INFORMETRICS 87/88
L. Egghe and R. Rousseau (Editors)

PREDICTIVE ASPECTS OF SOME BIBLIOMETRIC PROCESSES

Quentin L. BURRELL

Statistical Laboratory, Department of Mathematics, University of Manchester, Manchester M13 9PL, United Kingdom

Abstract

A statistical model for any bibliometric situation, whether it be a frequency-of-citation distribution over a population of authors, a frequency-of-circulation distribution for a library collection or a Bradford type analysis of journal relevance in a field, is of most value when it can be used to guide future action. The model can be used as a predictive tool only if a time parameter features in the model explicitly, in which case we are obliged to consider not bibliometric distributions but bibliometric processes.

We consider various models for such processes and suggest ways in which they can be used for predictive purposes. In many cases, since it will not be obvious, on the basis of available data, which is the *most appropriate model*, we need to compare the predictions arising from the various models. Consideration of the moment structure as a means of distinguishing between the models is also discussed.

1. INTRODUCTION

To quote Sichel [1], bibliometrics "rotates around books and journals, authors and readers, book circulation and journal usage, references and citations. In a wider sense it includes some linguistic topics such as word frequency, sentence length and author identification ...". In seeking to model mathematically the mechanism producing the data, a useful portmanteau description of various of these contexts is provided by imagining a population of "sources" producing "items" observed in time. For instance in the library loan situation described by Burrell [2] and Cane [3], the library collection constitutes the population with individual *monographs being the sources and* the observed "items" being the recorded borrowings of the particular monograph.

Mathematically the observation of the items produced by a particular source is equivalent to the observation of a *stochastic point process*, *i.e.* a *sequence* of "events" occurring randomly in time, while the modelling of such a process requires specification of the probabilistic mechanism producing the observed "events". The simplest such point process is the Poisson process which corresponds to the situation in which (in intuitive terms) the events *occur completely at* random in time with the overall average rate of occurrence remaining constant so that the expected number of events occurring increases linearly with time (see Fig. 1 (a)). This is the underlying process assumed in Sichel [1], Burrell [2, 4] and Burrell and Cane [3].

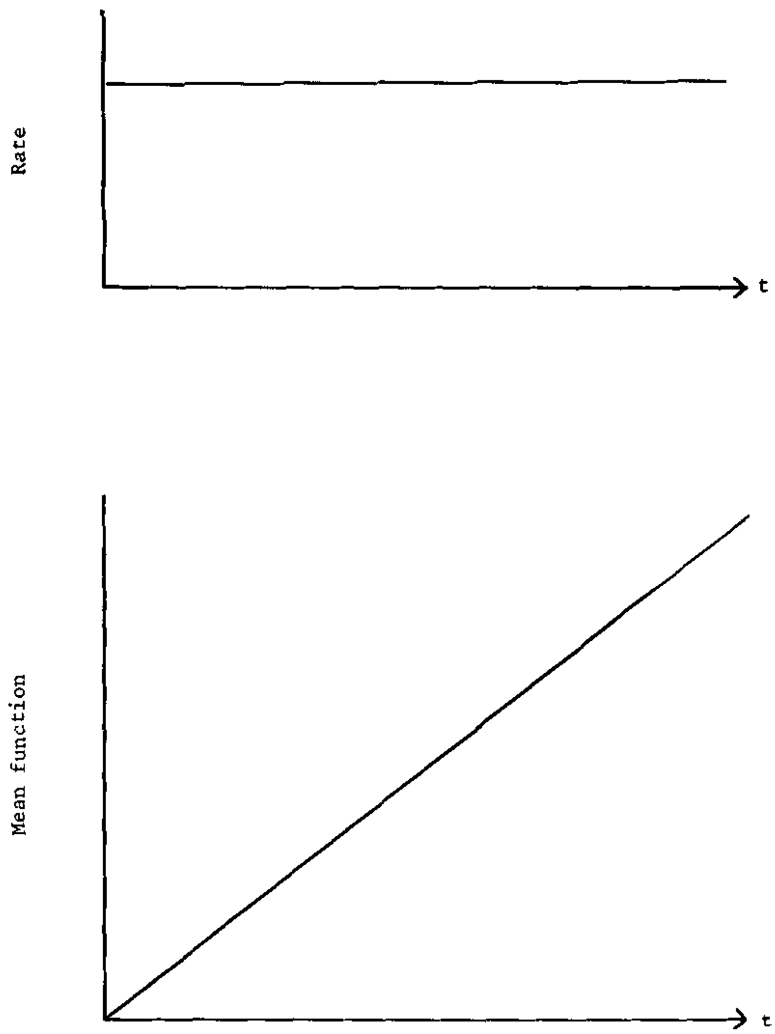


Fig. 1(a) Constant rate : Poisson process

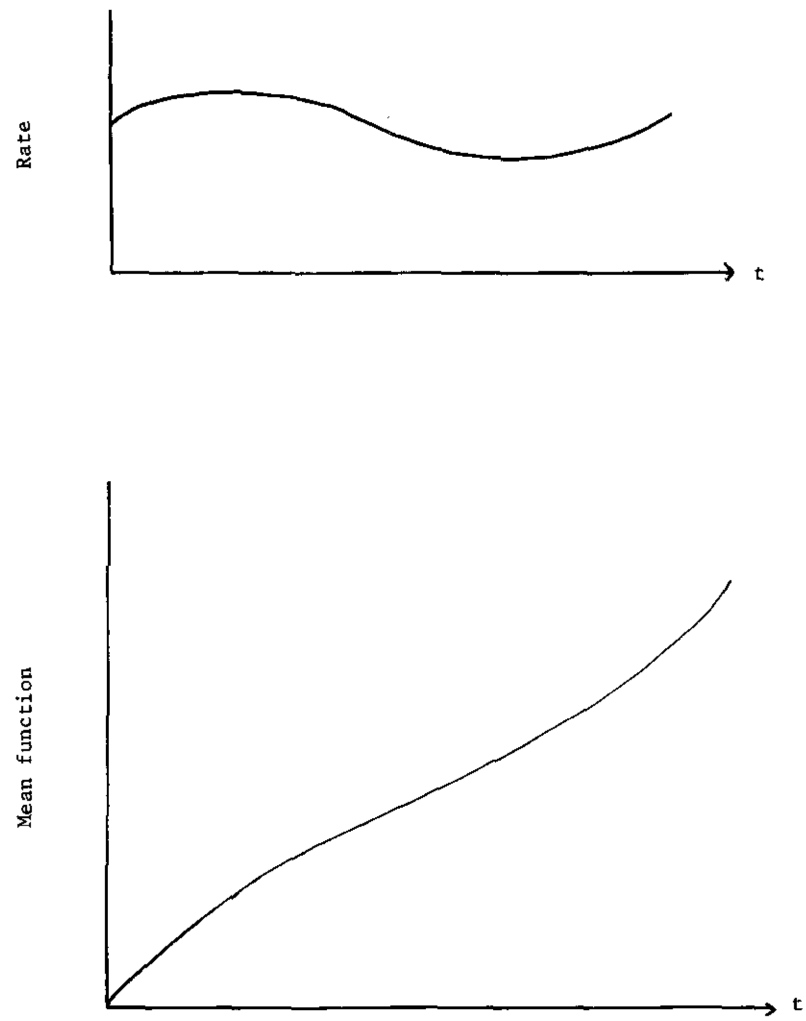


Fig. 1(b) Deterministically varying rate : Non-homogeneous Poisson process.
(i) Cyclically varying rate.

