MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **6**, Number **3**, July **2009**

doi:10.3934/mbe.2009.6.509

pp. **509–520**

APPLICATIONS OF OCCUPANCY URN MODELS TO EPIDEMIOLOGY

Carlos M. Hernández-Suárez

Facultad de Ciencias, Universidad de Colima
Apdo. Postal 25, Colima, Colima, México

Oliver Mendoza-Cano

Instituto de Innovación y Desarrollo Tecnológico, Universidad de Colima
Apdo. Postal 25, Colima, Colima, México

(Communicated by Carlos Castillo-Chavez)

Abstract. This paper shows how occupancy urn models can be used to derive useful results in epidemiology. First we show how simple epidemic models can be re-interpreted in terms of occupancy problems. We use this reformulation to derive an expression for the expected epidemic size, that is, the total number of infected at the end of an outbreak. We also use this approach to derive point and interval estimates of the *Basic Reproduction Ratio*, R_0. We show that this construction does not require that the underlying SIR model be a homogeneous Poisson process, leading to a geometric distribution for the number of contacts before removal, but instead it supports a general distribution. The urn model construction is easy to handle and represents a rich field for further exploitation.

1. **Introduction.** Epidemic models are a basic tool to understand the main factors driving an epidemic. Among these, two stochastic models stand for their simplicity and usefulness: the SIS (Susceptible – Infected – Susceptible) and the SIR (Susceptible – Infected – Recovered) models without vital dynamics (no births or deaths). In the first, an individual may become infected by infectious individual and after a period of time characteristic of the disease, become susceptible again. In the SIR model, permanent immunity follows the infectious period. This epidemic models are formulated via a system of differential equations that defines the rates at which the different events (infections, removals, births, deaths, etc.) occur. Most systems can only be solved numerically due to non-linearity terms or emphasis is put on the asymptotic behavior of the model. Efforts to evaluate the impact of changing parameters on the behavior of the epidemic are carried out in the search for efficient control polices (see [2], [1], [7], [23] and [10]).

The SIR model dynamics differ from those of the SIS in important ways. The progressive depletion of susceptible individuals (without replacement) implies that there is a small probability that the epidemic will affect the whole population. This is true even for epidemics with unusually high transmission parameters. Thus, among the most important asymptotic characteristics of this epidemic model is its final epidemic size, the total number of individuals that were infected over the

2000 *Mathematics Subject Classification.* Primary: 62P10; Secondary: 92B05.
Key words and phrases. Urn models; R_0; basic reproductive ratio; stochastic epidemic models.
The first author is supported by Fondo Mixto Conacyt-Colima Grant 2007-C02-74994.

outbreak. The final asymptotic epidemic size is closely related to the cost of the epidemic, which can be thought of a linear function of the summation of the duration of the infectious period of all those infected.

The estimation of R_0, the basic reproduction ratio is also critically important. R_0 is defined as *the average number of secondary infections caused by an infectious individual when introduced to a population of fully susceptibles* [12]. Point estimates of R_0 under the underlying assumption that all rates can be thought of as parameters of exponential distributions –due to the fact that stochastic SIR models are assumed to be homogeneous Poisson process– do exist, but confidence intervals for R_0 are more or less approximate [8].

Urn models provide a natural framework for analyzing epidemic models. Their simple construction implies that their essence is easy to capture. A rich source of information on the topic with applications to biology, engineering, medicine, physics etc. can be found in [21]. These models have recently provided a way to construct models of infections that allow to estimate the efficacy of vaccines under very robust settings [18], [19]. By placing balls at random in a set of N empty urns, and considering every ball as a *threat* of infection, empty and occupied urns can be thought of as susceptible and infected respectively. The process of infection can be reproduced under this simple paradigm, we focus on the number of occupied urns and from this information we will make inferences on the total number of balls thrown. From these last two pieces of information we will make inferences on the reproductive ability of the disease. Conversely, we will show that if what is known is the observed number of infections at the end of the outbreak, the epidemic size can be estimated.

The organization of this paper is as follows: first we review the basic stochastic SIR model and introduce some basic results from occupancy distributions. We use these results to find expressions for epidemic size and for the construction of point and interval estimates of R_0.

