tilde{z}}(z)$;

4. 对随机变量 $\tilde{x}$ 和 $\tilde{y}$ 做重复试验, 令 $\tilde{n}_x$ 是第一次使 $\tilde{y}$ 小于 $\tilde{x}$ 时以前的试验数, 也就是说, 做了 $\tilde{n}_x + 1$ 次试验, 前 $\tilde{n}_x$ 次都是 $\tilde{x} < \tilde{y}$, 第 $\tilde{n} + 1$ 次才有 $\tilde{y} < \tilde{x}$, 则

$$P[\tilde{n}_x = n] = \left(\frac{\alpha}{\alpha + \beta}\right)^n \left(\frac{\beta}{\alpha + \beta}\right) \quad (n \geq 0)$$

5. 令 $\{\tilde{x}_i, i = 1, 2, \dots\}$ 是相互独立的参数为 α 的指数分布随机变量序列, $\tilde{n}$ 是参数为 p 的几何分布随机变量, 与 $\{\tilde{x}_i, i = 1, 2, \dots\}$ 相互独立, 令

$$\tilde{u} = \sum_{i=1}^{\tilde{n}} \tilde{x}_i$$

则 $\tilde{u}$ 是参数为 $p\alpha$ 的指数分布。

证明

1. 因为 $P[\tilde{z} > z] = P[\min(\tilde{x}, \tilde{y}) > z] = P[\tilde{x} > z, \tilde{y} > z] = P[\tilde{x} > z]P[\tilde{y} > z]$ (由于 $\tilde{x}$ 和 $\tilde{y}$ 的独立性)。又因 $P[\tilde{x} > z] = e^{-\alpha z}$, $P[\tilde{y} > z] = e^{-\beta z}$, 故 $P[\tilde{z} > z] = e^{-(\alpha + \beta)z}$, 从而 $\tilde{z}$ 是参数为 $\alpha + \beta$ 的指数分布。

2. 因为 $P[\tilde{x} < \tilde{y}] = \int \int_{x < y} \alpha \beta e^{-\alpha x - \beta y} dx dy$, 所以

$$P[\tilde{x} < \tilde{y}] = \int_0^{\infty} \alpha e^{-\alpha x} \left(\int_x^{\infty} \beta e^{-\beta y} dy \right) dx = \int_0^{\infty} \alpha e^{-\alpha x} e^{-\beta x} dx = \frac{\alpha}{\alpha + \beta}$$

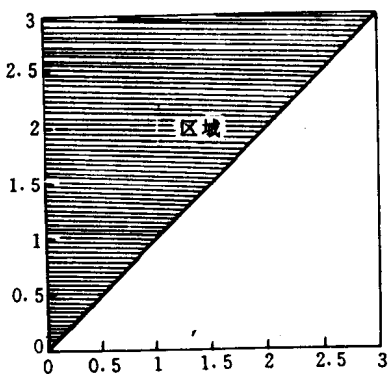


图 1.1 性质 2 中积分的积分区域

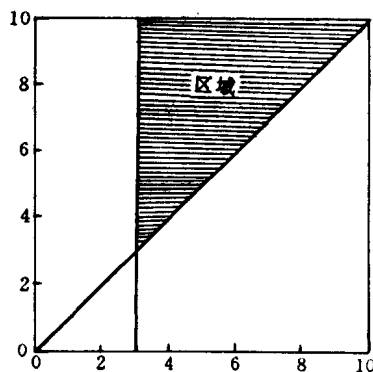


图 1.2 性质 3 中积分的积分区域

这里二重积分的积分区域如图1.1所示, 在二重积分化为二次积分时, 如果先对变量 y 积分, 则二次积分为 $\int_0^\infty (\int_x^\infty f(x, y) dy) dx$ 。

3. 这里的结论表明随机变量 $\tilde{z}$ 与随机事件 $\tilde{x} < \tilde{y}$ 相互独立, 所以只要证明 $P[\tilde{z} > z, \tilde{x} < \tilde{y}] = P[\tilde{z} > z]P[\tilde{x} < \tilde{y}]$, 就说明 $\tilde{z}$ 的分布与事件 $(\tilde{x} < \tilde{y})$ 独立, 从而性质3成立。为此计算

$$\begin{aligned} P[\tilde{z} > z, \tilde{x} < \tilde{y}] &= P[\tilde{y} > \tilde{x} > z] = \int_z^\infty \left(\int_x^\infty \beta e^{-\beta y} dy \right) \alpha e^{-\alpha x} dx \\ &= \int_z^\infty \alpha e^{-(\alpha+\beta)x} dx = \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)z} = P[\tilde{x} < \tilde{y}]P[\tilde{z} > z] \end{aligned}$$

