Stochastic CRNs

Lecture 1
Introduction

Grzegorz Rempala
The Ohio State University

December 4, 2019

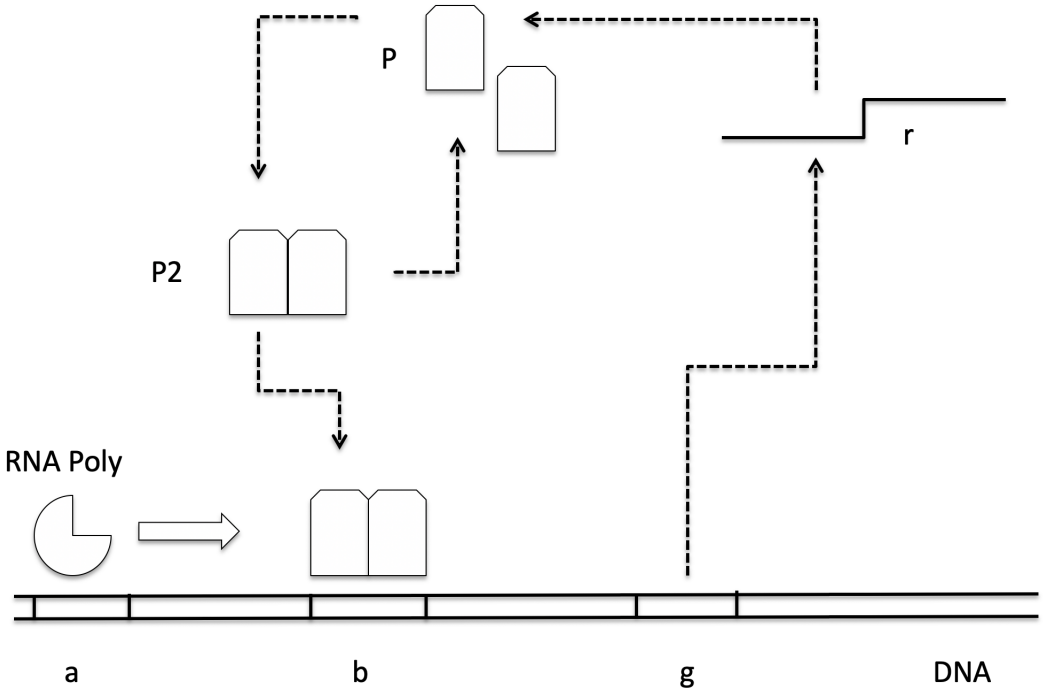
1 Systems biology models

- SB - studying of ways in which interactions between the components of a biological system give rise to its function and behavior
- Computational SB - building models of complex biological pathways, then validating and analyzing those models using a variety of methods, including time-course simulation
- CSB models represented as sets of (pseudo-)biochemical reactions: chemical reactions network (CRN)
- Classically, this leads to continuous and deterministic models (DMs: coupled ODE or PDE systems)
- However, stochastic models (SMs) are increasingly of interest
  - Much intra-cellular behavior (e.g., gene expression) is intrinsically stochastic
  - In macro models (eq. ID models) *probability* of disease transmission is more natural to study
  - Epidemic or chem. process initiation is a stochastic event (threshold thm)

1.1 SB and stochastic models

- *Law of large numbers (mean field theory)* may not apply so DMs may not offer good approximations
- Multi-scaling: some species may be present in much greater abundance than others
- The rate constants for the reactions may vary over several orders of magnitude
- Main tool: stochastic trajectory equations and Markov jump processes theory
    - Classical simulation techniques in chemical physics: *Gillespie direct algorithm (SSA) and its extensions*
    - Underlying assumptions: *law of mass action* (system in thermal equilibrium, well stirred)
    - CRNs offer a principled way of describing the system
- Examples: gene transcription kinetics and SIR epidemic

