Theorem 2.1 (Hall): The distribution of clump length consists of an atom of size $e^{-a\lambda}$ at a and an absolutely continuous component on (a, ∞) with density

$$f(x) = \frac{\lambda e^{-a\lambda}}{1 - e^{-a\lambda}} \left[1 + \sum_{j=1}^{[(x/a)-1]} \frac{(-1)^j}{j!} \{\lambda(x-(j+1)a)\}^{j-1} e^{-ja\lambda} \{\lambda(x-(j+1)a)+j\} \right]$$

The continuous component of the density above satisfies $\int_a^\infty f(x)dx = 1$, the normalized version that satisfies $\int_a^\infty \tilde{f}(x)dx = 1 - e^{-\lambda a}$ is then given by

$$f(x) = \lambda e^{-a\lambda} \left[1 + \sum_{j=1}^{[(x/a)-1]} \frac{(-1)^j}{j!} \{\lambda(x-(j+1)a)\}^{j-1} e^{-ja\lambda} \{\lambda(x-(j+1)a)+j\} \right]$$

An important function in all this is $p_n(u)$, the probability that when sampling n elements from $U(0,1)$, none of the resulting $n+1$ parts of the unit interval are longer than u.

$$\begin{aligned} p_n(u) &= \sum_{j=0}^{n+1} (-1)^j \binom{n+1}{j} (1-ju)_+^n \\ &= \sum_{j=0}^{[u^{-1}]} (-1)^j \binom{n+1}{j} (1-ju)^n \\ &= 1 - (n+1)(1-u)^n + \binom{n+1}{2} (1-2u)^n - \dots \end{aligned}$$

Elementary probability yields the conditional probability

$$(f(x)) \Pr(K = k | X = x) = \lambda e^{-a\lambda} \frac{e^{-\lambda(x-a)}}{(k-2)!} (\lambda(x-a))^{(k-2)} p_{k-2}(a/(x-a))$$

Result #1: for $a < x \leq 2a$, we have that $u = a/(x-a) > 1$ and $p_n(a/(x-a)) \equiv 1$ and

$$\Pr(K = k|X = x) = \frac{e^{-\lambda(x-a)}}{(k-2)!}(\lambda(x-a))^{(k-2)}$$

That is, $K - 2$ has a Poisson distribution with mean $\lambda(x-a)$.

Consider a Poisson process with rate lambda until time t. Let A_t denote the total number of arrivals by time t. Suppose that only clusters are observable. Let $n \leq A_t$ denote the number of clusters. Let the lengths and orders of these clusters be denoted by $X_1, \ldots, X_n$ and $K_1, \ldots, K_n$ respectively so that

$$A_t = \sum_i K_i.$$

We'll worry about the n^{th} clump later. Then for inference about the unobservable A_t consider

$$\hat{A}_t = E(K_i | X_i).$$

Likelihood

Pair the i^{th} spacing with the i^{th} clump. Let the spacing be denoted by Z_i and the clump by Y_i, with density function g. Let $X_i = Z_i + Y_i$ with density h. Let the number of renewals be denoted by N and the remnant of renewal life by $\tilde{X}_t = t - \sum_1^N X_i$. The joint density of $N, Y_1, \ldots, Y_N, Z_1, \ldots, Z_N$ and $\tilde{X}_t$ is proportional to

$$L(\theta) = \left\{\prod_1^N h(x_i;\theta)\right\}\left\{\prod_1^N \frac{\lambda}{\exp\{-\lambda Z_i\}}\right\}$$

blah blah
An approximate likelihood for the complete spacings and clumps is given by

$$\tilde{\mathcal{L}}(\lambda) = \underbrace{\lambda^{N_1} e^{-\lambda \sum Z_i}}_{\text{spacings}} \quad \underbrace{e^{-Ma\lambda}}_{\text{single segments}} \quad \underbrace{(1 - e^{-a\lambda})^{N_2 - M} \prod_1^{N_2 - M} f(y_i;\lambda)}_{\text{complete multisegment lengths}}$$

Frequencies

The order K of a cluster randomly sampled from all clusters is geometrically distributed:

$$\Pr(K = k) = e^{-a\lambda}(1 - e^{-a\lambda})^{k-1} k = 1, 2, \ldots$$

The mean order of a cluster randomly sampled from all clusters is

$$E(K) = \sum_{1}^{\infty} k e^{-a\lambda}(1 - e^{-a\lambda})^{k-1} = e^{a\lambda}$$

