Håkan Andersson
Tom Britton

DNO 2/04

NOTICE: THIS MATERIAL MAY BE PROTECTED BY COPYRIGHT LAW (Title 17 U.S. Code)

# Stochastic Epidemic Models and Their Statistical Analysis

Springer

Springer
*New York*
*Berlin*
*Heidelberg*
*Barcelona*
*Hong Kong*
*London*
*Milan*
*Paris*
*Singapore*
*Tokyo*

Håkan Andersson
Group Financial Risk Control
SwedBank
SE-105 34 Stockholm
Sweden
hakan.b.andersson@forenigssparbanken.se

Tom Britton
Department of Mathematics
Uppsala University
PO Box 480
SE-751 06 Uppsala
Sweden
tom.britton@math.uu.se

Library of Congress Cataloging-in-Publication Data

Stochastic epidemic models and their statistical analysis / Håkan Andersson, Tom Britton.
p. cm. -- (Lecture notes in statistics ; 151)
Includes bibliographical references and index.
ISBN 0-387-95050-8 (softcover : alk. paper)
1. Epidemiology--Mathematical models. 2. Epidemiology--Statistical methods. 3. Stochastic analysis. I. Andersson, Håkan. II. Britton, Tom. III. Lecture notes in statistics (Springer-Verlag) ; v. 151.

RA652.2.M3 S76 2000
614.4'01'5118--dc21

00-041911

Printed on acid-free paper.

© 2000 Springer-Verlag New York, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Camera ready copy provided by the authors.
Printed and bound by Sheridan Books, Ann Arbor, MI.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-95050-8 Springer-Verlag New York Berlin Heidelberg SPIN 10768919

# Preface

The present lecture notes describe stochastic epidemic models and methods for their statistical analysis. Our aim is to present ideas for such models, and methods for their analysis; along the way we make practical use of several probabilistic and statistical techniques. This will be done without focusing on any specific disease, and instead rigorously analyzing rather simple models. The reader of these lecture notes could thus have a two-fold purpose in mind: to learn about epidemic models and their statistical analysis, and/or to learn and apply techniques in probability and statistics.

The lecture notes require an early graduate level knowledge of probability and statistics. They introduce several techniques which might be new to students, but our intention is to present these keeping the technical level at a minimum. Techniques that are explained and applied in the lecture notes are, for example: coupling, diffusion approximation, random graphs, likelihood theory for counting processes, martingales, the EM-algorithm and MCMC methods. The aim is to introduce and apply these techniques, thus hopefully motivating their further theoretical treatment. A few sections, mainly in Chapter 5, assume some knowledge of weak convergence; we hope that readers not familiar with this theory can understand the these parts at a heuristic level.

The text is divided into two distinct but related parts: modelling and estimation. The first part covers stochastic models and their properties, often assuming a large community in which the disease is spread. The second part deals with statistical questions, that is, what can be said about the model and its parameters, given that an epidemic outbreak has been observed. The second part uses results from the first part, and is hence not suited for reading without having read the first part.

The lecture notes are self-instructive and may be read by anyone interested in the area. They are suited for a one-semester course of approximately 15 two-hour lectures. Most chapters may be presented in one such lecture. Chapters that need somewhat longer treatment are Chapters 5, 6, and 8. Each chapter ends with a few exercises giving extensions of the theory presented in the text.

These notes were written during the spring term 1999 when the authors gave a joint graduate course in the Departments of Mathematics at Stockholm and Uppsala Universities. We thank the participants in the course: Anders Björkström, Nestor Correia, Maria Deijfen, Peter Grenholm, Annika Gunnerhed, Allan Gut, Jemila Seid

# 2 The standard SIR epidemic model

In this chapter we present a simple model for the spread of an infectious disease. Several simplifying assumptions are made. In particular, the population is assumed to be closed, homogeneous and homogeneously mixing. Also, the effects of latent periods, change in behaviour, time varying infectivity and temporary or partial immunity are not taken into account. In later chapters we shall indicate ways of handling some of these complicating features of a real-life epidemic.