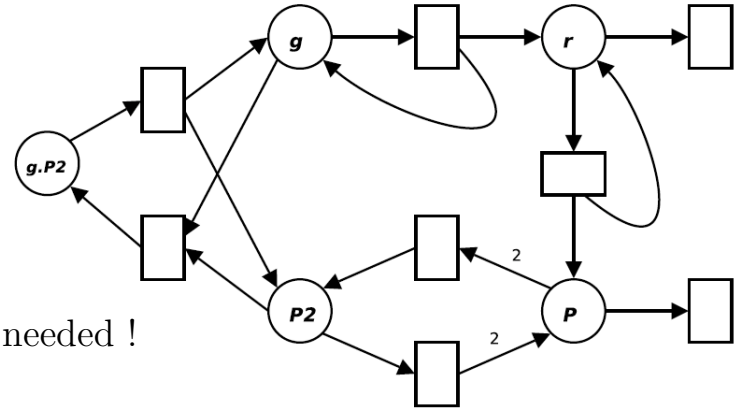
Example: genetic auto-regulation (Wilkinson 2011)



Reactions:

(1)	$g + P_2 \rightleftharpoons g \cdot P_2$	Repression
(2)	$g \rightharpoonup g + r$	Transcription
(3)	$r \rightharpoonup r + P$	Translation
(4)	$2P \rightleftharpoons P_2$	Dimerization
(5)	$r \rightharpoonup \emptyset$	mRNA degradation
(6)	$P \rightharpoonup \emptyset$	Protein degradation

Petri Net:



- Reaction rates needed !

1.2 Markov jump processes

- Assume X is a continuous time Markov chain in $E \subset Z^d$. The rate matrix, $Q = [q_{kl}]$ for the chain gives

$$P\{X(t+\Delta t) = l | X(t) = k\} \approx q_{kl}\Delta t, \quad k \neq l \in E,$$

and hence for f in an appropriate domain $\mathcal{D}(\mathbb{A})$

$$E[f(X(t+\Delta t)) - f(X(t))|\mathcal{H}_t^X] \approx \sum_l q_{X(t),l}(f(l) - f(X(t))\Delta t \equiv \mathbb{A}f(X(t))\Delta t$$

where $\mathcal{H}_t^X$ is the history of the process

- Alternative notation is via intensity. Define $\beta_l(k) = q_{k,k+l}$. Then

$$\mathbb{A}f(k) = \sum_l \beta_l(k)(f(k+l) - f(k)) \qquad \text{for any } f \in \mathcal{D}(\mathbb{A})$$

1.3 Martingale problem and master equation

- $\approx$ is made precise by the requirement that

$$M_f(t) = f(X(t)) - f(X(0)) - \int_0^t \mathbb{A}f(X(s))ds$$

be a $\{\mathcal{H}_t^X\}$-martingale for all f.

- X is called a *solution of the martingale problem* for $\mathbb{A}$.
- Defining $u(x,t) = E[f(X(t))|X(0) = x]$, one can derive the *backward equation*

$$\partial_t u(t,x) = \mathbb{A}u(t,x)$$

and setting $\nu_t(G) = P\{X(t) \in G\}$ and $\nu_t f = \int_E f d\nu_t$, the martingale property gives the *forward equation* (in weak form)

$$\nu_t f = \nu_0 f + \int_0^t \nu_s \mathbb{A} f ds, \quad f \in \mathcal{D}(\mathbb{A}).$$

- $f = \mathbf{1}_{\{k\}}$ and setting $p_k(t) = \nu_t(\{k\})$, gives the *master equation*

$$\dot{p}_k(t) = \sum_l p_{k-l}(t)\beta_l(k-l) - p_k(t)\sum_l \beta_l(k)$$

1.4 Rescaling into Poisson process

- Counting process $N(t)$ (i.e., increases by 1) adapted to filtration $\{\mathcal{H}_t\}$
- Process *intensity* (λ)

$$P(N(t+\Delta t) - N(t) > 0 | \mathcal{F}_t) = \lambda(t)\Delta t + o(\Delta t)$$

