Recovering Stellar Flare-Rate Distributions: Comparative Study of Methods, Application to Flare Stars in Galactic Field

A. A. Akopian*

NAS RA V.Ambartsumian Byurakan Astrophysical Observatory, Byurakan 0213, Aragatsotn Province, Armenia

Abstract

The revised and substantially extended pipeline/code for recovering the flare-rate distribution $\varphi(\nu)$ of a population of randomly flaring objects is presented. The capabilities of existing methods have been significantly expanded, and new ones have been added. Specifically, the new code allows *(i)* the use of the χ^2 test to significantly improve the estimation of the number of unknown (undiscovered) flare stars n_0, *(ii)* a refinement of the moment fit via a maximum likelihood estimate of a truncated Poisson distribution and a comparison of Pearson's types based on AIC/BIC metrics, *(iii)* an analysis of the temporal evolution of $\varphi(\nu)$ to account for accumulated observations, *(iv)* the application of modern numerical calculation methods, and *(v)* the use of numerical simulation to work with both synthetic and real observational data. Finally, two methods for recovering the flare-rate distribution $\varphi(\nu)$ of a stellar population are considered. The first method is the inverse-Laplace formulation from Ambartsumian (1978), applied to the discovery curve of first-flare events. The second is the method of moments (MoM, Akopian, 2003), improved and extended here to the full Pearson family of probability distributions. A comparative analysis of these methods is also presented.

In the second part of the study, the code was used to determine the distribution of flare rate among flare stars within the Kepler telescope's field of view. Typically, both methods yield identical results for synthetic data. However, when applied to a sample of flare stars from the Kepler field of view, their results diverge. The MoM yields a Pearson Type III distribution, or gamma distribution, with a negative bias in the minimum flare frequency, which is physically impossible. In contrast, Ambartsumian's approach yields a Type III distribution, which is physically plausible, with $\nu_0 > 0$. To identify the cause of the discrepancies, regardless of the specific form of the Pearson distribution, a numerical inverse Laplace transform was applied using modern algorithms. The inverse function $m_1(t)/m_1(0)$, where $m_1(t)$ is the frequency of the first flare, was in this case represented and approximated by its purely observational realization $n_1(t)/n(t)$, where $n_1(t)$ is the number of stars that have survived exactly one flare by time t, and $n(t)$ is the cumulative number of flares. This approach does not require knowledge of N_{tot} and provides the most reliable of the available cross-checks for reconstruction using the inverse Laplace transform. The nature of these discrepancies reveals the structure of the underlying distribution: a smooth density curve with a peak shifted towards values of $\nu \sim 2 \times 10^{-3}\ d^{-1}$, with a long tail in the region of high ν values.

Keywords: *stellar flares, Poisson mixtures, Pearson distributions, Laplace transform, method of moments, rate/frequency distribution function*

*aakopian57@gmail.com

doi: https://doi.org/10.52526/25792776-26.73.1-39

1. Introduction

In the second half of the last century, there was considerable interest in flare stars,and Ambartsumian's final series of scientific papers was devoted to the statistics of these stars. As is often the case, the new methods developed by Ambartsumian have found—and continue to find—applications far beyond the field of astrophysics. In particular, Ambartsumian's estimator of the unknown number of undiscovered flare stars (Ambartsumian, 1969) is now widely used in various fields of statistics, medicine and the social sciences, biology and ecology, linguistics, criminology, epidemiology, cryptography and other disciplines. In these fields, various names are used for this or similar methods: "capture-recapture, mark-recapture, capture-mark-recapture, observation-reobservation and mark-release-recapture, etc...".

Due to the specific characteristics of flaring stars and the relatively limited capabilities for their detection and study, interest in them gradually waned. However, over the past few decades, the Kepler and TESS space observatories have provided new, far more informative data on these objects. Significant advances in statistical and computational methods, combined with the new data, have revived interest in statistical studies of flaring stars. In this article, the statistical research methods developed at the Byurakan Observatory are examined in the light of these developments.

The first part of this paper presents a revised and significantly enlarged pipeline/code for recovering the flare-rate/frequency distribution $\varphi(\nu)$ of a population of randomly flashing objects from the observed distribution of flare counts n_k (the number of stars with k registered flares). That is critical for reconstructing the flare rate distribution function, thus we must establish the total flare-star count in the sample by calculating the number of undiscovered flare stars n_0. The new code allows the use of the χ^2 test and the maximum likelihood method to significantly improve the estimate of the number of undiscovered flare stars n_0 (Ambartsumian, 1969) and the statistical properties of this estimate.

Two methods for recovering the flare-rate distribution $\varphi(\nu)$ of a stellar population are considered. The first method is the inverse-Laplace formulation from Ambartsumian (1978), applied to the discovery curve of first-flare events. The second method (Akopian, 2003) is the method of moments (MoM), improved and extended here to the full Pearson family of probability distributions. A comparative analysis of previously proposed methods is also presented.

In the second part of the study, the code was used to determine the flare rate distribution of flare stars in the Kepler telescope's field of view (hereinafter - Kepler stars). Both methods produce the same results for synthetic data. However, for the Kepler stars sample, they diverge significantly.

To address this problem independently of any specific form of the Pearson distribution and resolve discrepancies that have arisen between the two methods, a new approach has been proposed, in which the use of the numerical inverse Laplace transform plays a key role.

We have endeavoured to ensure that the structure of this article, in terms of content, corresponds as closely as possible to the structure of the code. This explains the somewhat unusual style of presentation. The paper is structured as follows. Section 2 outlines the key assumptions and approaches used in this study. Section 3 sets up the moment-transformation formulas and reviews the Pearson family of densities. Section 4 describes the MoM pipeline including the estimation of the silent population and ML refinement. Section 5 introduces Ambartsumian's inverse-Laplace formulation. Section 6 applies both methods to the Kepler stars (Yang et al., 2019) and quantifies their disagreement.

Section 7 discuss results and identified discrepancies. Section 8 concludes.

doi: https://doi.org/10.52526/25792776-26.73.1-39

2. Basic assumptions and approaches

A photometric survey of N stars observed for a time T produces, for each star j, an integer count K_j of detected flares. The counts $\{K_j\}$ are the raw data of stellar-activity statistics. Behind each star is a hidden intrinsic flare rate ν_j, and the population of rates $\{\nu_j\}$ is itself a sample from an unknown population density $\varphi(\nu)$. Recovering $\varphi(\nu)$ from the observed flare counts $\{K_j\}$ is the central inference problem of this paper.

The Poisson-mixture model is the standard framework. Given the rate ν_j, the count K_j is Poisson-distributed with mean $\nu_j T$:

$$\Pr(K_j = k \mid \nu_j) \;=\; \frac{(\nu_j T)^k}{k!}\, e^{-\nu_j T}. \tag{1}$$

The rate ν_j itself is drawn from $\varphi(\nu)$, so the marginal distribution of K over the population is

$$\Pr(K = k) \;=\; \int_0^\infty \frac{(\nu T)^k}{k!}\, e^{-\nu T}\, \varphi(\nu)\, \mathrm{d}\nu. \tag{2}$$

Eq. (2) defines the burst-count distribution. Recovering φ from observations of K is a deconvolution problem: we observe a mixture and must infer the mixing density.

Two approaches

This problem can be addressed using two different approaches.

Inverse Laplace transform. Ambartsumian (1978) proposed a statistical method for determining $\varphi(\nu)$ based on inverse Laplace transform. Ambartsumian's main idea was to determine the desired function using the chronology of discoveries of new flare stars. From a mathematical point of view, the key observation is that the rate of first-flare events, $m_1(t)$, equals $\int_0^\infty \nu\, e^{-\nu t}\, \varphi(\nu)\, \mathrm{d}\nu = \mathcal{L}[\nu\varphi(\nu)](t)$.

Method of moments (MoM). This method was proposed in (Akopian, 2003). The basic idea is to compute the sample central moments μ_{k_r} of K from the observed histogram, transform them via Eq. (6) below into the central moments μ_{ν_r} of φ, and then fit a parametric density (from the Pearson family) to those moments. This approach uses all of the data, weighting it heavily toward the high-K tail through the third and fourth moments.

The two approaches agree on data that are well-described by their assumed parametric family. However, they can disagree on observational data that depart from these parametric assumptions. The character of this disagreement serves as a useful model-diagnostic tool.

3. Theoretical framework

From flare-count moments to rate moments

The Poisson-mixture model (2) relates the r-th sample central moment of the observed counts, μ_{k_r}, to the corresponding moment of the rate distribution $\varphi(\nu)$, μ_{ν_r}, as follows:

$$\mu_{\nu_1} = \mu_{k_1}/T, \tag{3}$$

$$\mu_{\nu_2} = (\mu_{k_2} - \mu_{k_1})/T^2, \tag{4}$$

$$\mu_{\nu_3} = (\mu_{k_3} - 3\mu_{k_2} + 2\mu_{k_1})/T^3, \tag{5}$$

$$\mu_{\nu_4} = (\mu_{k_4} - 6\mu_{k_3} + 11\mu_{k_2} - 6\mu_{k_1} - 6\mu_{k_1}\mu_{k_2} + 3\mu_{k_1}^2)/T^4. \tag{6}$$

doi: https://doi.org/10.52526/25792776-26.73.1-39

Pearson family of densities

The Pearson family of probability density functions is a comprehensive system of continuous probability distributions developed by Karl Pearson (Pearson, 1895) consisting of all probability density functions $\varphi(y)$ that satisfy the differential equation

$$\frac{\mathrm{d}\varphi(y)}{\mathrm{d}y} = \frac{(y-a)\,\varphi(y)}{b_0+b_1y+b_2y^2}. \tag{7}$$

The roots of the quadratic denominator $b_0+b_1x+b_2x^2=0$ govern the mathematical properties and shape of the density. The discriminant $\Delta = b_1^2 - 4b_0b_2$ classifies these distributions. The classical types are labelled I through VII. For our purposes, six are relevant: Types I, III, IV, V, VI, and the Normal as a boundary case. The coefficients of the ODE map onto the moments via

$$a = b_1 = -\frac{\mu_{\nu_3}(\mu_{\nu_4}+3\mu_{\nu_2}^2)}{A}, \quad b_0 = -\frac{\mu_{\nu_2}(4\mu_{\nu_2}\mu_{\nu_4}-3\mu_{\nu_3}^2)}{A}, \quad b_2 = -\frac{2\mu_{\nu_2}\mu_{\nu_4}-3\mu_{\nu_3}^2-6\mu_{\nu_2}^3}{A}, \tag{8}$$

with $A = 10\mu_{\nu_2}\mu_{\nu_4} - 18\mu_{\nu_2}^3 - 12\mu_{\nu_3}^2$.