## 2.1 Definition of the model

We assume that initially there are $m$ infectious individuals (that have just become infected) and $n$ susceptible individuals. The infectious periods of different infectives are independent and identically distributed according to some random variable $I$, having an arbitrary but specified distribution. During her infectious period an infective makes contacts with a given individual at the time points of a time homogeneous Poisson process with intensity $\lambda/n$. If a contacted individual is still susceptible, then she becomes infectious and is immediately able to infect other individuals. An individual is considered 'removed' once her infectious period has terminated, and is then immune to new infections, playing no further part in the epidemic spread. The epidemic ceases as soon as there are no more infectious individuals present in the population. All Poisson processes are assumed to be independent of each other; they are also independent of the infectious periods.

We call this model the *standard SIR epidemic model*, the letters S, **I**, R standing for the terms 'susceptible', 'infectious' and 'removed', respectively. Following Ball (1995), we denote the process by $E_{n,m}(\lambda, I)$. Also denote the mean and the variance of the infectious period $I$ by $\iota$ and $\sigma^2$, respectively. The rate of contacting a given individual is set to $\lambda/n$ in order to keep the rate at which a given infective makes contact with other (initially susceptible) individuals constant ($= \lambda$), independently of the population size. The special case where the infectious period has an exponential distribution will be discussed in some detail further on.

The basic reproduction number and the final epidemic size, already encountered in the Introduction, are two extremely important epidemiological quantities. The final size of the epidemic, $Z$, is simply defined as the number of initially susceptible individuals that ultimately become infected. Thus $Z$ is a finite random variable taking values between 0 and $n$. In Section 2.4 we shall derive a linear system of equations for the distribution of the final epidemic size.

The basic reproduction number, $R_0$, is a little more difficult to describe. For this simple model $R_0$ is conveniently defined as the expected number of infections generated by one infectious individual in a large susceptible population. For the

model presented above $R_0 = \lambda\iota$ since $\iota$ is the average length of the infectious period and during the infectious period an infectious individual has contact with initially susceptible individuals at rate $\lambda$. The branching approximation of Section 3.3 will in a rigorous way give the basic reproduction number the interpretation of a critical parameter indicating whether a large outbreak is possible or not. For epidemic models with different types of heterogeneities, it is not always obvious how $R_0$ should be defined. We shall return to this problem in Chapters 6 and 7.

## 2.2 The Sellke construction

The following alternative elegant construction of the standard SIR epidemic model is based on Sellke (1983). We keep track of the total 'infection pressure' generated by the infectious individuals. Each susceptible individual is associated with a critical level of 'exposure to infection', and as soon as the infection pressure reaches this level, the susceptible becomes infected. We call this level the *threshold* of the individual. This purely mathematical construction does not reflect any properties of a real-life epidemic but it will serve as an important tool in the derivation of several results in later chapters.

Label the initial infectives $-(m-1), -(m-2), \ldots, 0$ and the initial susceptibles $1, 2, \ldots, n$. Let $I_{-(m-1)}, I_{-(m-2)}, \ldots, I_n$ be independent and identically distributed random variables, each distributed according to $I$. Also, let $Q_1, Q_2, \ldots, Q_n$ be an independent sequence of independent and identically distributed exponential random variables, having mean 1. These are the individual thresholds. For $i = -(m-1), -(m-2), \ldots, 0$, the initial infective labelled $i$ remains infectious for a time $I_i$ and is then removed. Denote by $Y(t)$ the number of infectives at time $t$, and let

$$A(t) = \frac{\lambda}{n} \int_0^t Y(u)\, du \tag{2.1}$$

be the total infection pressure exerted on a given susceptible up to time $t$. Note that in $A(t)$ the infectives are weighted according to their infectious periods. For $i = 1, 2, \ldots, n$, the susceptible labelled $i$ becomes infected when $A(t)$ reaches $Q_i$. The $j$th susceptible who becomes infected (*not* necessarily the susceptible labelled $j$!) remains infectious for a time $I_j$ and is then removed. The epidemic ceases when there are no more infectives present.

In Figure 2.1 we have plotted the total infection pressure $A(t)$ against $t$ for an epidemic starting with one infectious individual ($m = 1$). On the $y$-axis we have indicated the smallest individual thresholds ($Q_{(i)}$ denotes the $i$th order statistic) and horizontally the corresponding infectious period translated in time to the instant when the individual becomes infected. Note that the slope of $A(t)$ is proportional to the number of infectious periods covering the time point $t$ (i.e. the number of infectives $Y(t)$!) as it should be according to the definition of $A$ given in (2.1).