- Let $\tau(t)$ be a stopping time such that $\int_0^{\tau(t)} \lambda(s)ds = t$
- Define $Y(t) = N(\tau(t))$ then one can show $(Y(t) - t)$ is $\{\mathcal{H}_{\tau(t)}\}$-martingale
- Watanabe's thm implies that $Y(t)$ is a *unit Poisson process*

$$N(r) = Y(\int_0^r \lambda(s)ds)$$

1.5 Time change equation (TCE)

- Consider the process evolution

$$X(t) = X(0) + \sum_l l N_l(t)$$

where $N_l(t)$ is the number of jumps of l at or before time t. N_l is a counting process with intensity (*propensity* in the chemical literature) $\beta_l(X(t))$, that is,

$$N_l(t) - \int_0^t \beta_l(X(s))ds$$

is a martingale. Consequently, we can write

$$N_l(t) = Y_l(\int_0^t \beta_l(X(s))ds),$$

where the Y_l are independent, unit Poisson processes, and

$$X(t) = X(0) + \sum_l l Y_l(\int_0^t \beta_l(X(s))ds)$$

1.6 Simulation – Gillespie direct method (SSA)

1. For given state of the system $X(t) = k$ compute $\beta_l(k)$ for all M possible vals of l.
2. Simulate the reaction time as the minimum of M independent exponential variables with means $1/\beta_l(k)$ for $k = 1 \ldots, M$.
3. Simulate the reaction type r by drawing from M point discrete distribution where $P(r = r_l) = \beta_l(k)/\bar{\beta}(k)$.
4. Update the system state to new k and repeat.

- Not very efficient for large systems
- Lots of work over last decade to speed up SSA
- Quasi-steady state and τ-leaping methods most popular

2 Reaction networks

- Standard notation for chemical reactions

$$A + B \overset{k}{\rightharpoonup} C$$

is interpreted as *"a molecule of A combines with a molecule of B to give a molecule of C in the presence of a catalyst k" (or at rate k).*

$$A + B \rightleftharpoons C$$

means that the reaction can go in either direction, that is, a molecule of C can dissociate into a molecule of A and a molecule of B

- We consider a *network* of reactions involving m chemical species, $A_1, \ldots, A_m$.

$$\sum_{i=1}^{m} \nu_{ik} A_i \rightharpoonup \sum_{i=1}^{m} \nu'_{ik} A_i \qquad k = 1, 2, \ldots$$

where the ν_{ik} and ν'_{ik} are nonnegative integers

Definition 2.0.1. *A chemical reaction network is a triple* $\{\mathscr{S}, \mathscr{C}, \mathscr{R}\}$ *where*

(i) $\mathscr{S} = \{S_1, \ldots, S_n\}$ *is the set of species,*

(ii) $\mathscr{C}$ *is the set of complexes, consisting of nonnegative linear combinations of the species,*

(iii) $\mathscr{R} = \{y_k \to y'_k : y_k, y'_k \in \mathscr{C} \text{ and } y_k \neq y'_k\}$ *is the set of reactions.*

- For instance for the reaction network

$$S + E \rightleftharpoons C \to S + P, \quad E \rightleftharpoons \emptyset$$

$\mathscr{S} = \{S, E, C, P\}$,

$\mathscr{C} = \{S + E, C, S + P, E, \emptyset\}$, and

$\mathscr{R} = \{S + E \to C, C \to S + E, C \to S + P, E \to \emptyset, \quad \emptyset \to E\}$

2.1 Markov chain models

- $X(t)$ number of molecules of each species in the system at time t.
- ν_k number of molec's of each chemical species consumed in the kth reaction.
- ν'_k number of molecules of each species created by the kth reaction.
- $\lambda_k(x)$ rate at which the kth reaction occurs.
- If the kth reaction occurs at time t, the new state becomes

$$X(t) = X(t-) + \nu'_k - \nu_k.$$

The number of times that the kth reaction occurs by time t is given by the counting process satisfying

$$R_k(t) = Y_k(\int_0^t \lambda_k(X(s))ds),$$

where the Y_k are independent unit Poisson processes

$$\begin{aligned} X(t) &= X(0) + \sum_k R_k(t)(\nu'_k - \nu_k) \\ &= X(0) + \sum_k Y_k(\int_0^t \lambda_k(X(s))ds)(\nu'_k - \nu_k) = (\nu' - \nu)R(t) \end{aligned}$$

2.2 Law of mass action (LMA)

- Rates obtained via *the stochastic law of mass action*

$$\lambda_k^N(x) = \kappa_k \frac{\prod_i \nu_{ik}!}{N^{|\nu_k|-1}} \prod_i \binom{x_i}{\nu_{ik}} = N\kappa_k \frac{\prod_i \nu_{ik}!}{N^{|\nu_k|}} \prod_i \binom{x_i}{\nu_{ik}},$$

where $|\nu_k| = \sum_i \nu_{ik}$ and N is a *scaling parameter* usually taken to be the volume of the system times Avogadro's number

$$x_1 \xrightarrow{\kappa_1} x_2 \qquad \lambda_1^N(x) = \kappa_1 x_1$$

$$x_1 + x_2 \xrightarrow{\kappa_2} x_3 \qquad \lambda_2^N(x) = \kappa_2 x_1\, x_2/N$$