The dimensionless ratios $\beta_1 = \mu_{\nu_3}^2/\mu_{\nu_2}^3$ (squared skewness) and $\beta_2 = \mu_{\nu_4}/\mu_{\nu_2}^2$ (kurtosis) determine the type. The diagnostic quantity

$$\kappa = \frac{\beta_1\,(\beta_2+3)^2}{4\,(4\beta_2-3\beta_1)\,(2\beta_2-3\beta_1-6)} \tag{9}$$

classifies the regions: $\kappa<0 \Rightarrow$ Type I (Beta on a finite interval); $0<\kappa<1 \Rightarrow$ Type IV (real-line density with complex roots of the denominator); $\kappa>1 \Rightarrow$ Type VI (Beta-prime on $[L,\infty)$, heavy upper tail). The line $2\beta_2-3\beta_1-6=0$ is the Type III boundary (shifted gamma); $\kappa=1$ is the Type V boundary (shifted inverse gamma); and the point $(\beta_1,\beta_2)=(0,3)$ is the Normal.

Figure 1 shows representative densities for each type. Figure 2 maps the regions in the (β_1,β_2) plane.

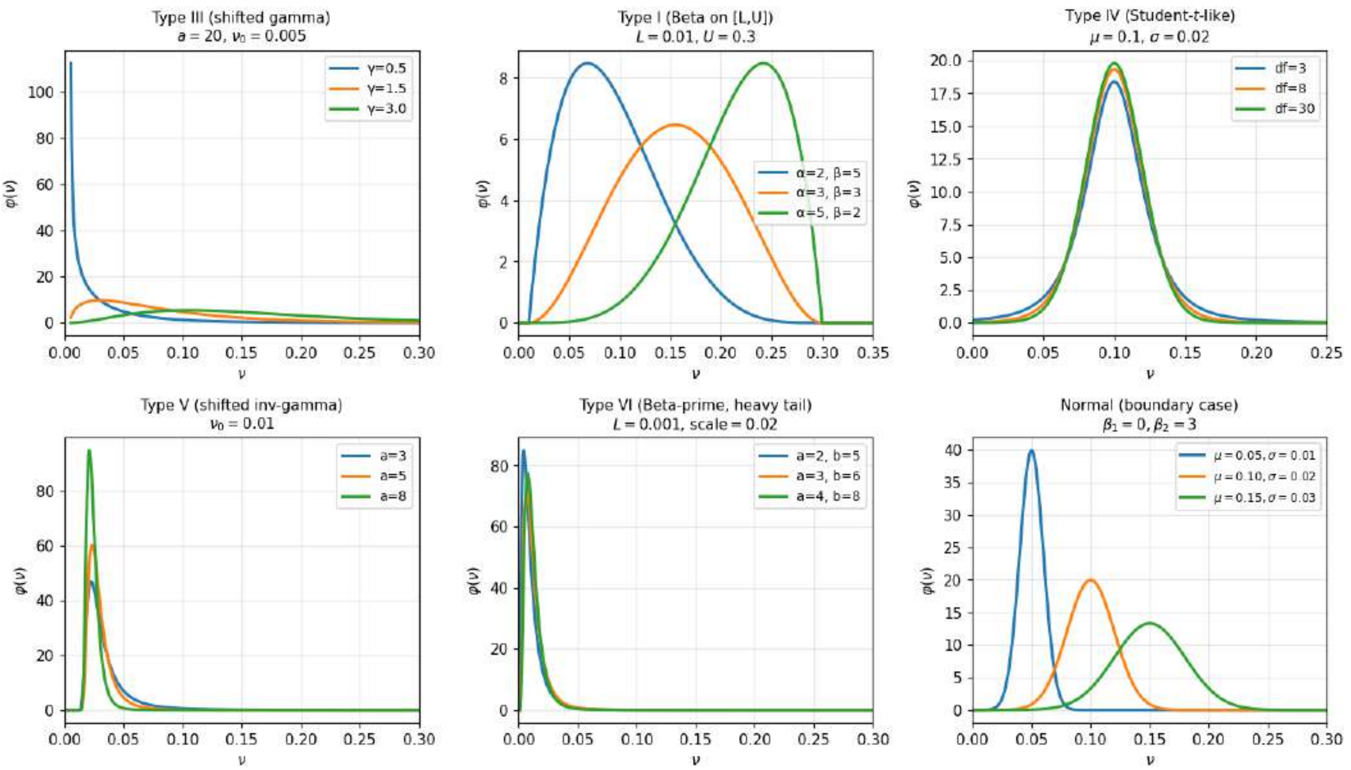

Figure 1: Representative densities of the six Pearson types plotted for parameters relevant to stellar flare rates (ν in day^{-1}). Top row: Type III (shifted gamma, monotone or unimodal depending on γ); Type I (bounded support $[L,U]$); Type IV (real-line symmetric or skew, Student-t when symmetric). Bottom row: Type V (shifted inverse gamma); Type VI (beta-prime, heavy upper tail); Normal as the boundary case.

doi: https://doi.org/10.52526/25792776-26.73.1-39

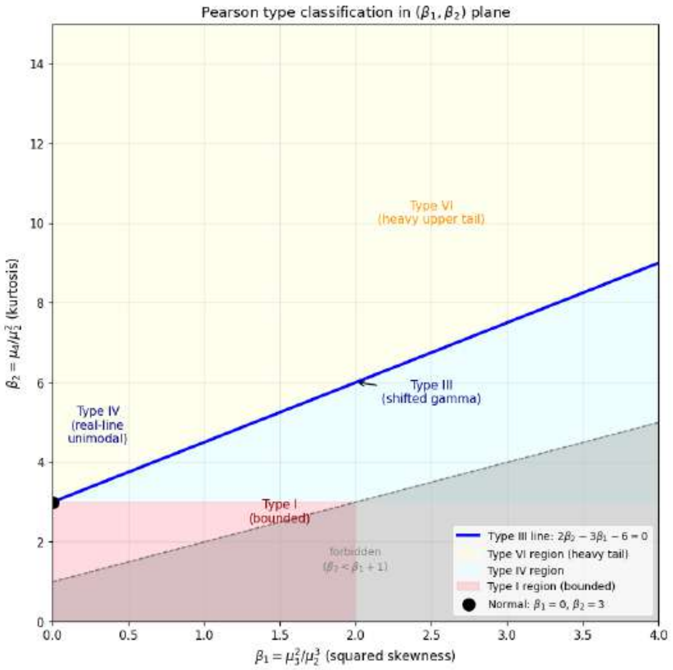

Figure 2: Pearson type classification in the (β_1, β_2) plane. The Type III line is $2\beta_2 - 3\beta_1 - 6 = 0$. Above it lies the Type VI region (heavy upper tail); below it the Type IV region; in the bounded-density part the Type I region. The Normal sits at $(0, 3)$. The forbidden region $\beta_2 < \beta_1 + 1$ is excluded by moment inequalities.

The special role of the Pearson Type III distribution

Type III (strictly speaking, a shifted gamma distribution) is

$$\varphi_{\mathrm{III}}(\nu) \;=\; \frac{a^\gamma}{\Gamma(\gamma)}\,(\nu - \nu_0)^{\gamma-1}\,e^{-a(\nu-\nu_0)}, \qquad \nu \geq \nu_0. \tag{10}$$

Its parameters are recovered directly from the first three moments:

$$a = \frac{2\mu_{\nu_2}}{\mu_{\nu_3}}, \qquad \gamma = a^2\,\mu_{\nu_2}, \qquad \nu_0 = \mu_{\nu_1} - \frac{\gamma}{a}. \tag{11}$$

The Laplace transform of Type III is elementary:

$$\mathcal{L}[\varphi_{\mathrm{III}}](t) \;=\; e^{-\nu_0 t}\left(\frac{a}{a+t}\right)^{\gamma}. \tag{12}$$

This makes Type III the only Pearson type for which both MoM recovery and inverse-Laplace recovery have closed analytical forms. Type V's Laplace transform involves the Bessel function K_a; Types I and VI involve confluent hypergeometric functions; The type IV distribution has no closed-form expression. For this reason, all our applications of Ambartsumian's formulation (Section 5) use Type III as the analytical smoothing function.

A reliable benchmark for checking the quality of any fitted density is the population mean rate,

$$\langle\nu\rangle \;=\; \int_0^\infty \nu\,\varphi(\nu)\,\mathrm{d}\nu \;=\; \mu_{\nu_1}. \tag{13}$$

For the Type III distribution, $\langle\nu\rangle = \gamma/a + \nu_0$. This quantity is robust against moderate misspecification of the parametric family — much more robust than the shape parameters themselves.

doi: https://doi.org/10.52526/25792776-26.73.1-39

4. MoM pipeline

Implementation and synthetic validation

The full implementation is available in the Jupyter notebook accompanying this paper. Given the numbers $\{n_k\}$ of stars with k flares (histogram), the pipeline performs the following steps:

1) Compute the sample moments μ_{k_r} from the histogram using standard unbiased estimators.
2) Transform them to the rate moments μ_{ν_r} via Eqs. (3)–(6).
3) Compute the Pearson coefficients via Eq. (8) and classify the distribution type via Eq. (9).
4) Apply the type-specific parameter recovery (e.g., Eq. (11) for Type III) and return $\varphi(\nu)$.

We have implemented closed-form parameter recovery for all six types.

Figure 3 shows the recovery accuracy for Type III on synthetic samples of increasing size N. The ground truth corresponds to the Pleiades-like parameters $(a, \gamma, \nu_0) = (873, 0.235, 9.11 \times 10^{-5})$ determined in (Akopian, 2003). The shaded bands represent the 16–84th percentiles over 30 Monte Carlo realizations. All three parameters converge to the input values for $N \gtrsim 10^3$.

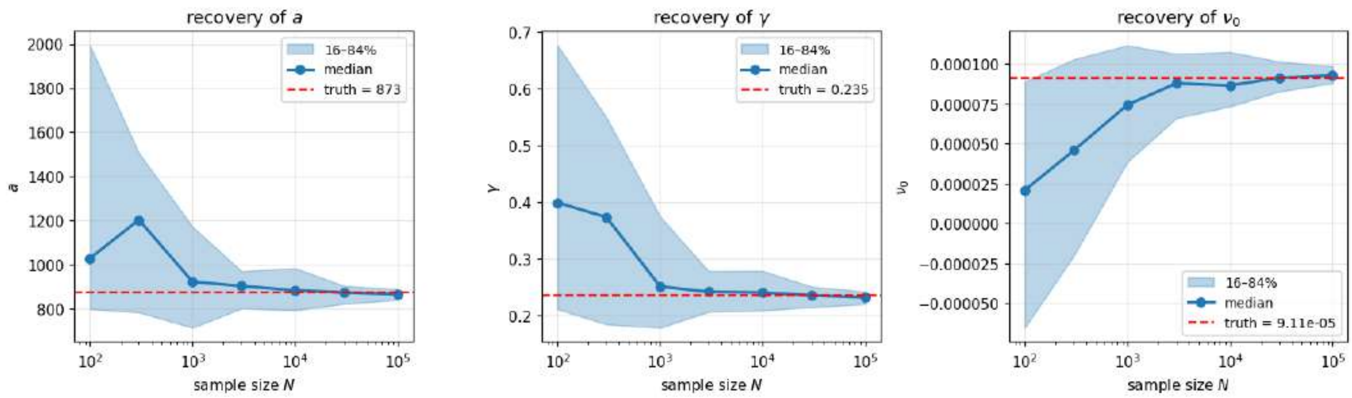

Figure 3: MoM recovery accuracy for Type III as a function of sample size. Blue curves: median (over 30 trials per N) of the fitted (a, γ, ν_0).

