Published in final edited form as:

Am Stat. 2016 ; 70(4): 413–423. doi:10.1080/00031305.2016.1200484.

Sojourning with the Homogeneous Poisson Process

Piaomu Liu[*] and **Edsel A. Peña**[†]

[*]PhD student, Department of Statistics, University of South Carolina, Columbia, SC 29208. liu256@email.sc.edu

[†]Professor, Department of Statistics, University of South Carolina, Columbia, SC 29208. pena@stat.sc.edu

Abstract

In this pedagogical article, distributional properties, some surprising, pertaining to the homogeneous Poisson process (HPP), when observed over a possibly random window, are presented. Properties of the gap-time that covered the termination time and the correlations among gap-times of the observed events are obtained. Inference procedures, such as estimation and model validation, based on event occurrence data over the observation window, are also presented. We envision that through the results in this paper, a better appreciation of the subtleties involved in the modeling and analysis of recurrent events data will ensue, since the HPP is arguably one of the simplest among recurrent event models. In addition, the use of the theorem of total probability, Bayes theorem, the iterated rules of expectation, variance and covariance, and the renewal equation could be illustrative when teaching distribution theory, mathematical statistics, and stochastic processes at both the undergraduate and graduate levels. This article is targeted towards both instructors and students.

Keywords

California earthquakes; Iterated Expectation, Variance, and Covariance Rules; HPP Model Validation; Normalized Spacings Statistics; Renewal Process; Renewal Function; Size-Biased Sampling; Sum-Quota Accrual Scheme; Teaching Statistics

1 Introduction and Motivation

The 2015 blockbuster movie *San Andreas* (Peyton (2015)) about a catastrophic earthquake (including its foreshocks and aftershocks) hitting the State of California has brought fear, fascination, and curiosity to many people, since apparently it is not a question of *if* but rather of *when* a catastrophic earthquake will hit California. The occurrence of earthquakes is one of those natural phenomena that we, humans, have no control over at all (cf., Hough (2010)), but it instills numerous beliefs, fantasies, and facts as evidenced in the United States Geological Survey (USGS) website:

http://earthquake.usgs.gov/learn/topics/megaqk_facts_fantasy.php.

Correspondence to: Edsel A. Peña.

Among US states, California is usually the main focus when the subject of earthquakes arises since beneath its beautiful landscape are some of the major faults (e.g., San Andreas, San Jacinto, Elsinore, and Imperial), and it is adjacent to the so-called Ring of Fire or the circum-Pacific belt to which the San Andreas fault belongs. Figure 1 is a plot of the occurrence times (in number of days since January 9, 1857) of the *mainshocks* of earthquakes with Richter magnitude at least 4.9 in California from January 9, 1857 until December 31, 2015. The full data set can be found in the link (Wikipedia Contributors (2016)):

https://en.wikipedia.org/wiki/List_of_earthquakes_in_California.

Only earthquakes after January 9, 1857, which is the date of the Fort Tejon earthquake, the strongest recorded earthquake in California, were included. *Doublets, swarms*, and *triggered* occurrences were excluded since mainshocks are of primary interest in our modeling. `EvTimes`, the variable denoting the number of days since January 9, 1857 of occurrences of mainshocks, and `GapTimes`, denoting the number of days between successive occurrences of mainshocks, are presented in Table 1. Note that the exact time of occurrence, denoted by the Pacific Time Zone (`PTZ`) variable in the `Wiki` page, was not used in determining `EvTimes` and `GapTimes` in Table 1. However, we ascertained that the impact of incorporating the `PTZ` information in the statistical analysis performed in Section 4 was negligible.

The occurrence of earthquake is an example of a recurrent event. Such events occur in many settings. Examples of recurrent events with a negative flavor are mass shooting, terrorist attack, re-occurrence of a tumor, machine failure, economic recession, non-life insurance claim, and Dow Jones Industrial Average Index (DJIA) decreasing by at least 5% on a trading day. On the other hand, recurrent events with a positive flavor are publication of a paper by a researcher, citing of a journal article, acquiring a new job, and DJIA increasing by at least 5% on a trading day. The probabilistic modeling and statistical analysis of recurrent events has been a highly active area of scholarly research. There are well-known recurrent event data sets that have been used in the literature such as the air-conditioning failure data set in Proschan (1963) (see also Follmann and Goldberg (1988) which uses this data set), the migratory motor complex (MMC) data set in Aalen and Husebye (1991), and the bladder tumor data set used in Wei et al. (1989). See the book by Cook and Lawless (2007) for other examples of recurrent events and for other real-life recurrent event data sets.

The earthquake data set presented in Figure 1 is the type of real and interesting data that instructors of probability, stochastic process, or mathematical statistics courses could use when introducing stochastic processes to undergraduate and graduate students. Specifically, such a data set could be used to motivate the first and the most basic stochastic process that students are exposed to, namely, the Homogeneous Poisson Process (HPP). See, for instance, the review article by Anagnos and Kiremidjian (1988) which discusses the use of the HPP model as well as other stochastic models for seismic hazard analysis. We recall the axiomatic definition of an HPP in Definition 1 (see Ross (1984); Resnick (1992)).

Deftnition 1 *A homogeneous Poisson process with rate* λ, *shortened HPP*(λ), *is a family of random variables* $\mathbf{N} = \{N(s),\ s \geq 0\}$ *on some probability space* $(\Omega, \mathfrak{F}, \mathbf{P})$ *satisfying the following conditions, where for every* $s, t \geq 0$, $N(s, s + t] = N(s + t) - N(s)$:

- **i.** $\mathbf{P}\{N(0) = 0\} = 1$;
- **ii.** *For* $0 \leq s_1 < s_2 \leq s_3 < s_4$, $N(s_1, s_2]$ *and* $N(s_3, s_4]$ *are independent;*
- **iii.** *For every s and* $ds > 0$, *as* $ds \downarrow 0$, $\mathbf{P}\{N(s, s + ds] \geq 1\} = \lambda ds + o(ds)$; *and*
- **iv.** *For every s and* $ds > 0$, *as* $ds \downarrow 0$, $\mathbf{P}\{N(s, s + ds] \geq 2\} = o(ds)$.

The practical interpretation is that $N(s)$ denotes the number of occurrences of a recurrent event (e.g., earthquakes) during the period $(0, s]$, so $N(s, s + t]$ is the number of occurrences in $(s, s + t]$. The "little oh" notation means $o(ds)/ds \to 0$ as $ds \downarrow 0$. It is well-known and always an instructive exercise for students to show, by setting appropriate differential equations using the properties in Definition 1, that $N(s, s + t]$ has a Poisson distribution with mean λt, that is,

$$\mathbf{P}\{N(s,s+t]=j\}=\frac{\exp(-\lambda t)(\lambda t)^j}{j!}, j=0,1,2,\ldots.$$

For the earthquake data, the process was observed from January 9, 1857, our time origin, until December 31, 2015. Thus, when monitoring such processes, there will usually be an end to the monitoring period, denoted by τ. In the earthquake data, τ was fixed to be $\tau =$ 58064, the number of days from 1/9/1857 to 12/31/2015. More generally, this τ may be random and it could represent the time to an event which causes cessation of the monitoring of the event. For instance, in order to relate, catch, and spark the interest of our young and starry-eyed students, the instructor could mention, tongue-in-cheek, that τ could be the time that the First Order's Starkiller Base superweapon annihilates a planet as in the movie *Star Wars: The Force Awakens* (Abrams (2015)). On the other hand, τ could represent the time to a relatively more benign event, such as an imposed deadline for a Statistics doctoral student to complete her dissertation research.

Associated with the process $\{N(s),\ s \geq 0\}$, we also define equivalent random variables. The time of the kth event occurrence will be denoted by S_k, so that we have $0 \equiv S_0 < S_1 < S_2 < S_3 < \ldots$. The inter-event times, also called gap-times, are $T_1, T_2, T_3, \ldots$, so that with $S_0 = 0$ and for each $k = 1, 2, \ldots$,

$$T_k=S_k - S_{k-1} \quad \text{and} \quad S_k=\sum_{j=1}^{k} T_j.$$

For the earthquake data, the S_k's and T_k's are the `EvTimes` and `GapTimes` respectively in Table 1. The probabilistic properties of the process $\{N(s),\ s \geq 0\}$ can be equivalently specified via the distribution of the family of event time random variables $\{S_k : k = 0, 1, 2, \ldots\}$ or the distribution of the family of gap-times $\{T_k : k = 1, 2, 3, \ldots\}$. Such distributional properties may involve other aspects of the observation process, such as environmental variables that could impact the event occurrences (cf., Peña (2006); Lindqvist (2006); Ross (2009)). However, in this paper, we restrict focus to the HPP(λ).

Under the HPP(λ), it is an excellent exercise for the students to show from Definition 1 that the gap-times T_k's are independent and identically distributed (IID) random variables with common exponential distribution with mean $1/\lambda$, denoted by Exp(λ). Thus the common probability density function (PDF) of the T_k's is $f_T(t) = \lambda \exp(-\lambda t)$ for $t \geq 0$. It follows, for $k \geq 1$, that the marginal distribution of $S_k = \sum_{j=1}^{k} T_j$ is a gamma distribution with shape parameter k and rate parameter λ, so its PDF is

$$f_{S_k}(s) = \frac{\lambda^k}{\Gamma(k)} s^{k-1} \exp(-\lambda s) I\{s \geq 0\},$$

where $\Gamma(\alpha) = \int_0^\infty w^{\alpha-1} \exp(-w) \mathrm{dw}$ for $\alpha > 0$ is the gamma function, and $I\{\cdot\}$ is the indicator function. However, the S_k's are not independent random variables.