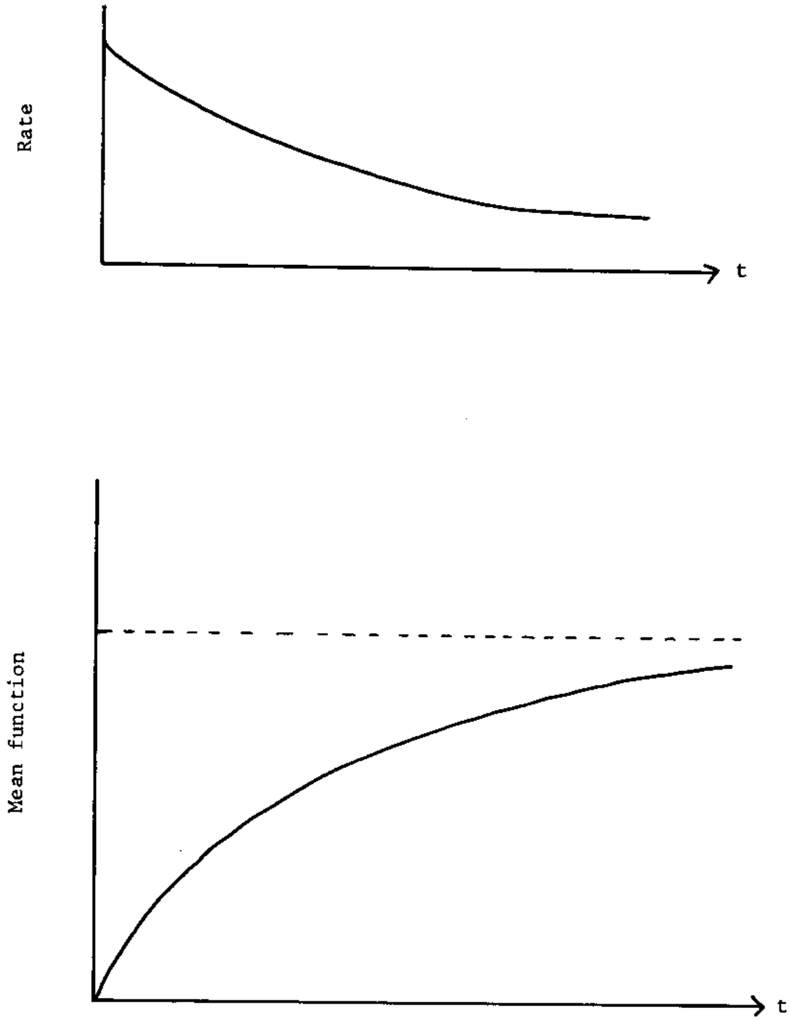
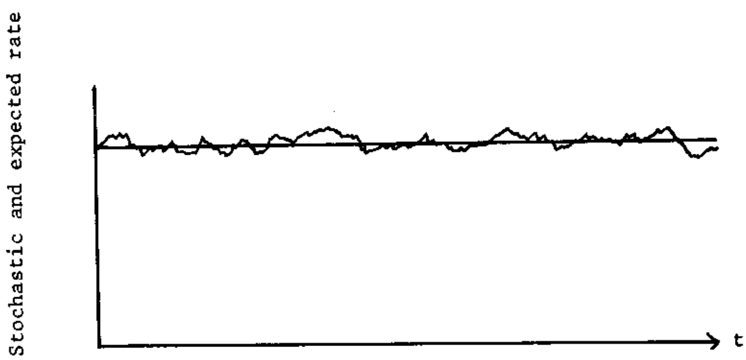


Fig. 1(b) Deterministically varying rate : Non-homogeneous Poisson process.
(ii) Exponentially decaying rate.



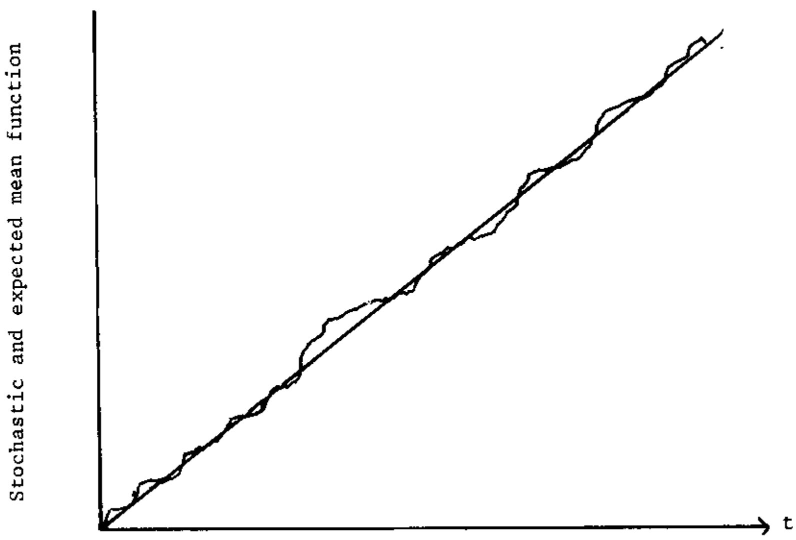


Fig. 1(c) Stochastically varying rate : doubly-stochastic Poisson process.

The next natural extension is to assume that the rate of the process varies with time in some predetermined fashion. For instance in the library-loan context one might wish to incorporate seasonal variation in loan frequency or, over a longer time scale, a gradual decline in average usage. This latter variant has been investigated in some detail by Burrell [5, 6, 7]. In this situation, where the rate of the underlying process is changing in a deterministic fashion, so that the expected number of events occurring increases with time in a non-linear fashion (see Fig. 1(b)), the process is termed a non-homogeneous Poisson process.

The final extension we mention is that in which the rate itself varies in some random fashion, the resultant process being termed in general a Cox process, after Cox [8], or a doubly-stochastic Poisson process. A fairly comprehensive mathematical treatment of the structure of such processes is given in Grandell [9]. In this case the expected number of events occurring increases with time in a stochastic fashion (see Fig. 1(c)). An application of this type of process to bibliometrics may be found in Burrell [10].

Each of these three types has independent increments but only the first and third have stationary increments, i.e. model situations in which the probability distribution of the number of events in a period of time depends only upon the length of the period and not upon the time at which it begins. We restrict our attention to such stationary or steady-state processes in what follows.

When we turn to consideration of the entire population of sources, it is natural to assume that there is some variability in the underlying rate of production between different items. The overall process observed is then a mixture of the individual processes and is conveniently modelled mathematically by "mixing" the parameter(s) determining the rates of production of the individual sources. It should be noted that the resulting mixed process still has stationary increments but that the mixing destroys the independence of increments. This observation will be crucial in later stages.