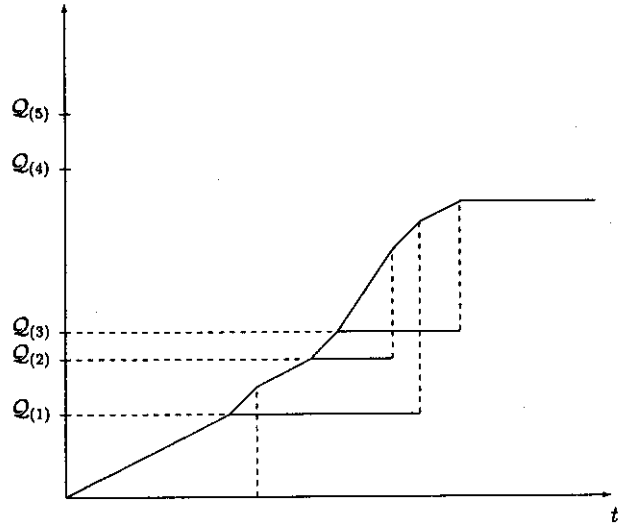


Figure 2.1: A typical realisation of the total infection pressure with $m = 1$ initially infectious individual. Note that the infection pressure never reaches $Q_{(4)}$ so the epidemic stops and the final size is $Z = 3$.

Let us check that this construction gives a process equivalent to the standard SIR epidemic. The infectious periods follow the correct distribution and, if $Y(t) = y$ and the individual labelled $i$ is still susceptible at time $t$, then she will become infected during $(t, t+\Delta t)$ with probability $\lambda y \Delta t/n + o(\Delta t)$. Indeed, owing to the lack-of-memory property of the individual threshold $Q_i$, the probability of the complementary event is given by

$$\begin{aligned} \mathbf{P}(Q_i > A(t+\Delta t) \,|\, Q_i > A(t)) &= \mathbf{P}(Q_i > A(t+\Delta t) - A(t)) \\ &= e^{-[A(t+\Delta t)-A(t)]} \\ &= \exp\left(-\frac{\lambda}{n} y \Delta t + o(\Delta t)\right) \\ &= 1 - \lambda y \Delta t/n + o(\Delta t), \end{aligned}$$

where the third equality follows from the definition of $A(t)$, since no infections will occur in a small enough time increment, making $Y(u)$ constant and equal to $y$ in $(t, t+\Delta t)$. In the original formulation of the model, our susceptible individual is contacted according to the superposition of $y$ independent Poisson processes, each of intensity $\lambda/n$, giving rise to the same infection probability.

## 2.3 The Markovian case

Consider the standard SIR model $E_{n,m}(\lambda, I)$ and denote by $X(t)$ and $Y(t)$ the number of susceptibles and the number of infectives, respectively, at time $t$. The process $(X, Y) = \{(X(t), Y(t)); t \geq 0\}$ will be a Markov process if and only if the infectious period has the lack-of-memory property. Assume therefore that $I$ is exponentially distributed with intensity $\gamma$. Then the process $(X, Y)$ is governed by the following transition table:

| from | to | at rate |
|---|---|---|
| $(i,j)$ | $(i-1, j+1)$ | $\lambda ij/n$ |
| | $(i, j-1)$ | $\gamma j$ |

which follows immediately from the definition of the model. This model, which originated with Bartlett (1949), is known as the *general stochastic epidemic*, a name that now seems inappropriate, since the model has over the years been generalized in an innumerable number of ways. The assumption of an exponentially distributed infectious period is certainly not epidemiologically motivated, although with this assumption the mathematical analysis becomes much simpler. Notably, using Markov process theory we can obtain deterministic and diffusion approximations for the whole trajectory, which are valid for large population sizes (see Chapter 5). This is usually hard to achieve when the stochastic process is not Markovian. On the other hand, the modern probabilistic methods used in this text to derive branching process approximations (Section 3.3) together with results for the final epidemic size (Section



2.4 and Chapter 4) do *not* rely on the Markov property, but can be carried out for all instances of the standard SIR epidemic defined in Section 2.1.