As has been pointed out for the earthquake data, the process $\{N(s), s \geq 0\}$ will only be observed over a possibly random period $[0, \tau]$. We assume that the distribution of τ is G, possibly degenerate as in the earthquake data, and τ is independent of the T_k's. During this monitoring period $[0, \tau]$, the random variable which denotes the number of *observed* events is

$$K \equiv K(\tau, \mathbf{T}) = N(\tau) = \max\{k \in \{0, 1, 2, ...\} : S_k \leq \tau\} = \max\left\{k \in \{0, 1, 2, \ldots\} : \sum_{i=1}^{k} T_i \leq \tau\right\}, \quad (1)$$

where $\mathbf{T} = (T_1, T_2, T_3, \ldots)$ is the sequence of gap-times. The gap-time T_{K+1} between the observed Kth and unobserved $(K+1)$th events, which covers or straddles τ, is unobserved and is right-censored by $\tau - S_K$. For the earthquake data where $K = 51$, this right-censoring value equals $\tau - S_{51} = 58064 - 57570 = 494$ days, so we only know that the gap-time between the 51st and the 52nd earthquakes exceeds 494 days.

In this pedagogical paper, which is targeted to both instructors and students, we demonstrate seemingly paradoxical results under the HPP(λ) model. We also address the statistical inference questions of how to estimate λ and how to assess if the HPP(λ) model is suitable based on the observed occurrence data during the monitoring period $[0, \tau]$. In particular, we shall address these questions with respect to the California earthquake occurrences using the observed data during the period from 1/9/1857 to 12/31/2015 as presented in Figure 1 and Table 1. To whet the appetite of the reader, at this point we pose the following questions: assume an HPP(λ) is monitored over $[0, \tau]$ and assume that τ, which has an exponential distribution, is independent of the gap-times T_j's. Let K be as defined in (1).

- **Question 1:** If you are informed that $K = k$ with $k \geq 2$, what do *you* think is the sign of the conditional correlation coefficient between T_1 and T_2?
- **Question 2:** *On average,* is the length of T_{K+1}, the gap-time straddling τ, equal to $1/\lambda$, the mean of T_1? If *you* think not, do *you* think it is larger or smaller than $1/\lambda$?

These are just two of the questions addressed in this paper. We reveal some surprising distributional results and demonstrate how well-known results such as the theorem of total probability, Bayes theorem, iterated expectation, variance, and covariance rules, together with renewal equation arguments can be utilized to answer such questions. We expect this paper to further establish the pedagogical utility of HPP-based models. The technical level of this paper is accessible to advanced undergraduate and beginning graduate students who are taking or have taken distribution theory, mathematical statistics, or stochastic process courses. Some of the results will demonstrate that some seemingly-intuitive statements are in fact mathematically false. This will further help enhance the allure and excitement of the study of probability and statistics among undergraduate and graduate students.

We outline the contents of this paper. Section 2 reviews some well-known properties of the HPP. Section 3 deals with an HPP that is observed over a random monitoring period. The impact of the sum-quota accrual scheme, equivalently of size-biased sampling, is demonstrated in this section. In particular, we examine the distributional properties of T_{K+1}. Section 4 discusses the estimation of λ and the assessment of the validity of the HPP model using data over the monitoring period. Section 5 provides some concluding remarks.

2 Some Properties of the HPP

In this section we recall some well-known properties of the HPP(λ), which are needed in later sections. For the proofs of these results, we refer the reader to Ross (1984), Resnick (1992), Ross (2009), or other probability and stochastic process books. Better still is for the curious and challenged reader to establish these results!

Theorem 1 *For an HPP(λ), (S_1, S_2, …, S_k), given that $N(s) = k$ with $k \geq 1$ and s fixed, has the joint distribution of the order statistics of a random sample of size k from a uniform distribution over* $[0, s]$*. As a consequence* $(V_1, V_2, \ldots, V_k) \equiv \frac{1}{s}(S_1, S_2, \ldots, S_k)$*, given $N(s) =$ k with $k \geq 1$, has joint density given by*

$$f_{(V_1,V_2,\ldots,V_k)|N(s)}(v_1, v_2, \ldots, v_k|k)=k!I\{0 \leq v_1 \leq v_2 \leq \ldots \leq v_k \leq 1\}.$$

Corollary 1 *For an HPP(λ),* $(W_1, W_2, \ldots, W_k) \equiv \frac{1}{s}(T_1, T_2, \ldots, T_k)$*, given that $N(s) = k$ with $k \geq 1$ and s fixed, has a Dirichlet distribution with parameter* $\mathbf{1}_{k+1}$*, a* $1 \times (k + 1)$ *vector of* 1*'s, so that its joint density function is*

$$f_{(W_1, W_2, \ldots, W_k)|N(s)}(w_1, w_2, \ldots, w_k|k) = k! I\left\{w_i \in [0, 1]: \sum_{i=1}^{k} w_i \le 1\right\}.$$

Thus, for $i, j \in \{1, 2, \ldots, k\}$ *with* $i \ne j$, $W_i | (N(s) = k)$ *has a beta (Dirichlet) distribution with parameter* $(1, k)$, *while* $(W_i, W_j) | (N(s) = k)$ *has a Dirichlet distribution with parameter* $(1, 1, k-1)$.

These results possess the intuitive content that under an HPP model, the occurrences of a known number of events over a fixed finite interval in the time axis are according to a *uniform* law. Using well-known results about the moments of the Dirichlet distribution (see, for instance, Kotz et al. (2000)), we have:

Corollary 2 *For a HPP*(λ), *when s is fixed, we have for* $i, j \in \{1, 2, \ldots, k\}$ *with* $i \ne j$ *and* $k \ge 1$, *(i)* $\mathbf{E}[T_i|N(s)=k] = s\frac{1}{k+1}$ *; (ii)* $\mathbf{Var}[T_i|N(s)=k] = s^2\frac{k}{(k+2)(k+1)^2}$ *; (iii)* $\mathbf{Cov}[T_i, T_j|N(s)=k] = -s^2\frac{1}{(k+2)(k+1)^2}$ *;and (iv)* $\mathbf{Cor}[T_i, T_j|N(s)=k] = -\frac{1}{k}$.

3 HPP over a Random Window

We now consider the situation when an HPP(λ) is monitored over a possibly random window $[0, \tau]$. Results will be obtained for the case where τ is fixed, and for the case where τ is independent of the process and has an exponential distribution with mean $1/\eta$. The latter setting will be referred to as a HPP with a random window, abbreviated HPPRW(λ, η). This is a special case of the model considered in Peña et al. (2001). We point out that it is not actually necessary to have the exponential distribution assumption on τ; however, imposing this condition leads to tractable expressions which help in the pedagogical aims of this paper. The random number of event occurrences observed over $[0, \tau]$ is $K = K(\tau, \mathbf{T})$ as defined in (1). We emphasize that K is both a function of $\mathbf{T} = (T_1, T_2, T_3, \ldots)$ and τ. From its definition we also obtain the inequality, called the sum-quota constraint, given by

$$S_K \le \tau < S_K + T_{K+1}. \quad (2)$$

The random observable over the monitoring period $[0, \tau]$ is therefore

$$\mathbf{D} = (\tau, K, T_1^*, T_2^*, \ldots, T_k^*, \tau - S_k), \quad (3)$$

where $T_j^* = T_j I\{j \le K\}$ for $j = 1, 2, 3, \ldots$. At this point we call attention to a change in notation compared to earlier works (e.g., Peña et al. (2001)) dealing with recurrent events where we now use a superscript asterisk on the T_j^*'s to indicate that the observable random variables T_j^*'s are dependent on the τ and the sequence $\mathbf{T}$ through the random variable K. This will make the situation more notationally appropriate and avoid some misconceptions.

Our first result relates to the notion of size-biased sampling. For any *fixed* $k \in \{0, 1, 2, \ldots\}$, we know that T_{K+1}, given τ, has an Exp(λ) distribution, so in particular, $\mathbf{E}(T_{K+1}|\tau) = 1/\lambda$. How about T_{K+1}, given τ, where T_{K+1} is the gap-time that covers or straddles τ? Examination of this random variable, which is always right-censored by $\tau - S_K$, is of importance both from a distributional viewpoint and statistical inference perspective as will be demonstrated in subsection 4.1 dealing with estimation. This innocent-looking random variable T_{K+1} is actually a complicated one since it depends on both τ and $\mathbf{T}$ through K. To see this, observe that

$$T_{K+1} \equiv T_{K(\tau,\mathbf{T})+1} = \sum_{j=1}^{\infty} T_j I\left\{\sum_{i=1}^{j-1} T_i \le \tau < \sum_{i=1}^{j} T_i\right\} \quad (4)$$