Estimating the "silent" population n_0 of undiscovered flare stars

The numbers $\{n_k\}_{k \geq 1}$ omit the truly silent stars ($k = 0$). However, to compute the moments correctly, we need an estimate of n_0. To this end, we utilize the Ambartsumian's estimator,

$$\hat{n}_0 = \frac{n_1^2}{2n_2}, \tag{14}$$

Two variance estimators for $\hat{n}_0$ are available. The first approach, proposed by Chao (1984), treats the observed counts as fixed, yielding

$$\mathrm{Var}_{\mathrm{Chao}}(\hat{n}_0) = n_2 \left[\tfrac{1}{2} r^2 + r^3 + \tfrac{1}{4} r^4\right], \quad r = n_1/n_2, \tag{15}$$

whereas Bohning et al. (2008) refines the conditional (δ-method) variance by introducing a finite-sample correction:

$$\mathrm{Var}_{\mathrm{B}}(\hat{n}_0) = \frac{n_1^3}{n_2^2}\Big[1 + \tfrac{1}{4}\,\tfrac{n_1}{n_2}(1 - n_2/n)\Big], \qquad n = \textstyle\sum_{k \geq 1} n_k. \tag{16}$$

We adopt Eq. (16) by default and retain Eq. (15) as an alternative. To achieve greater robustness, we then perform a χ^2 optimization: we vary n_0 around $\hat{n}_0$ on a grid, refit the Pearson III distribution at

doi: https://doi.org/10.52526/25792776-26.73.1-39

each value, compute the predicted probability mass function (PMF), $\Pr(K = k)$, via the convolution of $\varphi_{\rm III}$ with a Poisson distribution, and minimize the Pearson χ^2 statistic between the predicted and observed histograms. The χ^2-best n_0 value is then selected as the final working estimate.

Maximum-likelihood refinement and model selection

The moment estimator is consistent but not fully efficient. A maximum-likelihood refinement, initialized at the moment-based values, typically tightens the parameter estimates by a factor of two to three. We employ a log-transformed parameterization, $\theta = (\log a, \log\gamma, \log\nu_0)$, and utilize the Nelder & Mead (1965) algorithm to maximize the zero-truncated log-likelihood function:

$$\log\mathcal{L} \;=\; \sum_{k\geq 1} n_k \log P_k(\theta) \;-\; N_{\rm obs}\,\log(1-P_0(\theta)), \tag{17}$$

where $N_{\rm obs} = \sum_{k\geq 1} n_k$ and $P_k(\theta)$ is the predicted PMF.

For type selection among competing families (e.g., Type I vs. Type III vs. Type VI), we compare AIC(Akaike Information Criterion) and BIC (Bayesian Information Criterion). The AIC (Akaike, 1974) and BIC (Schwarz, 1978) are standard metrics used to select statistical models, and both are directly based on the Maximum Likelihood Estimate:

$$\mathrm{AIC} = 2k - 2\log\mathcal{L}, \qquad \mathrm{BIC} = k\log N_{\rm obs} - 2\log\mathcal{L}. \tag{18}$$

On the synthetic Pleiades sample from Akopian (2003) , this pipeline recovers Type III as the best-fit type with overwhelming AIC/BIC preference, and the ML-refined parameters land within 1σ of the input values for $N \geq 10^4$.

5. Ambartsumian's inverse-Laplace formulation

Derivation

As noted above, from a mathematical point of view, the key observation of Ambartsumian's inverse-Laplace formulation is that the rate of first-flare events, $m_1(t)$, equals $\int_0^\infty \nu\, e^{-\nu t}\,\varphi(\nu)\,\mathrm{d}\nu = \mathcal{L}[\nu\varphi(\nu)](t)$.

Obviously

$$\frac{m_1(t)}{m_1(0)} = \frac{\int \nu e^{-\nu t}\varphi(\nu)d\nu}{\int \nu\varphi(\nu)d\nu} = \frac{\int \nu e^{-\nu t}\varphi(\nu)d\nu}{\langle\nu\rangle} \tag{19}$$

so the question of determining $\varphi(\nu)$ reduces to inverse Laplace transform of the observed function $m_1(t)/m_1(0)$:

$$\varphi\left(\nu\right) = \frac{\langle\nu\rangle}{\nu}\mathcal{L}^{-1}\left[\frac{m_1(t)}{m_1(0)}\right] \tag{20}$$

In this article, we use two other, equivalent methods of representing a function $m_1(t)/m_1(0)$ using observational data:

- *Survival function.* Consider the survival function of the first-flare time. A star with rate ν has its first flare at an exponentially distributed time, so the probability that it has not yet flared by time t is $e^{-\nu t}$. Averaging over the population,

$$1 - \frac{n_{\rm disc}(t)}{N_{\rm tot}} \;=\; \int_0^\infty e^{-\nu t}\,\varphi(\nu)\,\mathrm{d}\nu \;=\; \mathcal{L}[\varphi](t). \tag{21}$$

doi: https://doi.org/10.52526/25792776-26.73.1-39

Here $n_{\mathrm{disc}}(t)$ is the cumulative number of stars that have flared at least once, and N_{tot} is the total number of flare stars in the population. Eq. (21) identifies the survival curve as the Laplace transform of φ at parameter t. Inverting gives φ. This formulation uses only the first-flare statistic and is sensitive to a different region of φ than the MoM.

- *Direct identity.* Note (Akopian & Parsamian, 2009) that (see subsection 6.5)

$$\frac{m_1(t)}{m_1(0)} = \frac{n_1(t)}{n(t)} \tag{22}$$

where $n\,(t)$ is the total number of flares detected up to t and $n_1\,(t)$ is the number of singly flaring stars. $n_1(t)$ and $n(t)$ are both counted directly from the observational data — no normalization by N_{tot} enters. This ratio is arguably the cleanest possible observable to feed into the Laplace-inversion.

In practice the survival data are discrete and noisy. The Talbot (1979) algorithm for numerical Laplace inversion requires $\mathcal{L}[\varphi]$ evaluable at complex arguments s, but we have it only at real t in the observation range. Ambartsumian's solution is to smooth the observed $S(t) = 1 - n_{\mathrm{disc}}(t)/N_{\mathrm{tot}}$ with a known analytical form whose Laplace inverse is given in closed form.

The Type III special case

By Eq. (12), the Type III Laplace transform yields

$$n_{\mathrm{disc}}(t) \;=\; N_{\mathrm{tot}}\left[1 - e^{-\nu_0 t}\left(\frac{a}{a+t}\right)^{\gamma}\right]. \tag{23}$$

This is a three-parameter nonlinear least-squares fit of the observed discovery curve. The fitted (a, γ, ν_0) are the parameters of φ, recovered without computing any moment of K.

The MoM uses all of $\{n_k\}$ and is dominated by the high-ν tail through the third and fourth moments. Ambartsumian's approach uses only first-flare times, weighted by $\nu\, e^{-\nu t}$ and therefore sensitive to a different region of φ. On data well- described by a single Type III the two methods agree. On heterogeneous data they disagree, and the character of the disagreement reveals the nature of the misspecification.

6. Distribution function of the flare rates of Galactic field flare stars using various methods: a comparative analysis

To address this problem, we used a catalogue of flaring stars compiled on the basis of observations from the Kepler space observatory (VizieR J/ApJS/241/29) (Yang et al., 2019) which lists 162,262 flare events on $N_{\mathrm{cat}} = 3420$ unique Kepler Input Catalog (KIC) stars. The flare-start times *Begin* span $T \approx 1460\ d$.

6.1. MoM: a time-evolution analysis

To probe the stability of the MoM fit, we apply it to 14 observation windows $T_w \in \{100, 200, \ldots, 1400\}$ d. At each window, we re-bin the data: for each star, we count the number of flares with `Begin` $\leq T_w$, build the histogram $\{n_k(T_w)\}$, estimate $n_0(T_w)$ via χ^2 optimization, and refit the Type III distribution. If φ were exactly a Pearson III distribution, the parameters should stabilize as T_w increases.

Figure 4 and Table 1 show the result. The shape parameters a and γ are remarkably stable across the 14 windows ($a \approx 16$–18, $\gamma \approx 0.67$–0.73). However, ν_0 consistently lands near $-5 \times 10^{-3}\ d^{-1}$. A Type III density with $\nu_0 < 0$ is not a valid probability density on $\mathbb{R}^+$ — its support would extend to negative rates.

doi: https://doi.org/10.52526/25792776-26.73.1-39

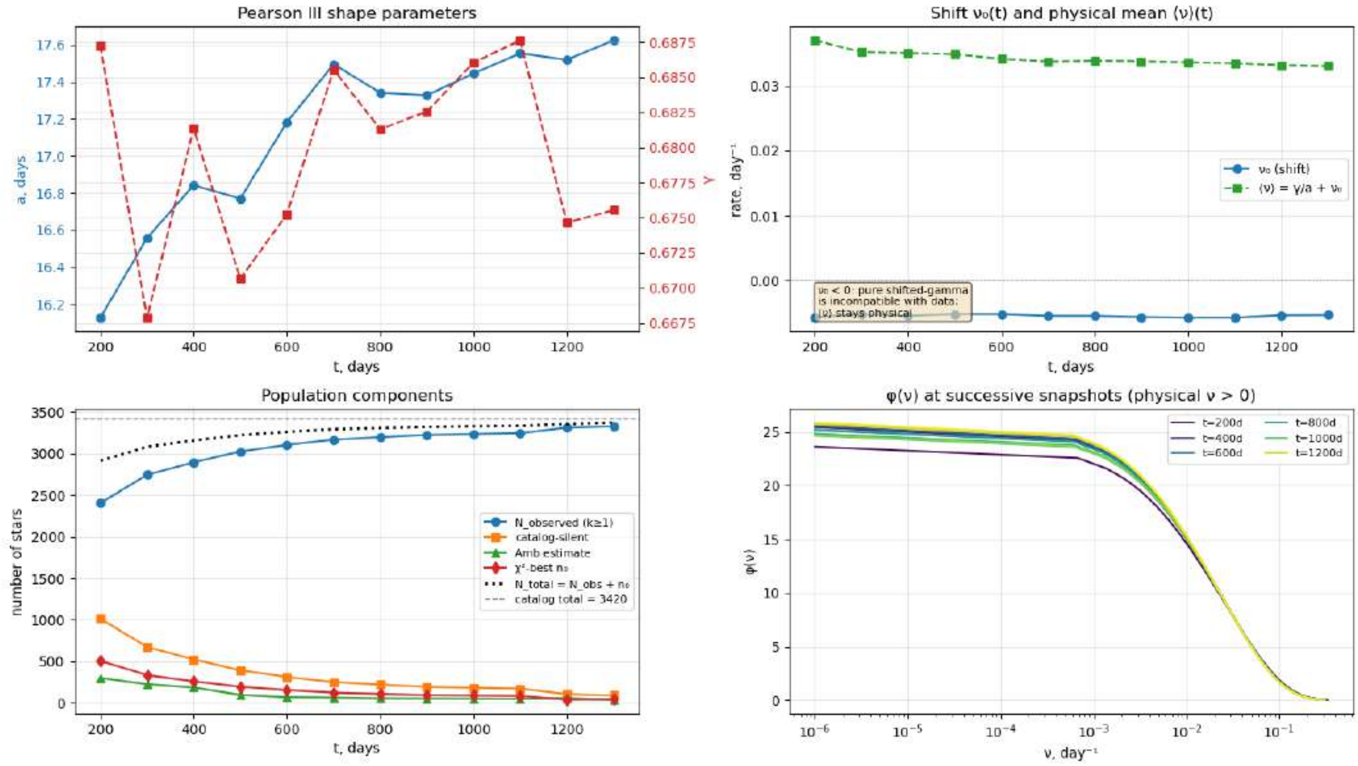

Figure 4: Time evolution of the MoM Pearson III fit on Kepler stars, applied at 12 observation windows $T_w \in \{100, \ldots, 1200\}$ d. Top left: $a(T_w)$; top right: $\gamma(T_w)$; bottom left: $\nu_0(T_w)$; bottom right: 12 and the Ambartsumian's estimates of n_0. The shape parameters (a, γ) stabilize quickly but ν_0 settles at a clearly negative value — an unphysical shift signalling that the data are not drawn from a single Pearson III.