- Basic assumption: the system is *uniformly mixed* and in *thermal equilibrium.*

2.3 First scaling limit – the reaction rate equation

If x gives the number of molecules of each species present, then $c = N^{-1}x$ gives the concentrations in moles per unit volume

Then

$$\lambda_k^N(x) \approx N\kappa_k \prod_i c_i^{\nu_{ik}} \equiv N\tilde{\lambda}_k(c).$$

The *law of large numbers* for the Poisson process implies $N^{-1}Y(Nu) \approx u$,

$$C(t) = N^{-1}X(t) \approx C(0) + \sum_k \int_0^t \kappa_k \prod_i C_i(s)^{\nu_{ik}}(\nu'_k - \nu_k)ds,$$

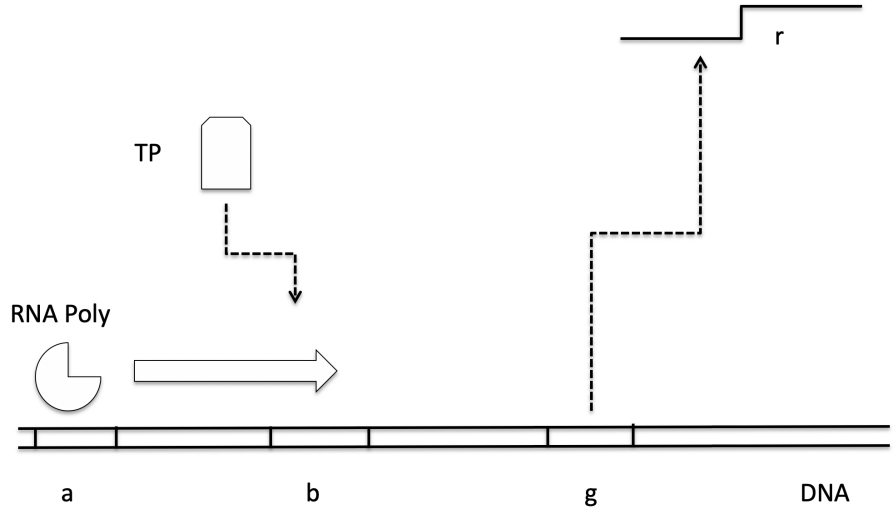
which in the large volume limit gives the classical *deterministic law of mass action* (RRE)

$$\boxed{\frac{dC(t)}{dt} = \sum_k \kappa_k \prod_i C_i(t)^{\nu_{ik}}(\nu'_k - \nu_k)}$$

2.4 Example: Speed of genetic transcription (SGT)

• Two chemical species, r (mRNA) and TP (transcription proteins) with the following simplified reactions

$$\begin{aligned} TP &\xrightarrow{\kappa_1} TP + r \\ r &\xrightarrow{\kappa_2} \emptyset \\ \emptyset &\xrightarrow{\kappa_3} TP \\ TP &\xrightarrow{\kappa_4} \emptyset \end{aligned}$$



Time change representation

- Denote the number of molecules of r by X_1 and the number of protein molecules (TP) by X_2.
- Let $X(t) = (X_1(t), X_2(t))$ be the state of the system at time t.

$$\begin{aligned} X_1(t) &= X_1(0) + Y_1\left(\int_0^t \kappa_1 X_2(s)ds\right) - Y_2\left(\int_0^t \kappa_2 X_1(s)ds\right) \\ X_2(t) &= X_2(0) + Y_3\left(\kappa_3 t\right) - Y_4\left(\int_0^t \kappa_4 X_2(s)ds\right) \end{aligned}$$

where Y_i for $i = 1, \ldots, 4$ are independent unit Poisson processes.