## 2.4 Exact results

Consider again the standard SIR epidemic $E_{n,m}(\lambda, I)$. We will derive a triangular linear system of equations for $P^n = (P_0^n, P_1^n, \ldots, P_n^n)$, where $P_k^n$ is the probability that $k$ of the initial susceptibles are ultimately infected.

Let $Z$ be the final size of the epidemic, and let $A = A(\infty) = \frac{\lambda}{n}\int_0^\infty Y(u)\,du$ be the total pressure of the epidemic. Recall the Sellke construction above. Both the final size and the total pressure can be expressed in terms of the infectious periods and the individual thresholds. First,

$$Z = \min\left\{ i : Q_{(i+1)} > \frac{\lambda}{n} \sum_{j=-(m-1)}^{i} I_j \right\},$$

where $Q_{(1)}, Q_{(2)}, \ldots, Q_{(n)}$ are the order statistics of $Q_1, Q_2, \ldots, Q_n$, since the epidemic stops as soon as the infection pressure generated by the previously infected individuals is insufficient to infect any more susceptibles. Also,

$$A = \frac{\lambda}{n} \sum_{j=-(m-1)}^{Z} I_j,$$

which is just another way of writing $A(\infty)$.

It is thus clear that the final size and the total pressure are intimately related. In fact, we have the following Wald's identity for epidemics (Ball, 1986):

**Lemma 2.1** *Consider the standard SIR epidemic $E_{n,m}(\lambda, I)$ and let $A$ be as above. Then*

$$E\left[e^{-\theta A}/\phi(\lambda\theta/n)^{Z+m}\right] = 1, \qquad \theta \geq 0,$$

*where $\phi(\theta) = E[\exp(-\theta I)]$ is the Laplace transform of $I$.*

*Proof.* To prove the identity, we note that

$$\begin{aligned} \left(\phi(\lambda\theta/n)\right)^{n+m} &= E\left[\exp\left(-\frac{\lambda\theta}{n}\sum_{j=-(m-1)}^{n} I_j\right)\right] \\ &= E\left[\exp\left(-\theta\left(A + \frac{\lambda}{n}\sum_{j=Z+1}^{n} I_j\right)\right)\right] \\ &= E\left[e^{-\theta A}\left(\phi(\lambda\theta/n)\right)^{n-Z}\right], \end{aligned}$$

where the last identity follows since the variables $I_j$, $j \geq Z+1$, are independent of both $Z$ and $A$. •

We are now in position to derive the system of equations for $P^n = (P_0^n, \ldots, P_n^n)$. For each $k \geq 1$, define $\mathbf{K}$ to be the set $\{1, 2, \ldots, k\}$; also, let $\mathbf{0}$ be the empty set. Recall that the initial susceptibles are labelled $1, 2, \ldots, n$. $P_k^n$ is the probability that $k$ initial susceptibles are infected in the $E_{n,m}(\lambda, I)$ epidemic, and $P_{\mathbf{K}}^n$ is the probability that precisely the set $\mathbf{K}$ is infected. By symmetry, $P_k^n = \binom{n}{k} P_{\mathbf{K}}^n$.

Now fix $k$ and choose $\ell$ such that $0 \leq k \leq \ell \leq n$, implying that $\mathbf{K} \subseteq \mathbf{L} \subseteq \mathbf{N}$. We use the notion of infection pressure to compare an epidemic within $\mathbf{N}$ with a sub-epidemic within $\mathbf{L}$. The event that an epidemic within $\mathbf{N}$ infects precisely the set $\mathbf{K}$ is the same as the event that a sub-epidemic within $\mathbf{L}$ infects precisely $\mathbf{K}$, and that these $k$ new infectives, together with the $m$ initial infectives, fail to infect any of the individuals in the set $\mathbf{N} \setminus \mathbf{L}$. We know from the Sellke construction that the probability of avoiding the infection is given by $\exp(-a)$, given that the sub-epidemic has generated the infection pressure $A^\ell = a$. It follows that

$$P_{\mathbf{K}}^n = P_{\mathbf{K}}^\ell E[\exp(-A^\ell(n-\ell)) \mid Z^\ell = k],$$

where $Z^\ell$ is the final size of the sub-epidemic. This equation is equivalent to

$$\frac{\binom{\ell}{k} P_k^n}{\binom{n}{k}} = P_k^\ell E[\exp(-A^\ell(n-\ell)) \mid Z^\ell = k]. \tag{2.2}$$

Can't just compute this directly by letting $Z > k$ and taking finite sum of iid rvs because the terms correlated with their size.