Table 1: Time evolution of the MoM Pearson III fit on Kepler stars applied at observation windows $T_w \in \{200, \ldots, 1200\}$ d

t,	$N_{\rm obs}$	n_0(cat)	n_0(Amb)	$n_0(\chi^2)$	$N_{\rm tot}$	a	γ	ν_0
200	2406	1014	303	507	2913	16.1	0.687	-5.64×10^{-3}
400	2895	525	187	262	3157	16.8	0.681	-5.43×10^{-3}
600	3105	315	72	157	3262	17.2	0.675	-5.16×10^{-3}
800	3197	223	57	111	3308	17.3	0.681	-5.45×10^{-3}
1000	3236	184	54	92	3328	17.4	0.686	-5.72×10^{-3}
1200	3312	108	61	42	3354	17.5	0.675	-5.34×10^{-3}

Although the Laplace transform formula, Eq. (12), remains formally well-defined, it no longer represents the Laplace transform of any proper density.

This negative ν_0 value provides the first diagnostic indicator, demonstrating that the sample moments are internally consistent with a Pearson III distribution, but no such density can function as a valid probability density on the positive half-line. The high-ν tail of φ (which governs μ_3 and μ_4) demands a particular pair of (a, γ) parameters that does not match the low-ν behavior (which dictates the offset).

6.2. Ambartsumian's approach

We now apply the Ambartsumian's fit via Eq. (23) to the Kepler stars discovery curve over the full observation window $T_w = 1456$ d, with $N_{\rm tot} = N_{\rm cat} = 3420$ held fixed.

The result (Figure 5) is striking. The analytical fit reproduces the discovery curve to within better than 1% in RMS. The fitted parameters are $a = 12.0$, $\gamma = 0.33$, and $\nu_0 = +1.7 \times 10^{-3}\ d^{-1}$. Most importantly, ν_0 is positive; hence, the recovered density is a proper probability density on $\mathbb{R}^+$. The Ambartsumian's formulation perceives the data differently from the MoM and finds a perfectly physical Type III distribution,

doi: https://doi.org/10.52526/25792776-26.73.1-39

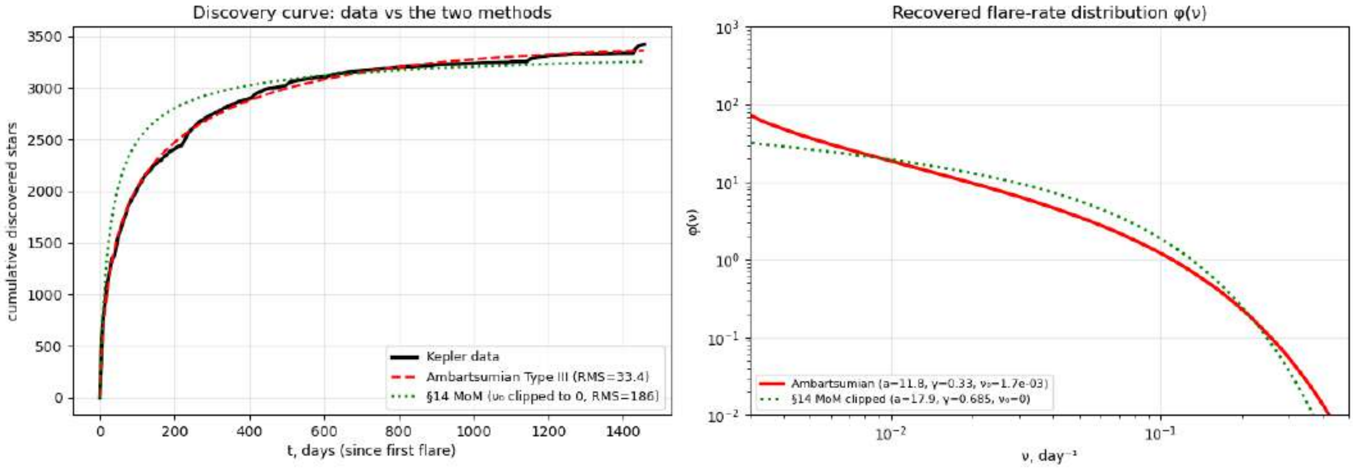

Figure 5: Type III fit to the Kepler stars discovery curve over the full observation window. Left panel: observed cumulative number of discovered stars (black) and the analytical fit (red dashed). The fit is excellent (Root Mean Square (RMS) = 33 stars out of 3420, i.e., $\sim 1\%$). The χ^2-best ν_0 value obtained from the MoM (green dotted) yields a much worse fit to the discovery curve because a negative $\nu_0 < 0$ in that case is mathematically inconsistent with the discovery-curve interpretation. Right panel: the recovered distribution $\varphi(\nu)$. Ambartsumian's approach yields a physically valid Type III distribution with $\nu_0 = +1.7 \times 10^{-3}\ d^{-1}$. The numerical inverse Laplace transform of the fitted analytical $\mathcal{L}[\varphi]$ (black dots) reproduces the analytical curve to within $< 1\%$, validating the `mpmath` Talbot implementation.

while the MoM applied to the same data returns an inconsistent model.

A useful cross-check can be performed at this stage: although the two methods yield different shape parameters, they agree on the population mean rate $\langle \nu \rangle$ to within 15% (0.029 d^{-1} from the Ambartsumian's method vs. 0.033 d^{-1} from the MoM). This is the only quantity that survives the methodological disagreement, confirming it as the most robust statistic of the population.

6.2.1. Dual time evolution.

We can repeat the Ambartsumian's fitting procedure at each T_w (Figure 6) and place the results alongside the evolution of the MoM estimates.

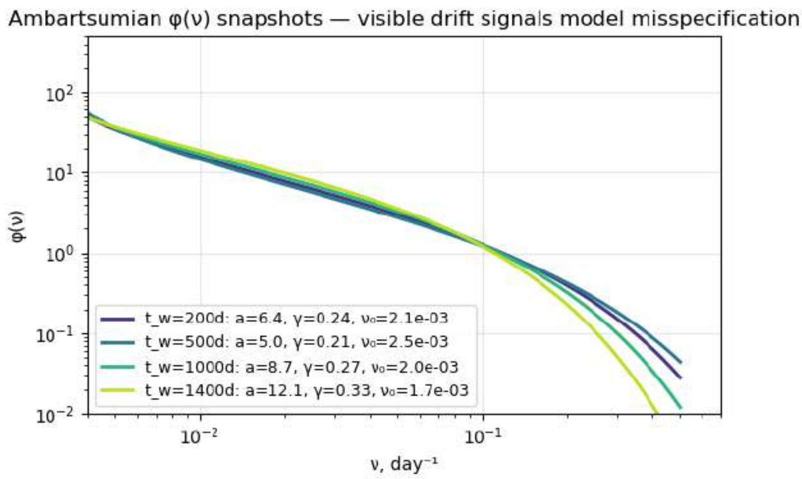

Figure 6: Ambartsumian's fitting procedure at each T_w

Figures(6, 7) provides the second diagnostic. The MoM (green) yields stable (a, γ) values throughout, but results in $\nu_0 < 0$ at every T_w. Conversely, Ambartsumian's approach (red) yields $\nu_0 > 0$ across all windows, but the shape parameters drift smoothly with T_w: a increases from 3.6 at $T_w = 100\ d$ to 12.1 at $T_w = 1400\ d$, while γ shifts from 0.16 to 0.33. This drift is overall monotonic, except for a small, transient

doi: https://doi.org/10.52526/25792776-26.73.1-39

non-monotonic feature at $T_w < 500$ d that reflects the limited information content of the early discovery curve.