Also of use is $E(K|K > 1)$:

$$E(K|K > 1) = \sum_{2}^{\infty} k e^{-a\lambda}(1 - e^{-a\lambda})^{k-2} = \frac{e^{\lambda a} - e^{-\lambda a}}{1 - e^{-\lambda a}}$$

$$\mathrm{Var}(K|K > 1) = e^{2\lambda a}(1 - e^{-\lambda a})$$

Note also, that the mean of a positive Poisson random variable W with parameter λ is

$$E(W|W > 0) = \sum_{1}^{\infty} w \frac{e^{-\lambda}}{1 - e^{-\lambda}} \lambda^w / w! = \frac{\lambda}{1 - e^{-\lambda}}$$

and the variance is

$$\mathrm{Var}(W|W > 0) = \sum_{1}^{\infty} (w - \frac{\lambda}{1 - e^{-\lambda}})^2 \frac{e^{-\lambda}}{1 - e^{-\lambda}} \lambda^w / w! = \frac{\lambda - \lambda^2 e^{-\lambda a}/(1 - e^{-\lambda a})}{1 - e^{-\lambda}}$$

Let $A(t)$ denote the number of arrivals by time t. $E(A(t)) = \lambda t$. Let $N(t)$ denote the number of clusters by time t. Let $M(t) = E(N(t))$. Then

$$\begin{aligned} M(t) &= \lambda \int_0^t \exp\{-\lambda \int_0^u \Pr(X > x) dx\} du \\ &= \lambda \int_0^t \exp\{-\lambda g(u)\} du \\ &= 1 - e^{-\lambda a} + (t - a)\lambda e^{-\lambda a} \end{aligned}$$

where

$$g(u) = \begin{cases} u & u < a \\ a & u > a \end{cases}$$

Note that for large t, $M(t) \approx \lambda t e^{-\lambda a}$.

Estimators

1. $\hat{A}_1(t) = N(t)e^{M_2/M_1}$ (Doesn't do well when λa large.)

2. $\hat{A}_2(t) = N(t)e^{a/\bar{Z}}$ where $\bar{Z}$ denotes the mean spacing. (Doesn't do well when λa large.)

3. $\hat{A}_3(t) = M + (N(t) - M_1)E(K_i|X_i > a) = M + (N - M)\frac{e^{\lambda a} - e^{-\lambda a}}{1 - e^{-\lambda a}}$

4. $\hat{A}_{\text{good}}(t) = M + \sum_{i:y_i>a}(s + \left(\frac{y_i - sa}{a}\right) e^{\hat{\lambda} a}$

where $M_1 = \#$ single-segment clumps and $M_2 = \#$ number of clumps with length in $(a, 2a]$.

Multinomial Distribution

If again $M_1 = \#$ single-segment clumps and $M_2 = \#$ number of clumps with length in $(a, 2a]$, then $(M_1, M_2)|N(t) = n$ have the trinomial distribution with probability parameter $p = (e^{-\lambda a}, \lambda a e^{-\lambda a})$.

For large $N(t) = n$, this trinomial can be approximated by the bivariate normal distribution. Given $N(t) = n$,

$$\sqrt{n}\left(\begin{pmatrix} M_1/n \\ M_2/n \end{pmatrix} - \begin{pmatrix} e^{-\lambda a} \\ \lambda a e^{-\lambda a} \end{pmatrix}\right) \xrightarrow{\mathcal{L}} N_2(0, \Sigma)$$

where $\sigma_{11} = e^{-\lambda a}(1 - e^{-\lambda a})$, $\sigma_{22} = \lambda a e^{-\lambda a}(1 - \lambda a e^{-\lambda a})$ $\sigma_{12} = -\lambda a e^{-2\lambda a}$.