Reaction rate equation

Let N be system volume. Set

$$C_1(t) = X_1(t)/N \text{ and } C_2(t) = X_2(t)/N$$

• When N is large, the above consideration gives us the following deterministic approximation to the system of the stochastic Poisson equations

$$C_1(t) = C_1(0) + \kappa_1 \int_0^t C_2(s)ds - \kappa_2 \int_0^t C_1(s)ds$$
$$C_2(t) = C_2(0) + \tilde{\kappa}_3 t - \kappa_4 \int_0^t C_2(s)ds$$

or in the equivalent form obtained by differentiating the above

$$\begin{aligned} C_1'(t) = &\ k_1 C_2(t) - \kappa_2 C_1(t) \\ C_2'(t) = &\ \tilde{\kappa}_3 - \kappa_4 C_2(t) \end{aligned}$$

where $\tilde{\kappa}_3 = \lim_{N\to\infty} \kappa_3/N$

• The *deterministic steady state* is attained at $t = \infty$

Stationary distribution

- The generator in SGT example has the form

$$\begin{aligned}\mathbb{A}f(x_1,x_2) = \kappa_1 x_2[f(x_1+1,x_2)-f(x_1,x_2)] + \kappa_2 x_1[f(x_1-1,x_2)-f(x_1,x_2)] \\ +\kappa_3[f(x_1,x_2+1)-f(x_1,x_2)] + \kappa_4 x_2[f(x_1,x_2-1)-f(x_1,x_2)]\end{aligned}$$

- May be used to determine the stationary distribution
- Assume that $X(0) = (0,0)$
- Note: for the stationary distribution $\pi = \pi(x_1,x_2)$ we need

$$\pi\mathbb{A}f(x_1,x_2) = 0$$

- Taking for $t > 0$

$$f(x_1,x_2) = \exp(tx_2)$$

gives the differential equation for the marginal m.g.f.

Laplace transforms of the stationary distribution

- $\psi(t) = e^{tX_2}$ and set $\lambda = \kappa_3/\kappa_4$ then

$$\lambda(e^t - 1)\psi(t) = (1 - \exp(-t))\psi'(t)$$

with $\psi(0) = 1$ this implies

$$\psi(t) = \exp(\lambda(e^t - 1))$$

- That is, the second marginal of π, say π_2, is Poisson(κ_3/κ_4)
- More generally let $\phi(s,t) = Ee^{sX_1}e^{tX_2}$.
- As before, using the form $\mathbb{A}$ we obtain PDE for $\phi(s,t)$

$$[\kappa_1(e^s - 1) + \kappa_4(e^{-t} - 1)]\partial_t\phi + \kappa_2(e^{-s} - 1)\partial_s\phi = \kappa_3(1 - e^t)\phi \qquad (*)$$

with boundary conditions $\phi(0,t) = \psi(t)$.

- This linear PDE may be solved numerically using the method of *characteristic curves.*

Moments of the stationary distribution

- Taking $f(x_1, x_2) = x_1$ and using the form of $\mathbb{A}$

$$EX_1 = \kappa_1 EX_2/\kappa_2 = \frac{\kappa_1\kappa_3}{\kappa_2\kappa_4}$$

Similarly, taking $f(x_1, x_2) = x_1x_2$ and $f(x_1, x_2) = x_1^2$

$$EX_1X_2 = \frac{\kappa_1\frac{\kappa_3}{\kappa_4}\left[1+\frac{\kappa_3}{\kappa_4}\right] + \kappa_3\frac{\kappa_1\kappa_3}{\kappa_2\kappa_4}}{\kappa_2+\kappa_4}$$

and

$$2\kappa_2\, EX_1^2 = 2\kappa_1 EX_1X_2 + \kappa_1 EX_2 + \kappa_2 EX_1$$

etc

2.5 Stochastic SIR (McKendrick 1926)

- Three types of molecules: susceptibles (S), infectives (I), removed (R)
- Molecules combine after exponential holding time according to the current reaction rates (Gillespie algorithm)
- Rates given by *the law of mass action*

$$\begin{aligned} S + I &\xrightarrow{\lambda_1} 2I; \quad \lambda_1 = \beta SI \\ I &\xrightarrow{\lambda_2} R; \qquad \lambda_2 = \gamma I \end{aligned}$$

- The *basic reproduction number* $\mathcal{R}_0 = \beta/\gamma$.