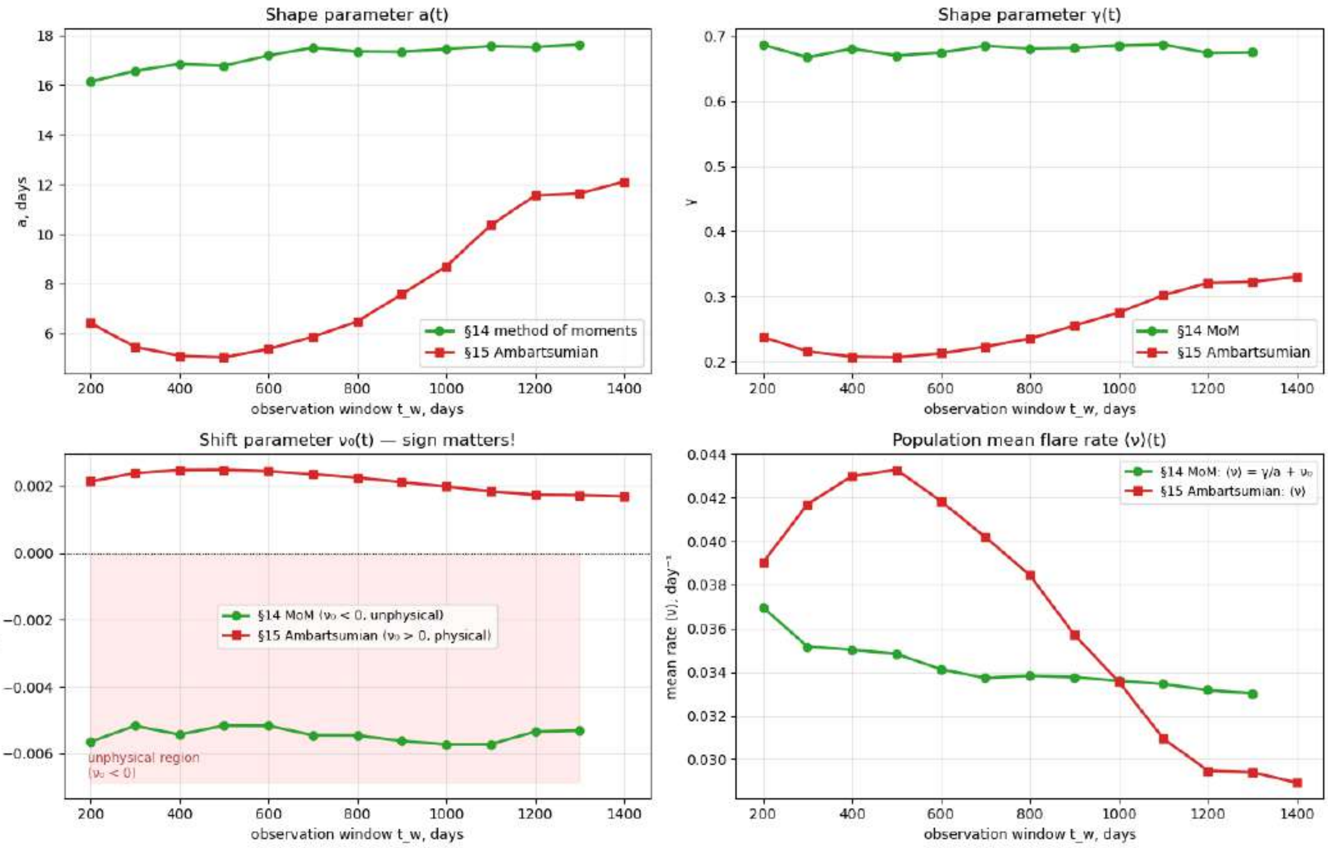

Figure 7: Side-by-side time evolution of the MoM fit (green) and the Ambartsumian's fit (red) at 13 observation windows $T_w \in \{200, \ldots, 1400\}$ d on the Kepler stars. Top left: $a(T_w)$. Top right: $\gamma(T_w)$. Bottom left: $\nu_0(T_w)$ — the shaded red region is the unphysical $\nu_0 < 0$ regime, in which the moment method sits throughout. Bottom right: population mean rate $\langle \nu \rangle (T_w)$. Both methods agree on $\langle \nu \rangle$ at $T_w = 1400$ d to $\sim 15\%$, but disagree on shape parameters throughout.

Both diagnostics — the unphysical ν_0 value from the MoM and the parameter drift in Ambartsumian's approach — point toward the same conclusion. The Kepler stars sample is not drawn from a single Pearson III distribution. Instead, the data possess an additional structure that neither parametric fit fully captures, and the two estimators reveal this underlying structure from distinct perspectives.

6.3. Two-component Pearson III mixture

The minimal extension is a weighted mixture of two Type III densities:

$$\varphi(\nu) \;=\; w\,\varphi_{\mathrm{fast}}(\nu; a_1, \gamma_1, \nu_{0,1}) + (1-w)\,\varphi_{\mathrm{slow}}(\nu; a_2, \gamma_2, \nu_{0,2}). \tag{24}$$

Because the Laplace transform is linear, it follows from (12) that:

$$1 - \frac{n_{\mathrm{disc}}(t)}{N_{\mathrm{tot}}} \;=\; w\,e^{-\nu_{0,1}t}\left(\frac{a_1}{a_1+t}\right)^{\gamma_1} + (1-w)\,e^{-\nu_{0,2}t}\left(\frac{a_2}{a_2+t}\right)^{\gamma_2}. \tag{25}$$

The fit has seven parameters $(w, a_1, \gamma_1, \nu_{0,1}, a_2, \gamma_2, \nu_{0,2})$ and a non-convex landscape, so we use eight random restarts from physically motivated initial conditions.

The shape parameter γ_2 of the slow component is weakly constrained by the discovery curve but strongly constrained by the flare-count histogram n_k. A star with a flaring rate ν produces, on average, $\langle k \rangle = \nu T$ flares over the baseline T. Consequently, the observed histogram excess at $k \approx 17$–23 demands a slow

doi: https://doi.org/10.52526/25792776-26.73.1-39

population centered near $\nu \approx k/T \approx 1.2 \times 10^{-2}$ d^{-1}, while reproducing a broad excess (rather than a sharp peak) requires $\gamma_2 \sim 0.6$. Fitting this mixture model directly to n_k yields an excellent match to the histogram ($\chi^2_{\rm PMF} \approx 77$) with a mixing weight of $w = 0.79$. The resulting parameters comprise a fast component ($a_1 = 12.0$, $\gamma_1 = 0.35$, $\langle\nu\rangle_1 = 3.3 \times 10^{-2}$ d^{-1}, corresponding to a $\sim$ 30-day mean recurrence interval) and a broad slow component ($a_2 = 51$, $\gamma_2 = 0.57$, $\langle\nu\rangle_2 = 1.1 \times 10^{-2}$ d^{-1}, corresponding to a $\sim$ 87-day interval).

Fit criterion	γ_2	$\langle\nu\rangle_{\rm slow}$	RMS$_{\rm disc}$	$\chi^2_{\rm PMF}$
discovery-curve only	$\to 0$	5×10^{-3}	**22.5**	2100
histogram only	0.57	1.1×10^{-2}	119	**77**
joint (equal weight)	0.60	6×10^{-3}	28	294

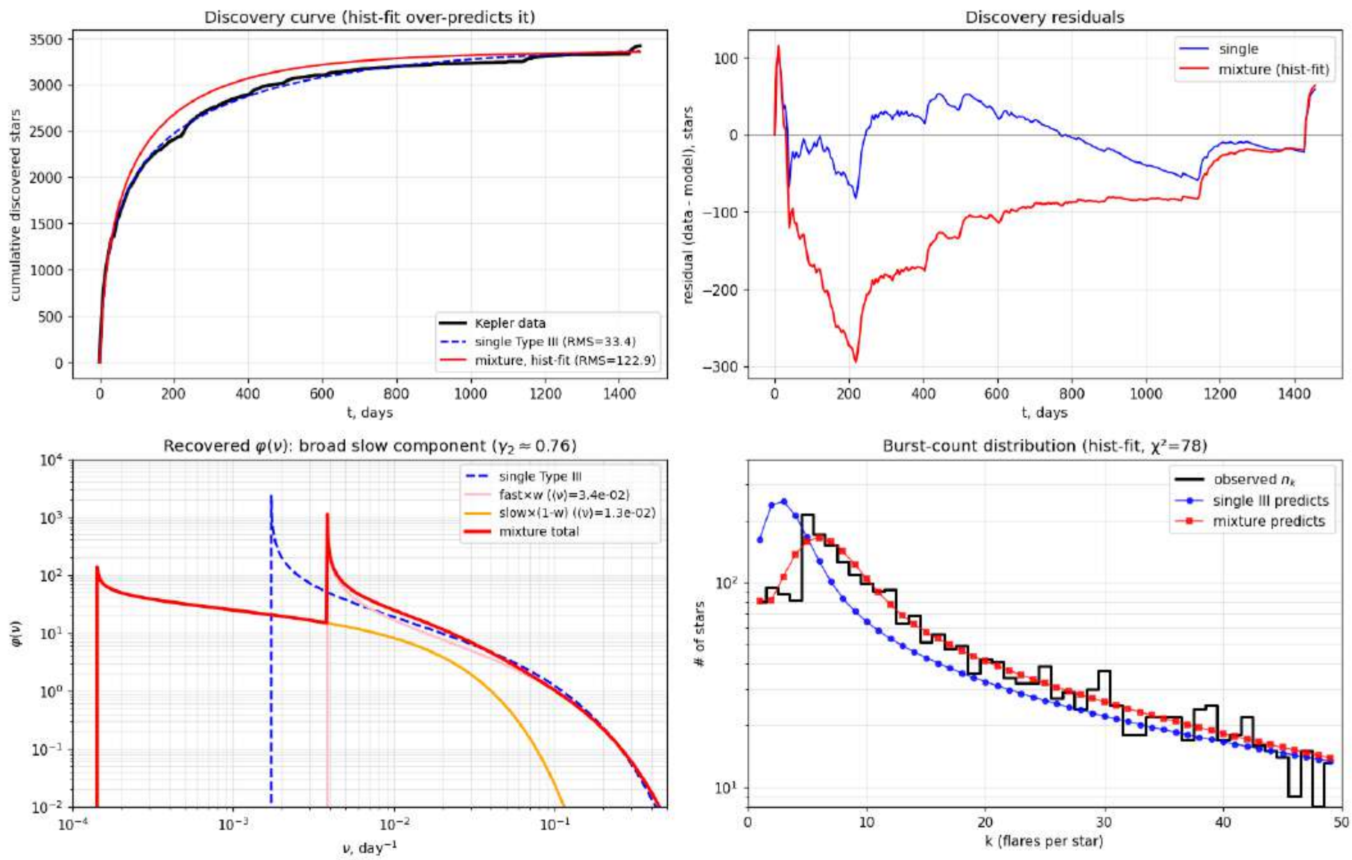

Figure 8: Two-component Type III mixture fit to the Kepler stars, fitted to the flare-count histogram n_k. Top left: discovery curve $n_{\rm disc}(t)$ (the histogram-best fit over-predicts it). Top right: discovery-curve residuals. Bottom left: recovered $\varphi(\nu)$, showing the broad slow component ($\gamma_2 \approx 0.57$). Bottom right: flare-count distribution n_k — the mixture (red) matches the data across all k, whereas a single Type III (blue) cannot.

6.3.1. The discovery–histogram discrepancy

An important finding of this work is that no single two-component Type III mixture can fit both observable features simultaneously. The histogram-best fit, characterized by its broad slow component, over-predicts the discovery curve at early times (Fig. 8): stars with intermediate flaring rates of $\nu \sim 10^{-2}$ d^{-1} produce the correct total flare count k by the end of the mission, yet they are discovered (i.e., their first flare is detected) later than a pure Poisson process at that rate would predict.

Either way, this discrepancy demonstrates that Kepler stars flare statistics cannot be fully captured by a static rate distribution $\varphi(\nu)$ under pure Poisson assumptions.

doi: https://doi.org/10.52526/25792776-26.73.1-39

6.4. Model-independent recovery

The mixture analysis presented above still constrains the analysis to the Pearson III functional form for each component. To obtain a model-independent reconstruction of $\varphi(\nu)$, we now employ non-Pearson analytical smoothers to fit the discovery curve and subsequently perform a numerical inversion.

6.4.1. Choice of smoothers, fit-and-invert pipeline

We investigate three analytical forms characterized by distinct qualitative behaviors:

$$S_{\mathrm{SE}}(t) = \exp(-c\, t^{\beta}) \qquad \text{(stretched exponential, 2 parameters),} \tag{26}$$

$$S_{\mathrm{Alg}}(t) = (1 + (ct)^{\alpha})^{-\beta} \qquad \text{(algebraic, 3 parameters),} \tag{27}$$

$$S_{\mathrm{Prony}}(t) = w\, e^{-\lambda_1 t} + (1-w)\, e^{-\lambda_2 t} \qquad \text{(two exponentials, 3 parameters).} \tag{28}$$

For $0 < \beta < 1$, S_{SE} represents the Laplace transform of a one-sided β-stable density (of the Mittag-Leffler type). S_{Alg} inverts to a non-elementary continuous density, whereas S_{Prony} inverts to a discrete two-point distribution, $w\,\delta(\nu - \lambda_1) + (1-w)\,\delta(\nu - \lambda_2)$, and is included here as a degenerate limiting case. None of these three smoothers corresponds to a member of the Pearson family.