Consider $h(M_1, M_2) = \exp\{\frac{M_2}{M_1}\}$. Then, conditionally on $N(t) = n$, for large n,

$$\begin{aligned} E(h(M_1, M_2)) &= \lambda a + O(n^{-1}) \\ \text{Var}(h(M_1, M_2)) &= n^{-1} \nabla h(M_1, M_2)' \Sigma \nabla h(M_1, M_2)|_{\mu} \end{aligned}$$

If my algebra is without error, this works out to

$$\text{Var}(e^{M_2/M_1}|N(t) = n) = n^{-1} e^{3\lambda a} \lambda a (1 + \lambda a)$$

This can be used to obtain $\text{Var}(\hat{A}_1(t)) = \text{Var}(N(t) e^{M_2/M_1})$:

$$\begin{aligned} \text{Var}\left(N(t) e^{M_2/M_1}\right) &= E[\text{Var}(N(t) e^{M_2/M_1}|N(t)] + \text{Var}[E(N(t) e^{M_2/M_1}|N(t)] \\ &= E[N(t) e^{3\lambda a} \lambda a (1 + \lambda a)] + \text{Var}[N(t) e^{\lambda a}] \\ &= e^{3\lambda a} \lambda a (1 + \lambda a) E[N(t)] + e^{2\lambda a} \text{Var}[N(t)] \end{aligned}$$

Now we can use moments for $N(t)$:

$$\begin{aligned} E[N(t)] &= \lambda t e^{-\lambda a} \\ \text{Var}[N(t)] &= \lambda t \left(e^{-\lambda a} - 2a\lambda e^{-2a\lambda}\right) \end{aligned}$$

The proof for these two moments eludes me, but I believe it follows from elementary renewal theory. If the mean and variance of a renewal period is denoted by μ and σ, then $N(t)$ is asymptotically normal as $t \to \infty$:

$$\frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} \longrightarrow N(0, 1)$$

Substitution of these moments for $N(t)$ into the expression for $\mathrm{Var}(\hat{A}_1(t)$ yield

$$\mathrm{Var}(\hat{A}_1(t)) = \lambda a(1+\lambda a)e^{2\lambda a} + \lambda t e^{\lambda a} - 2\lambda a \lambda t.$$

Additionally, the distribution of $\hat{A}_1(t)$ is normal with approximate mean λt.

So, $\hat{A}_1(t)$ is simply a method-of-moments estimator for λt: since

$$E(N(t)) = \lambda t e^{-\lambda a}$$

we have that

$$E(N(t)e^{\lambda a}) = \lambda t$$

and since $E(M_2/M_1|N(t)) = \lambda a$, substitution of M_2/M_1 for λa yields the estimator

$$\hat{A}_1(t) = N(t)e^{M_2/M_1}$$

which is approximately unbiased.

The estimator $\hat{A}_2(t) = N(t)e^{a/\bar{Z}}$ has approximate mean and variance

$$\begin{aligned} E(\hat{A}_2(t)) &= \lambda t \\ \mathrm{Var}(\hat{A}_2(t)) &= \lambda t((1+(\lambda a)^2)e^{\lambda a} - 2\lambda a) \end{aligned}$$

These follow from moments for functions of $\bar{Z}$:

$$\begin{aligned} E(e^{a/\bar{Z}}|N(t)=n) &= e^{a\lambda} + O(n^{-1}) \\ \mathrm{Var}(e^{a/\bar{Z}}|N(t)=n) &= (\lambda a)^2 e^{2\lambda a}/n \text{wrong} \\ \mathrm{Var}(e^{a/\bar{Z}}|N(t)=n) &= (\lambda a)^2 e^{2\lambda a}/n + \frac{1}{n^2}\left((\lambda a e^{\lambda a})^3 + 1/2(\lambda a e^{\lambda a})^4\right) \end{aligned}$$

Higher order moments are given below:

$$\mathrm{Var}(e^{a/\bar{Z}}|N(t)=n) = \frac{1}{n}(\lambda a)^2 e^{2\lambda a} + \frac{1}{n^2}e^{2\lambda a}\left[\frac{1}{2}(\lambda a)^4 - 2(\lambda a)^2\right]$$