with the convention that an empty summation equals zero. In the sequel, at the cost of sacrificing some mathematical rigor, we do not mention measurability issues so as not to detract from the more elementary and pedagogical foci of this paper. When dealing with HPP's, a very important notion is that of *regeneration* (see, for instance, subsection 3.7.1 of Resnick (1992)). For an infinite sequence $\mathbf{t} = (t_1, t_2, t_3, \ldots)$, let $\mathbf{t}[-1] = (t_2, t_3, t_4, \ldots)$. For our notation, recall that two random vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ are said to be stochastically equal or are equal-in-distribution if, for every $\mathbf{v}$, $F_{\mathbf{V}_1}(\mathbf{v}) = F_{\mathbf{V}_2}(\mathbf{v})$. In such a case we write $\mathbf{V}_1 \stackrel{\text{st}}{=} \mathbf{V}_2$ or $\mathbf{V}_1 \stackrel{d}{=} \mathbf{V}_2$. Now, for an HPP($\lambda$), its associated gap-times $\mathbf{T} = (T_1, T_2, T_3, \ldots)$, which are IID Exp($\lambda$), possess the *regenerative* property

$$\mathbf{T} \stackrel{d}{=} \mathbf{T}[-1], \quad (5)$$

so that for any mapping $l : \Re_+^{\infty} \to \mathscr{L}$ where $\mathscr{L}$ is some space, it follows that

$$l(\mathbf{T}) \stackrel{d}{=} l(\mathbf{T}[-1]). \quad (6)$$

Lemma 1 *For an HPP*(λ) *with gap-times* $\mathbf{T} = (T_1, T_2, T_2, \ldots)$, *we have that:*

i. *if* $T_1 = v$ *with* $v > \tau$, *then* $T_{K(\tau,\mathbf{T})+1} = T_1$; *while*

ii. *if* $T_1 = v$ *with* $0 \le v \le \tau$, *then* $T_{K(\tau,\mathbf{T})+1} \stackrel{d}{=} T_{K(\tau-v,\mathbf{T})+1}$.

Proof: Result (i) is obvious since if $T_1 = v > \tau$ then $K(\tau, \mathbf{T}) = 0$ so $T_{K(\tau,\mathbf{T})+1} = T_1$. To prove (ii) where $T_1 = v \le \tau$, we have by using the representation in (4) the following relations, where we also use the notation $\mathbf{T}^\dagger \equiv (T_1^\dagger, T_2^\dagger, T_3^\dagger, \ldots) = (T_2, T_3, T_4, \ldots) \equiv \mathbf{T}[-1]$:

$$
\begin{aligned}
T_{K(\tau,\mathbf{T})+1} &= \sum_{j=1}^{\infty} T_j I\left\{\sum_{i=1}^{j-1} T_i \leq \tau < \sum_{i=1}^{j} T_i\right\} = \sum_{j=1}^{\infty} T_j I\left\{v + \sum_{i=2}^{j-1} T_i \leq \tau < v + \sum_{i=2}^{j} T_i\right\} \\
&= \sum_{j=2}^{\infty} T_j I\left\{\sum_{i=2}^{j-1} T_i \leq \tau - v < \sum_{i=2}^{j} T_i\right\} = \sum_{l=1}^{\infty} T_{l+1} I\left\{\sum_{m=1}^{l-1} T_{m+1} \leq \tau - v < \sum_{m=1}^{l} T_{m+1}\right\} \\
&= T^{\dagger}_{K(\tau-v,\mathbf{T}^{\dagger})+1} \stackrel{d}{=} T_{K(\tau-v,\mathbf{T})+1}
\end{aligned}
$$

with the last equality arising from the regeneration properties (5) and (6).

Another important mathematical notion is that of convolution of two functions. Let F_1 and F_2 be two nonnegative and nondecreasing functions on $\Re_+$. The convolution of F_1 and F_2 is the function $F_1 * F_2$ on $\Re_+$ defined according to

$$(F_1 * F_2)(t) = \int_0^t F_2(t-v) F_1(dv), t \in \Re_+.$$

If F_i is the distribution function of the positive-valued random variable V_i for $i = 1, 2$, and if the V_i's are independent, then $F_1 * F_2$ is the distribution function of the sum $V = V_1 + V_2$ since, for $t \geq 0$,

$$
\begin{aligned}
\mathbf{P}\{V \leq t\} &= \mathbf{P}\{V_1 + V_2 \leq t\} = \int_0^t \mathbf{P}\{V_1 + V_2 \leq t | V_1 = v\} F_1(dv) \\
&= \int_0^t \mathbf{P}\{V_2 \leq t - v | V_1 = v\} F_1(dv) = \int_0^t F_2(t-v) F_1(dv) = (F_1 * F_2)(t),
\end{aligned}
$$

where we use the independence between V_1 and V_2 to get the second-to-last equality. If $F_1 = F_2 = F$, we write $F^{*2} = F * F$. It is a nice and simple exercise for the student to verify the associative property of the convolution operator, that is, for F_1, F_2, F_3 we have the identity $[(F_1 * F_2) * F_3] = [F_1 * (F_2 * F_3)]$. As such, if $F_1 = F_2 = \ldots = F_n = F$, it is appropriate to write $F^{*n} = F * F * \cdots * F$. We have mentioned earlier that for an HPP(λ), the distribution of $S_n = \sum_{j=1}^{n} T_j$ is a gamma distribution with shape parameter n and rate parameter λ, which is the nth convolution of the exponential distribution with parameter λ. By convention, $F^{*0}(t) = I\{t \geq 0\}$, the degenerate distribution at zero. Finally, for a distribution function F, its associated renewal function is defined to be

$$U(t) \equiv \sum_{n=0}^{\infty} F^{*n}(t). \quad (7)$$

It will be a simple, but instructive, exercise for the student to verify that for the HPP(λ), for which the gap-time distribution F is exponential with parameter λ, the renewal function is

$$U(t;\lambda) = [1 + \lambda t] I\{t \geq 0\}. \quad (8)$$

A practical interpretation of $U(t; \lambda) - 1$ is that it is the mean number of events that occur in $(0, t]$. See also Cox (1962) for a treatment of renewal theory.

We are now in a position to present and prove Theorem 2, which contains a general result concerning the mean of a function of T_{K+1}, conditional on τ. We opted for this more general result since it will enable us to obtain other results by simply specializing the function h. Furthermore, this will illustrate the power and utility of renewal or regenerative arguments.

Theorem 2 *For an HPP*(λ) *monitored over* $[0, \tau]$ *with* τ *fixed, let* $h : \Re_+ \to \Re$ *be a function with* $\mathbf{E}[|h(T_1)|] < \infty$ *and define the mapping* $\tau \mapsto A(\tau) \equiv A_h(\tau) \equiv \mathbf{E}[h(T_{K(\tau,\mathbf{T})+1})|\tau]$. *Then*

$$A(\tau)=\exp\{-\lambda\tau\}\mathbf{E}[h(T_1+\tau)|\tau]+\int_0^\tau\mathbf{E}[h(T_1+v)]\lambda\exp\{-\lambda v\}dv. \quad (9)$$

Proof: We first establish that when $\mathbf{E}[|h(T_1)|] < \infty$, then $\mathbf{E}[|h(T_{K(\tau,\mathbf{T})+1})||\tau] < \infty$. To prove this, we have by using (4), the iterated expectation rule, and, with $g(\cdot; j, \lambda)$ being the gamma density function with parameters (j, λ), that

$$A(\tau)=\mathbf{E}[|h(T_{K+1})||\tau]=\sum_{j=1}^{\infty}\mathbf{E}\{|h(T_j)|I\{S_{j-1}\le\tau<S_{j-1}+T_j\}|\tau\}$$
$$=\sum_{j=1}^{\infty}\int_0^\tau\mathbf{E}[|h(T_j)|I\{T_j>\tau-v\}]g(v;j-1,\lambda)dv\le\mathbf{E}[|h(T_1)|]U(\tau;\lambda)=\mathbf{E}[|h(T_1)|](1+\lambda\tau)<\infty.$$

Our next step is to establish a renewal equation for $A(\tau)$. By conditioning on T_1, we obtain using the theorem of total probability or the iterated expectation rule that

$$A(\tau)=\int_0^\tau\mathbf{E}[h(T_{K(\tau,\mathbf{T})+1})|\tau,T_1=v]F(dv;\lambda)+\int_\tau^\infty\mathbf{E}[h(T_{K(\tau,\mathbf{T})+1})|\tau,T_1=v]F(dv;\lambda),$$

where $F(\cdot) = F(\cdot; \lambda) = Exp(\lambda)$ is the distribution function of T_1 so that $F(dv; \lambda) = \lambda \exp(-\lambda v)dv$. Using Lemma 1 and since T_1 is independent of $\mathbf{T}[-1]$, we obtain from the preceding equation

$$\begin{aligned}A(\tau)&=\int_0^\tau\mathbf{E}[h(T_{K(\tau-v,\mathbf{T})+1})|\tau]F(dv;\lambda)+\int_\tau^\infty h(v)F(dv;\lambda)\\&=\int_0^\tau A(\tau-v)\lambda\exp\{-\lambda v\}dv+\int_\tau^\infty h(v)\lambda\exp\{-\lambda v\}dv\\&=\int_0^\infty h(v+\tau)\lambda\exp\{-\lambda(v+\tau)\}dv+\int_0^\tau A(\tau-v)\lambda\exp\{-\lambda v\}dv\\&=z(\tau)+(F*A)(\tau)\end{aligned}$$

with $z(\tau) = \exp\{-\lambda\tau\}\mathbf{E}[h(T_1 + \tau)|\tau]$. Thus, we get the renewal equation $A = z + F * A$. At this point we may simply invoke a theorem pertaining to renewal equations such as that in Resnick (1992) to obtain the solution for $A(\tau)$. However, to fulfill the pedagogical nature of this paper, we provide a brief, though not fully rigorous, derivation. First, note that by successive convolution with F and using the associative property of the convolution operator, we get the sequence of equations

$$\begin{aligned} A &= z+F*A \\ F*A &= F*z+F^{*2}*A \\ F^{*2}*A &= F^{*2}*z+F^{*3}*A \\ &\vdots \\ F^{*n}*A &= F^{*n}*z+F^{*(n+1)}*A \\ &\vdots \end{aligned}$$

From this sequence, we obtain for each $n \geq 1$ that $A=\left(\sum_{j=0}^{n} F^{*j}\right)*z+F^{*(n+1)}*A$. Letting $n \to \infty$, and noting that, for $\tau < \infty$, $F^{*(n+1)}(\tau) = \mathbf{P}\{S_{n+1} \leq \tau\} \to 0$ since $S_{n+1} \xrightarrow{\text{pr}} \infty$ by the weak law of large numbers, then we get the solution $A(\tau)=(U*z)(\tau)=z(\tau)+\int_0^{\tau} z(\tau-v)\lambda \mathrm{dv}$, where U is the renewal function of F given in (8). Thus, it follows that $A(\tau)=\exp\{-\lambda\tau\}\mathbf{E}[h(T_1+\tau)|\tau]+\int_0^{\tau}\mathbf{E}[h(T_1+v)]\lambda\exp\{-\lambda v\}\mathrm{dv}$.