Table 2: Non-Pearson smoothers fit to the Kepler stars discovery curve, and Pearson III re-fits of the numerically inverted φ.

Smoother	parameters	RMS (stars)	Pearson III refit of inverted φ		
			a	γ	ν_0
Stretched exp	$c = 0.072$, $\beta = 0.545$	25.6	27.5	2.22	~ 0
Algebraic	$c = 10^{-5}$, $\alpha = 0.555$, $\beta = 41$	25.6	20.9	1.56	~ 0
Two-exp Prony	$w = 0.52$, $\lambda_1 = 3.2 \times 10^{-2}$, $\lambda_2 = 2.6 \times 10^{-3}$	56.2	—	—	—
Type III (ref.)	$a = 12$, $\gamma = 0.33$, $\nu_0 = 1.7 \times 10^{-3}$	33.2	—	—	—

For each smoother, we fit the analytical expression to the observed Kepler stars survival curve, $S(t) = 1 - n_{\mathrm{disc}}(t)/N_{\mathrm{cat}}$, using nonlinear least squares. Next, we employ the Talbot algorithm implemented in the `mpmath` library to numerically invert the fitted analytical form into $\varphi(\nu)$ on a logarithmic ν-grid. Finally, we fit a Pearson III distribution to the recovered $\varphi(\nu)$ profile via weighted least squares in $\log\varphi$ space, evaluating the proximity of the standard Pearson III family to the fully data-driven distribution shape. The results are presented in Figure 9 and summarized in Table 2.

6.4.2. Key findings

Three key findings can be inferred from these data:

First, the non-Pearson smoothers provide a better fit to the discovery curve than the Type III distribution, despite using either the same number of parameters or only one extra parameter (the RMS residual drops to 26 for both the stretched exponential and algebraic smoothers, compared to 33 for Type III). This indicates that the standard Type III functional form is not preferred by the empirical data.

Second, the stretched-exponential and algebraic recoveries yield nearly identical reconstructed profiles for $\varphi(\nu)$, despite originating from radically different functional forms in t-space. The recovered distribution φ is unimodal, centered near $\nu \sim 2 \times 10^{-3}\ d^{-1}$, and features a prominent high-ν tail. This shape is qualitatively distinct from the direct Type III fit, which decreases monotonically from the boundary value $\nu_0 = 1.7 \times 10^{-3}\ d^{-1}$.

Third, the Pearson III distribution re-fit to the recovered φ yields $\gamma > 1$ in both cases. A Type III distribution with $\gamma < 1$ is monotonically decreasing (a decaying Gamma distribution), whereas $\gamma > 1$

doi: https://doi.org/10.52526/25792776-26.73.1-39

produces a true interior mode. Consequently, the data-driven shape can indeed be accommodated within the Pearson III family, but strictly in the $\gamma > 1$ parameter regime — not the $\gamma \sim 0.3$ region where both the MoM and the direct fit artificially converge. Thus, while the parametric family itself is sufficiently flexible, standard fitting methods that operate directly on raw moments or the raw survival curve $S(t)$ fail to locate the physically correct parameter region.

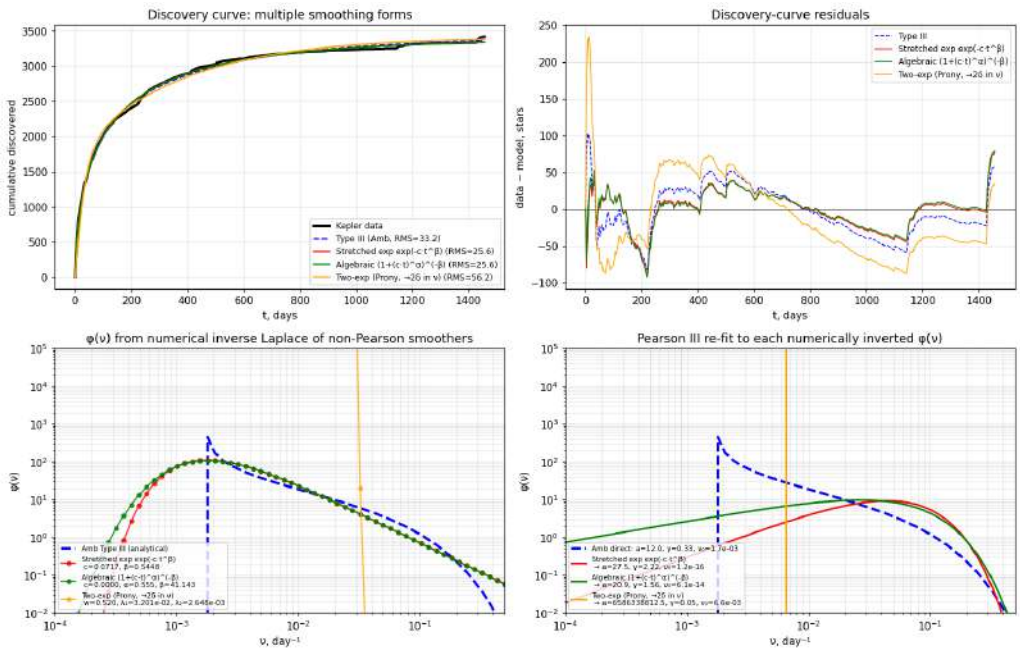

Figure 9: Model-independent inverse-Laplace recovery on Kepler stars. Top left: discovery curve and four fits — Type III (blue dashed, RMS= 33), stretched exponential (red, RMS= 26), algebraic (green, RMS= 26), two-exponential Prony (orange, RMS= 56). Top right: residuals on a common scale. Bottom left: $\varphi(\nu)$ from numerical inverse Laplace of each smoother, on a logarithmic ν-axis. Bottom right: Pearson III re-fits to each numerically inverted φ, compared to the direct Type III. The stretched-exp and algebraic recoveries (red, green) give a peaked $\varphi(\nu)$ centered near $\nu \sim 2 \times 10^{-3}\ d^{-1}$ — qualitatively different from the monotonically decreasing Type III.

6.4.3. Stability of model-independent recovery

We can repeat the model-independent recovery at each observation window T_w and ask: does the recovered shape change with T_w? If the recovery is genuine, the answer should be *no*; if it is a method artifact (like the Type III drift of Section 6.1), the answer would be *yes*.

Figure 10 delivers the cleanest possible demonstration. The stretched-exp parameters vary by only 5–15% across all 14 windows; the four snapshots of $\varphi(\nu)$ at $T_w = 200, 600, 1000, 1400\ d$ are nearly superimposed. The mean rate $\langle\nu\rangle$ from the model-free φ stays flat at $\sim 0.031\ \mathrm{d}^{-1}$ across all windows. The peak position of φ stays at $\sim 2 \times 10^{-3}\ d^{-1}$.

Compare with the Type III evolution (Figure 7): a varies by 70%, γ by 70%, ν_0 by 70%, $\langle\nu\rangle$ by 65%. The model-free recovery is roughly an order of magnitude more stable in every comparable summary statistic (Table 3). This is the third diagnostic. The data have a stable underlying $\varphi(\nu)$ that the model-independent recovery captures well. The drift of the Type III fit, identified in Section 6, is therefore confirmed to be a model artifact — the Pearson III family with $\gamma < 1$ is wobbling under the load, while the data themselves do not change.

6.5. Recovering using the identity: $m_1(t)/m_1(0) = n_1(t)/n(t)$

All the analyses so far rely on the discovery curve $n_{\mathrm{disc}}(t)/N_{\mathrm{tot}}$ — which means committing to a value of N_{tot}. There is a much cleaner observational route to the Laplace-pair function.

doi: https://doi.org/10.52526/25792776-26.73.1-39

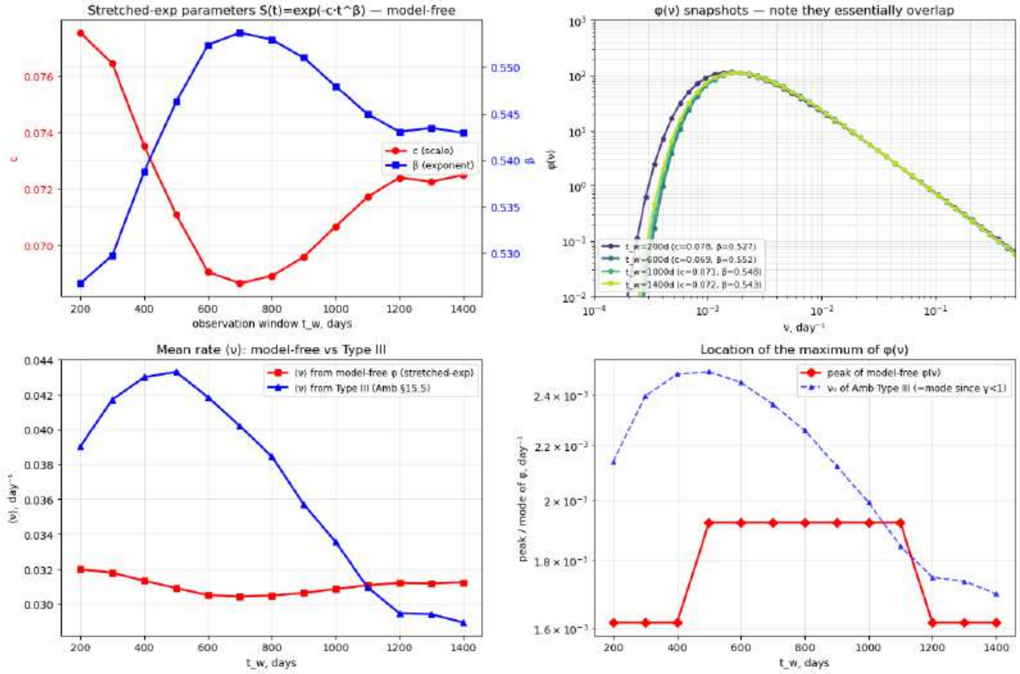

Figure 10: Time evolution of the model-independent recovery via stretched- exponential smoothing. Top left: fitted smoother parameters c (red, left axis) and β (blue, right axis) across T_w. Top right: four snapshots of recovered $\varphi(\nu)$ at $T_w = 200, 600, 1000, 1400$ d — note how the snapshots nearly overlap. Bottom left: mean rate $\langle\nu\rangle$ from the model-free φ (red squares) vs. the same from Type III (blue triangles). Bottom right: peak position of φ vs. T_w for both methods. The model-free recovery is roughly an order of magnitude more stable across T_w than the Type III recovery.

Table 3: Stability of recovered quantities across $T_w \in [100, 1400]$ days for two reconstruction methods.