Stochastic SIR (II)

- $(S(t), I(t), R(t))$ state of the collection of units at time $t > 0$
- $S(0) = n$; $I(0) = m$; $R(0) = 0$;
- β overall transmission rate; γ recovery rate

$$
\begin{aligned}
S(t) &= S(0) - Y_1\left(\frac{\beta}{n}\int_0^t S(u)I(u)du\right)\\
I(t) &= I(0) + Y_1\left(\frac{\beta}{n}\int_0^t S(u)I(u)du\right) - Y_2\left(\gamma\int_0^t I(u)du\right)\\
R(t) &= Y_2\left(\gamma\int_0^t I(u)du\right)
\end{aligned}
$$

- Y_1 and Y_2 are two independent unit Poisson processes.
- Recall basic assumption: the entire population is *uniformly mixed*.

Non-Markovian Case: the Sellke Construction

• Label the initial infectives $-(m-1), -(m-2), \ldots, 0$ and the initial susceptibles $1, 2, \ldots, n$. Let $\mathcal{I}_{-(m-1)}, \mathcal{I}_{-(m-2)}, \ldots, \mathcal{I}_n$ be iid random variables, each distributed according to same law $\mathcal{I}$.

• Let $Q_1, Q_2, \ldots, Q_n$ be an independent sequence of iid 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 $\mathcal{I}_i$ and is then removed.

• Let $Y(t)$ be the number of infectives at time t, and let

$$A(t) = \frac{\beta}{n} \int_0^t Y(u)du$$

be the total infection pressure exerted on a given susceptible up to time t.

• In $A(t)$ the infectives are weighted according to their infectious periods. For $i = 1, \ldots, n$, the susceptible labeled i becomes infected when $A(t)$ exceeds Q_i. The j-th susceptible infected who becomes infected (not one labeled j) remains infectious for a time $\mathcal{I}_j$ and is then removed.

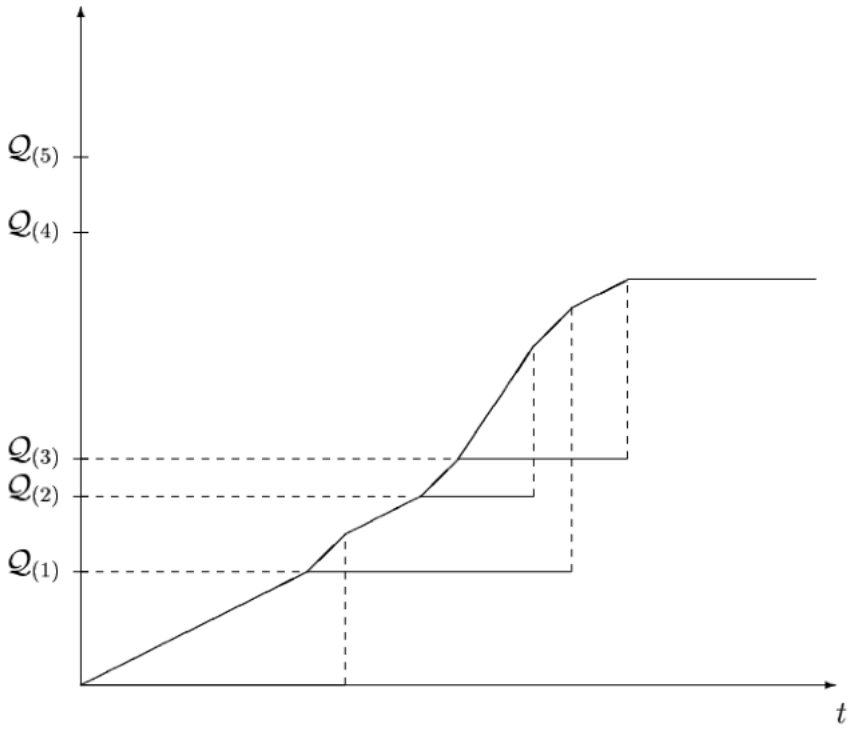


Figure 2.1: A typical realisation of the total infection pressure with $m = 1$ initially infectious individual. Note that the infection pressure never reaches $Q_{(4)}$ so the epidemic stops and the final size is $Z = 3$.