Theorem 3 *Under the conditions of Theorem 2, the conditional distribution of T_{K+1}, given τ, is*

$$\mathbf{P}\{T_{K+1} \leq w|\tau\}=\{1-\exp(-\lambda w)[1+\lambda \min(w,\tau)]\}I\{w \geq 0\}.$$

Its conditional mean and variance are, respectively,

$$\mathbf{E}[T_{K+1}|\tau]=\frac{1}{\lambda}\{1+[1-\exp(-\lambda\tau)]\};\ \mathbf{Var}[T_{K+1}|\tau]=\frac{1}{\lambda^2}\{1+[1-\exp(-2\lambda\tau)]-2(\lambda\tau)\exp(-\lambda\tau)\}.$$

As a consequence, for each $\tau > 0$ and $w > 0$, $\mathbf{P}\{T_{K+1} \leq w|\tau\} < \mathbf{P}\{T_1 \leq w\}$, $\mathbf{E}[T_{K+1}|\tau] > \mathbf{E}[T_1]$, and $\mathbf{Var}[T_{K+1}|\tau] > \mathbf{Var}[T_1]$. In addition,

$$\lim_{\tau\downarrow 0}\mathbf{E}[T_{K+1}|\tau]=\lim_{\tau\downarrow 0}\sqrt{\mathrm{Var}[T_{K+1}|\tau]}=\frac{1}{\lambda},\ \lim_{\tau\uparrow\infty}\mathbf{E}[T_{K+1}|\tau]=\frac{2}{\lambda},\ \text{and} \lim_{\tau\uparrow\infty}\sqrt{\mathrm{Var}[T_{K+1}|\tau]}=\frac{\sqrt{2}}{\lambda}.$$

Proof: The first result follows from Theorem 2 by taking $h(t) = I\{t > w\}$ to get $\mathbf{P}\{T_{K+1} > w| \tau\}$. To obtain the conditional mean, take $h(t) = t$ to get $\mathbf{E}[T_{K+1}|\tau]$, while to obtain the conditional variance first take $h(t) = t^2$ to obtain $\mathbf{E}[T_{K+1}^2|\tau]$. Straightforward simplifications lead to the expressions in the theorem. Clearly, for $w > 0$, $\mathbf{P}\{T_{K+1} \leq w|\tau\} < 1-\exp(-\lambda w) = \mathbf{P}\{T_1 \leq w\}$. Also, since $\mathbf{E}[T_1] = 1/\lambda$, then it is clear that $\mathbf{E}[T_{K+1}|\tau] > \mathbf{E}[T_1|\tau] = \mathbf{E}[T_1]$. In addition, we have that $\mathbf{Var}[T_1] = 1/\lambda^2$, so that since the mapping $t \mapsto g(t) = 1-\exp(-2t) - 2t\exp(-t)$ for $t \geq 0$ is increasing in t with $g(0) = 0$, then it follows that $\mathbf{Var}[T_{K+1}|\tau] > \mathbf{Var}[T_1|\tau] = \mathbf{Var}[T_1]$. The limiting results are immediate.

The next theorem pertains to results which are *not conditional* on τ, but under the HPPRW(λ, η) model wherein $\tau \sim Exp(\eta)$.

Theorem 4 *Under a HPPRW(λ, η) model, the unconditional distribution of T_{K+1} is*

$$\mathbf{P}\{T_{K+1} \le w\} = \left[1 - \exp(-\lambda w)\left\{1 + \frac{\lambda}{\eta}(1 - \exp(-\eta w))\right\}\right] I\{w \ge 0\};$$

and its unconditional mean and variance are, respectively,

$$\mathbf{E}[T_{K+1}] = \frac{1}{\lambda}\left[1 + (\frac{\lambda}{\lambda+\eta}\right) \quad \text{and} \quad \mathbf{Var}[T_{K+1}] = \frac{1}{\lambda^2}\left\{1 + \left(\frac{\lambda}{\lambda+\eta}\right)^2\right\}.$$

In addition, for $w > 0$, $\mathbf{P}\{T_{K+1} \le w\} < \mathbf{P}\{T_1 \le w\}$, $\mathbf{E}[T_{K+1}] > \mathbf{E}[T_1]$, *and* $\mathbf{Var}[T_{K+1}] > \mathbf{Var}[T_1]$.

Proof: The results follow from Theorem 3, the iterated expectation and variance rules, and straightforward simplifications. The stated inequalities are immediate.

Recall that for random variables X and Y, we say that X is stochastically larger than Y, denoted by $X \overset{\text{st}}{\geq} Y$, if for each $x \in \Re$, $\mathbf{P}\{X \le x\} \le \mathbf{P}\{Y \le x\}$. Thus, variables X and Y are stochastically equal, that is $X \overset{d}{=} Y$, if $X \overset{\text{st}}{\geq} Y$ and $Y \overset{\text{st}}{\geq} X$. From Theorem 3 we therefore see that, conditionally on $\tau, T_{K+1} \overset{\text{st}}{\geq} T_1$; while from Theorem 4 we see that this stochastic dominance of T_{K+1} over T_1 still holds *unconditionally* on τ. In layman's terms, this means that even though each of the gap-times T_{K+1}'s for $k \ge 0$ in an HPP(λ) are exponentially distributed with parameter λ, when we focus on the gap-time T_{K+1} that straddles or covers τ, with τ viewed as the fixed or random time of an inspection of the HPP, this gap-time tends to be longer than each of the T_{K+1}'s! This answers the second motivating question and appears to be a paradoxical and counter-intuitive result. This phenomenon arises because of the sum-quota constraint which makes T_{K+1} to become a function of τ and $\mathbf{T}$. This could also be explained by the so-called size-biased sampling phenomenon [also called the inspection paradox (cf., Ross (2009))] through the representation in (4). Imagine generating the infinite, but countable, event times from an HPP(λ), so that $0 \equiv s_0 < s_1 < s_2 < s_3 < \ldots$ are the realized event times. View these times as fixed, but transform each of them to $w_i = 1 - \exp(-\eta s_i)$, $i = 0, 1, 2, 3, \ldots$. Thus, $0 \equiv w_0 < w_1 < w_2 < w_3 < \ldots$ is a sequence in [0, 1]. Generate τ according to an Exp(η) distribution, and convert this to $V = 1 - \exp(-\eta\tau)$, so that V has a uniform distribution on [0, 1]. Then the probability that V is contained in the interval $[w_{i-1}, w_i)$ is $w_i - w_{i-1}$, so that longer intervals will have higher probabilities of catching V. This is equivalent to the longer intervals $[s_{i-1}, s_i)$'s having higher probabilities of covering τ, implying that the length of the gap-time that covers τ tends to be longer. This is one explanation of the size-biased sampling in this context. Another way to see why T_{K+1} tends to be longer than an Exp(λ) random variable is by examining the backward and forward recurrence times. Given τ, the backward and forward recurrence times are given, respectively, by $B(\tau) = \tau - S_K$ and $R(\tau) = S_{K+1} - \tau$. See, for instance, section 3.5 of Resnick (1992). T_{K+1} can be expressed as $T_{K+1} = B(\tau) + R(\tau)$. The results in Theorem 3 could then be alternatively obtained by utilizing the interesting distributional properties of $(B(\tau), R(\tau))$ contained in Theorem 5 below.

