ON THE ASYMPTOTIC DISTRIBUTION OF THE
SIZE OF A STOCHASTIC EPIDEMIC

BY

THOMAS SELLKE

TECHNICAL REPORT NO. 17
MAY 1982

PREPARED UNDER CONTRACT
NO0014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE ASYMPTOTIC DISTRIBUTION OF THE
SIZE OF A STOCHASTIC EPIDEMIC

by

THOMAS SELLKE

TECHNICAL REPORT NO. 17
MAY 1982

Prepared Under the Auspices
of
Office of Naval Research Contract
N00014-77-C-0306(NR-042-373)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE ASYMPTOTIC DISTRIBUTION OF
THE SIZE OF A STOCHASTIC EPIDEMIC

by
Thomas Sellke
Stanford University

ABSTRACT

For a stochastic epidemic of the type considered by Bailey [1] and
Kendall [3], Daniels [2] showed that "when the threshold is large but
the population size is much larger, the distribution of the number remain-
ing uninjected in a large epidemic has approximately the Poisson form."
A simple, intuitive proof is given for this result without use of Daniels'
assumption that the original number of infectives is "small". The proof
is based on a construction of the epidemic process which is more explicit
than the usual description.

Key words: limiting Poisson distribution, stochastic epidemic
ON THE ASYMPTOTIC DISTRIBUTION OF THE SIZE OF
A STOCHASTIC EPIDEMIC

1. Introduction

Consider a population which at time $t = 0$ consists of $X(0) = n$ healthy individuals and $Y(0) = m$ individuals with a contagious infection. An epidemic in such a population is often modeled by a continuous-time Markov process as follows. (See, for example, Bailey [1] or Kendall [3].) If $X(t)$ and $Y(t)$ are the numbers of healthy individuals and infectious individuals, respectively, present at time $t$, then the transition probabilities are given by

$$P\{X(t+\delta), Y(t+\delta) = (x', y') | (X(t), Y(t)) = (x, y)\}$$

$$= xy\delta + o(\delta) \text{ for } (x', y') = (x-1, y+1)$$

$$= \rho y \delta + o(\delta) \text{ for } (x', y') = (x, y-1)$$

$$= 1 - xy\delta - \rho y \delta + o(\delta) \text{ for } (x', y') = (x, y) .$$

The transitions listed represent the infection of a healthy individual, the removal of an infectious individual from the population, and "no change", respectively. All other possible transitions in $[t, t+\delta]$ are assumed to have collective probability $o(\delta)$. The positive constant $\rho$ is variously called the threshold or the relative removal rate of the epidemic.

Note that states of the form $(x, 0)$ are absorbing, so that no more transitions occur after the last infectious individual has been removed. Absorption at $(x, 0)$ means that $x$ individuals have escaped infection at the end of the epidemic and that $X(\infty) = x$, where $X(\infty) = \lim_{t \to \infty} X(t)$. It is easy to see that, with probability one, some such absorbing state is
reached eventually. Most of the work on this stochastic epidemic model has been directed toward finding the distribution of the number of new infections occurring during the course of the epidemic. This is equivalent to finding the distribution of the absorbing state. See Nagaev and Startsev [4] for an asymptotic analysis of limiting cases other than the one considered in this paper.

Under the assumption that \( Y(0) = m \) is "small", Daniels [2] shows that "when the threshold is large but the population size is much larger, the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form with deterministic mean \( ne^{-n/\rho} \)." However, the asymptotic approximations used by Daniels in obtaining this result are not very enlightening, and he speculates that "there must be a direct argument in terms of the epidemic process itself to explain this Poisson-like behavior."

This paper gives a rather intuitive proof of Daniels' result. In addition, the assumption that \( m \) is small is dispensed with. The proof is based on a construction of the epidemic process which is more explicit than the usual description given above.

2. Construction of the Epidemic Process

Let the \( n \) originally healthy individuals be indexed by \( i, 1 \leq i \leq n \), and let the \( m \) originally infectious individuals be indexed by \( j, 1 \leq j \leq m \). Let \( \{ \hat{r}_j \}_{j=1}^m \) and \( \{ r_i \}_{i=1}^n \) be i.i.d. random variables with density \( \rho e^{-\rho t} \) on \([0, \infty)\). Individual \( j \) in the original infectious group will remain infectious for \( \hat{r}_j \) time units before removal from the population. Individual \( i \) in the original healthy group will remain infectious for \( r_i \) time units if he becomes infected.
Let \( \{ \lambda_i \}_{i=1}^n \) be i.i.d. random variables with density \( e^{-t} \) on \([0, \infty)\), independent of the \( \hat{\lambda}_j \)'s and \( r_i \)'s. The variable \( \lambda_i \) will be thought of as the "resistance to infection" of individual \( i \) in the original healthy group. Let \( \{ \lambda^{(k)} \}_{k=1}^n \) be the associated order statistics, so that \( \lambda^{(1)} < \lambda^{(2)} < \cdots < \lambda^{(n)} \). Let \( r^{(k)} = r_i \) if \( \lambda^{(k)} = \lambda_i \).