2. CONSTRUCTION AND EXAMPLES

In what follows X_t denotes the number of items produced by a source in $[0,t]$, $t \geqslant 0$. We assume that for a particular source, the point process is specified by a parameter θ, say, and that for given θ the process $\{X_t | \theta \; ; \; t \geqslant 0\}$ is of known form. It is further assumed that θ varies over the population according to some probability distribution given by a density function f_θ. Then for a source chosen at random from the population we have a point process $\{X_t \; ; \; t \geqslant 0\}$ having stationary, but not independent, increments and

$$\begin{aligned} P(X_t = r) &= E_\theta P(X_t = r | \theta) \\ &= \int P(X_t = r | \theta = x) \, f_\theta(x) dx \qquad (1). \end{aligned}$$

We therefore have to specify (i) the form of the conditional processes $\{X_t | \theta \; ; \; t \geqslant 0\}$ and (ii) the probability density function f_θ.

2.1. The Gamma-Poisson Process

In this model, first discussed by Greenwood and Yule [11] in 1920 in the context of accident data, we assume that each source produces items as a Poisson process of rate λ, where λ has a gamma distribution with index ν and scale parameter β.

Thus $P(X_t = r | \lambda) = e^{-\lambda t} \frac{(\lambda t)^r}{r!}$, $r = 0,1,\ldots,$ (2)

and $$f_\lambda(x) = \frac{\beta^{-\nu} x^{\nu-1}}{\Gamma(\nu)} \exp(-x\beta^{-1}), \text{ for } x > 0 \qquad (3).$$

Substitution in (1) yields

$$P(X_t = r) = \frac{t^r \beta^{-\nu}}{r!\Gamma(\nu)} \int_0^\infty x^{r+\nu-1} \exp[-x(t + \beta^{-1})]\,dx$$

$$= \binom{r+\nu-1}{r} \left(\frac{1}{1+\beta t}\right)^\nu \left(\frac{\beta t}{1+\beta t}\right)^r, \; r = 0,1,2,\ldots \qquad (4).$$

Thus X_t has a negative binomial (NB) distribution of index ν and parameter $p_t = (1 + \beta t)^{-1}$ and the resulting process $\{X_t \,;\, t \geqslant 0\}$ we call the gamma-Poisson (GP) process of index ν and parameter β. (Note that some authors term this a negative binomial process. We prefer our terminology not only because it hints at the construction of the process but also because in 2.3 we introduce another process having a more legitimate claim to the title). We shall write $X_t \sim GP(\nu,\beta t)$ for the distribution given by (4).

2.2. The Generalized Inverse Gaussian-Poisson Process

This is again a mixture of Poisson processes, much discussed and applied by Sichel [1, 12, 13, 14]. The mixing distribution is the generalized inverse Gaussian distribution, for details of which the reader is referred to Jorgensen [15], given by

$$f_\lambda(x) = c(\alpha,\gamma,\theta)\, x^{\gamma-1} \exp\{-x(\theta^{-1}-1) - \frac{\alpha^2\theta}{4x}\}, \text{ for } x > 0 \qquad (5)$$

where $-\infty < \gamma < \infty$, $0 \leqslant \theta \leqslant 1$, $\alpha \geqslant 0$ and the normalizing constant is given by

$$c(\alpha,\gamma,\theta) = \frac{(1-\theta)^{\gamma/2}}{(\frac{1}{2}\alpha\theta)^\gamma\, 2K_\gamma\{\alpha(1-\theta)^{\frac{1}{2}}\}}$$

where $K_\gamma(\cdot)$ denotes the modified Bessel function of the second kind of order γ (see Abramowitz and Stegun [16, p. 374 ff]). (Note that Jorgensen [15] talks of $K_\gamma(\cdot)$ as a modified Bessel function of the <u>third</u> kind). This is a very flexible three parameter family of distributions, which includes the gamma distribution by taking $\gamma > 0$ and letting $\alpha \to 0$.

Substitution of the density (5) in (1) leads to

$$P(X_t = r) = \frac{(1-\theta_t)^{\gamma/2}}{K_\gamma\{\alpha(1-\theta)^{\frac{1}{2}}\}} \cdot \frac{(\frac{1}{2}\alpha_t\theta_t)^r}{r!} K_{r+\gamma}(\alpha_t), \quad r = 0,1,2,\ldots \qquad (6)$$

where $\alpha_t = \alpha[1 + (t-1)\theta]^{\frac{1}{2}}$ and $\theta_t = \dfrac{t\theta}{1+(t-1)\theta}$. If $t = 1$, so that $\alpha_t = \alpha$ and $\theta_t = \theta$, this distribution is called the generalized inverse Gaussian-Poisson (GIGP) distribution with parameters α and θ and index γ. We shall refer to the process $\{X_t \,;\, t \geqslant 0\}$ as a GIGP process and write $X_t \sim GIGP(\alpha_t, \theta_t, \gamma)$ for the distribution (6). In many practical studies Sichel [1, 12, 13, 14] has found the version of (6) with $\gamma = -\frac{1}{2}$ to be particularly successful and it is this form that we shall consider. In this case, i.e. $X_t \sim GIGP(\alpha_t, \theta_t, -\frac{1}{2})$, it can be shown that (6) simplifies to

$$P(X_t = r) = \left(\frac{2\alpha_t}{\pi}\right)^{\frac{1}{2}} \exp\{\alpha(1-\theta)^{\frac{1}{2}}\} \frac{(\frac{1}{2}\alpha_t\theta_t)^r}{r!} K_{r-\frac{1}{2}}(\alpha_t), \; r = 0,1,2,\ldots \quad (6')$$

with α_t, θ_t as before.

2.3. The Waring Process

Perhaps the simplest example of a doubly stochastic Poisson process is that where the mean function of the process corresponds to a gamma process, in which case the resulting point process may properly be termed a negative binomial process of parameter p and index Ψ, say, and then

$$P(X_t = r|p) = \binom{r + \Psi t - 1}{r} p^{\Psi t} (1 - p)^r, \quad r = 0,1,2,\ldots \quad (7).$$

If we now suppose that the parameter p has a beta distribution with parameters a and b, so that

$$f_p(x) = \frac{1}{B(a,b)} x^{a-1}(1 - x)^{b-1}, \text{ for } 0 \leqslant x \leqslant 1 \quad (8)$$

then substitution in (1) leads to

$$P(X_t = r) = \frac{\Gamma(\Psi t + a)}{B(a,b)\Gamma(\psi t)} \cdot \frac{\Gamma(r+\Psi t)\Gamma(r+b)}{\Gamma(r+\psi t+a+b)r!}, \quad r = 0,1,2,\ldots \quad (9).$$

This family of distributions was first introduced and applied by Irwin [17, 18] and termed by him the Generalized Waring (GW) distribution. The stochastic process constructed above we shall call the GW process, for further details of which we refer to Burrell [19]. (The particular case a = 1 has an inverse-square form for the tail of the distribution and has been called the Lotka-Bradford process by Burrell [10]). We shall write $X_t \sim GW(a,b,\Psi t)$ for (9).

2.4. Remarks

(i) When considering a Poisson process, the parameter λ corresponds to the mean number of events per unit time, i.e. $\lambda = E[X_1|\lambda]$. Thus in 2.1 and 2.2 the mixing distribution relates to the way in which this mean is distributed over the population of sources. On the other hand in 2.3, where the conditional distribution of X_t is NB, we have

$$E[X_1|p] = \psi \frac{(1-p)}{p}.$$

Hence if we want to compare the mixing distributions of the rates of production we should in 2.3 consider the distribution of $\lambda = \psi(1-p)p^{-1}$. It is straightforward to derive the corresponding probability density function in this case as

$$f_\lambda(x) = \frac{1}{B(a,b)} \frac{\psi^a x^{b-1}}{(x+\psi)^{a+b}}, \quad x \geqslant 0 \quad (10).$$

(ii) The GIG density function (5) with $\gamma = -\frac{1}{2}$ can be strongly reverse J-shaped. The other mixing densities for the rates may also be reverse J-shaped, provided in (3) we have $\nu < 1$ and in (10) we have $b < 1$. Hence although they have different analytic forms they may have similar shapes, at least in general terms. In what follows we shall implicitly be restricting attention to situations in which such comparable forms are appropriate.