2.6 Exact Results for SIR $E_{n,m}(\beta, \mathcal{I})$

- Derive a triangular system of equations for the distribution of Z, say $P^n = (P_0^n, \ldots, P_n^n)$, where $P_k^n = P(Z = k)$.
- Let $A = A(\infty) = \frac{\beta}{n}\int_0^\infty Y(u)du$ be the total pressure of the epidemic. Note

$$Z = \min\{i : Q_{(i+1)} > \frac{\beta}{n}\sum_{j=-(m-1)}^{i} \mathcal{I}_j\},$$

where $Q_{(1)}, \ldots Q_{(n)}$ are the order statistics of $Q_1, \ldots, Q_n$. Additionally, note

$$A = \frac{\beta}{n}\sum_{j=-(m-1)}^{Z} \mathcal{I}_j.$$

Lemma 2.6.1 (Wald's Identity)**.** *Consider the standard SIR epidemic $E_{n,m}(\beta, \mathcal{I})$ and let A be as above. Then*

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

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

Proof.

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

□

- Using this result and the exchangeability of susceptibles we obtain

Theorem 2.6.1. *Consider standard SIR model $E_{n,m}(\beta,\mathcal{I})$ and let $P(Z=k)=P_k^n$. Then*

$$
\sum_{k=0}^{\ell}\left(\begin{array}{c} n-k \\ \ell-k \end{array}\right) P_k^n/[\phi(\lambda(n-\ell)/n)]^{k+m} = \left(\begin{array}{c} n \\ \ell \end{array}\right), \quad 0\le\ell\le n
$$

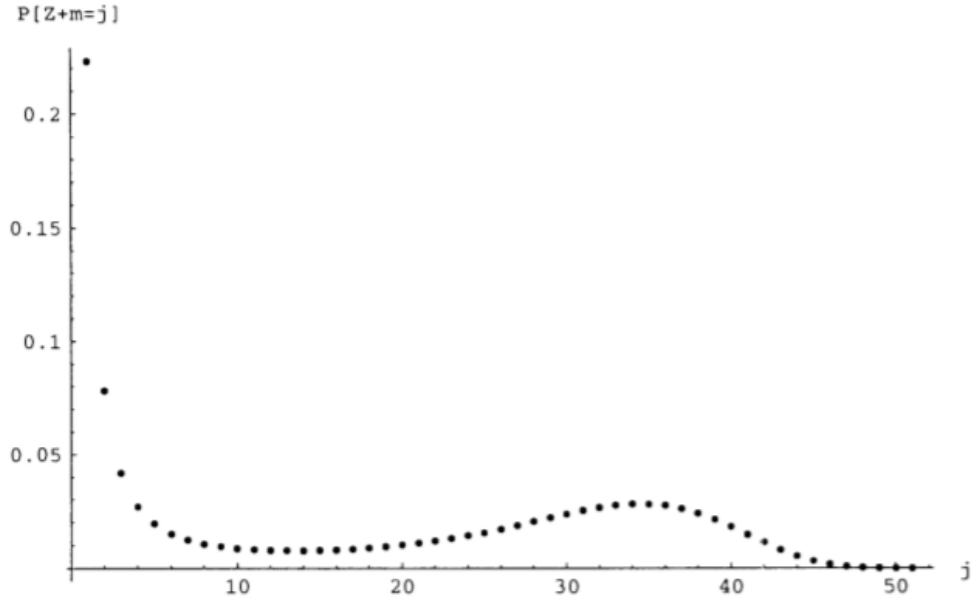


Figure 2.2: The exact distribution of $Z + m$ for $m = 1$, $n = 50$, $\lambda = 1.5$ and $I \equiv 1$, i.e. the infectious period is constant and equal to 1.

- Note the bimodality; This is know as *the threshold effect*.

References

[1] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*, volume 1. Springer, 2015.

[2] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.

[3] Grzegorz A Rempala, Kenneth S Ramos, and Ted Kalbfleisch. A stochastic model of gene transcription: an application to l1 retrotransposition events. *J Theor Biol*, 242(1):101–16, Sep 2006.

[4] Grzegorz A Rempala, Kenneth S Ramos, Ted Kalbfleisch, and Ivo Teneng. Validation of a mathematical model of gene transcription in aggregated cellular systems: application to l1 retrotransposition. *J Comput Biol*, 14(3):339–49, Apr 2007.

[5] Darren J Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2011.