Theorem 5 *Under the conditions of Theorem 2, (i) $B(\tau)$ and $R(\tau)$ are stochastically independent; (ii) $R(\tau)$ is stochastically equal to T_1, so it has an exponential distribution with rate λ; and (ii) $B(\tau)$ is stochastically equal to* $\min(T_1, \tau)$, *so its distribution function is*

$$\mathbf{P}\{B(\tau) \le w|\tau\}=[1-\exp(-\lambda w)]I\{w<\tau\}+I\{w \ge \tau\}.$$

Proof: Fix $\tau > 0$, $w > 0$, and $v > 0$. Conditionally on τ, we have, with $g(\cdot; k, \lambda)$ being the gamma density function with shape parameter k and rate parameter λ,

$$\begin{aligned}
\mathbf{P}\{B(\tau)>w, R(\tau)>v|\tau\} &= \sum_{k=0}^{\infty}\mathbf{P}\{K=k, B(\tau)>w, R(\tau)>v|\tau\} \\
&= \sum_{k=0}^{\infty}\mathbf{P}\{S_k \le \tau<S_k+T_{k+1}, \tau-S_k>w, S_k+T_{k+1}-\tau>v|\tau\} \\
&= \mathbf{P}\{\tau<T_1, \tau>w, T_1-\tau>v|\tau\}+\sum_{k=1}^{\infty}\mathbf{P}\{S_k<\tau-w, T_{k+1}>v+\tau-S_k|\tau\} \\
&= I\{w<\tau\}\left\{\exp\{-\lambda(\tau+v)\}+\sum_{k=1}^{\infty}\int_0^{\tau-w}\mathbf{P}\{T_{k+1}>v+\tau-z|S_k=z, \tau\}g(z;k,\lambda)dz\right\} \\
&= I\{w<\tau\}\{\exp\{-\lambda(\tau+v)\}+\int_0^{\tau-w}\sum_{k=1}^{\infty}\exp\{-\lambda(v+\tau-z)\}\tfrac{\lambda^k}{\Gamma(k)}z^{k-1}\exp\{-\lambda z\}dz\right\} \\
&= I\{w<\tau\}\left\{\exp\{-\lambda(\tau+v)\}+\exp\{-\lambda(v+\tau)\}\lambda\int_0^{\tau-w}\sum_{k=1}^{\infty}\tfrac{(\lambda z)^{k-1}}{(k-1)!}dz\right\} \\
&= I\{w<\tau\}\exp\{-\lambda(\tau+v)\}(1+\exp(\lambda z)|_{z=0}^{z=\tau-w}) \\
&= I\{w<\tau\}\exp(-\lambda w)\exp(-\lambda v)
\end{aligned}$$

Note that the independence between T_{K+1} and S_k, given τ, is used to obtain the fifth equality. Also, observe that $B(\tau)$ is a *mixed* random variable with jump point at τ. All three results in the statement of the theorem now follow from the end-result that $\mathbf{P}\{B(\tau) > w, R(\tau) > v|\tau\} = I\{w < \tau\}\exp(-\lambda w)\exp(-\lambda v)$. We remark that a renewal argument-based proof is also available for establishing the results (cf., Resnick (1992)).

We examine the impact of τ on the conditional distribution of T_{K+1} and the impact of η on the unconditional distribution function of T_{K+1}. From Theorem 3, it is clear that $\mathbf{P}\{T_{K+1} \le w|\tau\}$ is decreasing in τ, indicating that T_{K+1}, given τ, is stochastically increasing in τ. Note that this stochastic ordering of the conditional distribution of T_{K+1}, given τ, as τ changes could also be deduced immediately from the results of Theorem 5. On the other hand, from Theorem 4, $\mathbf{P}\{T_{K+1} \le w\}$ is an increasing function (proof is left as an exercise for the student or the instructor) of η, indicating that T_{K+1} is stochastically decreasing in η, equivalently, stochastically increasing in $1/\eta$. These results are pictorially illustrated in Figure 2. Panel 1 in Figure 2 contains overlaid plots of $\mathbf{P}\{T_{K+1} \le w|\tau\}$ for $\lambda = 1$ and $\tau \in [0, 10]$ where we see that the distribution functions are decreasing in τ. The case with $\tau = 0$ corresponds to the distribution of T_1, which is exponential with parameter λ. Panel 2 in Figure 2 contains the distribution functions of T_{K+1} unconditionally on τ for different values of $\eta \in [.2, 10]$. Here we observe that as η increases, which shortens the monitoring period, the distribution functions increase. We also call attention to a subtle result that undergraduate students will usually *not* encounter. In the representation $T_{K+1} = B(\tau) + R(\tau)$, for fixed τ, $B(\tau)$ is a mixed random variable with a jump point at τ with jump probability $\exp(-\lambda\tau)$, whereas $R(\tau)$ is a continuous random variable. The sum of these mixed and continuous random variables, which is T_{K+1}, is now a purely continuous random variable, a manifestation of the notion of smoothing.

To help students visualize the aforementioned theoretical results, we could exploit the computing resources in our classrooms such as the ubiquitous `R` Statistical Computing platform (R Core Team (2013)). In this context it is enlightening and exciting to demonstrate to the students the agreement between theory and the results of computer experiments in real-time. To accomplish such a demonstration for the T_{K+1} variable, we may suppose that we have an HPP(λ = 1) and a τ having an exponential distribution with η = .2. Denoting by ξ_q the qth quantile of the distribution of τ, we have that $\xi_q = -\log(1-q)/\eta$, so we get $\xi_{.25}$ = 1.438, $\xi_{.5}$ = 3.466, and $\xi_{.75}$ = 6.931. For the demonstration, we can then do the following computer simulations, each with M = 5000 replications.

i. Generate M Exp(λ) random variates, representing the times to the first event in the HPP(λ).

ii. For fixed $\tau = \xi_{.25}$, generate M realizations of an HPP(λ) and determine the associated T_{K+1}. These M values represent sample realizations from the conditional distribution of T_{K+1}, given $\tau = \xi_{.25}$. Repeat the same experiment for $\tau = \xi_{.50}$ and $\tau = \xi_{.75}$.

iii. For each of M replications, generate a τ according to Exp(η) and then an HPP(λ) realization to determine T_{K+1}. These M values represent sample realizations from the unconditional distribution of T_{K+1}.

For the five sets of sample realizations, each with M values, construct the comparative boxplots to show the empirical realizations. For instance, we performed these simulations using simple `R` programs and the results are presented in Figure 3. Through such simulations the instructor is then able to empirically illustrate differences in distributional properties of T_1, T_{K+1} given τ, for different values of τ, and T_{K+1}, thereby demonstrating in a concrete manner the theoretical results. In particular, from our simulation runs, we note that the sample realizations from the conditional distributions of T_{K+1}, given τ, and from its unconditional distribution are consistent with the stochastic ordering obtained theoretically. Also note the increasing stochastic ordering of the conditional distributions of T_{K+1}, given τ, as τ increases, but that the unconditional distribution of T_{K+1} is not necessarily stochastically larger than the conditional ones. The `R` codes we used in our simulations are available upon request from the authors, though we strongly encourage students to create their own `R` simulation codes since it will improve their programming skills.

Next, we investigate properties of K and $(T_1^*, T_2^*, \ldots, T_K^*)$ under the HPPRW. Immediately, conditionally on τ, K has a Poisson distribution with rate $\lambda\tau$. Unconditionally on τ, it is also easy to show that K has a geometric distribution with success probability $\eta/(\lambda + \eta)$, hence $\mathbf{E}[K]=\frac{\lambda}{\eta}$ and $\mathbf{Var}[K]=\frac{\lambda}{\eta}\left[1+\frac{\lambda}{\eta}\right]$.

Theorem 6 *Under the HPPRW(λ, η) model,* $(T_1^*, T_2^*, \ldots, T_k^*)$, *given $K = k$ with $k \geq 1$, are IID from Exp($\lambda + \eta$).*

Proof: From Corollary 1, the conditional density of $(T_1^*, \ldots, T_k^*)$, given ($K = k$, τ), is

$$f_{(T_1^*,\ldots,T_k^*)|(K,\tau)}(t_1^*,\ldots,t_k^*|k,\tau)=\frac{k!}{\tau^k}I\left\{t_i^*\in[0,\tau],i=1,\ldots,k;\sum_{i=1}^k t_i^*\le\tau\right\}.$$

The conditional density function of τ, given $K = k$, satisfies, via Bayes Theorem,

$$f_{\tau|K}(\tau|k)\propto f_\tau(\tau)f_{K|\tau}(k|\tau)\propto\tau^{(k+1)-1}\exp\{-(\lambda+\eta)\tau\}.$$

Consequently, $\tau|(K = k)$ has a gamma distribution with shape parameter $k + 1$ and rate parameter $\lambda + \eta$. Invoking the theorem of total probability, the conditional joint density of $(T_1^*, \ldots, T_k^*)$, given $K = k$, is

$$\begin{aligned} &f_{(T_1^*,\ldots,T_k^*)|K}(t_1^*,\ldots,t_k^*|k)=\int_0^\infty f_{(T_1^*,\ldots,T_k^*)|(K,\tau)}(t_1^*,\ldots,t_k^*|k,\tau)f_{\tau|K}(\tau|k)d\tau\\ =&\int_{\sum_{i=1}^k t_i^*}^\infty\frac{k!}{\tau^k}\frac{(\lambda+\eta)^{k+1}}{\Gamma(k+1)}\tau^{(k+1)-1}\exp\{-(\lambda+\eta)\tau\}d\tau=(\lambda+\eta)^k\exp\{-(\lambda+\eta)\sum_{i=1}^k t_i^*\}\\ =&\prod_{i=1}^k[(\lambda+\eta)\exp\{-(\lambda+\eta)t_i^*\}I\{t_i^*\ge0\}].\end{aligned}$$

This completes the proof that, given $K=k, T_1^*, T_2^*, \ldots, T_k^*$ are IID Exp($\lambda + \eta$).