(iii) In both theoretical and empirical work in bibliometrics there is often a difficulty with the zero category, i.e. with the non-producing items in a period of study. In the library circulation context Burrell [2] and Burrell and Cane [3] built into their model the assumption that there was in most collections a set of books which for one genuine reason or another (lost, stolen, on permanent reserve, etc.) could not contribute to the circulation statistics. In the discussion following Burrell and Cane [3], Chatfield doubted the necessity for such an assumption, at least for reverse J-shaped circulation distribution, while Bagust [20], in an unfortunately flawed paper, was contemptuously scornful of the whole idea.

In other bibliometric situations a source can only announce its membership of the population by producing an item, e.g. in citation studies, non-contributors cannot be observed. There may well, however, be many potential contributors who would be observed if the period of observation was to be extended. This population of potential contributors is not, unfortunately, well-defined.

Certainly in empirical studies, therefore, one may be obliged to restrict attention to known productive sources which must therefore be modelled by zero-truncated forms of mixed point processes. However, in the present study we shall be mainly concerned with theoretical aspects and so in what follows we assume that our population of sources is well-defined, that every source is a potential producer and that the suggested mixing distribution is a mixture of the entire population. This means in particular that we can identify the non-productive sources.

3. THE DEVELOPMENT IN TIME

Let us suppose that we are able to observe our population of sources over an extending period of time. At some convenient time t, say, we can then find the observed production frequency distribution, i.e. for $r = 0,1,2,\ldots$ we find $f_r(t)$ = No. of sources producing r items in $(0,t]$.

Let us suppose that this distribution, as is very often the case in bibliometric studies, is reverse J-shaped so that members of each of the three families discussed in §2 might reasonably describe the general form of the observed distribution, for suitable choices of parameter values. Because of the presence of a time parameter in each case, once the basic defining parameters have been determined or estimated, we can find the (theoretical) production distribution for any other value of t. One can then ask what are the similarities and differences between these predicted distributions. If the predictions are very similar then the choice of model is not crucial and we may employ the simplest. If there are great differences then the actual behaviour of the observed production distribution will suggest how to distinguish between appropriate and inappropriate models.

3.1. Moments

For the GP and GIGP distributions moments of all orders exist while for the GW distribution certain restrictions have to be placed on allowable parameter values in order that moments exist. We shall assume in what follows that whatever may be the true population point process, at least its first two moments (i.e. mean and variance) are finite.

(a) If $X_t \sim GP(\nu, \beta t)$ then from the well known form of the mean and variance of the NB distribution we have, for the mean,

$$E[X_t] = \frac{\nu(1-p_t)}{p_t} = \nu\beta t,$$

and for the variance,

$$V(X_t) = \frac{\nu(1-p_t)}{p_t^2} = \nu\beta t(1+\beta t).$$

We shall also make use of the index of dispersion

$$I(X_t) = \frac{V(X_t)}{E[X_t]} = 1 + \beta t.$$

(The index of dispersion is a useful quantity in the study of point processes, see Cox and Isham [21], in comparing a given process with the Poisson process, for which the index is constant.)

(b) If $X_t \sim GIGP(\alpha_t, \theta_t, -\frac{1}{2})$ then, following Sichel [13] we have

$$E[X_t] = \frac{\alpha_t\theta_t}{2(1-\theta_t)^{\frac{1}{2}}} = \frac{\alpha\theta t}{2(1-\theta)^{\frac{1}{2}}},$$

$$V(X_t) = \frac{\alpha_t\theta_t(2-\theta_t)}{4(1-\theta_t)^{3/2}} = \frac{\alpha\theta t}{4(1-\theta)^{3/2}}[2(1-\theta)+t\theta],$$

and hence

$$I(X_t) = \frac{2-\theta_t}{2(1-\theta_t)} = 1 + \frac{\theta t}{2(1-\theta)}.$$

(c) If $X_t \sim GW(a, b, \Psi t)$ the moments are given in Irwin [18, Part I] and Burrell [19] so that we can write

$$E[X_t] = \frac{\Psi bt}{a-1}, \text{ provided } a > 1,$$

$$V(X_t) = \frac{\Psi b(a+b-1)}{(a-1)^2(a-2)}(a-1+\Psi t), \text{ provided } a > 2,$$

and then

$$I(X_t) = \frac{(a+b-1)(a-1+\Psi t)}{(a-1)(a-2)} = \frac{a+b-1}{a-2} + \frac{(a+b-1)\Psi t}{(a-1)(a-2)}, \text{ provided } a > 2$$

Note that in all three cases the mean increases linearly with time, as is necessarily the case when we have a process *with stationary increments*, while the variance increases quadratically. In each case the index of dispersion increases as a linear function of t, but note that the intercept (i.e. when t = 0) equals 1 in (a) and (b), but is $1 + \frac{b+1}{a-2} > 1$ in (c).

3.2. An example

In order to illustrate the applicability of all three models to the description of data collected over a single time period, and bearing in mind our previous remarks (2.4, (ii) and (iii)), we given in Table 1 a frequency-of-circulation distribution compiled at the University of Sussex Library. These empirical data have been presented previously by Burrell [5] to which paper the reader is referred for further details. Note that in the situation where the data have been collected over a single period of time, we can use the observational period as determining the unit of time, i.e. we can take t = 1. (In our example the time period is in fact the (academic) year 1976-77).

Table 1 - Observed and fitted frequency-of-circulation distributions

Number of circulations, r	Data from University of Sussex, 1976-77	(i) GP	Fitted values (ii) GIGP	(iii) GW
0	160 978	160 991.4	159 141.0	162 098.7
1	41 782	43 545.0	49 031.3	43 226.3
2	19 264	18 691.2	17 114.4	17 934.3
3	10 137	9 012.2	7 450.3	8 540.2
4	5 304	4 583.8	3 767.4	4 395.7
5	2 606	2 404.2	2 092.4	2 388.6
6	1 132	1 286.4	1 236.0	1 353.5
7	456	698.1	762.2	793.8
8	199	382.8	485.1	479.3
9	97	211.6	316.3	296.8
10	43	117.7	210.2	188.0
11	12	65.8	141.9	121.5
12	19	36.9	97.0	79.9
13	9	20.8	67.0	53.5
14	6	11.8	46.7	36.3
15	2	6.7	32.8	25.0
>15	29	8.6	83.0	63.6