2. Review of the SIR model.

In a stochastic SIR model, every infectious individual has contacts with individuals chosen at random at the rate λ. Contact is any activity that will result in a susceptible becoming infected by an infectious individual. Infected individuals recover at the rate μ. In stochastic models, both parameters, λ and μ are assumed to be parameters of exponential distributions, that is, both events, infection and recovery are modeled as a homogeneous Poisson process. The total population size $N = I + S + R$ is assumed to be constant. Then the state of the process at time t can be identified with $X(t) = \{S_t, I_t\}$. That is, the number of infected and recovered individuals at time t. When there are I_t infected and S_t susceptible, the transition probabilities are:

$$\begin{aligned} P(X_{t+\delta} = \{S_t - 1, I_t + 1\} \mid X_t = \{S_t, I_t\}) &= \lambda I_t S_t / N + o(\delta) \\ P(X_{t+\delta} = \{S_t, I_t - 1\} \mid X_t = \{S_t, I_t\}) &= \mu I_t + o(\delta). \end{aligned} \tag{1}$$

The main parameter driving the epidemic is the *basic reproductive ratio*, R_0. Since the number of infections that an individual will cause is clearly influenced by the number of susceptibles available, it is customary to define R_0 as the expected number of infections that an infected individual causes when it is introduced in a population of susceptibles [12]. There is a continuous depletion of susceptibles in this model and thus the first infectious individual has a larger expected number of secondary infections than the second one, and so on. Thus, if the expected

number of infections caused by the first individual is less or equal to one, from the theory of branching processes, we conclude that the epidemic will die out with probability one. In fact, when $N \to \infty$, by approximating the epidemic process with a branching process, the probability of eventual extinction starting with n initially infected individuals and large N is approximately equal to the $\text{Min}(1, R_0^{-n})$.

While the number of infections caused by an infected individual during the length of its infectious period depends on the number of susceptibles available at the moment of infection, the number of contacts of an individual follows the same distribution for all individuals regardless of the time of infection. This distribution is easy to derive and as a requirement to construct an epidemic model using occupancy urn models, we will will use here the definition of R_0 adopted in [20]:

R_0 is the expected number of contacts that an infectious individual has during its entire infectious period, where a contact between two individuals, one susceptible and one infected, is any act that would cause the infection of the susceptible.

A review on R_0, its properties and general methods for calculation can be found in [13], [17]. A simple algorithm for its computation based on Markov chains that can be applied to complicated stochastic epidemic models is presented in [20].

In the SIR stochastic model (1), it can be shown that the number of contacts of an infective individual during its entire infectious life is a geometric random variable with parameter $\mu/(\lambda+\mu)$, that is,

$$P(X=k) = \left(\frac{\lambda}{\lambda+\mu}\right)^k \frac{\mu}{\beta+\mu}, \qquad k = 0, 1, 2, ... \tag{2}$$

with expected value λ/μ, which is the definition of R_0. Therefore, in terms of R_0, the probability mass function of the number of infected can be written as

$$P(X=k) = \left(\frac{R_0}{1+R_0}\right)^k (1+R_0)^{-1}, \qquad k = 0, 1, 2, ...$$

An additional important result that will complete our brief account of R_0 is the following: in a seminal paper, [24] derived an expression for the epidemic size of a deterministic SIR epidemic. They found it to be the unique root of the equation of

$$x = N(1 - e^{-R_0 x/N}). \tag{3}$$

For stochastic SIR models, it is known that epidemics can be "small " or "large" with a given probability. For instance, there is a $(R_0+1)^{-1}$ probability that the initial infected individual will recover before infecting anyone else, hence, for $R_0 = 3$, approximately 25 percent of "outbreaks" will end with only one infected. Figure 1 shows the distribution of the epidemic size for two different values of R_0. We can observe small and large outbreaks. Several versions of stochastic models lead to a Poisson distribution for the number of ultimate susceptibles with a mean given by the deterministic solution of (3) (see [11] and [26]), or to a normal distribution with that mean [5], [6]. These results are based in the stochastic process given by system (1). See also [3] and [4].