Now let the epidemic process proceed as follows. The originally infected individual \( j \) remains in the population for \( r_j \) time units, after which he is removed. The healthy individual \( i \) accumulates "exposure to infection" at a rate equal to the number of infected individuals present. When the total exposure to infection of healthy individual \( i \) reaches \( \lambda_i \), individual \( i \) becomes infected and then remains in the population for an additional \( r_i \) time units before removal.

It remains to be shown that the resulting process is Markov with the correct transition probabilities. Suppose \( (X(t), Y(t)) = (x, y) \). The probability that a particular infected individual is removed in the time interval \([t, t+\delta]\) is \( \rho \delta + o(\delta) \) because the distribution of the \( \hat{\lambda}_j \)'s and \( r_i \)'s has constant hazard rate \( \rho \). The probability that exactly one of the \( y \) infected individuals is removed in \([t, t+\delta]\) is therefore \( \rho y \delta + o(\delta) \). The probability that a particular one of the \( x \) healthy individuals will become infected in \([t, t+\delta]\) is \( y \delta + o(\delta) \), so that the probability that exactly one of the healthy individuals becomes infected is \( x y \delta + o(\delta) \). It follows that the transition probabilities are as desired. The Markov property follows from the memoryless property of the exponential distribution.

Let \( \nu \) be the number of new infections occurring during the course of the epidemic. If \( \lambda^{(1)} > \sum_{j=1}^m \hat{\lambda}_j \), then all originally infectious individuals are removed before the resistance to infection of any healthy individual has
been exceeded, so that \( \nu = 0 \). Otherwise, the originally healthy individual associated with \( l(1) \) becomes the first new infection, and \( \nu \geq 1 \). An easy induction argument shows that \( \nu + 1 \) is the smallest \( k \), \( 1 \leq k \leq n \), for which

\[
l(k) > \sum_{j=1}^{m} \hat{r}_j + \sum_{i=1}^{k-1} r(i).
\]

If this inequality does not hold for any \( k \), \( 1 \leq k \leq n \), then \( \nu = n \).

Define

\[
R = \sum_{j=1}^{m} \hat{r}_j + \sum_{i=1}^{\nu} r(i).
\]

Then \( R \) is the amount of "exposure to infection" withstood by those individuals who remain healthy at the end of the epidemic, and \( X(\infty) = n - \nu \) is the number of \( l_i \)'s greater than \( R \).

3. Statement of the Theorem and an Outline of the Proof

Consider a sequence of epidemics with parameters \( n_k, m_k, \) and \( \rho_k \), \( 1 \leq k < \infty \). Let \( \nu_k \) be the number of new infections in a realization of the \( k \)-th epidemic, so that \( X_k(\infty) = n_k - \nu_k \) is the number of individuals who escape infection.

**Theorem.** If \( n_k \to \infty \), \( \rho_k \to \infty \), and

\[
n_k \exp\left(-\frac{n_k + m_k}{\rho_k}\right) \to b, \quad 0 < b < \infty,
\]

then \( X_k(\infty) \) converges in distribution to a Poisson random variable with mean \( b \).
Suppressing the subscript $k$, we can summarize the proof as follows. The fact that $\rho$ is $o(n+m)$ implies that, with high probability, all but a tiny fraction of the population becomes infected. Thus, $R$ will be close to $\sum_{j=1}^{m} \hat{X}_j + \sum_{i=1}^{n} \hat{Y}_i$, which is in turn close to $(m+n)/\rho$ with high probability. If $R$ is close enough to $(m+n)/\rho$, then $X(\omega)$ will equal the number of $\hat{Y}_i$'s which are greater than $(m+n)/\rho$. The number of $\hat{Y}_i$'s greater than $(m+n)/\rho$ has the distribution Binomial $(n, e^{-(m+n)/\rho})$, which converges in distribution to Poisson $(b)$.

4. Proof of the Theorem

The subscript $k$ will again be suppressed.

Taking logarithms in $n e^{-(m+n)/\rho} \to b$ yields

$$\log n - \frac{n+m}{\rho} \to \log b,$$

so that $\rho \sim \frac{m+n}{\log n}$. Thus, $\rho$ is $o(m+n)$, but $(m+n)\gamma$ is $o(\rho)$ for $0 < \gamma < 1$.