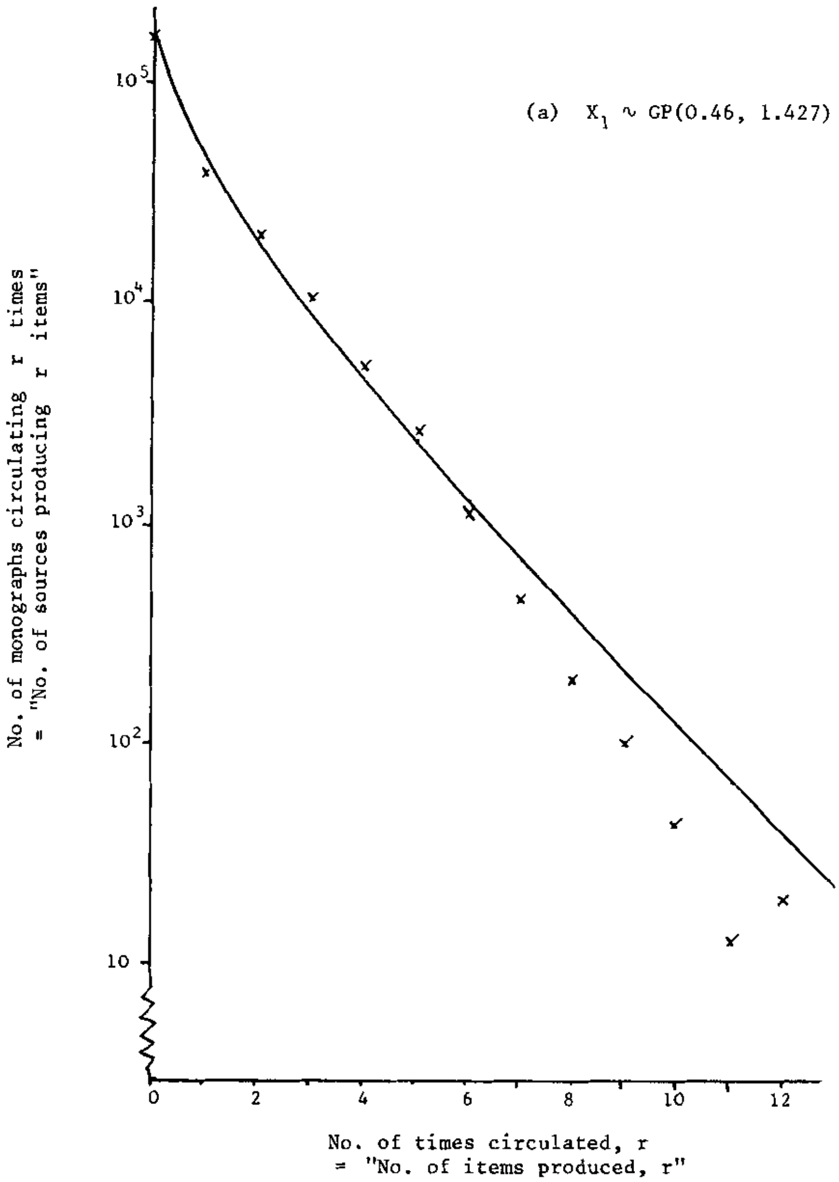
Parameter values : $\nu = 0.460$ $\alpha = 0.79$ $a = 10$
$\beta = 1.427$ $\theta = 0.78$ $b = 0.5$
$\Psi = 12$

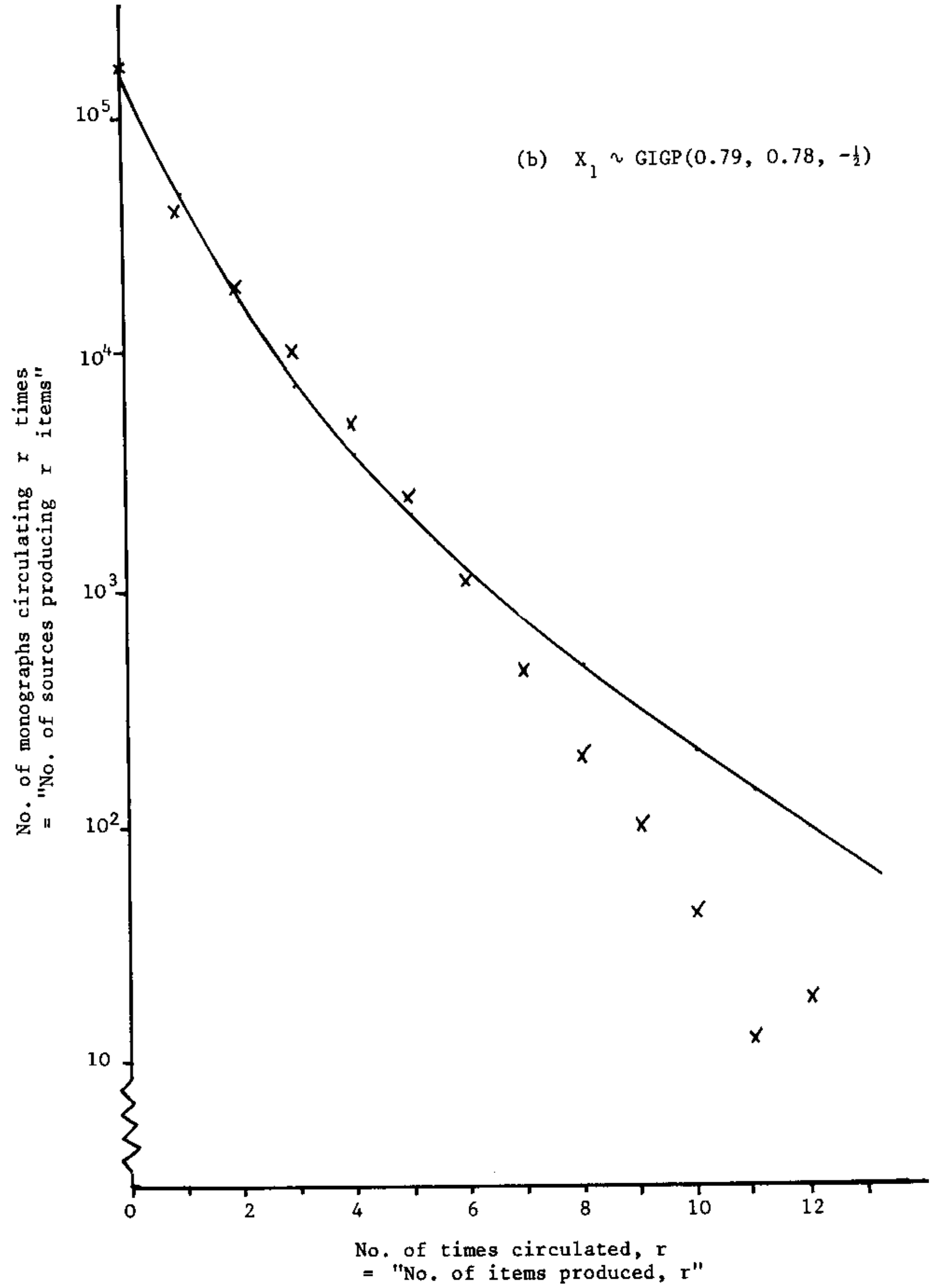
It is not our intention to discuss the relative merits of different statistical methods of estimation of parameters so we have chosen parameter values more for computational convenience than for statistical respectability. The resulting distributions fitted to the data of Table 1 are illustrated in Figure 2(a, b, c), the aim being to stress the similarities of the three fitted distributions rather than their relative goodness (or badness) of fit.