3. **Occupancy distributions.** An extensive account of urn models is found in [21] and in [22]. Basic occupancy models relate to the distribution of some random variables associated with the placement of a fixed number of balls b on a fixed number of urns N at random, meaning that every ball can occupy any particular urn with probability N^{-1}. Perhaps one of the most simple random variables is the

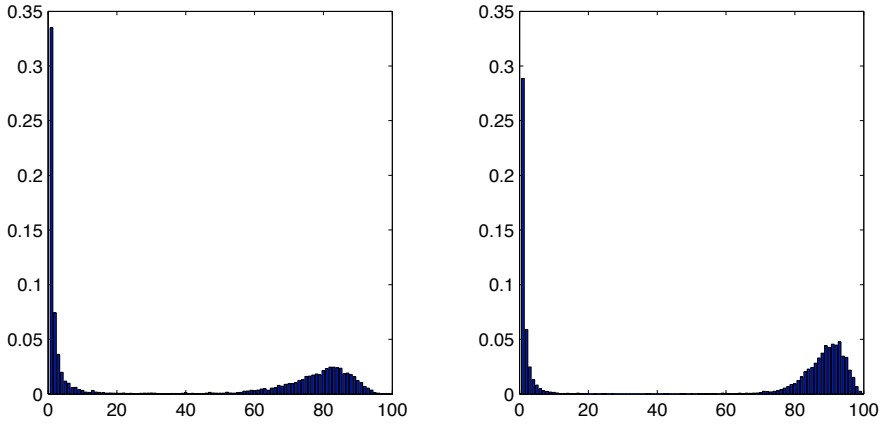

FIGURE 1. Distribution of the epidemic size for $R_0 = 2.0$ and $R_0 = 2.5$, showing the relative frequency of large and small epidemics. Population size is $N = 100$.

distribution of empty urns. The probability that *exactly* x urns are empty when b balls are thrown to N urns is

$$P(X = x) = \sum_{j=x}^{N} (-1)^{j-x} \binom{j}{x} \binom{N}{j} \left(\frac{x}{j}\right) \left(\frac{N-j}{N}\right)^b. \tag{4}$$

The factorial moments for the number of empty cells are given elsewhere (see [22], p. 415). In particular

$$E[X] = (N-1)^b N^{1-b}. \tag{5}$$

which tends to $Ne^{-b/N}$ for moderate N and b. In [25] it is shown that the distribution tends to a Poisson distribution with parameter $Ne^{-b/N}$ when both b and N tend to infinity as long as $Ne^{-b/N}$ remains bounded.

4. **An alternative construction of an SIR epidemic using urn models.** Here we show how to construct an SIR epidemic model via an occupancy urn process. Begin with N urns, labeled $1, 2, 3, ..., N$ that can accommodate an unlimited number of balls each. We start with a single occupied urn. This urn will throw to the remaining $N-1$ urns a random number of balls described by the random variable Y_i, where Y_i is a geometric random variable as in (2). Every ball can fall in any one of the remaining $N-1$ urns with the same probability. Thus, urns will end up empty or occupied, depending on whether they contain no balls or at least one. When an empty urn becomes occupied –which occurs when it receives a ball for the first time–, it throws in turn a random number of balls Y_i on the complementary $N-1$ urns. The process repeats until there are no newly occupied urns. In this construction, we have incorporated all the elements of an SIR model. The fact that a newly occupied urn only throws a random number of balls when it was first occupied accounts for the fact that subsequent attempts of infection on an infected or recovered individual do not affect the infectiousness of individuals. As in an SIR epidemic model, the epidemic halts when the last ball thrown falls in an already occupied urn.

5. **Applications.**

5.1. **Epidemic size.** When the outbreak is over, one only identifies the number of occupied urns (infected individuals) and not the number of balls they contain. Therefore, there is no information on the total number of balls thrown. Hereafter the term infected and occupied are used indistinctly. The same applies to the terms susceptible and empty. Assume that the outbreak ends up with X infected individuals. Hence, the total number of balls thrown is given by the random sum

$$b = \sum_{i=1}^{X} Y_i, \qquad i = 1, 2, 3, .., X. \tag{6}$$

If we denote X as the number of occupied urns when b balls are thrown in N urns, then using

$$\begin{aligned} E[X] &= E[E[X|b]] \\ &= E[N(1 - e^{-b/N})] \end{aligned}$$

which can be approximated with $N(1 - e^{-E[b]/N})$. Since b is a random sum, its expected value equals $E[X]\ E[Y_i]$ if Y_i is independent of X, which is not the case. For instance, if we observe a small X –value, then it is more likely that this is due to low values of Y_i, $i = 1, 2, 3, ..., X$. However, for moderately large values of X, X is independent of Y_i and thus we can use the approximation