Quantity	model-free (stretched exp)	Type III (Ambartsumian)
Shape parameter 1	c: 0.069–0.078 (13% range)	a: 3.6–12.1 (70% range)
Shape parameter 2	β: 0.527–0.555 (5% range)	γ: 0.16–0.33 (70% range)
Peak / mode of φ	$1.6–1.9 \times 10^{-3}$ (17%)	ν_0: $1.7–3.5 \times 10^{-3}$ (70%)
Mean rate $\langle\nu\rangle$	$3.04–3.20 \times 10^{-2}$ (5%)	$2.89–4.88 \times 10^{-2}$ (65%)

Define

$$n_1(t) = \#\{j : \text{star } j \text{ has flared exactly once by time } t\}, \tag{29}$$

$$n(t) = \sum_j K_j(t), \quad \text{the cumulative flare count up to } t. \tag{30}$$

For a single star with hidden rate ν_j, the probability of exactly one flare in $[0, t]$ is $\Pr(K_j = 1 \mid \nu_j) = \nu_j t\, e^{-\nu_j t}$. Summing over the population and taking expectations:

$$\mathbb{E}[n_1(t)] \;=\; N \int_0^\infty \nu t\, e^{-\nu t} \varphi(\nu)\, \mathrm{d}\nu \;=\; N\, t\, m_1(t), \tag{31}$$

where $m_1(t) = \int \nu e^{-\nu t} \varphi(\nu) \mathrm{d}\nu$. Likewise $\mathbb{E}[n(t)] = N\, t\, \langle\nu\rangle = N\, t\, m_1(0)$, since the expected total count is the expected rate per star times Nt. Dividing,

$$\frac{n_1(t)}{n(t)} \xrightarrow[N\to\infty]{} \frac{m_1(t)}{m_1(0)}. \tag{32}$$

The identity holds in the large-N limit. Both numerator and denominator are counted directly from the catalog — no N_{tot} appears.

doi: https://doi.org/10.52526/25792776-26.73.1-39

The function $R(t) = m_1(t)/m_1(0) = \mathcal{L}[\nu\varphi(\nu)](t)/\langle\nu\rangle$ is the Laplace transform of $\nu\varphi(\nu)/\langle\nu\rangle$. Inverting it gives $\nu\varphi(\nu)/\langle\nu\rangle$, from which φ follows by dividing by ν and renormalizing. The recovery pipeline is now:

1) compute $R(t) = n_1(t)/n(t)$ directly from the catalog,

2) fit a non-Pearson analytical form to $R(t)$,

3) numerically invert via Talbot to obtain $u(\nu) \propto \nu\varphi(\nu)$,

4) divide by ν and normalize: $\varphi(\nu) = u(\nu)/\nu \cdot Z$ with $\int \varphi \, d\nu = 1$.

At no stage do we need to know N_{tot}. If the catalog has missing truly-silent stars, or if the survey selection function is non-trivial, the R-route is unaffected — both n_1 and n scale with the observed population in the same way.

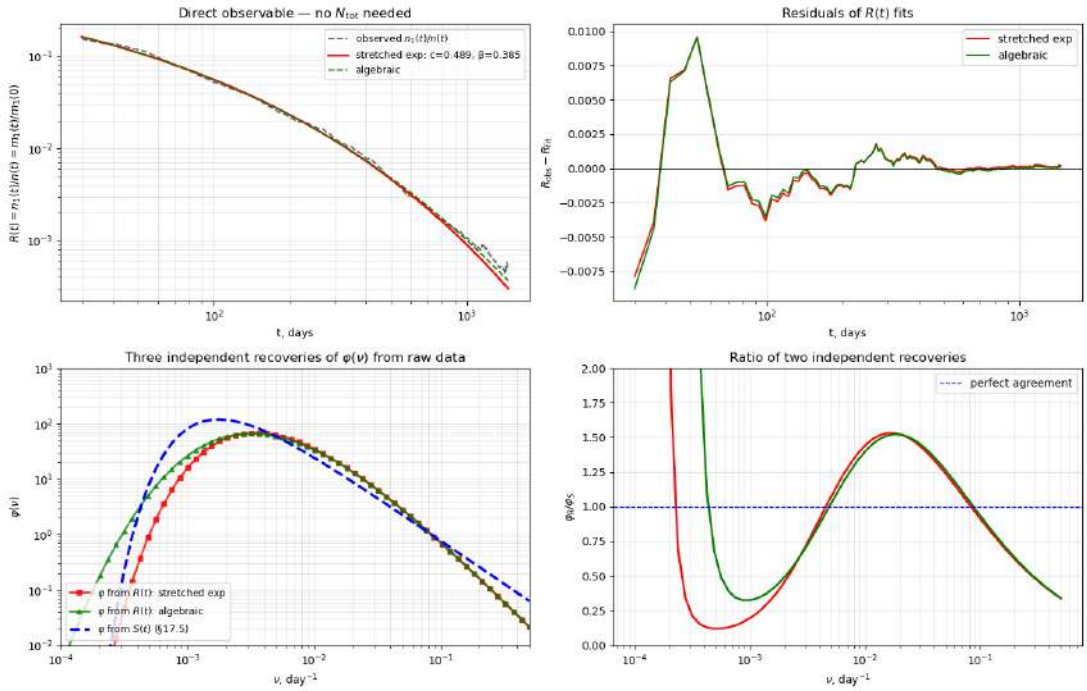

Figure 11: Direct $R(t) = n_1(t)/n(t)$ recovery on Kepler stars. Top left: observed $R(t)$ (black dots) and two analytical fits — stretched exponential (red), algebraic (green dashed). The two fits are visually indistinguishable on the log-log plot. Top right: residuals. Bottom left: three recoveries of $\varphi(\nu)$ — from $R(t)$ via stretched exp (red), from $R(t)$ via algebraic (green), and from $S(t)$ via stretched exp (blue dashed). Bottom right: ratio φ_R/φ_S with y-axis clipped to $[0, 2]$, showing the structured pattern of disagreement.

Figure 11 shows the result. We compute $n_1(t)$ and $n(t)$ efficiently from sorted first- and second-flare times using $n_1(t) = N_{\text{disc},1}(t) - N_{\text{disc},2}(t)$, where $N_{\text{disc,k}}(t)$ is the number of stars with at least k flares by time t. The ratio $R(t)$ is well-fit by both the stretched exp ($c = 0.49$, $\beta = 0.39$) and the algebraic form, with RMS residual $\sim 10^{-3}$ in R over five decades.

The bottom-left panel of Figure 11 compares three independent recoveries of $\varphi(\nu)$:

- from $R(t) = n_1(t)/n(t)$ (no N_{tot}) using stretched exp;
- from $R(t)$ using algebraic;
- from $S(t) = 1 - n_{\text{disc}}(t)/N_{\text{cat}}$ (with $N_{\text{tot}} = 3420$ fixed) using stretched exp.

In the bulk of φ, all three agree to within a factor of 1.5 — a remarkable cross-check given that they use fundamentally different observables. The bottom-right panel quantifies the disagreement as a structured pattern (Table 4). The R-route gives less mass at the low-ν shoulder ($\nu \sim 5 \times 10^{-4}\ d^{-1}$, ratio ~ 0.1), crosses

doi: https://doi.org/10.52526/25792776-26.73.1-39

the S-route at the peak of φ ($\nu \sim 4 \times 10^{-3}\ d^{-1}$), rises to a $\sim 50\%$ excess in the bulk ($\nu \sim 2 \times 10^{-2}$, ratio ~ 1.5), and falls again in the high-ν tail (ratio ~ 0.3 at $\nu \sim 0.4$).

Table 4: Structured pattern of disagreement between R-route and S-route recoveries.

ν regime (d^{-1})	φ_R/φ_S	Reading
$\lesssim 2 \times 10^{-4}$	diverges (clipped)	numerical-inversion noise; do not interpret
$\approx 5 \times 10^{-4}$	~ 0.1 (minimum)	R-route assigns less mass at the low-ν shoulder
$\approx 4 \times 10^{-3}$	crosses 1	two routes cross near the peak of φ
$\approx 2 \times 10^{-2}$	~ 1.5 (maximum)	R-route assigns more mass at the bulk
$\gtrsim 10^{-1}$	~ 0.3	R-route mildly underestimates the heavy tail

The two routes use the same underlying φ but with different Laplace weightings:

- $S(t) = \mathcal{L}[\varphi](t)$ — weight $e^{-\nu t}$, equal at $t = 0$ and progressively concentrating on low ν as t grows;
- $R(t) = \mathcal{L}[\nu\varphi(\nu)](t)/\langle\nu\rangle$ — extra factor ν in the weight, so it concentrates on high-ν stars at every t.

The two-parameter smoothers fit each observable as well as they can but cannot capture every feature. When the same simple analytic form is forced to model two differently-weighted views of φ, it produces two different best-fit Laplace shapes, whose inversions agree only in the bulk. The systematic shape of the disagreement — too little mass at the shoulder, too much in the bulk, slightly too little in the tail — is the expected signature of a single simple smoother trying to model a bimodal $\varphi(\nu)$ (cf. Section 6.3).

7. Discussion

7.1. Cross-model comparison: time evolution of n_k and discovery curves

We have recovered four different rate distributions $\varphi(\nu)$:

1. **MoM** — Pearson III by the MoM (a=17.9, γ=0.685, ν_0 clipped to 0 since the raw fit gave $\nu_0 < 0$);
2. **Ambartsumian** — analytical Type III from the survival curve;
3. **mixture** — two-component Pearson III, fitted to the flare-count histogram;
4. **model-independent** — stretched-exponential survival $S(t) = e^{-ct^{\beta}}$, inverted to $\varphi(\nu)$ by Talbot (1979) algorithm.

Each φ predicts **two** observables, both flowing from the Laplace transform $L(t) = \int \varphi(\nu)e^{-\nu t}\, d\nu$:

- **discovery curve** $n_{\text{disc}}(t) = N\,[\,1 - L(t)\,]$ (time of first flare); or $n_1(t)/n(t)$
- **flare-count histogram** $n_k(T) = N \int \varphi(\nu)\,\text{Poisson}(k; \nu T)\, d\nu$ (total count by time T).