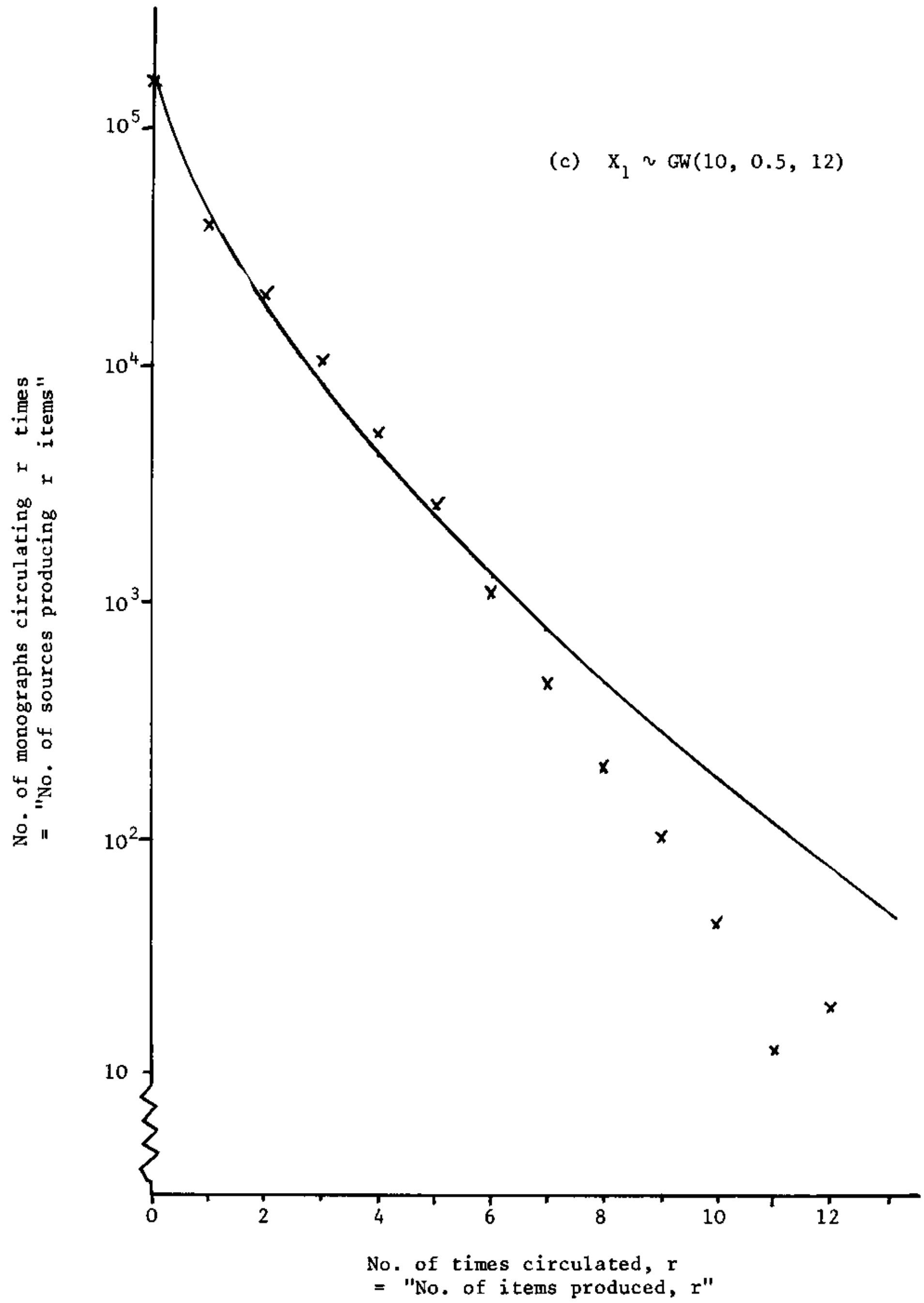
3.3. The Zero-Category

In each of (4), (6') and (9), $P(X_t = r)$ gives a probability distribution on 0,1,2,... for each fixed t. If alternatively we fix r and consider $P(X_t = r)$ as t increases

Fig. 2 Frequency-of-circulation distribution of monographs in the University of Sussex long loan library, academic year 1976-77, together with fitted theoretical distributions.




(b) X1 ∿ GIGP(0.79, 0.78, -½)
10^5
10^4
10^3
10^2
10
No. of monographs circulating r times
= "No. of sources producing r items"
0
2
4
6
8
10
12
No. of times circulated, r
= "No. of items produced, r"


(c) $X_1 \sim GW(10, 0.5, 12)$
No. of monographs circulating r times
= "No. of sources producing r items"
10^5
10^4
10^3
10^2
10
0
2
4
6
8
10
12
No. of times circulated, r
= "No. of items produced, r"

we find how the size of each of the productivity categories varies with time. The simplest to consider, and in many circumstances the one of most interest is the zero category, i.e. $P(X_t = 0)$ which gives the (theoretical) proportion of the population which has produced no items by time t.

If we write $p_0(t) = P(X_t = 0)$ then from (4), (6') and (9) we find

(a) if $X_t \sim GP(\nu, \beta t)$ then $p_0(t) = (1 + \beta t)^{-\nu}$,

(b) if $X_t \sim GIGP(\alpha_t, \theta_t, -\frac{1}{2})$ then

$$p_0(t) = \left(\frac{2\alpha_t}{\pi}\right)^{\frac{1}{2}} \exp\{\alpha(1-\theta)^{\frac{1}{2}}\} K_{-\frac{1}{2}}(\alpha_t)$$

$$= \exp\{\alpha(1-\theta)^{\frac{1}{2}} - \alpha(1-\theta + t\theta)^{\frac{1}{2}}\},$$

making use of $K_{-\frac{1}{2}}(z) = K_{\frac{1}{2}}(z) = \left(\frac{\pi}{2z}\right)^{\frac{1}{2}} e^{-z}$ (see e.g. Abramowitz and Stegun [16, p. 444]),

(c) if $X_t \sim GW(a, b, \psi t)$ then

$$p_0(t) = \frac{\Gamma(a + \Psi t)\Gamma(a + b)}{\Gamma(a)\Gamma(a + b + \Psi t)} .$$

These are, of course, necessarily decreasing functions of t and having explicit analytic forms for them allows them to be plotted as functions of t.

Again for purposes of illustration we make use of the various parameter estimates given in Figure 2 in order to plot the graphs of the theoretically predicted decline in the proportion of non-producers as in Figure 3.

4. CORRELATION STRUCTURE

4.1. Prediction of Future Behaviour

As we have already seen, several different models may provide similar descriptions of a given set of data collected over a single time period while differences between the models are revealed by their predictions of general features (moments and/or frequencies) over an extending period of time. In this respect, what we have been considering is, in a sense, long-term forecasting. In practice, it is of great importance to detect serious differences between the models at an early stage. Indeed, in a world of rapidly changing budgets and demands, it is perhaps the accuracy of short-term forecasting which is paramount. Hence we need to consider what the models predict will happen, given the current data, in the immediate future.

Let us suppose, then, that we have observed our population of sources during $[0,t_1]$ and using the data collected i.e. the observed productivity distribution, we wish to predict the subsequent distributions during $[t_1, t_1 + t_2]$. If we write, for a randomly chosen source, Y_1 and Y_2 for the numbers of items produced during $[0,t_1]$ and $[t_1,t_1+t_2]$ respectively then really what we want is the conditional distribution of Y_2, given the value of Y_1. (As remarked earlier, mixing destroys the independence of increments so that Y_1 and Y_2 are not independent).