Corollary 3 *Under a HPPRW(λ, η) model, given $K = k$ with $k \ge 1$, then for $i \in \{1, 2, \ldots, k\}$, T_i^* has an Exp($\lambda + \eta$) distribution, so that $\mathbf{E}[T_i^*|K=k]=1/(\lambda+\eta)$ and $\mathbf{Var}[T_i^*|K=k]=1/(\lambda+\eta)^2$; and for $i, j \in \{1, 2, \ldots, k\}$ with $k \ge 2$ and $i \ne j$, $\mathbf{Cor}(T_i^*,T_j^*|K=k)=0$.*

In Corollary 3, the conditional moments and correlation of the T_i^*s are immediate from the joint conditional distribution of $(T_1^*, \ldots, T_k^*)$ given $K = k$. For pedagogical purposes, we may also derive them via the iterated expectation, variance, and covariance rules. We demonstrate this approach using the conditional covariance in Corollary 2. For a fixed $k \ge 2$, using the iterated expectation rule, we have for $i, j \in \{1, 2, \ldots, k\}$ with $i \ne j$,

$$\begin{aligned} &\mathbf{Cov}[T_i^*,T_j^*|K=k]\\ =&\mathbf{E}[\mathbf{Cov}[T_i^*,T_j^*|K=k,\tau]|K=k]+\mathbf{Cov}[\mathbf{E}[T_i^*|K=k,\tau],\mathbf{E}[T_j^*|K=k,\tau]|K=k]\\ =&\mathbf{E}\{\tau^2(\tfrac{-1}{(k+2)(k+1)^2})|K=k\}+\mathbf{Cov}\{\tau(\tfrac{1}{k+1}),\tau(\tfrac{1}{k+1})|K=k\}\\ =&(\tfrac{-1}{(k+2)(k+1)^2})[\mathbf{Var}(\tau|K=k)+(\mathbf{E}(\tau|K=k))^2]+(\tfrac{1}{k+1})^2(\tfrac{k+1}{(\lambda+\eta)^2})=0.\end{aligned}$$

It should be noted that the distributional result regarding $(T_1^*, T_2^*, \ldots, T_k^*)$, given $K = k$, contained in Theorem 6 is more general than just obtaining the moments via the iterated rules. Nevertheless, it is instructive to see the use of the iterated rules. An important advice to students who want to utilize the iterated rules when there is conditioning is *not* to forget the conditioning event in the outside expectation, variance, or covariance operations. In the above derivations, every operation is conditioned on the event $K = k$. Note that the last result of Corollary 3 answers the first motivating question.

4 Some Inference Issues

4.1 Estimating the Rate Parameter of the HPP

Assume that we have an HPP(λ) process. If we fix a $k \in \{1, 2, 3, \ldots\}$ and observe the gap-times $T_1, T_2, \ldots, T_k$, then since these are IID from Exp(λ), the likelihood function for λ is

$$L_0(\lambda | t_1, t_2, \ldots, t_k) = \lambda^k \exp\{-\lambda \sum_{i=1}^{k} t_i\}.$$

As a consequence, the maximum likelihood estimator (MLE) of λ is

$$\check{\lambda}(T_1, T_2, \ldots, T_k) = \frac{k}{\sum_{i=1}^{k} T_i} = \frac{k}{S_k} = \frac{1}{\overline{T}}.$$

On the other hand, when the HPP(λ) process is observed over the window $[0, \tau]$ with τ fixed, the number of events observed, $K = N(\tau)$, is a random variable which depends on λ. Specifically, K has a Poisson distribution with mean $\lambda\tau$. If we only know K and not the event times S_i's or the gap-times T_i's, then we could estimate λ via

$$\widehat{\lambda} = \frac{K}{\tau} = \frac{K}{\sum_{i=1}^{K} T_i + (\tau - S_K)},$$

which could be viewed as an occurrence-exposure rate. Could we improve on this estimator if we had observed the vector $\mathbf{D} = (\tau, K, T_1^*, T_2^*, \ldots, T_K^*, \tau - S_K)$, such as in the California earthquake data? Or, should we simply use the estimator

$$\stackrel{\approx}{\lambda} = \frac{K}{\sum_{i=1}^{K} T_i^*} = \frac{K}{S_K}$$

analogously to the case where k was fixed? Let us determine the likelihood function for λ given $\mathbf{D} = \mathbf{d} = (\tau, k, t_1^*, t_2^*, \ldots, t_k^*, \tau - s_k)$, where τ is fixed. Using the event equivalence that, for $(t_1^*, t_2^*, \ldots, t_k^*)$ with $s_k = \sum_{i=1}^{k} t_i^* \le \tau$, we have

$$(\tau, K = k, T_1^* = t_1^*, T_2^* = t_2^*, \ldots, T_K^* = t_k^*, \tau - S_K = \tau - \sum_{i=1}^{k} t_i^*) = (\tau, T_1 = t_1^*, T_2 = t_2^*, \ldots, T_k = t_k^*, T_{k+1} > \tau - \sum_{i=1}^{k} t_i^*)$$

then, since $T_1, T_2, \ldots, T_k, T_{k+1}, \ldots$ are IID Exp(λ), we obtain the likelihood function

$$L(\lambda|\mathrm{d})=\left[\prod_{i=1}^{k}\lambda\exp(-\lambda t_i^*)\right]\exp\left(-\lambda\left[\tau-\sum_{i=1}^{k}t_i^*\right]\right)=\lambda^k\exp(-\lambda\tau).$$

From this likelihood, it follows that the MLE of λ based on **D** is, also, $\hat{\lambda} = K/\tau$, which coincides with the estimator obtained when we only observe K. At this point, it would be an opportune time to ask students who had already taken mathematical statistics courses to *immediately* surmise why the MLE of λ based on data **D** *only* involved K. The instructor could then reveal (or confirm) that this is because of the Sufficiency Principle (cf., Wackerly et al. (2008)), and to leave it as an exercise for the students to show that when the observable is **D**, with τ fixed, the sufficient statistic for λ is in fact just K, hence it contains all the information about λ. Regarding the properties of the estimators, $\hat{\lambda} = K/\tau$ is the easier one to deal with since, for fixed τ, $K \sim \mathrm{POI}(\lambda\tau)$, hence we immediately see that $\mathbf{E}[\hat{\lambda}] = \lambda\tau/\tau = \lambda$ (it is unbiased) and $\mathbf{Var}[\hat{\lambda}] = \lambda/\tau$. The estimator $\tilde{\lambda}$ is harder to deal with since we require the joint distribution of (K, S_K). But alas, via the iterated rule of expectation, it can be shown that $\mathbf{E}[\tilde{\lambda}] = \infty$. This is a very surprising result, and it is an instructive exercise for the student to demonstrate that this in fact is the case. As such, the variance of $\tilde{\tilde{\lambda}}$ does *not* exist. Furthermore, the superiority of $\hat{\lambda}$ over $\tilde{\tilde{\lambda}}$ illustrates the danger of ignoring T_{K+1}, specifically its observed component $\tau - S_K$. This also highlights the peculiar behavior of T_{K+1} from the other observed gap-times, as we have demonstrated in Section 3.

We now go back to the California earthquake data. Suppose that an HPP(λ) model applies for the earthquake occurrences (but see subsection 4.2 which demonstrates that this is *not* actually the case), we would have obtained two estimates of λ from the above formulas given by

$$\tilde{\lambda}=\frac{51}{S_{51}}=\frac{51}{57570}=.00088588 \quad \text{and} \quad \hat{\lambda}=\frac{51}{\tau}=\frac{51}{58064}=.00087834.$$

The difference between these two estimates appears minuscule, but when we convert them to the mean of the gap-times between successive earthquakes, they become, respectively, 1128.824 days and 1138.51 days. In the context of earthquake occurrences, the difference between these mean estimates could be impactful to people as it could lead to heightened anxiety and anticipation of an earthquake.

In the preceding analysis we only dealt with estimation based on one observed process or unit. For more materials and discussions pertaining to statistical inference in these recurrent event models with many units, we refer students and instructors to Wang and Chang (1999), Peña et al. (2001), Peña (2006), Lindqvist (2006), Cook and Lawless (2007), Aalen et al. (2008), and Rahman et al. (2014).

4.2 Assessing Adequacy of the HPP Model

There is also the important question of assessing if an HPP model is appropriate in light of observed data over a window of observation. We are able to use the uniformity properties in Section 2 to develop a procedure for addressing this issue. We then use the procedure to see if an HPP model is suitable for California earthquake occurrences based on the data in Figure 1 and Table 1. Our procedure is based on the following result.