Now let us use Wald's identity (Lemma 2.1) applied to the sub-epidemic and with $\theta = n - \ell$ to get

$$E\left[e^{-A^\ell(n-\ell)} / [\phi(\lambda(n-\ell)/n)]^{Z^\ell+m}\right] = 1,$$

or, conditioning on the final size $Z^\ell$,

$$\sum_{k=0}^{\ell} \frac{P_k^\ell E\left(\exp(-A^\ell(n-\ell)) \mid Z^\ell = k\right)}{[\phi(\lambda(n-\ell)/n)]^{k+m}} = 1. \tag{2.3}$$

Equations (2.2) and (2.3) immediately give us

$$\sum_{k=0}^{\ell} \frac{\binom{\ell}{k} P_k^n}{\binom{n}{k} [\phi(\lambda(n-\ell)/n)]^{k+m}} = 1.$$

Finally, noting that $\binom{\ell}{k} / \binom{n}{k} = \binom{n-k}{\ell-k} / \binom{n}{\ell}$, we arrive at the following result:

**Theorem 2.2** *Consider the standard SIR epidemic $E_{n,m}(\lambda, I)$. Denote by $P_k^n$ the probability that the final size of the epidemic is equal to $k$, $0 \leq k \leq n$. Then*

$$\sum_{k=0}^{\ell} \binom{n-k}{\ell-k} P_k^n / [\phi(\lambda(n-\ell)/n)]^{k+m} = \binom{n}{\ell}, \qquad 0 \leq \ell \leq n. \tag{2.4}$$



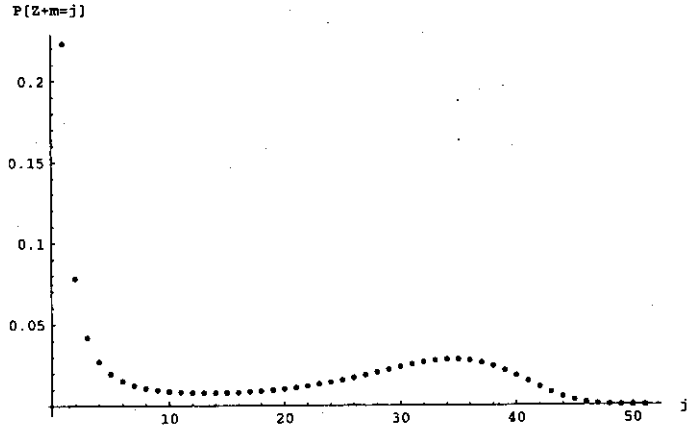


Figure 2.2: The exact distribution of $Z + m$ for $m = 1$, $n = 50$, $\lambda = 1.5$ and $I \equiv 1$, i.e. the infectious period is constant and equal to 1.

Note that, since the system of equations is triangular, the final-size-probabilities can be solved recursively. The proof of the theorem depends on the *infection pressure* generated by the various infectives, rather than the actual infectious periods. This indicates that we may allow for latent periods and time-dependent infection rates in the modelling, and still get the same type of final size results, as long as the required infection pressure can be calculated. That this is indeed the case will become even clearer in Section 7.1 where the concept of random graphs is used to describe the epidemic flow through a homogeneous and uniformly mixing population. In Figure 2.2 the probabilities $P_0^n, P_1^n, \ldots, P_n^n$, are plotted for a specific choice of community and parameter values (the figure actually shows the plot of $P(Z + m = k)$ for different values of $k$ but since $m = 1$ this corresponds to $P_{k-1}^n$). When the infectious period is constant, $I \equiv c$ say, an infectious individual infects susceptible individuals independently (with probability $p = 1 - e^{-\lambda c}$); the distribution of the final size is then equivalent to the Reed-Frost model defined for discrete time dynamics in Section 1.2. It is seen in the figure that the distribution is bimodal: either a few individuals are infected or else a fairly large number are infected. This qualitative behavior becomes more and more evident as $n$, the initial number of susceptibles, increases. A mathematical proof ofthis is given by the threshold limit theorem of Chapter 4.