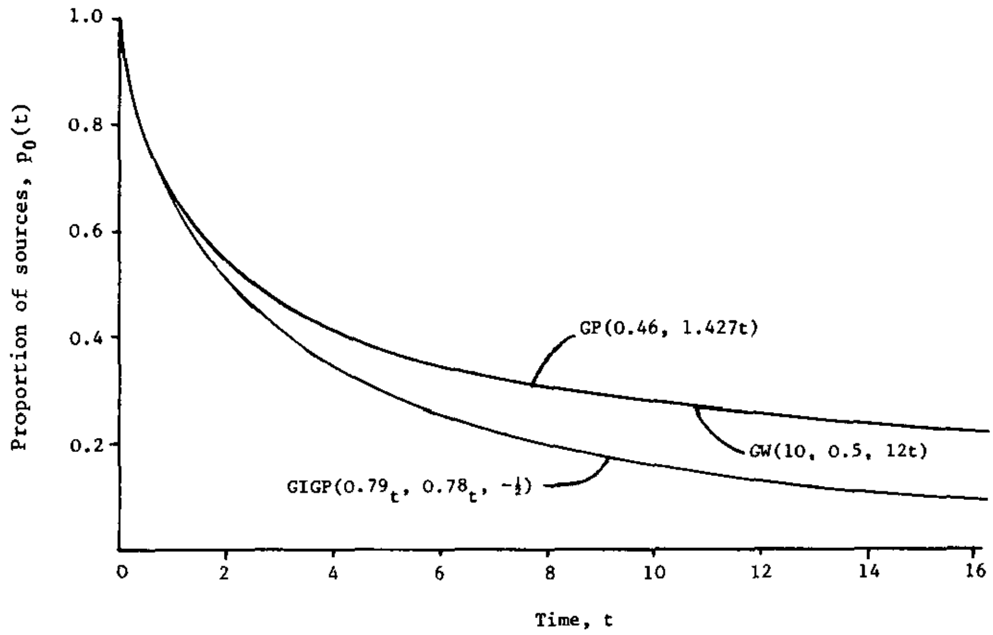


Fig. 3 Proportion of sources producing no items by time t
(Note that the graphs of $p_0(t)$ for GP(0.46, 1.427t) and GW(10, 0.5, 12t) are practically indistinguishable, differing by less than 0.013 on $0 < t < 20$).

Now $$P(Y_2 = r \mid Y_1 = k) = \frac{P(Y_1 = k,\ Y_2 = r)}{P(Y_1 = k)}$$

$$= \frac{E_\theta\,[P(Y_1 = k,\ Y_2 = r \mid \theta\,)]}{P(Y_1 = k)}$$

$$= \frac{E_\theta\,[P(Y_1 = k \mid \theta\,)P(Y_2 = r \mid \theta\,)]}{P(Y_1 = k)}$$

since, given θ, we are looking at one of the individual point processes, for which we do have independence of increments.

If we have a mixture of Poisson processes, so that θ is the Poisson parameter λ then

$$E_\theta\,[P(Y_1 = k \mid \theta)\ P(Y_2 = r \mid \theta\,)]$$

$$= \int \frac{e^{-xt_1}(xt_1)^k}{k!} \cdot \frac{e^{-xt_2}(xt_2)^r}{r!}\ f_\lambda(x)\ dx$$

$$= \frac{t_1^k\,t_2^r}{k!r!} \int e^{-x\,(t_1+t_2)}\, x^{k+r}\ f_\lambda(x)\ dx$$

$$= \binom{k+r}{k} \frac{t_1^k\,t_2^k}{(t_1+t_2)^{k+r}} \int \frac{e^{-x\,(t_1+t_2)}}{(k+r)!}\ [x(t_1+t_2)]^{k+r} f_\lambda(x)dx$$

$$= \binom{k+r}{k}\left(\frac{t_1}{t_1+t_2}\right)^k \left(\frac{t_2}{t_1+t_2}\right)^r\ P(X_{t_1+t_2} = k + r)$$

$$= \binom{k+r}{k}\left(\frac{t_1}{t_1+t_2}\right)^k \left(\frac{t_2}{t_1+t_2}\right)^r\ P(Y_1 + Y_2 = k + r).$$

Hence we have

$$P(Y_2 = r \mid Y_1 = k) = \binom{k+r}{k}\left(\frac{t_1}{t_1 + t_2}\right)^k \left(\frac{t_2}{t_1 + t_2}\right)^r \frac{P(Y_1 + Y_2 = k + r)}{P(Y_1 = k)}$$

or, since $Y_1 = X_{t_1}$ and $Y_1 + Y_2 = X_{t_1+t_2}$,

$$P(Y_2 = r \mid Y_1 = k) = \binom{k+r}{k}\left(\frac{t_1}{t_1+t_2}\right)^k\left(\frac{t_2}{t_1+t_2}\right)^r \frac{P(X_{t_1+t_2} = k+r)}{P(X_{t_1} = k)} \tag{11}$$

In the case of a mixture of negative binomial processes an exactly analogous calculation yields

$$P(Y_2 = r \mid Y_1 = k) = \frac{\binom{k+\Psi t_1-1}{k}\binom{r+\Psi t_2-1}{r}}{\binom{k+r+\Psi(t_1+t_2)-1}{k+r}} \cdot \frac{P(X_{t_1+t_2} = k+r)}{P(X_{t_1} = k)} \tag{12}$$

For our three processes we know the distribution of X_t for each t so that we have expressions for the probabilities on the right hand side of (11), in the case of the GP and GIGP processes, and (12) in the case of the GW process. In each case the expressions can be simplified, after a certain amount of algebraic manipulation, into recognisable form which we summarise as follows.

(i) If $X_t \sim GP\ (\nu, \beta t)$, as in 2.1, then

$$(Y_2 \mid Y_1 = k) \sim GP\ \left(\nu + k, \left(\frac{\beta}{1+\beta t_1}\right) t_2\right). \tag{13}$$

(ii) If $X_t \sim GIGP\ (\alpha_t, \theta_t, \frac{-1}{2})$ as in 2.2, then

$$(Y_2 \mid Y_1 = k) \sim GIGP\ (\alpha^*_{t_2}, \theta^*_{t_2}, k - \tfrac{1}{2}), \tag{14}$$

where $\alpha^\bullet = \alpha(1+\theta t_1)$ and $\theta^\bullet = (1+\theta_{t_1})^{-1}$.

(iii) If $X_t \sim GW\ (a, b, \Psi t)$, as in 2.3, then

$$(Y_2 \mid Y_1 = k) \sim GW\ (a + \Psi t_1\ , b + k, \Psi t_2) \tag{15}$$

Hence we have in each case the pleasing result that the predicted (conditional) distribution is of the same family as the original unconditional distribution. In practice this means that, on the basis of the data collected during $[0,t_1]$ we can estimate the basic parameters of whichever model is adopted and use these to calculate numerical estimates of probabilities of various levels of future usage.

As a particular application, let us mention the problem of relegation of library material as presented by, for instance, Fussler and Simon [22], Trueswell [23] and Burrell [2,7], where zero usage over a period of time is suggested as a reasonable indicator of a book's candidacy for relegation. Hence what we need to consider is the conditional distribution of Y_2, given that $Y_1 = 0$. Insofar as possible recall from some relatively inaccessible store is concerned the probability of interest is

$$P(Y_2 \neq 0 \mid Y_1 = 0) = 1 - P(Y_2 = 0 \mid Y_1 = 0)$$

which is readily calculable.