$$E[b] \approx E[X]\ E[Y_i] = E[X]\ R_0,$$

that is,

$$E[X] = N(1 - e^{-E[X]R_0/N}), \tag{7}$$

which is (3), the solution for the deterministic model. This expression has no closed (analytical) solution for $E[X]$, but can be solved numerically. Kermack et al [24] provided approximations for its solution. This construction makes it evident that the number of contacts of an infected individual, Y_i, does not have to follow a geometric distribution, and in principle, it could be any discrete distribution as long as its expected value is R_0. This is an appealing result because the underlying stochastic process relies on homogeneous Poisson process with constant hazard rates, an assumption that is generally to restrictive in epidemiology in general. Whereas the inter-event times for the contact process may be assumed to be exponentially distributed, this is a more restricted assumption for the time to removal (duration of the infectious period) which in general has an increasing hazard rate.

In Table (1) we compare the observed epidemic size obtained from solving (7) numerically with the values obtained from stochastic simulations of the epidemic averaging over the simulations of "large" outbreaks only, for four different distributions with the same expected value R_0. There is no consensus on what is a "large" outbreak. Here we define it as any outbreak that ends up with more than 10 percent of the population infected. For geometric distributions, the probability of a small epidemic is R_0^{-1}, and thus the expected proportion of large epidemics is $1 - R_0^{-1}$ which is very close to the observed values of large epidemics for this distribution. It can be seen that the expected epidemic size conditional on the occurrence of a major outbreak is very similar for the four distributions tested: Geometric, Poisson, Binomial and Hypergeometric. It is clear that the numerical solution of (7) fits very well the observed values for "large" outbreaks for the four distributions.

TABLE 1. Epidemic size for several distributions of the number of contacts. All distributions have the same mean. Results from 1×10^4 simulations. $N = 1000$

R_0	Geometric $1/(R_0+1)$		Poisson R_0		Binomial $(10, R_0/10)$		Hypergeometric $(100, 10R_0, 10)$		Expected
	(∗)	(†)	(∗)	(†)	(∗)	(†)	(∗)	(†)	(‡)
1.2	0.15	319.3	0.29	311.2	0.33	312.5	0.34	312.4	313.7
1.5	0.33	577.8	0.57	581.2	0.62	581.3	0.64	581.4	582.8
2.0	0.49	794.5	0.80	795.5	0.85	796.1	0.86	796.3	796.8
2.5	0.60	891.9	0.89	892.3	0.93	892.2	0.94	892.6	892.6

(∗) Proportion of 'large' epidemics (exceeding 10 percent of the population size).

(†) Average observed epidemic size over 'large' epidemics.

(‡) Numerical solution of (7).

5.2. **Point and interval estimation of R_0.** In general, methods for estimating R_0, the key parameter governing an SIR epidemic are based on the analysis of the underlying stochastic model with transitions (1) (see [8]). These methods rely on asymptotic distributions for N large.

The exact distribution of the number of empty urns after placing b balls in N urns is given in (4). Observe that since all balls thrown (if any) by an infected individual must fall on the remaining $N-1$ urns, we must make a correction on the expected number of occupied urns after placing b balls at random, namely

$$\hat{\theta} = M\ e^{-b/M},$$

with $M = N - 1$. After observing k infections, the MLE of theta is $s = M - k$, the observed number of remaining susceptibles. By the invariance properties of the MLE's, that of b is $\hat{b} = -M\ log(s/M)$, and the MLE of the mean number of balls thrown per infected individual is

$$\widehat{R_0} = -M\ log(s/M)/(M-s), \tag{8}$$

which is the classical point estimate of R_0 obtained with deterministic or stochastic models (see [8]).

The urn model construction allows to see that the distribution of the number of balls produced by each newly infected should affect the variance of the epidemic size. Table (2) is complementary to Table (1) and shows the observed variances of the epidemic sizes obtained from the same distributions used in Table (1). In Table 2 all distributions have the same mean for each R_0, but the variances are such that $Var(X_G) > Var(X_P) > Var(X_B) > Var(X_H)$ where X_G, X_P, X_B and X_H are the random variables of the number of contacts per individual (balls) during the infectious period for the distributions: geometric, Poisson, binomial and hipergeometric respectively.