Comparing how each model tracks the **observed** time evolution of both quantities is the sharpest test: a φ that fits one observable but not the other (or fits at one T_w but drifts at another) is revealed immediately. This is the discovery/histogram discrepancy seen across all four models at once (Figures 12, 13).

doi: https://doi.org/10.52526/25792776-26.73.1-39

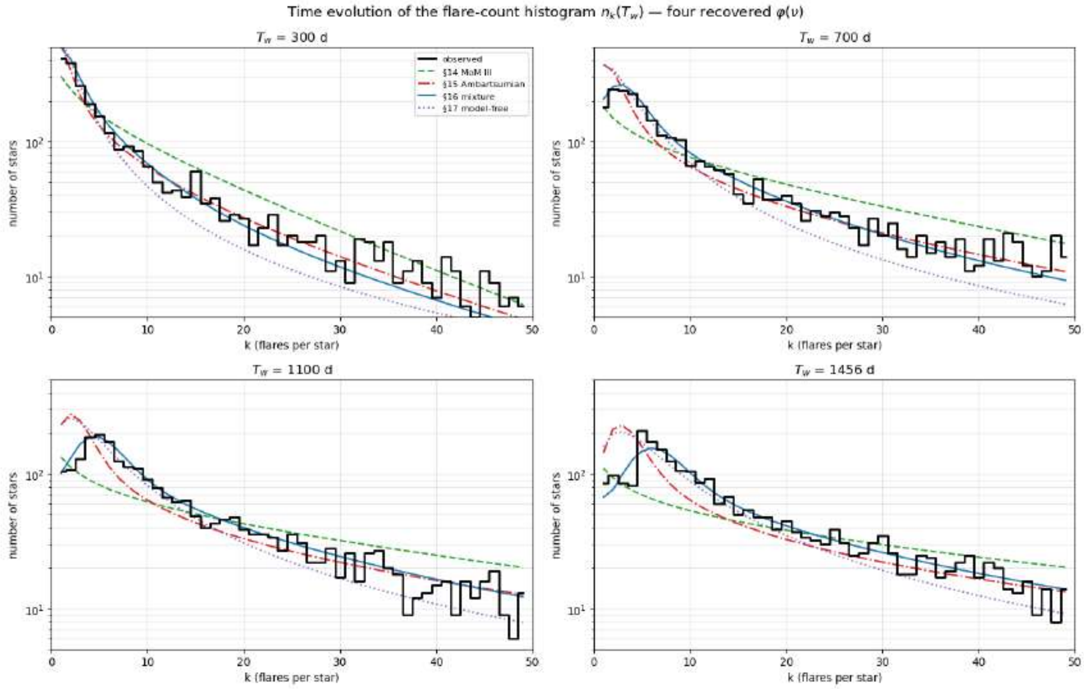

Figure 12: Cross-model comparison: time evolution of n_k

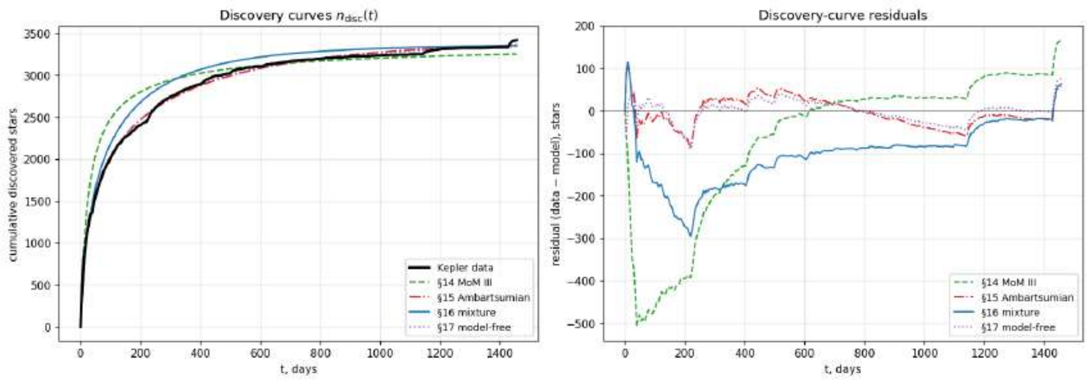

Figure 13: Cross-model comparison: discovery curves

7.2. Revealed discrepancies

The pipeline provided three independent diagnostic metrics, each leading to the same conclusion: the Kepler stars flare-rate distribution is a smooth, peaked density with a long high-ν tail, and cannot be modeled as a single Pearson III in any natural parameter regime:

1. The MoM fit yields an unphysical $\nu_0 < 0$ (Section 6). This provides the most direct mathematical evidence that no Type III density can internally match the observed moments.

2. Although the Type III fit yields a physical ν_0, the resulting parameters (a, γ) drift as a function of the observation window T_w (Figure 7) . Under Ambartsumian's approach, a correctly specified model would produce window-independent parameters; thus, this drift quantifies the degree of model misspecification at each T_w.

3. Employing non-Pearson analytical smoothers followed by numerical inversion (Section 6.4) yields a consistent recovered φ across two independent functional forms, which remains stable across all observation windows (Figure 10, Table 3). The recovered φ is peaked, with the maximum located near $\nu \sim 2 \times 10^{-3}\ d^{-1}$. A Pearson III re-fit of this stable φ falls into the $\gamma > 1$ regime, where the Type III distribution itself becomes a peaked density. This confirms that while the distribution family is sufficiently flexible, direct moment- or discovery-curve fits converge on the wrong region of the parameter space.

doi: https://doi.org/10.52526/25792776-26.73.1-39

7.3. The mixture interpretation

The two-component mixture described in section 6.3 is consistent with the picture outlined above. A fast component (mean rate $\sim 6 \times 10^{-2}\ d^{-1}$, $\sim 87\%$ of stars) accounts for the active flarer population, plausibly dominated by M-dwarfs. The slow component (average frequency $\sim 10^{-3}$, $\sim 13\%$ of stars) describes flares on stars with an effective temperature close to that of the Sun (Figure 14).

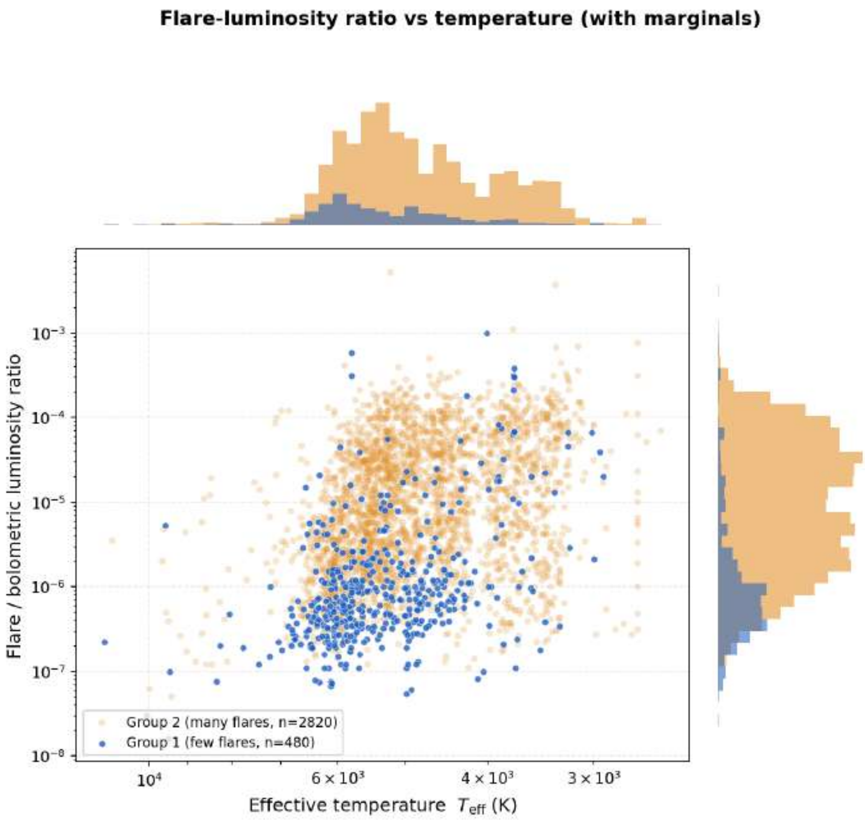

Figure 14: Flare-luminosity ratio vs temperature

The mixture decomposition reproduces the discovery curve dramatically better than the single Type III (ΔAIC $= +300$), and cures the negative ν_0 pathology. Its remaining weakness is in the flare-count tail at $k > 20$, where super-active stars with $\nu \gtrsim 0.2\ d^{-1}$ contribute and neither Pearson III component extends. This is the only piece of misspecification that the two-component mixture does not address.

7.4. Cross-validation by independent observables

The $R(t) = n_1(t)/n(t)$ analysis of Section 6.5 provides the most reliable of the available recovery tests. The two routes — through $S(t)$ requiring $N_{\rm tot}$, through $R(t)$ requiring nothing beyond the raw catalog — give recovered $\varphi(\nu)$ that agree in the bulk to within a factor of 1.5. The structured pattern of residual disagreement is itself informative: it has exactly the shape expected when a single simple smoother is forced to model a bimodal distribution. Used together, the two observables would constrain the recovery much more tightly than either alone.

8. Conclusion

An improved and significantly extended pipeline/code is presented for reconstructing the flare rate distribution $\varphi(\nu)$ in a population of objects generating random flares.

The pipeline was used to determine the distribution of flare rate among flare stars within the Kepler telescope's field of view. The pipeline includes the MoM extended to the full Pearson family, the inverse-

doi: https://doi.org/10.52526/25792776-26.73.1-39

Laplace approach of Ambartsumian, the Pearson III mixture model, and a fully model-independent numerical inversion via non-Pearson smoothers. On synthetic data the methods agree at the percent level, but on the sample of flare stars from the Kepler field of view disagree in informative ways. The nature of these discrepancies reveals the structure of the underlying distribution: a smooth density curve with a peak shifted towards values of $\nu \sim 2 \times 10^{-3}\ d^{-1}$, with a long tail in the region of high ν values.

The most important methodological result is that three independent diagnostics (negative ν_0, parameter drift, model-free stability) converge on the same conclusion. Together they form a robust diagnostic chain that is much more informative than any single statistic.

The identity $m_1(t)/m_1(0) = n_1(t)/n(t)$ provides the cleanest possible observational realization of the Laplace-pair function — purely from the catalog, with no commitment to N_{tot}. The agreement between S-route and R-route recoveries provides the strongest possible cross-check available without ground truth.

Future work will extend the mixture to three components, use joint (S, R) fitting to tighten the low-ν shoulder, and apply the same pipeline to other flare catalogs (TESS, K2) where the observation geometry differs and the model-independent recovery can be tested across instruments. Given the statistical basis of the methods presented, it is to be hoped that they will also prove useful in other fields of natural science and statistics.

References

Akaike H., 1974, IEEE Transactions on Automatic Control, 19, 716

Akopian A. A., 2003, Astrophysics, 46, 58

Akopian A. A., Parsamian E. S., 2009, Astrophysics, 52, 322

Ambartsumian V. A., 1969, Stars, Nebulae, and Galaxies. Publishing House of the Armenian Academy of Sciences

Ambartsumian V. A., 1978, Astrophysics, 14, 209

Bohning D., Suppawattanabodee B., Kusolvisitkul W., Viwatwongkasem C., 2008, Journal of Statistical Planning and Inference, 138, 1017

Chao A., 1984, Scandinavian Journal of Statistics, 11, 265

Nelder J. A., Mead R., 1965, Computer Journal, 7, 308

Pearson K., 1895, Philosophical Transactions of the Royal Society of London Series A, 186, 343

Schwarz G., 1978, The Annals of Statistics, 6, 461

Talbot A., 1979, IMA Journal of Applied Mathematics, 23, 97

Yang H., Liu J., Qiao E., Zhang H., Gao Q., Zhang Y., 2019, The Astrophysical Journal Supplement Series, 241, 29

doi: https://doi.org/10.52526/25792776-26.73.1-39