In a similar vein, although computationally not so attractive, there is no reason why we should not construct prediction intervals i.e. determine values $r_1(k)$ and $r_2(k)$ such that, given $Y_1 = k$, Y_2 will lie between $r_1(k)$ and $r_2(k)$ with (approximately) some prescribed probability, e.g. 0.9, 0.95 etc.

4.2. Regression

It was Philip Morse [24] who first noted (and made use of) the empirical observation that if one plots the mean number of borrowings in a given year of books which had been borrowed k times during the previous year then the resulting graph is (approximately) a linear function of k. This phenomenon has been subsequently reported in experimental studies by e.g., Chen [25], Burrell and Cane [3] and Beheshti & Tague [26]. When we turn to our theoretical models, using the conditional distributions (13) and (15) and the moments given in 3.1 (a) and (c) we can straightaway write down:

(i) If $X_t \sim GP\ (\nu, \beta t)$ then

$$E[Y_2 | Y_1 = k] = (\nu + k)\frac{\beta t_2}{(1 + \beta t_1)} \tag{16}$$

and (iii) If $X_t \sim GW\ (a, b, \Psi t)$ then

$$E[Y_2 | Y_1 = k] = \frac{\Psi t_2 (b + k)}{a + \Psi t_1 - 1} \tag{17}$$

For (ii), the case where $X_t \sim GIGP\ (\alpha_t, \theta_t, -\frac{1}{2})$, we require for (14) the mean of a GIGP distribution for which $\gamma \neq -\frac{1}{2}$. This may be found in Sichel [13] and we merely state here that

$$E[Y_2 | Y_1 = k] = \frac{\alpha_{t_2}{}^{*}\theta_{t_2}}{2(1-\theta_{t_2})^{\frac{1}{2}}} \quad \frac{K_{k+\frac{1}{2}}(\alpha_{t_2}{}^{*}(1-\theta_{t_2}{}^{*})^{\frac{1}{2}})}{K_{k-\frac{1}{2}}(\alpha_{t_2}{}^{*}(1-\theta_{t_2}{}^{*})^{\frac{1}{2}})}$$

Hence (16) and (17) do result in linear functions of k for k = 0, 1, 2,... and so are candidates for modelling Morse's phenomenon, while (18) is not in general. (It can be shown that in the special case of a GIGP with = 0 we do get that $E[Y_2 \; Y_1 = k]$ is linear in k for k = 1, 2,...). Indeed it can be shown that the only mixture of Poisson processes having this linear regression property is the GP process while the only mixture of negative binomial processes having the property is the GW process. For the first case see Johnson [27] or Burrell [28] and for both see a somewhat more general approach in Diaconis and Ylvisaker [29].

As a final point, since linear regression of the conditional means. does not distinguish between the GP and GW processes, note that the conditional variance, i.e. $V(Y_2 \; Y_1 = k)$, is a linear function of k for the GP process and a quadratic in k for the GW process.

REFERENCES

[1] Sichel, H.S., A bibliometric distribution which really works, J. Amer Soc. for Info. Sci. 36 (1985) pp. 314-321.

[2] Burrell, Q.L., A simple stochastic model for library loans, J. Documentation, 36 (1980) pp. 115-132.

[3] Burrell, Q.L. & Cane, V.R., The analysis of library data (with discussion), J. Royal Statistical Soc. (A), 145 (1982) pp. 431-471.

[4] Burrell, Q.L., Alternative models for library circulation data, J. Documentation, 38 (1982) pp. 1-13.

[5] Burrell, Q.L., A note on ageing in a library circulation model, J. Documentation, 41 (1985) pp. 100-115.

[6] Burrell, Q.L., A second note on ageing in a library circulation model: the correlation structure, J. Documentation, 42 (1986) pp. 114-128.

[7] Burrell, Q.L., A third note on ageing in a library circulation model: applications to future use and relegation, J. Documentation, 43 (1987) pp. 24-45.

[8] Cox, D.R., Some statistical methods connected with series of events (with discussion), J. Royal Statistical Soc. (B), 17 (1955) pp. 129-164.

[9] Grandell, J., Doubly stochastic Poisson processes. Lecture Notes in Mathematics, Vol. 529, (Springer-Verlag, Berlin, 1976).

[10] Burrell, Q.L., Modelling the Bradford phenomenon. Research Report 108/QLB, (Statistical Laboratory, University of Manchester, 1987).

[11] Greenwood, M. & Yule, G.U., An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, J. Royal Statistical Soc., 83 (1920) pp. 255-279.

[12] Sichel, H.S., On a distribution law for word frequencies, J. Amer. Statistical Assoc., 70 (1975) pp. 542-547.

[13] Sichel, H.S., On a distribution representing sentence-length in written prose, J. royal Statistical soc. (A), (1975) pp. 25-34.

[14] Sichel, H.S., Repeat-buying and the generalized inverse Gaussian-Poisson distribution, Applied Statistics, 37 (1982) pp. 193-204.

[15] Jorgensen, B., Statistical Properties of the Generalized Inverse Gaussian Distribution, Lecture Notes in Statistics, Vol. 9 (Springer-Verlag, Berlin, 1982).

[16] Abramowitz, M.& Stegun, I.A. (eds.), Handbook of Mathematical Functions, (Dover, New York, 1965).

[17] Irwin, J.O., The generalized Waring distribution applied to accident data, J. Royal Statistical soc. (A), 131 (1968) pp. 205-225.

[18] Irwin, J.O., The generalized Waring Distribution., J. Royal Statistical Soc. (A), 138 (1975) pp. 18-31 (Part I), pp. 204-227 (Part II) pp. 374-384 (partIII).

[19] Burrell, Q.L., Waring processes and some applications, (in preparation).

[20] Bagust, A., A circulation model for busy public libraries, J. Documentation, 39, (1983), pp. 24-37.

[21] Cox, D.R., and Isham, V., Point Processes. (Chapman •- Hall, London, 1980).

[22] Fussler, H.H. & Simon, J.L., Patterns in the Use of books in Large research Libraries. (University of Chicago Press, Chicago 1969).

[23] Trueswell, R.W., Growing libraries: Who needs them ? A statistical basis for the no-growth collection., In : D. Gore (ed.), *Farewell to Alexandria: solutions to space, growth and performance problems of libraries*, (Greenwood Press, London, 1976), pp. 72-104.

[24] Morse, P.M., Library Effectiveness: A systems Approach. (MIT Press, Cambridge, Mass., 1968).

[25] Chen, C-C., Applications of Operations Research Models to Libraries. (MIT Press, Cambridge, Mass., 1976).

[26] Beheshti, J.& Tague, J.M., Morse's model of book use revisited., Journal of the American society for Information Science, 35, (1984), pp. 259-267.

[27] Johnson, N.L., Uniqueness of a result in the theory of accident proneness., Biometrika, 44, (1957), pp. 530-531.

[28] Burrell, Q.L., A footnote to a result of Johnson., Research Report 99, Statistical Laboratory, University of Manchester, (1985).

[29] Diaconis, P & Ylvisaker, D., Conjugate priors for exponential families., Annals of Statistics, 7, (1979), pp. 269-281.