As we can see, the variances of the epidemic sizes maintain the same relationship among them than those of the distribution for the number of balls thrown by each individual, with the variance for the geometric distribution almost doubling the variance of the others. This implies that while point estimation of R_0 does not require assumptions on the distribution of the number of contacts per individual, interval estimation should be sensible to this distribution. In what follows we

TABLE 2. Variance of the epidemic size of large outbreaks for several distributions of the number of contacts. Results from 1×10^4 simulations. $N = 1000$

R_0	Geometric	Poisson	Binomial	Hypergeom.
	$1/(R_0+1)$	R_0	$(10, R_0/10)$	$(100, 10R_0, 10)$
1.2	10530.5	6908.5	5917.6	5388.7
1.5	3774.5	1843.3	1607.0	1539.4
2.0	837.1	458.4	423.7	416.2
2.5	310.8	178.4	171.1	167.9

introduce a method to derive a confidence interval (CI) for R_0 that results from occupancy models, based on the system of equations (1). This is equivalent to assuming that the number of contacts during the infectious period of an individual follows a geometric distribution.

Once the outbreak is over, we observe X individuals that were infected at some time. Every one of the infections was caused by the same infected individuals, and using the urn model construction we can think of X individuals that throw an unknown number of balls b to a fixed number of urns N, filling exactly X urns. This allows to establish our strategy to derive a CI for R_0: using the Probability Mass Function of the number of occupied cells when each of X individuals throw balls according to (2).

We begin by letting Y be the number of urns filled at the end of an occupancy process where each of X individuals throw balls according to the distribution (2). Letting $p = (1 + R_0)^{-1}$ and using (4) and (2) we have that:

$$P(Y = y; X, p) = \sum_{j=0}^{\infty} P(Y = y|W = j)P(W = j), \tag{9}$$

where W follows a negative binomial distribution with parameters X and p, and $P(Y = y|W = j)$ is the distribution of occupied urns when j balls are randomly placed in N urns. After some simplifications, we obtain

$$P(Y = y; X, p) = \sum_{j=N-y}^{N} (-1)^{j+y-N} p^X \left(\frac{j + Np - jp}{N}\right)^{-x} \binom{N-y}{j} \binom{N}{j} \binom{j}{N-y}. \tag{10}$$

In practice we are interested in evaluating $P(Y = X)$, the observed number of infected. A confidence interval (p_L, p_U) of size $1 - \alpha$ for p can be obtained (see [9]) using p_L and p_U such that

$$\sum_{j=0}^{X} P(Y = j; X, p_L) = \alpha; \sum_{j=0}^{X} P(Y = j; X, p_U) = 1 - \alpha, \tag{11}$$

From here a CI of the same size for R_0 is $((1-p_U)/p_U, (1-p_L)/p_L)$. Figures 2 and 3 show the shape of a 0.95 CI for R_0 for $N = 200$ and $N = 2 \times 10^4$ respectively, as a function of the epidemic size. At first glance it seems odd that the upper limit of the CI is higher for a very low epidemic size than that of an epidemic that decimated half the population. The explanation is simple: even high R_0 values can result in an epidemic of size 5, but as epidemic sizes increases, it is less likely that those

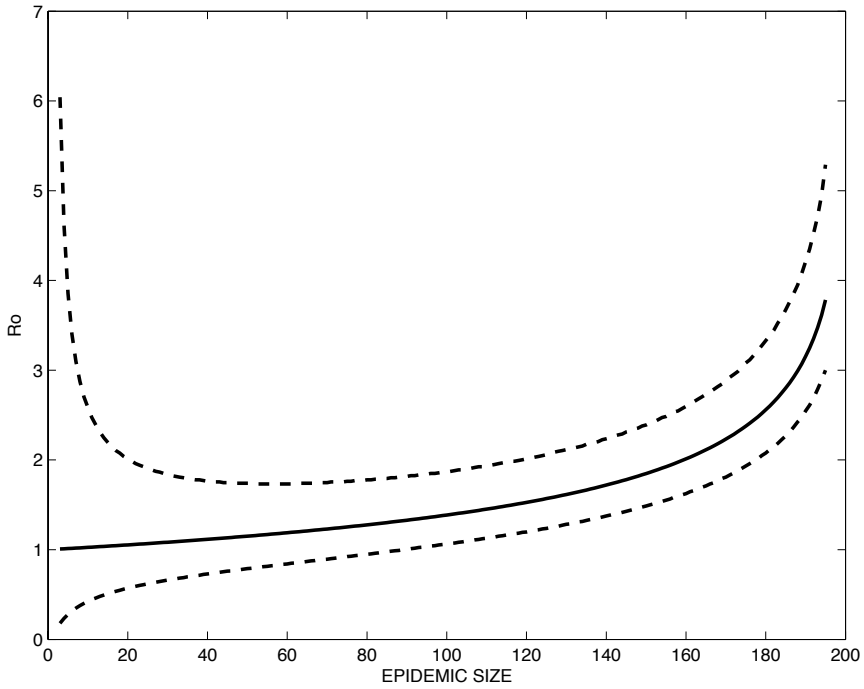

FIGURE 2. 95 percent confidence intervals for R_0 (dashed lines) at different values of epidemic size, for N=200. The solid line is the point estimate.