Theorem 7 *Let* $\{N(s),\ s\ \geq 0\}$ *be an HPP*(λ) *and assume that it has been observed over* $[0,\ \tau]$ *with* τ *known. Suppose* $K = k$ *events were observed and let* $S_0 = 0 < S_1 < S_2 < \ldots < S_k \leq \tau$ *be the event times. Then, conditional on*$(\tau,\ K) = (\tau,\ k)$*, the random variable* $V = -2\sum_{i=1}^{k} \log(1 - S_i/\tau)$ *has a chi-squared distribution with* $2k$ *degrees-of-freedom.*

Proof: All probability statements below are conditional on $(\tau,\ K) = (\tau,\ k)$ and under an HPP(λ) model. From Theorem 1, $(S_1, S_2, \ldots, S_k)/\tau \stackrel{d}{=} (U_{(1)}, U_{(2)}, \ldots, U_{(k)})$, where $(U_{(1)}, U_{(2)}, \ldots, U_{(k)})$ are the order statistics of size k from a standard uniform $U[0,\ 1]$. Since $U \sim U[0,\ 1]$ if and only if $W = -\log(1 - U) \sim Exp(1)$, then

$$(-\log(1-S_1/\tau), -\log(1-S_2/\tau), \ldots, -\log(1-S_k/\tau)) \stackrel{d}{=} (W_{(1)}, W_{(2)}, \ldots, W_{(k)}) \quad (10)$$

where $(W_{(1)},\ W_{(2)},\ \ldots,\ W_{(k)})$ are the order statistics of size k from a unit exponential distribution Exp(1). The associated normalized spacings statistics of the right-hand-side vector in (10) are

$$D_i = (k - i + 1)(W_{(i)} - W_{(i-1)}), i = 1, 2, \ldots, k,$$

with $W_{(0)} = 0$. It is well-known (cf., Barlow and Proschan (1975)), and is instructive for students to show, that $(D_1,\ D_2,\ \ldots,\ D_k)$ are IID from Exp(1). It then follows that $2\sum_{i=1}^{k} D_i$ has a chi-squared distribution with $2k$ degrees-of-freedom, abbreviated χ^2_{2k}, hence applying on the left-hand-side vector in (10) we get

$$V = 2\sum_{i=1}^{k} (k - i + 1)\{-\log(1 - S_i/\tau) + \log(1 - S_{i-1}/\tau)\} \sim \chi^2_{2k}.$$

Simplifying the expression for V, we obtain $V = -2\sum_{i=1}^{k} \log(1 - S_i/\tau)$, thus completing the proof.

Given the observed event occurrence data over $[0,\ \tau]$, we could test the null hypothesis (H_0) that the process is an HPP versus the alternative hypothesis (H_1) that the process is not an HPP, at level of significance α, using the decision rule which rejects H_0 whenever

$$V = -2\sum_{i=1}^{k} \log(1 - S_i/\tau) < \chi^2_{1-\alpha/2, 2k} \quad \text{or} \quad V > \chi^2_{\alpha/2, 2k},$$

where $\chi^2_{\alpha;2k}$ is the $100(1 - \alpha)$th quantile of χ^2_{2k}, that is, $\mathbf{P}\{\chi^2_{2k} \leq \chi^2_{\alpha;2k} = 1 - \alpha$. Observe, in particular, that when the gap-times are stochastically becoming shorter, which could be due to an increasing rate of event occurrences over time, the statistic V will tend to be large. So the test procedure will have good power towards such a departure from an HPP(λ).

For the California earthquake data, we have $k = 51$, $\tau = 58064$, and $V = 143.3053$. At 5% significance level, the critical values from a $\chi^2_{102}-\text{distribution}$ are $\chi^2_{.975;102}=75.9457$ and $\chi^2_{.025;102}=131.8375$, so that we reject the null hypothesis that the occurrences of earthquakes (of magnitude at least 4.9) follows an HPP! This result is informally indicated in Figure 1 where the gap-times in later years appear to be getting shorter. The formal test procedure performed provides an unequivocal test showing that the HPP model is not appropriate for modeling California earthquake occurrences of magnitude at least 4.9 in light of the observed data from 1/9/1857 to 12/31/2015. It would be of interest for geologists and seismologists to provide plausible explanations for this apparent departure, possibly with an increasing rate of earthquake occurrences over time, from an HPP model (see Anagnos and Kiremidjian (1988) for alternative models and chapter 18 of Hough (2010) for discussions of possible explanations).

5 Concluding Remarks

In this pedagogically-oriented paper, we examine the HPP as a model for the occurrences of a recurrent event over a possibly random monitoring window. Properties of the number of observed events in the monitoring period, the gap-time that covers the random termination time, and the observed gap-times were obtained. Some surprising properties, consequences of the sum-quota accrual scheme and size-biased sampling phenomenon, are highlighted and discussed. Procedures for estimating the HPP rate and for assessing the viability of the HPP model, based on the observed data, are also provided. The discussions and illustrations are aided by using data of California earthquake occurrences of Richter magnitude at least 4.9 during the period from January 9, 1857 (the day of the Fort Tejon earthquake) until December 31, 2015. The results and proofs in this paper could serve as excellent additions to topics covered in advanced undergraduate-level and beginning graduate-level distribution theory, mathematical statistics, and stochastic processes courses. In particular, the use of the theorem of total probability, the iterated expectation, variance, and covariance rules, Bayes theorem, and the renewal equation will be quite instructive in such courses.

Acknowledgments

This research was partially supported by NSF Grant DMS1106435 and NIH Grants R01CA154731 and P30GM103336-01A1. We are very grateful to the referees, associate editor, and the editor, Professor Nicole Lazar, for their thorough reading of the manuscript and their comments, suggestions, and criticisms which helped in considerably improving the paper. We also thank Professor James Lynch, Dr. AKM Fazlur Rahman, and graduate students Taeho Kim, Bereket Kindo, Beidi Qiang, Shiwen Shen, Jeff Thompson, Lillian Wanda, and Lu Wang for their comments, insights, and feedbacks on the paper.

References

Aalen O, Husebye E. Statistical analysis of repeated events forming renewal processes. Statistics in Medicine. 1991; 10:1227–1240. [PubMed: 1925154]

Aalen, OO., Borgan, Ø., Gjessing, HK. Survival and event history analysis, Statistics for Biology and Health. New York: Springer; 2008. A process point of view

Abrams, J. Star Wars: The Force Awakens. Walt Disney Studios, Film; 2015.

Anagnos T, Kiremidjian A. A review of earthquake occurrence models for seismic hazard analysis. Probabilistic Engineering Mechanics. 1988:3–11.

Barlow, RE., Proschan, F. Statistical theory of reliability and life testing. Holt, Rinehart and Winston, Inc.; New York-Montreal, Que.-London: 1975. Probability models, International Series in Decision Processes, Series in Quantitative Methods for Decision Making

Cook, RJ., Lawless, JF. The Statistical Analysis of Recurrent Events. New York: Springer; 2007.

Cox, DR. Renewal Theory. Methuen & Co. Ltd.; London: John Wiley & Sons, Inc.; New York: 1962.

Follmann DA, Goldberg MS. Distinguishing Heterogeneity from Decreasing Hazard Rates. Technometrics. 1988; 30:389–396.

Hough, S. Predicting the Unpredictable: The Tumultuous Science of Eathquake Prediction. Princeton: NJ: Princeton University Press; 2010.

Kotz S, Balakrishnan N, Johnson NL. Continuous multivariate distributions (Second). 2000; 1 Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York, Models and applications.

Lindqvist BH. On the statistical modeling and analysis of repairable systems. Statist. Sci. 2006; 21:532–551.

Peña EA. Dynamic modeling and statistical analysis of event times. Statist. Sci. 2006; 21:487–500.

Peña EA, Strawderman RL, Hollander M. Nonparametric estimation with recurrent event data. J. Amer. Statist. Assoc. 2001; 96:1299–1315.

Peyton, B. San Andreas. Warner Home Video, Film; 2015.

Proschan F. Theoretical explanation of observing decreasing failure rate. Technometrics. 1963; 5:375–383.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.

Rahman F, Lynch JD, Peña EA. Nonparametric Bayes estimation of gap-time distribution with recurrent event data. J. Nonparametr. Stat. 2014; 26:575–598.

Resnick, S. Adventures in Stochastic Processes. Boston, MA: Birkhäuser Boston Inc.; 1992.

Ross, S. A First Course in Probability. Second. New York: Macmillan Co.; 1984.

Ross, S. Introduction to Probability Models. 10. New York: Academic Press; 2009.

Wackerly, D., Mendenhall, W., III, Scheaffer, R. Mathematical Statistics with Applications. 7. Belmont, CA: Brooks/Cole Cengage Learning; 2008.

Wang M-C, Chang S-H. Nonparametric estimation of a recurrent survival function. J. Amer. Statist. Assoc. 1999; 94:146–153.

Wei L, Lin D, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J. Amer. Statist. Assoc. 1989; 84:1065–1073.

Wikipedia Contributors. List of Earthquakes in California. Wikipedia, The Free Encyclopedia. 2016