这里二重积分的积分区域如图1.2所示, 在二重积分化为二次积分时, 如果先对变量 y 积分, 则二次积分为 $\int_z^\infty (\int_x^\infty f(x, y) dy) dx$ 。

4. 因为 $(\tilde{n}_x = n) = (\tilde{y}_1 > \tilde{x}_1, \tilde{y}_2 > \tilde{x}_2, \dots, \tilde{y}_n > \tilde{x}_n, \tilde{y}_{n+1} < \tilde{x}_{n+1})$, 由于独立试验, 故

$$P[\tilde{n}_x = n] = P[\tilde{x}_{n+1} > \tilde{y}_{n+1}] \cdot \prod_{i=1}^n P[\tilde{x}_i < \tilde{y}_i] = \left(\frac{\alpha}{\alpha+\beta} \right)^n \left(\frac{\beta}{\alpha+\beta} \right)$$

5. 因为 $\tilde{x}$ 分布的Laplace变换为 $\frac{\alpha}{s+\alpha}$, $\sum_{i=1}^n \tilde{x}_i$ 的Laplace变换为 $\left(\frac{\alpha}{s+\alpha} \right)^n$, 故 $\tilde{u}$ 分布的Laplace变换为

$$\sum_{n=1}^{\infty} p(1-p)^{n-1} \left(\frac{\alpha}{s+\alpha} \right)^n = \frac{\alpha p}{s+\alpha p}$$

从而 $\tilde{u}$ 服从参数为 αp 的指数分布。

证毕。

1.2.2 Poisson 过程

定义 若用 $\tilde{n}(t)$ 表示从0开始到时刻 t 为止已经发生的事件的数日, 则称随机过程 $\{\tilde{n}(t), t \geq 0\}$ 为计数过程。一般地说, 在计数过程的记号上还应标以计数的起始时刻, 如果过程的统计特性与起始时刻无关, 则称过程为平稳的。下面我们讨论的, 除非特别指明都是平稳过程。显然, 计数过程 $\tilde{n}(t)$ 取非负的整数值, 并且 $\tilde{n}(t)$ 是 t 的非降函数。

对于 $s < t$, $\tilde{n}(t) - \tilde{n}(s)$ 是在区间 $(s, t]$ 内发生的事件数, 量 $\tilde{n}(t) - \tilde{n}(s)$ 称为 $\tilde{n}(t)$ 在区间 $(s, t]$ 的增量。

Poisson 过程定义(1): 一计数过程 $\tilde{n}(t)$ 如果满足

1. $\tilde{n}(0) = 0$;
2. $\{\tilde{n}(t), t \geq 0\}$ 有独立增量;
3. 发生在任何长为 t 的区间中的事件数服从参数为 λt 的Poisson分布

$$P[\tilde{n}(t+s) - \tilde{n}(s) = n] = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad (\forall s) \quad (1.3)$$

则称计数过程 $\tilde{n}(t)$ 是率(参数)为 $\lambda(>0)$ 的Poisson过程。

Poisson 过程定义(2): 一计数过程 $\tilde{n}(t)$ 如果满足

1. $\tilde{n}(0) = 0$;
2. $\{\tilde{n}(t), t \geq 0\}$ 是平稳的, 且有独立增量;
3. $P[\tilde{n}(h) = 1] = \lambda h + o(h)$;

4. $P[\tilde{n}(h) \geq 2] = o(h)$ 。则称计数过程 $\tilde{n}(t)$ 是参数为 $\lambda (> 0)$ 的 Poisson 过程。这两种定义是等价的。因为从定义(1) 的(1.3) 式有

$$P[\tilde{n}(h+s) - \tilde{n}(s) = 1] = \lambda h e^{-\lambda h} = \lambda h(1 - \lambda h + o(h)) = \lambda h + o(h)$$

$$P[\tilde{n}(h+s) - \tilde{n}(s) > 1] = 1 - P[\tilde{n}(h+s) - \tilde{n}(s) = 0] - P[\tilde{n}(h+s) - \tilde{n}(s) = 1] = 1 - e^{-\lambda h} - \lambda h e^{-\lambda h} = o(h)$$