high R_0's could cause such sizes. The upper bound for CI tends also to increase dramatically when the epidemic size is close to N.

Computing confidence intervals for large values of N is computationally intensive, since expression (10) has to be iteratively computed. The intervals of Figure 2 were calculated using approximation to a normal distribution for both the total number of balls thrown and the number of empty urns. This approach yields approximate intervals that are computationally less intensive to calculate. According to this, we approximate the total number of balls thrown by the observed X infected, that is W in (9) with $h_W(p, X; w)$ a Normal distribution with mean $X(1-p)/p$ and variance $X(1-p)/p^2$, whereas the distribution of the number of occupied urns conditioning on a given number of balls thrown, that is $P(Y|W)$ in (9) can be approximated by $f_Y(N, w; y)$ a Normal distribution with mean $N(1-e^{-w/N})$ and variance $Ne^{-w/N}$. If we let

$$g_X(N, p; x) = \int_0^{-\infty} f_Y(N, w; y)\, h_W(X, p; w) dw \tag{12}$$

then the confidence intervals for R_0 are obtained using $(1-p_U)/p_U$, $(1-p_L)/p_L)$ with p_L and p_U such

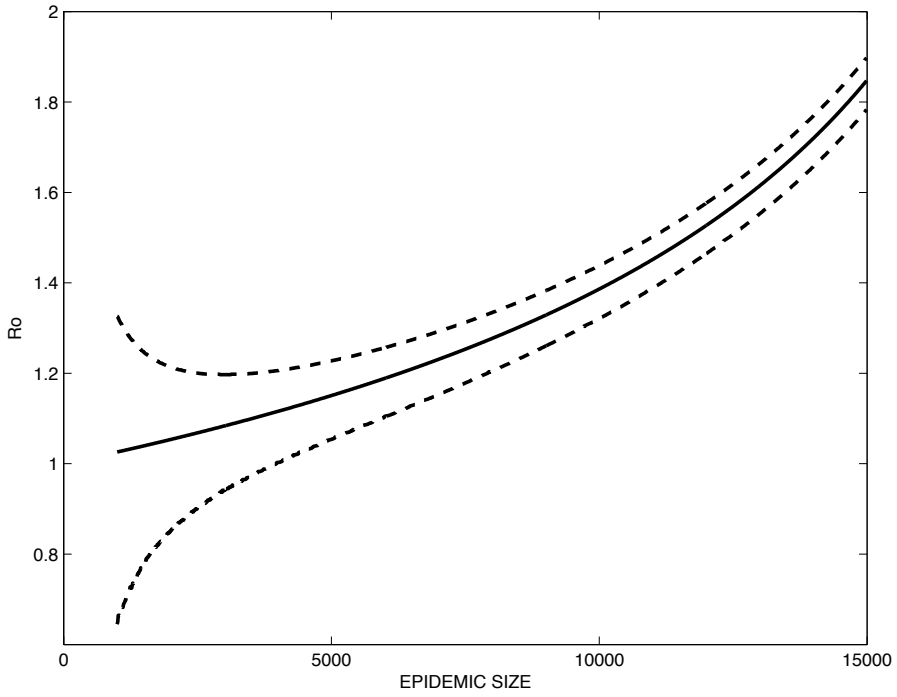

FIGURE 3. 95 percent confidence intervals for R_0 (dashed lines) at different values of epidemic size, for $N = 2 \times 10^4$. The solid line is the point estimate.

$$\int_0^X g_X(N, p_L; X) = \alpha; \int_0^X g_X(N, p_U; X) = 1 - \alpha. \tag{13}$$