Some summary

Estimation of the quantity $e^{\lambda a}$ given $N(t) = n$:

$$\begin{aligned}
E(e^{M_2/M_1}|N(t) = n) &= e^{\lambda a} \\
&+ n^{-1}\frac{1}{2}\left[(\lambda a)^2(e^{\lambda a} - 1) + 2(\lambda a)^2 e^{2\lambda a} + \lambda a e^{3\lambda a}(1 - \lambda a e^{-\lambda a})\right] \\
&+ O(n^{-2}) \\
&= e^{\lambda a} + \text{bias}_1 + O^(n^{-1}) \\
\text{Var}(e^{M_2/M_1}|N(t) = n) &= n^{-1}e^{3\lambda a}\lambda a(1 + \lambda a) + O(n^{-2}) \\
E(e^{a/\bar{Z}}|N(t) = n) &= e^{\lambda a} + n^{-1}\frac{1}{2}\left[e^{\lambda a}((\lambda a)^2 + 2\lambda a)\right] + O(n^{-2}) \\
&= e^{\lambda a} + \text{bias}_2 + O^(n^{-1}) \\
\text{Var}(e^{a/\bar{Z}}|N(t) = n) &= n^{-1}a^2 e^{\lambda a} + O(n^{-2})
\end{aligned}$$

Both of these results can be established by taking expectations of Taylor expansions about the means of the random vector (M_1, M_2) for (1) and the random variable $\bar{Z}$ for (2). Unconditional moments are given by

$$\begin{aligned}
E[N(t)e^{M_2/M_1}] &= E[E(N(t)e^{M_2/M_1}|N(t))] \\
&= \lambda t + e^{-\lambda a}\text{bias}_1 + O(t^{-1})\text{wrong} \\
&= \lambda t + \frac{1}{2}e^{\lambda a}[(\lambda a)^2 + 2\lambda a] + O(t^{-1})\text{wrong} \\
&= \lambda t + \frac{1}{2}e^{\lambda a}[(\lambda a)^2 + 2\lambda a] + O(t^{-1}) \\
E[N(t)e^{a/\bar{Z}}] &= E[E(N(t)e^{a/\bar{Z}}|N(t))] \\
&= \lambda t + e^{-\lambda a}\text{bias}_2 + O(t^{-1})
\end{aligned}$$

Multinomial likelihood

$$l(\lambda; M_1, M_2, M_+|N(t) = n) = -\lambda a(M_1+M_2)+\log\lambda+(n-M_1-M_2)*\log\left(1 - e^{-\lambda a}(1+\lambda a)\right)$$

The expected information is given by

$$\begin{aligned} l_{\lambda\lambda}(\lambda; m_1, m_2, n) &= \frac{-m_2}{\lambda^2} + (n - m_1 - m_2)\left[e^{-\lambda a}\frac{(a^2 - \lambda a^3)}{1 - e^{-\lambda a}(1+\lambda a)} - \frac{(\lambda a^2 e^{-\lambda a})^2}{(1 - e^{-\lambda a}(1+\lambda a))^2}\right] \\ E[-l_{\lambda\lambda}(\lambda; m_1, m_2, n)] &= n\left(\frac{ae^{-\lambda a}}{\lambda} - [e^{-\lambda a}(a^2 - \lambda a^3) - \frac{(\lambda a^2 e^{-\lambda a})^2}{1 - e^{-\lambda a}(1+\lambda a)}]\right) \end{aligned}$$

This can be inverted to obtain the asymptotic variance of the mle $\hat{\lambda}$:

$$E(-l_{\lambda\lambda})^{-1} = n^{-1}\frac{\lambda(1 - e^{\lambda a}(1+\lambda a))}{a(1 - e^{-\lambda a}) - \lambda a^2(1 - \lambda a)}$$

By the way, the asymptotic variance of $\sqrt{n}\frac{M_2}{M_1}$ in estimation of λa is

$$\mathrm{Var}(\sqrt{n}\frac{M_2}{M_1}) = e^{\lambda a}[\lambda a(1+\lambda a)] + 2(\lambda a)^2$$

so that

$$\mathrm{Var}(\sqrt{n}\frac{M_2}{aM_1}) = e^{\lambda a}[\lambda^2 + \frac{\lambda}{a}] + 2\lambda^2$$