这就从定义(1) 推得了定义(2)。现在, 我们从定义(2) 出发去推导定义(1)。

设在区间 $(0, t]$ 发生的事件数为 n 。将区间 $(0, t]$ 等分为充分小的小区, 每个小区间的长为 h , 使每个小时间区间内至多发生一个事件, 记这种小区间的数目是 M (即 $h = \frac{t}{M}$)。对于每个小区间, 记在其中发生 k 个事件的概率为 $v_k(h)$, 由定义(2)

$$v_0(h) = 1 - \lambda h + o(h), \quad v_1(h) = \lambda h + o(h), \quad v_k(h) = o(h) \quad (k \geq 2)$$

现在 M 个小区间中有 n 个小区间内有事件发生, $M - n$ 个小区间内无事件发生, 所以

$$P[\tilde{n}(t) = n] = \binom{M}{n} [v_1(h)]^n [v_0(h)]^{M-n}$$

因为

$$(1 - \lambda h + o(h))^{M-n} = \frac{1}{(1 - \lambda h + o(h))^n} \left[(1 - \lambda h + o(h))^{\frac{1}{\lambda h}} \right]^{\lambda t}$$

而

$$\text{当 } h \rightarrow 0 \text{ 时, } (1 - \lambda h + o(h))^{\frac{1}{\lambda h}} \rightarrow e^{-1}, \quad (1 - \lambda h + o(h))^n \rightarrow 1$$

所以

$$[v_0(h)]^{M-n} = (1 - \lambda h + o(h))^{M-n} = e^{-\lambda t} [1 + o(h)]$$

又

$$[v_1(h)]^n = (\lambda h + o(h))^n = \frac{(\lambda t)^n}{M^n} [1 + o(h)]$$

最后有

$$P[\tilde{n}(t) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \frac{M(M-1)(M-2) \cdots (M-n+1)}{M^n} [1 + o(h)]$$

令 $h \rightarrow 0 (M \rightarrow \infty)$ 对上式取极限, 便得

$$P[\tilde{n}(t) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

证毕。

尚有其他定义方法, 限于篇幅我们不再介绍。

Poisson 过程的性质如下:

1. 令 $\{\tilde{n}_1(t), t \geq 0\}$ 和 $\{\tilde{n}_2(t), t \geq 0\}$ 分别是具有参数 α 和 β 的独立 Poisson 过程, 定义 $\tilde{n}(t) = \tilde{n}_1(t) + \tilde{n}_2(t)$, 则 $\{\tilde{n}(t), t \geq 0\}$ 是率为 $\alpha + \beta$ 的 Poisson 过程。

2. 对于参数为 λ 的 Poisson 过程, 假设发生的每个事件独立地以概率 p 做了记录, 未做记录的概率为 $1 - p$ 。令 $\tilde{n}_1(t)$ 是到 t 为止被做了记录的事件数, 而 $\tilde{n}_2(t)$ 是未被记录的事件数, 则 $\{\tilde{n}_1(t), t \geq 0\}$ 和 $\{\tilde{n}_2(t), t \geq 0\}$ 分别是参数为 $p\lambda$ 和 $(1-p)\lambda$ 的 Poisson 分布, 而且它们相互独立。

这两条性质说的是: (1) 独立Poisson 过程的和仍是Poisson 过程, 而且“和过程”的参数为加项过程的参数之和; (2) Poisson 过程的分支也是Poisson 过程, 而且各分支过程的参数是分支概率乘以原过程的参数。现在我们证明这两个性质。

用母函数方法。参数为 λ 的Poisson 过程的概率母函数为

$$G(z) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} z^n = e^{-\lambda t} e^{\lambda t z} = e^{-\lambda t(1-z)} \quad (1.4)$$