Lemma 1. Let $0 < \varepsilon < 1$ be given. If $\rho < \varepsilon^2(m+n)$, then

$$\Pr[X(\omega) > \varepsilon(m+n)] < \left[\frac{\rho}{\varepsilon(m+n)}\right]^m \leq \varepsilon^m \leq \varepsilon.$$

Proof of Lemma 1.

As in Kendall [3], we can view the population of infected individuals as a continuous-time birth-and-death process with a variable birth-rate. The ratio of death-rate to birth-rate is $\rho/X(t)$, which is less than $\rho/(\varepsilon(m+n))$, until $X(t) \leq \varepsilon(m+n)$. The probability that a birth-and-death
process starting at \( m \) and with a death-rate to birth-rate ratio \( q < 1 \) is ever absorbed at zero is \( \frac{m}{q} \).

**Lemma 2.** Let \( 0 < \varepsilon < 1 \). For \( n \) sufficiently large,

\[
P\{R < \frac{(1-2\varepsilon)(m+n)}{\rho} \} < 2\varepsilon .
\]

**Proof of Lemma 2.**

By Lemma 1, \( R \) is greater than the sum of the first \((1-\varepsilon)(m+n)\) terms of \( \{\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m, r^{(1)}, r^{(2)}, \ldots, r^{(n)}\} \) with probability greater than \((1-\varepsilon)\) for sufficiently large \( n \). The sum of the first \((1-\varepsilon)(m+n)\) terms has mean \((1-\varepsilon)(m+n)/\rho\) and variance \((1-\varepsilon)(m+n)/\rho^2\). The Chebyshev inequality now implies the lemma.

**Lemma 3.** Let \( 0 < \varepsilon < 1 \). For \( n \) sufficiently large,

\[
P\{X(\infty) > n^{3\varepsilon} \} < 3\varepsilon .
\]

**Proof of Lemma 3.**

The number of \( \lambda_1 \)'s which are greater than \((1-2\varepsilon)(m+n)/\rho\) is Binomial \((n, e^{-(1-2\varepsilon)(m+n)/\rho})\). This distribution has mean

\[
e^{-\frac{(1-2\varepsilon)(m+n)}{\rho}} \cdot n \cdot \frac{b_{n}}{n} (1-2\varepsilon) = b^{1-2\varepsilon} n^{2\varepsilon} .
\]

Application of Lemma 2 and the Markov inequality finishes the proof of Lemma 3.

From Lemma 3, it is easy to see that \( X(\infty) \) is \( o(\sqrt{n}) \) in probability. Thus, except on a set of small probability, \( R \) is greater than the sum of
the first $m + n - \sqrt{n}$ terms of $\{\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m, r^{(1)}, r^{(2)}, \ldots, r^{(n)}\}$. 

$R$ is of course less than or equal to the sum of all the terms. An argument like the proof of Lemma 2 shows that, with probability approaching one,

$$\frac{m+n}{\rho} - \frac{(m+n)^{2/3}}{\rho} < R < \frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho}. $$

The number of $\lambda_i$'s greater than

$$\frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho}$$

is distributed as a Binomial $(n, \exp\{ - \frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho} \})$, which has mean

$$n \exp\{ - \frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho} \} \sim b \exp\{ \frac{(m+n)^{2/3}}{\rho} \} + b.$$ 

Thus, with probability approaching one, $X(\infty)$ is less than a

Binomial $(n, \exp\{ - \frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho} \})$ random variable

and greater than

Binomial $(n, \exp\{ - \frac{m+n}{\rho} - \frac{(m+n)^{2/3}}{\rho} \})$ random variable. 

Since both of these distributions converge in law to a Poisson with mean $b$, the theorem is proved.


References


**Title:** On the Asymptotic Distribution of the Size of a Stochastic Epidemic

**Author:** Thomas Sellke

**Performing Organization Name and Address:**
Dept. of Statistics
Stanford University
Stanford, Calif.

**Controlling Office Name and Address:**
Statistics & Probability Program Code 436
Office of Naval Research
Arlington, VA 22217

**Distribution Statement:**
Approved for public release: Distribution unlimited

**Keywords:** Limiting Poisson distribution
Stochastic epidemic

**Abstract:**
Please see other side.
ON THE ASYMPTOTIC DISTRIBUTION OF
THE SIZE OF A STOCHASTIC EPIDEMIC

by

Thomas Sellke
Stanford University

ABSTRACT

For a stochastic epidemic of the type considered by Bailey [1] and Kendall [3], Daniels [2] showed that "when the threshold is large but the population size is much larger, the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form." A simple, intuitive proof is given for this result without use of Daniels' assumption that the original number of infectives is "small". The proof is based on a construction of the epidemic process which is more explicit than the usual description.