Finally we mention briefly the very elegant theory developed by Lefèvre and Picard in a series of papers (see e.g. Lefèvre and Picard, 1990). They work with quite general classes of stochastic epidemic models, both single-type and multitype (cf. Chapter 6), and derive e.g. equations for the final size distribution and the total force of infection

$A(\infty)$, using a non-standard family of polynomials initially introduced by Gontcharoff (1937). However, we have decided not to include any presentation of their work, since the approach is rather algebraic in nature, and hardly increases the intuitive understanding of the models treated here.

## Exercises

**2.1.** Compute $P_0^n$, $P_1^n$ and $P_2^n$ numerically using the recursive formula given by (2.4) assuming $n = 10$, $m = 1$, $\lambda = 2$ and that the infectious period $I$ is:

a) exponentially distributed (the Markovian case) with mean 1 time unit.

b) $\Gamma(2, 2)$-distributed (i.e. with mean 1).

c) constant and equal to 1.

**2.2.** Assume $m = 1$, $\lambda = 1$ and that the infectious period is constant with mean 1 time unit. Use your favorite computer and mathematical software to see when the recursive formula of Section 2.4 breaks down numerically by computing and plotting $P_1^n, \ldots, P_n^n$ for $n = 10, 20, 30, \ldots$ (for most computers negative probabilities start appearing around $n \approx 70 - 90$).

**2.3.** Modify the standard SIR epidemic so that the infectious period is $= \infty$. (This model is denoted the SI model since individuals never get removed. It also has applications in sociology for the spread of rumours/knowledge. Infectious then corresponds to knowing, and spreading, the rumour.) For this model $X(t) + Y(t) = n + m$ for all $t$. Assume $m = 1$. Describe the random process $Y = \{Y(t); t \geq 0\}$, of infectious individuals. Calculate the expected waiting time until everyone in the population becomes infected. (Hint: Consider the consecutive waiting times between infections.) What happens with this expression when $n$ gets large?

# 3 Coupling methods

Let us assume that we are interested in comparing two or more random elements with each other. It is sometimes possible to construct versions of these random elements on the same probability space, in such a way that the comparison suddenly becomes easy (indeed, often trivial) to carry out. This procedure is called *coupling*, the term referring to the fact that the random elements so constructed are often highly dependent. The coupling method has found many important applications in various fields of probability theory, including Markov processes, renewal processes and Poisson approximation. The book by Lindvall (1992) provides a nice introduction to the subject.

Here we introduce some classical coupling ideas by providing simple examples. Then, after presenting the formal definition, we describe some applications of the coupling method to the standard SIR epidemic model $E_{n,m}(\lambda, I)$. First, it is shown that the number of infectious individuals in a large population initially behaves like a branching process. By coupling $E_{n,m}(\lambda, I)$ with a branching process, we justify the approximation of the epidemic by the simpler and thoroughly analysed branching process. This result indicates at the same time the significance of the basic reproduction number $R_0$. Second, by coupling two epidemics with different contact parameter $\lambda$, we prove the intuitively obvious fact that the accumulated number of infected individuals at a given time grows (in a sense yet to be defined) with $\lambda$, the infectiousness of the disease.

## 3.1 First examples

*Gambler's ruin problem*

Consider first the standard gambler's ruin problem. An individual with an initial capital of $m$ units of money goes to a casino. He plays a series of independent games, in each of which he wins one unit with probability $\theta$ and loses one unit with probability $1 - \theta$. He continues until either his capital reaches $n$ ($n > m$) or he goes bankrupt. We wish to use coupling to prove that $P(\theta)$, the probability of reaching the capital $n$ given $\theta$, is increasing in $\theta$ (as would be expected).

To this end, let $U_1, U_2, \ldots$ be independent and identically distributed random variables, each uniformly distributed on the interval $(0, 1)$. Then, for a given $\theta$, define

$$Y_\theta(i) = \begin{cases} +1 & \text{if } U_i \leq \theta, \\ -1 & \text{otherwise,} \end{cases}$$

$i = 1, 2, \ldots$, and

$$X_\theta(\nu) = m + \sum_{i=1}^{\nu} Y_\theta(i),$$