因为 $\tilde{n}_1(t)$ 分布的母函数为 $G_1(z) = e^{-\alpha t(1-z)}$, 同样 $\tilde{n}_2(t)$ 分布的母函数为 $G_2(z) = e^{-\beta t(1-z)}$, 又因 $\tilde{n}_1(t)$ 与 $\tilde{n}_2(t)$ 相互独立, 因此 $\tilde{n}(t)$ 分布的母函数为 $G(z) = G_1(z)G_2(z) = e^{-(\alpha+\beta)t(1-z)}$, 从而是参数为 $\alpha+\beta$ 的Poisson 过程(这里用到独立随机变量之和的分布母函数等于各加项的分布母函数之积)。这就证明了性质1, 再证性质2。因为

$$\begin{aligned} P[\tilde{n}_1(t) = m] &= \sum_{n=m}^{\infty} P[\tilde{n}_1(t) = m / \tilde{n}(t) = n] P[\tilde{n}(t) = n] \\ &= \sum_{n=m}^{\infty} \binom{n}{m} p^m (1-p)^{n-m} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda p t} \frac{(\lambda p t)^m}{m!} \end{aligned}$$

故而 $\{\tilde{n}_1(t), t \geq 0\}$ 是参数为 $p\lambda$ 的Poisson 分布。同理

$$\begin{aligned} P[\tilde{n}_2(t) = m] &= \sum_{n=m}^{\infty} P[\tilde{n}_2(t) = m / \tilde{n}(t) = n] P[\tilde{n}(t) = n] \\ &= \sum_{n=m}^{\infty} \binom{n}{m} (1-p)^m p^{n-m} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^m}{m!} \end{aligned}$$

故而 $\{\tilde{n}_2(t), t \geq 0\}$ 是参数为 $(1-p)\lambda$ 的Poisson 分布。因为 $\tilde{n}_1(t)$ 的分布母函数与 $\tilde{n}_2(t)$ 的分布母函数之积正好是 $\tilde{n}(t)$ 的分布母函数, 因此 $\tilde{n}_1(t)$ 与 $\tilde{n}_2(t)$ 独立。

证毕。

定理 设 $\tilde{n}(t)$ 是Poisson 过程, 在 $\tilde{n}(s+t) - \tilde{n}(s) = n$ 的条件下, 在时间区间 $(s, s+t]$ 内发生事件时刻的联合分布与在区间 $(s, s+t]$ 内 n 个按坐标增加次序排列的独立的均匀分布的随机点的联合分布相同。

证明 令 $s = t_0 < t_1 < t_2 < \cdots < t_k = s+t$ 是区间 $(s, s+t]$ 的一个分割。有 $n_1, n_2, \dots, n_k$ (它们之和为 n)个事件分别发生在不相交的时间区间 $(t_0, t_1], (t_1, t_2], \dots, (t_{k-1}, t_k]$ 内的概率是

$$\frac{\prod_{j=1}^k e^{-\lambda(t_j - t_{j-1})} \frac{[\lambda(t_j - t_{j-1})]^{n_j}}{n_j!}}{e^{-\lambda(t_k - t_0)} \frac{[\lambda(t_k - t_0)]^n}{n!}} = \frac{n!}{n_1! n_2! \cdots n_k!} \prod_{j=1}^k \left(\frac{t_j - t_{j-1}}{t_k - t_0} \right)^{n_j} \quad (1.5)$$

证毕。

附录: Laplace 变换和母函数

由于在排队分析中经常应用Laplace变换和母函数作为工具,故而在此占用少许篇幅对它们作一简要介绍:

定义 实变量的复函数 $f(t)$ 的Laplace变换(LT)为如下 s 的函数

$$F^*(s) = \int_0^{\infty} f(t)e^{-st} dt$$

函数 $f(t)$ 如满足以下条件,则它的Laplace变换 $F^*(s)$ 一定存在,而且 $F^*(s)$ 是半平面 $\operatorname{Re}(s) > 0$ 上的解析函数:

1. $f(t)$ 和 $f'(t)$ 在 $t \geq 0$ 上的任何有限区间上除了至多有第一类间断点外是连续的;
2. 当 $t < 0$ 时 $f(t) = 0$;
3. $f(t)$ 是有限阶的,即可以找到常数 a 和 A ,使 $|f(t)| \leq Ae^{at}$ (这里的数 a 称为 $f(t)$ 的增长指数,当 $f(t)$ 是有界函数时可取 $a = 0$).