The possibility of $M = 0$

The larger the so-called "occupation rate" λa, the greater the chance of $M = 0$. Note that

$$\begin{aligned}
\Pr(M = 0) &= \sum_{n=0}^{\infty} \Pr(M = 0|N(t) = n) \Pr(N(t) = n) \\
&= E(e^{N(t) \log(1-e^{-\lambda a})}) \\
&= M_{N(t)}(\log(1 - e^{-\lambda a})) \\
&\approx \exp\{\mu_{N(t)} \log(1 - e^{-\lambda a}) + \frac{1}{2}(\log(1 - e^{-\lambda a}))^2 \sigma^2_{N(t)}
\end{aligned}$$

where $M_{N(t)}, \mu_{N(t)}$, and $\sigma_{N(t)}$ denote the moment generating function, and asymptotic mean and variance of $N(t)$, respectively. Substitution of the asymptotic moments yields the expression below:

$$\Pr(M = 0) \approx \exp\{\lambda t e^{-\lambda a} \log(1-e^{-\lambda a})+\frac{1}{2}\lambda t(e^{-\lambda a}-2\lambda a e^{-2\lambda a})(\log(1-e^{-\lambda a}))^2$$

Total clump length Let $Y_1, Y_2, \ldots, Y_{N(t)-M}$ denote the lengths of the multi-segment clumps. Consider estimation of $A(t)$ based on $N(t), M, \sum Y_i$.

Let $\tilde{\mu}_Y = E(Y|Y > a)$ then $\tilde{\mu}_Y$ satisfies

$$\mu_Y = \frac{e^{\lambda a} - 1}{\lambda} = a \Pr(Y = a) + \tilde{\mu}_Y \Pr(Y > a)$$

a little alegbra yields

$$\tilde{\mu}_Y = \left(\frac{e^{\lambda a} - 1}{\lambda - a e^{-\lambda a}} \right) (1 - e^{-\lambda a})^{-1}$$

Similarly, let $\tilde{\sigma_Y}^2 = \mathrm{Var}(Y|Y > a)$ and $I = I(Y > a)$. Then $\tilde{\sigma_Y}^2$ satisfies

$$\begin{aligned}
\mathrm{Var}(Y) &= E[\mathrm{Var}(Y|I)] + \mathrm{Var}[E(Y|I)] \\
&= \tilde{\sigma}_Y^2 (1 - e^{-\lambda a}) + (\tilde{\mu}_Y - a)^2 e^{-\lambda a} (1 - e^{-\lambda a})
\end{aligned}$$

so that

$$\tilde{\sigma}_Y^2 = \left(\frac{e^{2\lambda a} - 2\lambda a e^{\lambda a} - 1}{\lambda^2} - (\tilde{\mu}_Y - a)^2 e^{-\lambda a} (1 - e^{-\lambda a}) \right) (1 - e^{-\lambda a})^{-1}$$

Conditional mean order given length, $E(K|Y)$

For $a < y \leq 2a$, $K - 1$ has the Poisson distribution, exactly, with mean $\lambda(y - a)$. However, there is a jump discontinuity in the mean function at $y = 2a$. Thereafter

The function $p_n(u)$:

For $u > 1, p_n(u) = 1$.

To evaluate $p_n(\frac{a}{y-a})$, note that

$$ja < y < (j+1)a \Longrightarrow \frac{1}{j} < \frac{a}{y-a} < \frac{1}{j-1}$$

So, we might expect $p_n(\frac{a}{y-a}) \approx p_n(\frac{1}{j})$ in these cases.

An expression for $E(K|Y=y)$ is available, but it is not easy to evaluate. Suppose $s = [\frac{y}{a}]$

$$\begin{aligned}
E(K|Y=y) &= \sum_{k=s+1}^{\infty} k \Pr(K=k|Y=y) \\
&= \frac{\lambda e^{-\lambda y}}{f(y)} \sum_{k=s+1}^{\infty} k \frac{(\lambda(y-a))^{k-2}}{(k-2)!} p_{k-2}(\frac{a}{y-a}) \\
&= \frac{\lambda e^{-\lambda y}}{f(y)} \sum_{k=s+1}^{\infty} k \frac{(\lambda(y-a))^{k-2}}{(k-2)!} \sum_{j=0}^{s-1} (-1)^j \left(\begin{array}{c} k-1 \\ j \end{array} \right) \left(1 - \frac{ja}{y-a}\right)^{k-2}
\end{aligned}$$

This double sum can be simplified a little bit, at least into a single sum, by following the arguments for deriving $f(y)$ as on p. 86 of Hall.

Let $\lambda_j = \lambda(y-(j+1)a)$.