6. **Discussion.** Becker [8] suggested a method for constructing confidence intervals that can be applied when final epidemic size data are available. For $N = 120$ and an observed epidemic size of $x = 30$ the point estimate is the same as (8), and a 95 percent confidence interval for R_0 is (1.02, 1.18). Using the method in this paper for small populations using (10) and (11) we get (0.71, 1.96). There is a huge difference in the size of both intervals not only in its width but more importantly, our interval includes 1, the threshold value. We believe that the interval constructed with this data should include 1 and support our argument as follows: Let Z be the final epidemic size of an SIR epidemic that started with one inital infected among a population of size N. Let $F_Z(z; R_0)$ the cdf of the epidemic size. Clearly, $F_Z(z; R_0)$ is decreasing with R_0 for every $z, 1 \leq z \leq N$, thus, after observing an epidemic size of x a lower bound for R_0 would be R_U such that

$$F_Z(x; R_U) = 1 - \alpha, \tag{14}$$

however, for $N = 120$ the population size is small and the density of the epidemic size can be obtained analytically. For $x = 30$ we have that $F_Z(30; 0.78) \approx 0.975$, that is, the exact lower value for an equal tail CI is less than 1.

R_0 will be in general underestimated: epidemic models do not include the natural decrease in the infectious activity of those infected in the later stages of the epidemic as well as in the exposure to infection, mainly due to an increase in the information available to the population. This implies that the expected number of balls thrown by an individual decreases with time by natural reasons. This has to be considered seriously since control polices are then based on an underestimated R_0 [8].

There are known models that include the possibility that a ball may *escape* (or may be *rejected*) from the urn with some constant probability. These models are called *leaky urns* models (see [22], p. 417 and references therein). These models may be useful to account for the efficacy of some protective measure against infection on an individual basis, the essence of *leaky* vaccines [15], [14]. If the probability that a ball will be retained is θ, then the expected number of empty cells tends to a Poisson distribution with mean $N\exp(-b\,\theta/N)$ as N and b become large under the assumption that the mean remains bounded (see [22], p. 418 and references therein). On the other hand, there are situations where a vaccine may fully protect a fraction of those vaccinated while leaving unprotected the remaining individuals [15], [14]. These are called *all/nothing* vaccines and urn models for these effects can be easily constructed. If b balls are placed at random in N urns and a fraction $1-\theta$ of the urns is covered whereas the remaining fraction θ is left 'susceptible', then the expected number of empty urns is $N\theta\exp(-b\,\theta/(N\theta)) = N\theta\exp(-b\,/N)$. This result is obtained from assuming that the total number of urns has been reduced to $N\theta$ and that every ball will have a probability θ of falling in a 'susceptible' urn.

Vaccines not only have a direct effect on the protection of an individual, but also have indirect effects [16]. One of the most important measures of the efficacy of a vaccine is the Population Vaccination Effectiveness (PVE) which is defined as one minus the ratio of two expectations: the expected number of infections under a given vaccination policy to the expected number of infections under no policy. Thus, PVE attempts to measure the reduction in the number of infections due to vaccination campaign, a reduction which is difficult to estimate since such reduction is the result of both the efficacy of the vaccine on an individual basis and *herd immunity*, the indirect protection gained by the fact that there will be less infectious pressure on individuals. Thus, once a vaccination policy has been applied it is very difficult to estimate what would be the number of infections under no vaccination at all. Urn models provide a suitable framework to differentiate both effects and to estimate the denominator of the PVE.

Distributions associated with placing at random m different types of balls have also been studied (see [21], [22]). These distributions include among others, the number of urns containing exactly (or at least) k types of balls, $k = 0, 1, 2, ..., m$. These distributions may be useful to model the spread of a disease that can be caused by different strains were immunity is strain specific (e.g. dengue fever). Sampling individuals to make inferences on the distribution of immunity generated by different strains may be useful in the development of control policies aimed to reduce the impact of seasonal outbreaks.

The fact that the expression for the epidemic size for large epidemics (7) and that of the point estimate of R_0 (8) do not require the assumption of exponential duration for both the time between contacts and the duration of the infectious state deserves special attention: while the former assumption may work, the assumption that time to recovery (or to removal) from an infected state has constant hazard rate is generally unrealistic. Unfortunately, the construction of confidence intervals

for R_0 does require assumptions on the distribution of the number of contacts made by an infectious individual.

It is true that (7) is the expected epidemic size conditioning on a large epidemic, and that the true epidemic size should weight this by considering the probability of a small epidemic, nevertheless, its importance should not be underestimated since control policies to reduce the cost of an epidemic should be conservative and hence probably based on a worst case scenarios.