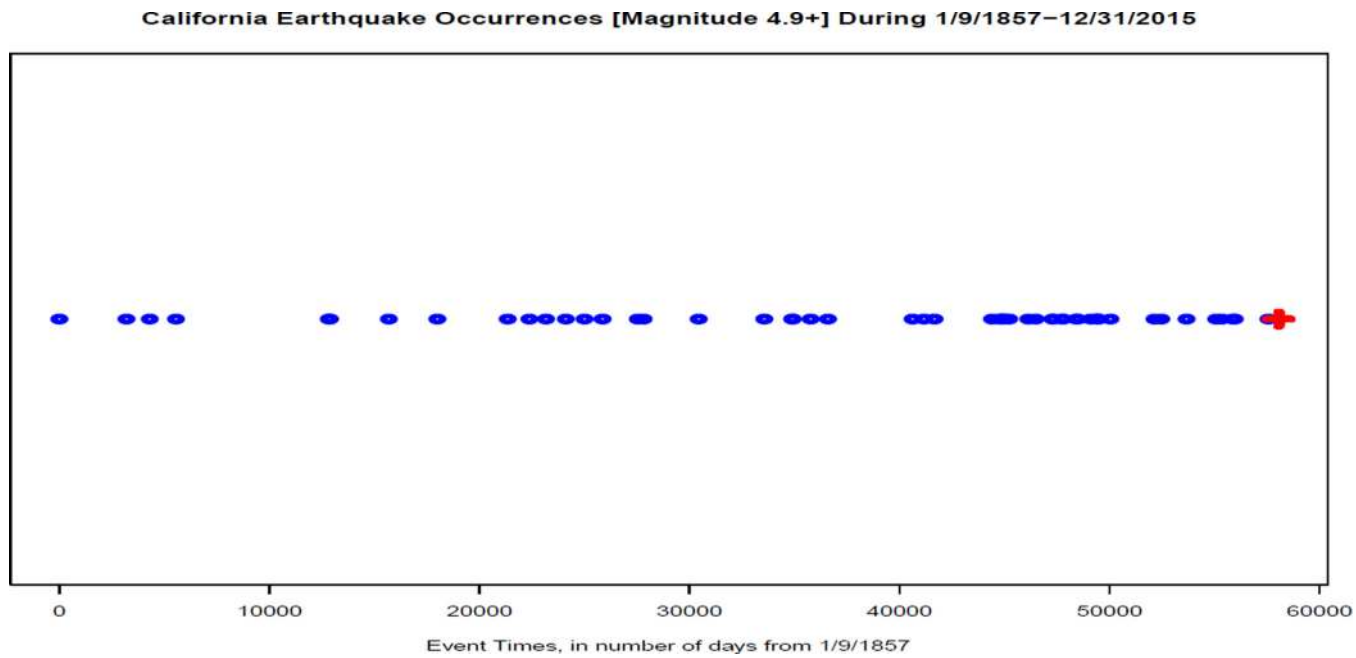

Figure 1.

Plot of the times (in number of days since January 9, 1857) of occurrences of the mainshocks of earthquakes of Richter magnitude at least 4.9 in the State of California from January 9, 1857 to December 31, 2015. Excluded are doublets, swarms, and triggered occurrences. The red cross is the end of monitoring period which is December 31, 2015, and no earthquake occurred during that day.

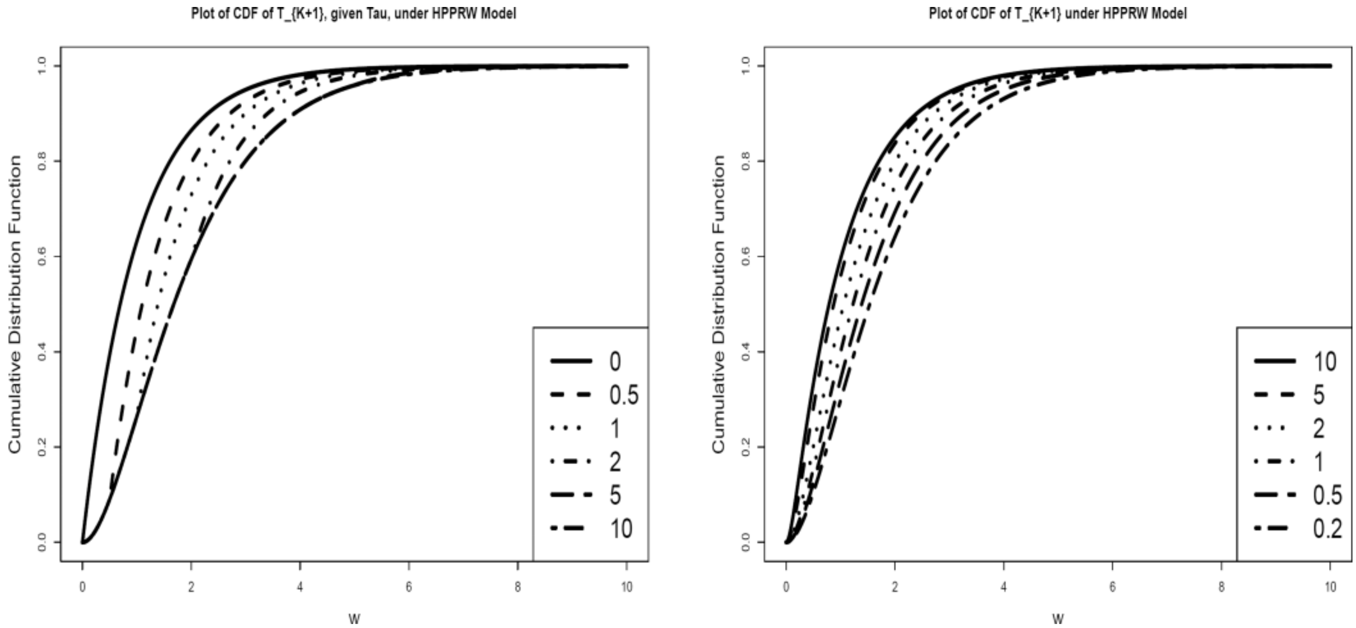

Figure 2.

Overlaid plots of the cumulative distribution functions of T_{K+1}, given τ, for $\lambda = 1$ and different values of τ [panel 1] and unconditional on τ for different values of η [panel 2] under the HPPRW(λ, η) model.

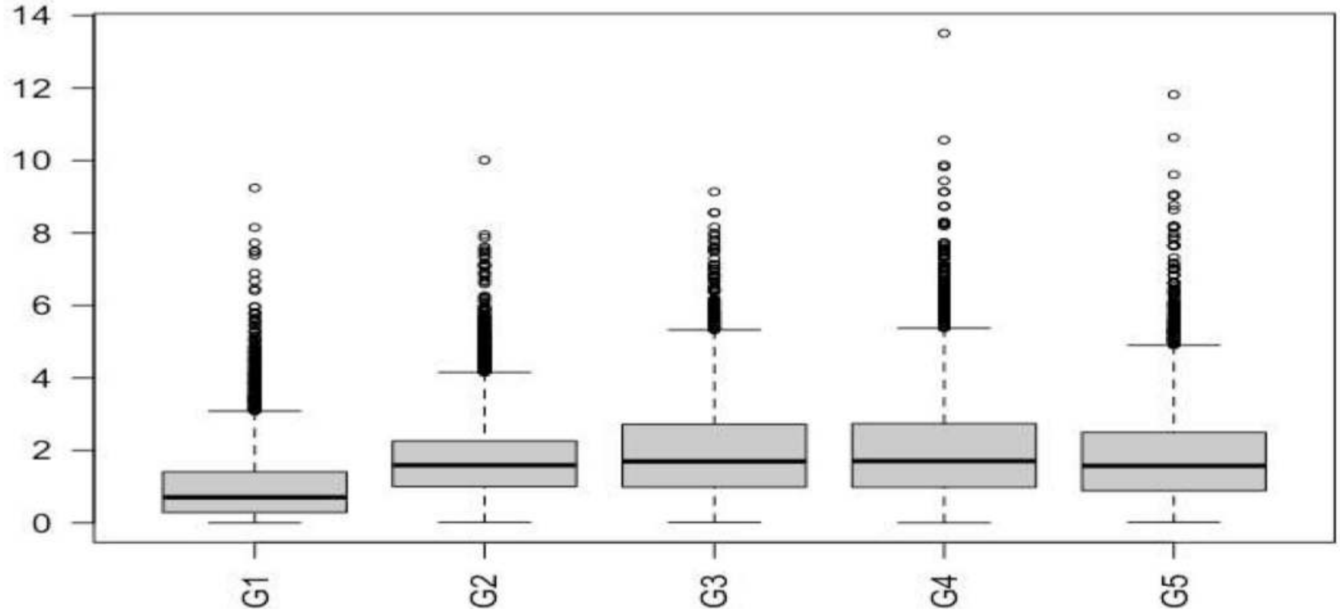

Figure 3.
Comparative boxplots arising from 5000 sample realizations of five random variables arising from the HPPRW($\lambda = 1,\ \eta = .2$). [LEGEND: $G1 \overset{d}{=} T_1$, so has unit exponential distribution; $G2 \overset{d}{=} [T_{K+1} | (\tau = \xi_{.25})]; G3 \overset{d}{=} [T_{K+1} | (\tau = \xi_{.50})]; G4 \overset{d}{=} [T_{K+1} | (\tau = \xi_{.75})$; and $G5 \overset{d}{=} T_{K+1}$.]

Table 1

The occurrence times (`EvTimes`) of 51 earthquakes (mainshocks) of Richter magnitude at least 4.9 and the gap-times (`GapTimes`) in California during 1/9/1857 to 12/31/2015. Event times are in number of days from 1/9/1857. There were 58064 days during the monitoring period and the number of days from the last one, which occurred on 8/24/2014, until 12/31/2015 was 494.

`EvTimes`: 3194, 4303, 5555, 12828, 12884, 15690, 17995, 21347, 22381, 23173, 24118, 25007, 25865, 27541, 27818, 30444, 33566, 34891, 34923, 35774, 36596, 40631, 41172, 41668, 44410, 44768, 44838, 44939, 45061, 45228, 46133, 46491, 47296, 47309, 47746, 47799, 48423, 48493, 49112, 49411, 49414, 49478, 50046, 52144, 52467, 53672, 55080, 55353, 55882, 55967, 57570.
`GapTimes`: 3194, 1109, 1252, 7273, 56, 2806, 2305, 3352, 1034, 792, 945, 889, 858, 1676, 277, 2626, 3122, 1325, 32, 851, 822, 4035, 541, 496, 2742, 358, 70, 101, 122, 167, 905, 358, 805, 13, 437, 53, 624, 70, 619, 299, 3, 64, 568, 2098, 323, 1205, 1408, 273, 529, 85, 1603.

Author Manuscript
Author Manuscript
Author Manuscript