Conjecture:

$$\sum_{k=s+1}^{\infty} \left(\begin{array}{c} k-1 \\ j \end{array} \right) \frac{1}{(k-2)!} \lambda_j^{k-j-1} = \frac{e^{-\lambda a j}}{j!} (\lambda_j + j) e^{\lambda(y-a)}$$

Non-conjecture:

$$\sum_{k=s+1}^{\infty} \frac{1}{(k-2)!} \sum_{j=0}^{s-1} (-1)^j \left(\begin{array}{c} k-1 \\ j \end{array} \right) \lambda_j^{k-2} = e^{\lambda(y-a)} \left(1 + \sum_{j=1}^{s-1} \frac{(-1)^j}{j!} \lambda_j^{j-1} (\lambda_j + j) e^{\lambda a j} \right)$$

Note that $f(y)$ can be written

$$f(y) = \lambda e^{-\lambda y} \left(e^{\lambda(y-a)} + \sum_{j=1}^{s-1} \frac{(-1)^j}{j!} \lambda_j^{j-1} e^{\lambda_j} (\lambda_j + j) \right)$$

The conditional distribution of the order of k is then

$$f(k|y) = \frac{\lambda e^{-y}}{f(y)} \left(\frac{(\lambda(y-a))^{k-2}}{(k-2)!} + \sum_{j=1}^{s-1} (-1)^j \left(\begin{array}{c} k-1 \\ j \end{array} \right) \frac{1}{(k-2)!} \lambda_j^{k-2} \right)$$

Summing this distribution over k and setting to unity yields the identity

$$\sum_{k=2}^{s} \frac{(\lambda(y-a))^{k-2}}{(k-2)!} + \sum_{j=1}^{s-1} \sum_{k=j+1}^{s} (-1)^j \left(\begin{array}{c} k-1 \\ j \end{array} \right) \frac{1}{(k-2)!} \lambda_j^{k-2} = 0$$

which I believe will be most helpful in simplifying $E(K|Y=y)$.

$$E(K|Y=y) = \sum_{k=s+1}^{\infty} k f(k|y)$$
$$= \frac{\sum_{k=s+1}^{\infty} k \frac{(\lambda(y-a))^{k-2}}{(k-2)!} + \sum_{k=s+1}^{\infty} \sum_{j=1}^{s-1} k(-1)^j \begin{pmatrix} k-1 \\ j \end{pmatrix} \frac{1}{(k-2)!} \lambda_j^{k-2}}{e^{\lambda(y-a)} + \sum_{j=1}^{s-1} \frac{(-1)^j}{j!} \lambda_j^{j-1} e^{\lambda_j} (\lambda_j + j)}$$

By adding and subtracting $k : 2 \leq k < s+1$ terms, we can obtain the following expression for the numerator of this mean:

$$num =$$
$$\left(\sum_{k=2}^{\infty} k \frac{(\lambda(y-a))^{k-2}}{(k-2)!} + \sum_{k=2}^{\infty} \sum_{j=1}^{s-1} k(-1)^j \begin{pmatrix} k-1 \\ j \end{pmatrix} \frac{1}{(k-2)!} \lambda_j^{k-2} \right) -$$
$$\left(\sum_{k=2}^{s} k \frac{(\lambda(y-a))^{k-2}}{(k-2)!} + \sum_{k=2}^{s} \sum_{j=1}^{s-1} k(-1)^j \begin{pmatrix} k-1 \\ j \end{pmatrix} \frac{1}{(k-2)!} \lambda_j^{k-2} \right)$$

Consider the first term, with the complete summation over k. Call it I.

$$I = 2 + \lambda(y-a) + + \sum_{j=1}^{s-1} \frac{(-1)^j}{j!} e^{\lambda_j} \left[\lambda_j^{j+1} + (2j+1)\lambda_j^j + j^2 \lambda_j^{j-1} \right]$$
$$=$$

To get this, another key identity, which I can't immediately prove, is

$$\sum_{j=1}^{k-1} (-1)^j \begin{pmatrix} k-1 \\ j \end{pmatrix} \frac{\lambda_j^{k-2}}{(k-2)!} = -\frac{\lambda_0^{k-2}}{(k-2)!}$$

or

$$\sum_{j=1}^{m+1} (-1)^j \begin{pmatrix} m+1 \\ j \end{pmatrix} \frac{\lambda_j^m}{m!} = -\frac{\lambda_0^m}{m!}$$

or

$$\sum_{j=1}^{k-1} (-1)^j \begin{pmatrix} k-1 \\ j \end{pmatrix} (1 - j\frac{a}{y-a})^{k-2} = -1$$

for any $k < \frac{y}{a}$.