Laplace变换常被用于连续型随机变量 X 的分布密度函数上,对分布函数 $F(t)$ 则使用以下Laplace-Stieltjes变换(LST)

$$F^*(s) = \int_0^{\infty} e^{-st} dF(t) (= E[e^{-sX}])$$

对于指数分布, $f(t) = \lambda e^{-\lambda t}, t \geq 0$, 则

$$F^*(s) = \int_0^{\infty} \lambda e^{-\lambda t} e^{-st} dt = \int_0^{\infty} \lambda e^{-(s+\lambda)t} dt = \frac{\lambda}{s+\lambda}$$

对于Gamma分布, $f(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x}, x \geq 0$, 则

$$F^*(s) = \int_0^{\infty} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} e^{-sx} dx = \left(\frac{\lambda}{s+\lambda}\right)^n$$

对于单点分布, $f(t) = \delta(t-D)$, 则

$$F^*(s) = \int_0^{\infty} \delta(t-D) e^{-st} dt = e^{-sD}$$

随机变量 X 的分布的Laplace变换记为 $F^*(s)$, 因为

$$\frac{d^m F^*(s)}{ds^m} = (-1)^m \int_0^{\infty} t^m f(t) e^{-st} dt$$

所以

$$\frac{d^m F^*(s)}{ds^m} \Big|_{s=0} = (-1)^m \int_0^{\infty} t^m f(t) dt$$

由此得

$$E[X^m] = (-1)^m \frac{d^m F^*(s)}{ds^m} \Big|_{s=0}$$

特别

$$E[X] = (-1) \frac{dF^*(s)}{ds} \Big|_{s=0}$$

$$D[X] = E[X^2] - (E[X])^2 = \frac{d^2 F^*(s)}{ds^2} \Big|_{s=0} - \left(\frac{dF^*(s)}{ds} \Big|_{s=0}\right)^2$$

定义 对于离散型随机变量 X , 若其概率分布为 $p_n, n = 0, 1, 2, 3, \dots$, 则称

$$G(z) = \sum_{n=0}^{\infty} p_n z^n$$

为概率分布 $\{p_n, n = 0, 1, 2, 3, \dots\}$ 的母函数。显然, $G(z) = E[z^X]$ 。

对于几何分布, $p_n = (1-p)p^n, n = 0, 1, 2, 3, \dots$, 有

$$G(z) = \sum_{n=0}^{\infty} (1-p)p^n z^n = \frac{1-p}{1-pz}$$

对于二项分布, $p_n = \binom{N}{n} p^n (1-p)^{N-n}, n = 0, 1, 2, 3, \dots, N$, 则有

$$G(z) = \sum_{n=0}^N \binom{N}{n} p^n z^n (1-p)^{N-n} = [pz + (1-p)]^N$$

对于Poisson 分布, $p_n = \frac{\lambda^n}{n!} e^{-\lambda}, n = 0, 1, 2, 3, \dots$, 则有

$$G(z) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} z^n = e^{-\lambda(1-z)}$$

随机变量 X 分布的母函数记为 $G(z)$, 则由母函数的定义得

$$E[X] = G'(1)$$

$$D[X] = G''(1) + G'(1) - (G'(1))^2$$

1.3 Markov 链

具有离散状态空间的Markov 过程称为Markov 链, 它分为时间连续与时间离散两种情况。

1.3.1 离散时间的Markov 链(DTMC)

一个随机序列 $\{X_n\}$, 如果下一个状态(值) 是 x_{n+1} 的概率仅依赖于现在状态的取值而与以前的状态取何值无关, 那么这个随机序列 $\{X_n\}$ 就称为离散时间的Markov 链。它的数学表达是

$$P[X_{n+1} = x_{n+1} / X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] = P[X_{n+1} = x_{n+1} / X_n = x_n] \quad (1.6)$$

离散时间的Markov 链有一个特点, 那就是每隔一个单位时间都要发生一次状态转移, 转移概率是任意的。当然在转移中也可回到同一状态, 此时叫做逗留在该状态上。离散时间的Markov 链逗留在某一状态上的时间服从几何分布。同样, 连续时间的Markov 链逗留在某一状态的时间服从指数分布。这两个结论我们将在下面证明。在离散时间的Markov 过程中, 如果去掉状态的逗留时间是几何分布的这个限制, 这种随机过程叫离散时间的半Markov 过程。连续时间的半Markov 过程是连续时间Markov 过程放弃逗留时间为指数分