We hope that the construction in this paper will increase the understanding of basic, yet useful epidemic models to professionals and applied researchers of those disciplines which are traditionally more exposed to the field of statistics than to dynamical systems or stochastic processes.

REFERENCES

[1] R. M. Anderson and R. M. May, "Infectious Diseases of Humans: Dynamics and Control," Oxford University Press, Oxford, 1992.

[2] M. T. Bailey, "An introduction to Stochastic Process," Charles Griffi, London, 1975.

[3] F. Ball, *A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models*, Adv. Appl. Probab., **20** (1986), 289–310.

[4] F. Ball and D. Clancy, *The final size and severity of a generalized stochastic multitype epidemic model*, Adv. Appl. Probab., **25** (1993), 721–736.

[5] A. D. Barbour, *The principle of diffusion of arbitrary constants*, J. Appl. Probab., **9** (1972), 519–541.

[6] A. D. Barbour, *On a functional central limit theorem for Markov population processes*, Adv. Appl. Probab., **6** (1974), 21–39.

[7] M. S. Bartlett, "An Introduction to Stochastic Process, with Special Reference to Methods and Applications," Cambridge University Press, New York, 1955.

[8] N. D. Becker, "Analysis of Infectious Disease Data," Chapman and Hall, London, 1989.

[9] G. Casella and R. L. Berger, "Statistical Inference," Wadsworth and Brooks/Cole, Pacific Grove, CA, USA, 1990.

[10] D. J. Daley and J. Gani, "Epidemic Modelling: an Introduction," Cambridge University Press, Cambridge, London, 1999.

[11] H. E. Daniels, *The distribution of the total size of an epidemic*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, (1967), 281–293.

[12] O. Diekmann, J. A. Heesterbeek and J. A. Metz, *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol., **28** (1990), 365–382.

[13] O. Diekmann, and J. A. Heesterbeek, "Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation," John Wiley and Sons, New York, 2000.

[14] M. Haber, I. M. Longini and M. E. Halloran, *Measures of the effects of vaccination in a randomly mixing population*, Int. J. Epidemiol., **20** (1999), 300–300.

[15] M. Haber, I. M. Longini and M. E. Halloran, *Estimation of vaccine efficacy in outbreaks of acute infectious diseases*, Int. J. Epidemiol., **10** (1991), 1573–1584.

[16] M. E. Halloran, M. Haber, I. M. Longini and C. J. Struchiner, *Direct and indirect effects in vaccine efficacy and effectiveness*, Am. J. Epidemiol., **133** (1991), 323–331.

[17] J. A. Heesterbeek, *The concept of R_0 in epidemic theory*, Stat. Neerl., **50** (1996), 89–110.

[18] C. M. Hernández-Suárez, *Urn models and vaccine efficacy estimation*, Stat. Med., **19** (2000), 827–835.

[19] C. M. Hernández-Suárez, *A note on the distribution of the number of vaccinated infected under non-random mixing conditions*, Stat. Med., **20** (2001), 1983–1986.

[20] C. M. Hernández-Suárez, *A Markov chain approach to calculate R_0 in stochastic epidemic models*, J. Theor. Biol., **215** (2002), 83–93.

[21] N. L. Johnson and S. Kotz, "Urn Models and Their Application. An Approach to Modern Discrete Probability Theory," John Wiley and Sons, Cambridge, New York, 1977.

[22] N. L. Johnson, A. W. Kemp, and S. Kotz "Univariate Discrete Distributions," John Wiley and Sons, New York, 1992.

[23] D. G. Kendall, *Deterministic and stochastic epidemics in closed populations*, in "Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability," University of California Press, (1956), 149–165.
[24] W. O. Kermack and A. G. McKendrick, *Contributions to the mathematical theory of epidemics, Part 1. (Reprinted from Proceedings of the Royal Society, Vol 115A, pp.700–21,1927*, Bull. Math. Biol., **53** (1991), 33–55.
[25] R. Von Mises, "Selected papers of R. Von Mises," American Mathematical Society, Providence, R.I. , 1963.
[26] T. Sellke, *On the asymptotic distribution of the size of a stochastic epidemic*, J. Appl. Probab., **20** (1983), 300–300.

Received February 19, 2008; revised October 22, 2008.

E-mail address: carlosmh@mac.com
E-mail address: oliverm@me.com