Exact mean of K when $s = 2$:

$$E(K|Y = y) = \frac{e^{\lambda_0}(\lambda_0 + 2) - e^{\lambda_1}(\lambda_1^2 + 4\lambda_1 + 2)}{e^{\lambda_0}(1 - e^{-\lambda a}(\lambda_1 + 1)}$$

where $\lambda_j = \lambda(y - (j + 1)a)$. For small λ_1,

$$\begin{aligned} E(K|Y = y) &= \frac{\lambda_0 + 2 - 2e^{-\lambda a}}{1 - e^{-\lambda a}(\lambda_1 + 1)} \\ &= \frac{\lambda_0}{1 - e^{-\lambda a}} + 2 \\ &= \frac{\lambda a}{1 - e^{-\lambda a}} + 2 \\ &= \frac{1 - \lambda a + e^{-\lambda a}}{1 - e^{-\lambda a}} + 3 \end{aligned}$$

Conversely, for y near $3a$, and $\lambda_0 \approx 2\lambda a$ and $\lambda_1 \approx \lambda a$, we have

$$E(K|Y = y) \approx 2 + 2\lambda a + \frac{(\lambda a)^2 e^{-\lambda a}}{1 - e^{-\lambda a}(1 + \lambda a)}$$

$$E(K|Y = y) \approx 3 + \frac{2\lambda a + -1 + e^{-\lambda a}(1 - \lambda a - (\lambda a)^2)}{1 - e^{-\lambda a}(1 + \lambda a)}$$

(Note: the gamma cdf may come into play.) For $s = 3$ and $3a < y < 4a$, we have

$$E(K|Y = y) = \frac{e^{\lambda_0}(2 + \lambda_0) - e^{\lambda_1}(\lambda_1^2 + 4\lambda_1 + 2) + e^{\lambda_2}(\frac{\lambda_2^3}{2} + 3\lambda_2^2 + 3\lambda_2) - 3\lambda_0 + 6\lambda_1 - 3\lambda_2}{e^{\lambda_0}\left(1 - e^{-\lambda a}(\lambda_1 + 1) + \frac{1}{2}e^{-2\lambda a}\lambda_2(\lambda_2 + 2)\right)}$$

When y is near $3a$, $\lambda_2 \approx 0, \lambda_1 \approx 0, \lambda_0 \approx 2\lambda a$ and

$$\begin{aligned} E(K|Y = y) &\approx \frac{(2 + 2\lambda a) - e^{-\lambda a}((\lambda a)^2 + 4\lambda a + 2)}{1 - e^{-\lambda a}(1 + \lambda a)} \\ &\approx 2 + 2\lambda a + (\lambda a)^2 \frac{e^{-\lambda a}}{1 - e^{-\lambda a}(1 + \lambda a)} \\ &= 4 + \frac{2(e^{-\lambda a}(1 - \frac{(\lambda a)^2}{2}) - 1 + \lambda a)}{1 - e^{-\lambda a}(1 + \lambda a)} \end{aligned}$$

At long last, the mean of clump order K given length y is approximately linear in y (for $y > 2a$.) The slope can be approximated by $E(K|Y = 3a) -$

$E(K|Y = 2a + \epsilon)$. (The mean as a function of y has a jump discontinuity of $\lambda a(1-e^{-\lambda a})^{-1}$ at $y = 2a$, but is otherwise continuous.) The exact expressions for $E(K|Y)$ given above can be used to approximate the slope as

$$\widehat{slope} = E(K|Y = 3a) - E(K|Y = 2a) = \lambda a[\frac{\lambda a e^{-\lambda a}}{1 - e^{-\lambda a}} - (1 - \lambda a e^{-\lambda a})]$$

From, this, the mean can be approximated by

$$E(K\widehat{|Y} = y) = 2 + \frac{\lambda a}{1 - e^{-\lambda a}} + \widehat{slope}(y - 2a).$$